

# PAC 1 - Anàlisi de dades òmiques

Alba Farràs Llobet

2024-10-27

## Descàrrega de les dades

He cercat un dataset relacionat amb el meu camp d'interès (Obstetrícia i Ginecologia) a Metabolomics Workbench.

### Dataset

**Metabolomics Workbench project:** PR000049

**Study ID:** ST000051

**Títol:** Biomarkers of Exposure to ARsenic (BEAR) study

**Autora:** Rebecca Fry, 1213 Michael Hooker Research Center, 135 Dauer Drive, Campus Box 7431, Chapel Hill, NC 27599

### Resum del projecte

The BEAR study has been established to assess the health effects associated with prenatal arsenic exposure during neonatal and early life. The study area is in Gómez Palacio, Durango, Mexico where women are exposed to levels of iAs in drinking water up to 236.0 µg As/L.

## Creació del contenidor SummarizedExperiment

```
# Carregar els paquets necessaris
library(readxl)
library(SummarizedExperiment)
library(dplyr)

# Llegir dades metabolòmiques
data_expr <- read_excel("ST000051/4a. Fetal Metabolomics of iAs exposure Phenotypic and Normalized Binn
  slice(1:50) %>%
  mutate(Exposure = as.factor(Exposure)) %>%
  rename(Subject_ID = `Subject ID`) %>%
  mutate(Subject_ID = as.factor(Subject_ID))

# Les primeres columnes són dades fenotípiques i les restants són dades d'expressió
phenotypic_data <- data_expr %>% select(1:2)
expression_data <- data_expr %>% select(3:ncol(data_expr)) %>% as.matrix()
expression_data <- t(expression_data)

# Assigna els Subject IDs com a noms de les columnes de expression_data
```

```
colnames(expression_data) <- phenotypic_data$`Subject_ID`

se <- SummarizedExperiment(
  assays = list(counts = expression_data),
  colData = phenotypic_data
)

se

## class: SummarizedExperiment
## dim: 221 50
## metadata(0):
## assays(1): counts
## rownames(221): [0.50 .. 0.52] [0.52 .. 0.56] ... [9.39 .. 9.45] [9.45
##   .. 9.50]
## rowData names(0):
## colnames(50): C-045 C-053 ... C-189 C-200
## colData names(2): Subject_ID Exposure
```

## Exploració de les dades

### Anàlisi exploratòria

```
# Resum de les dades fenotípiques
summary(colData(se)$Exposure)

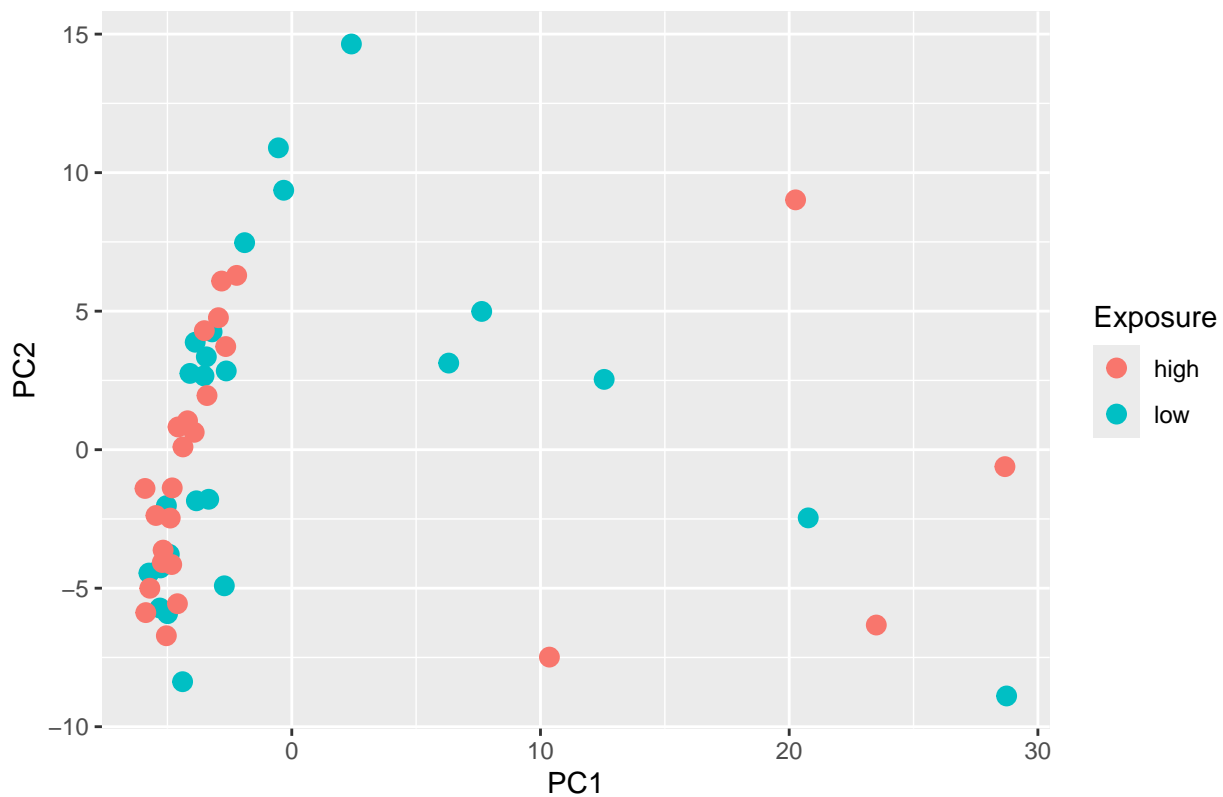
## high low
##    25   25

# PCA
library(ggplot2)

## Warning: package 'ggplot2' was built under R version 4.3.3
pca <- prcomp(t(assay(se, "counts")), scale. = TRUE)
pca_data <- as.data.frame(pca$x)
pca_data$Exposure <- colData(se)$Exposure

ggplot(pca_data, aes(PC1, PC2, color = Exposure)) +
  geom_point(size = 3) +
  labs(title = "PCA de les dades de Metabolòmica", x = "PC1", y = "PC2")
```

## PCA de les dades de Metabolòmica



```
# T-test entre grups d'exposició per cada variable
t_tests <- apply(assay(se, "counts"), 1, function(x) {
  t.test(x ~ colData(se)$Exposure)$p.value
})

# Mostra els resultats més significatius
t_tests_significant <- which(t_tests < 0.05)
rownames(se)[t_tests_significant]

## [1] "[1.48 .. 1.52]" "[1.69 .. 1.75]" "[1.90 .. 1.92]" "[3.63 .. 3.68]"

# Dades d'expressió dels metabòlits significatius
expr_significant <- assay(se, "counts")[t_tests_significant, ]

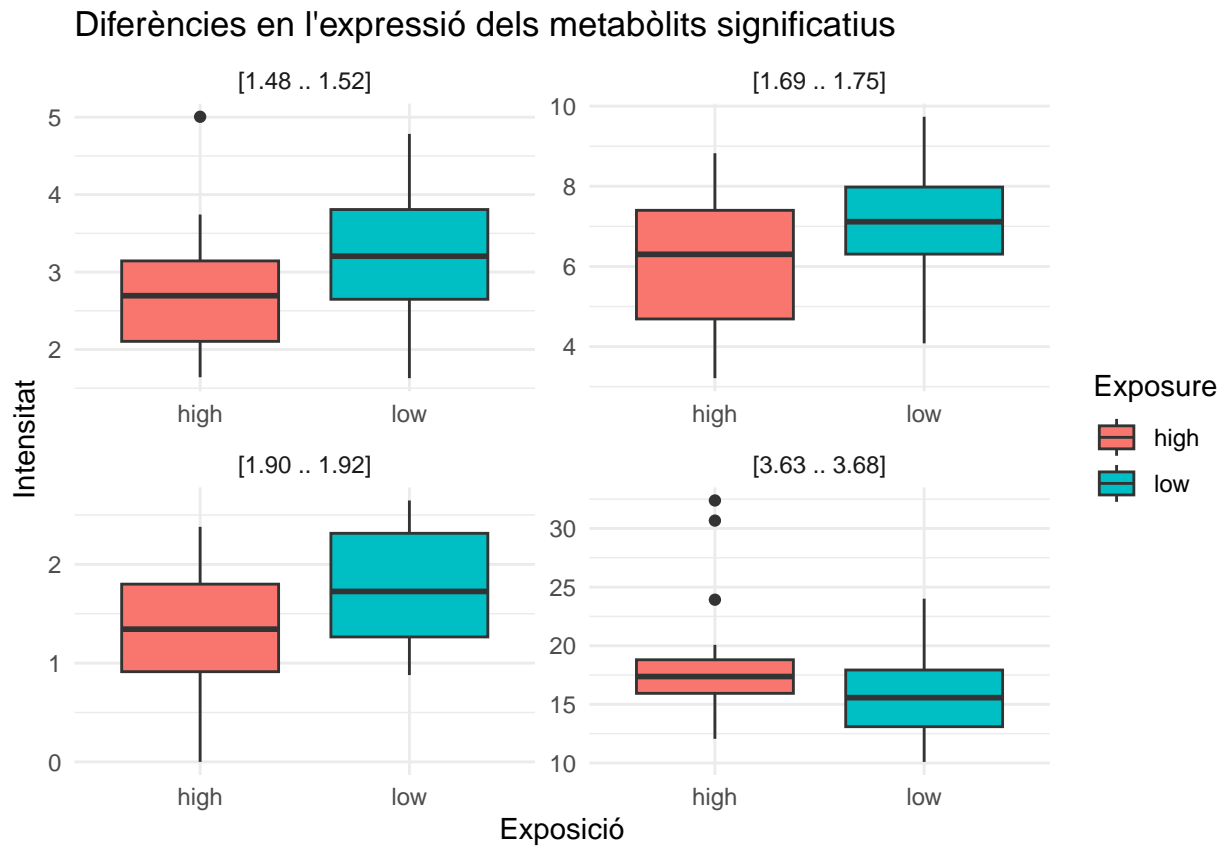
# Convertim les dades a format llarg
library(reshape2)

## Warning: package 'reshape2' was built under R version 4.3.1
expr_long <- melt(expr_significant)
colnames(expr_long) <- c("Metabolit", "Subject_ID", "Valor")

# Afegim la informació de l'exposició
expr_long <- merge(expr_long, phenotypic_data, by = "Subject_ID")

# Representem els boxplots
ggplot(expr_long, aes(x = Exposure, y = Valor, fill = Exposure)) +
  geom_boxplot() +
  facet_wrap(~ Metabolit, scales = "free") +
```

```
theme_minimal() +
labs(title = "Diferències en l'expressió dels metabòlits significatius", x = "Exposició", y = "Intensitat")
```



## Desat de les dades

### Desar l'objecte SummarizedExperiment en un arxiu .Rda

```
# Guarda l'objecte SummarizedExperiment en format .Rda
save(se, file = "SummarizedExperiment.Rda")
```

### Desar dades en format text .txt

```
# Exporta dades d'expressió a un fitxer .txt
write.table(assay(se, "counts"), file = "expression_data.txt", sep = "\t", row.names = TRUE, col.names = TRUE)

# Exporta dades fenotípiques a un fitxer .txt
write.table(colData(se), file = "phenotypic_data.txt", sep = "\t", row.names = TRUE, col.names = NA)
```

### Desar metadades en format .md

```
# Crea un fitxer Markdown amb les metadades del dataset
metadata_md <- "
# Metadades del Dataset
```

```
## Informació del Projecte
- Projecte: PR000049
- Study ID: ST000051
- Títol: Biomarkers of Exposure to ARsenic (BEAR) study
- Autora: Rebecca Fry, 1213 Michael Hooker Research Center, Chapel Hill, NC 27599

## Resum
L'estudi BEAR avalua els efectes en la salut associats amb l'exposició prenatal a l'arsènic a Gómez Pal...

## Dades Fenotípiques
- Subject_ID: Identificador únic del subjecte
- Exposure: Nivell d'exposició a l'arsènic ('high' o 'low')
"

writeLines(metadata_md, "metadata.md")
```

## Repositori de GitHub

He creat un repositori públic al meu compte de GitHub (@albafl) on hi he desat tots els fitxers de la PAC.  
<https://github.com/albafl/FARRAS-Llobet-Alba-PEC1>