PAC 1 Dades omiques

Balma Albalat Oliver

#INFORME ##1.DESCRIPCIÓ I JUSTIFICACIO DE LA BASE DE DADES ## He triat la database associada al projecte Metabolomics of Psoriasis. ## Aquesta pot consular-se en el repositori “Metabolomics workbench” amb l’identificador de projecte PR000239. ## He triat aquest tema perquè pateixo aquesta malaltia i m’ha causat curiositat #aprofundir en aquesta qüestió. ## Es tracta d’un estudi desenvolupat pel Departament de Dermatologia de la Universitat de Michigan, dirigit per l’investigador Johann Gudjonsson. L’objectiu de l’estudi és estudiar els metabolits esteroides en els teixits de pacients amb diferents estats de psoriasis. Per assolir-lo es van realitzar biòpsis per obtenir 9 mostres de pell que es van categoritzar en 3 subgrups:

#-Pell sana (3) -Pell afectada per psoriasis (3) -Pell no afectada en pacients amb psoriasis (3)

#A través de la cromatografía líquida acoplada a l’espectrometria de masses (LC-MS) es van analitzar les concentracions d’esteroides, expressades en pg/mg de teixit, caracteritzant d’aquesta manera els perfils metabòlics de cada mostra de pell.

if(!require(jsonlite)){  
 install.packages("jsonlite")  
 library(jsonlite)}

## Cargando paquete requerido: jsonlite

if (!require("BiocManager", quietly = TRUE))  
 install.packages("BiocManager")  
if (!require("ggtext", quietly = TRUE))  
 install.packages("ggtext")  
if (!require("magrittr", quietly = TRUE))  
 install.packages("magrittr")  
if (!require("dplyr", quietly = TRUE))  
 install.packages("dplyr")

##   
## Adjuntando el paquete: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

if (!require("tidyr", quietly = TRUE))  
 install.packages("tidyr")

##   
## Adjuntando el paquete: 'tidyr'

## The following object is masked from 'package:magrittr':  
##   
## extract

if (!require("tibble", quietly = TRUE))  
 install.packages("tibble")  
if (!require("purrr", quietly = TRUE))  
 install.packages("purrr")

##   
## Adjuntando el paquete: 'purrr'

## The following object is masked from 'package:magrittr':  
##   
## set\_names

## The following object is masked from 'package:jsonlite':  
##   
## flatten

if (!require("readr", quietly = TRUE))  
 install.packages("readr")  
if (!requireNamespace("SummarizedExperiment", quietly = TRUE)) {  
 install.packages("SummarizedExperiment")  
}  
if (!requireNamespace("ggplot2", quietly = TRUE)) {  
 install.packages("ggplot2")  
}  
BiocManager::install("Biobase")

## Bioconductor version 3.20 (BiocManager 1.30.25), R 4.4.3 (2025-02-28 ucrt)

## Warning: package(s) not installed when version(s) same as or greater than current; use  
## `force = TRUE` to re-install: 'Biobase'

## Installation paths not writeable, unable to update packages  
## path: C:/Program Files/R/R-4.4.3/library  
## packages:  
## cluster, foreign, MASS, Matrix

## Old packages: 'httr2', 'lme4', 'MatrixModels', 'recipes'

BiocManager::install("POMA")

## Bioconductor version 3.20 (BiocManager 1.30.25), R 4.4.3 (2025-02-28 ucrt)

## Warning: package(s) not installed when version(s) same as or greater than current; use  
## `force = TRUE` to re-install: 'POMA'

## Installation paths not writeable, unable to update packages  
## path: C:/Program Files/R/R-4.4.3/library  
## packages:  
## cluster, foreign, MASS, Matrix  
## Old packages: 'httr2', 'lme4', 'MatrixModels', 'recipes'

BiocManager::install("SummarizedExperiment")

## Bioconductor version 3.20 (BiocManager 1.30.25), R 4.4.3 (2025-02-28 ucrt)

## Warning: package(s) not installed when version(s) same as or greater than current; use  
## `force = TRUE` to re-install: 'SummarizedExperiment'

## Installation paths not writeable, unable to update packages  
## path: C:/Program Files/R/R-4.4.3/library  
## packages:  
## cluster, foreign, MASS, Matrix  
## Old packages: 'httr2', 'lme4', 'MatrixModels', 'recipes'

library(POMA)

## Welcome to POMA!  
## Version 1.16.0  
## POMAShiny app: https://github.com/pcastellanoescuder/POMAShiny

library(ggtext)  
library(magrittr)  
library(SummarizedExperiment)

## Cargando paquete requerido: MatrixGenerics

## Cargando paquete requerido: matrixStats

##   
## Adjuntando el paquete: 'matrixStats'

## The following object is masked from 'package:dplyr':  
##   
## count

##   
## Adjuntando el paquete: 'MatrixGenerics'

## The following objects are masked from 'package:matrixStats':  
##   
## colAlls, colAnyNAs, colAnys, colAvgsPerRowSet, colCollapse,  
## colCounts, colCummaxs, colCummins, colCumprods, colCumsums,  
## colDiffs, colIQRDiffs, colIQRs, colLogSumExps, colMadDiffs,  
## colMads, colMaxs, colMeans2, colMedians, colMins, colOrderStats,  
## colProds, colQuantiles, colRanges, colRanks, colSdDiffs, colSds,  
## colSums2, colTabulates, colVarDiffs, colVars, colWeightedMads,  
## colWeightedMeans, colWeightedMedians, colWeightedSds,  
## colWeightedVars, rowAlls, rowAnyNAs, rowAnys, rowAvgsPerColSet,  
## rowCollapse, rowCounts, rowCummaxs, rowCummins, rowCumprods,  
## rowCumsums, rowDiffs, rowIQRDiffs, rowIQRs, rowLogSumExps,  
## rowMadDiffs, rowMads, rowMaxs, rowMeans2, rowMedians, rowMins,  
## rowOrderStats, rowProds, rowQuantiles, rowRanges, rowRanks,  
## rowSdDiffs, rowSds, rowSums2, rowTabulates, rowVarDiffs, rowVars,  
## rowWeightedMads, rowWeightedMeans, rowWeightedMedians,  
## rowWeightedSds, rowWeightedVars

## Cargando paquete requerido: GenomicRanges

## Cargando paquete requerido: stats4

## Cargando paquete requerido: BiocGenerics

##   
## Adjuntando el paquete: 'BiocGenerics'

## The following objects are masked from 'package:dplyr':  
##   
## combine, intersect, setdiff, union

## The following objects are masked from 'package:stats':  
##   
## IQR, mad, sd, var, xtabs

## The following objects are masked from 'package:base':  
##   
## anyDuplicated, aperm, append, as.data.frame, basename, cbind,  
## colnames, dirname, do.call, duplicated, eval, evalq, Filter, Find,  
## get, grep, grepl, intersect, is.unsorted, lapply, Map, mapply,  
## match, mget, order, paste, pmax, pmax.int, pmin, pmin.int,  
## Position, rank, rbind, Reduce, rownames, sapply, saveRDS, setdiff,  
## table, tapply, union, unique, unsplit, which.max, which.min

## Cargando paquete requerido: S4Vectors

##   
## Adjuntando el paquete: 'S4Vectors'

## The following object is masked from 'package:tidyr':  
##   
## expand

## The following objects are masked from 'package:dplyr':  
##   
## first, rename

## The following object is masked from 'package:utils':  
##   
## findMatches

## The following objects are masked from 'package:base':  
##   
## expand.grid, I, unname

## Cargando paquete requerido: IRanges

##   
## Adjuntando el paquete: 'IRanges'

## The following object is masked from 'package:purrr':  
##   
## reduce

## The following objects are masked from 'package:dplyr':  
##   
## collapse, desc, slice

## The following object is masked from 'package:grDevices':  
##   
## windows

## Cargando paquete requerido: GenomeInfoDb

##   
## Adjuntando el paquete: 'GenomicRanges'

## The following object is masked from 'package:magrittr':  
##   
## subtract

## Cargando paquete requerido: Biobase

## Welcome to Bioconductor  
##   
## Vignettes contain introductory material; view with  
## 'browseVignettes()'. To cite Bioconductor, see  
## 'citation("Biobase")', and for packages 'citation("pkgname")'.

##   
## Adjuntando el paquete: 'Biobase'

## The following object is masked from 'package:MatrixGenerics':  
##   
## rowMedians

## The following objects are masked from 'package:matrixStats':  
##   
## anyMissing, rowMedians

library(readr)  
library(purrr)  
library(ggplot2)

##### PREPARACIÓ I IMPORTACIÓ DE LES DADES ####  
  
# Importar l'arxiu de dades '.json'  
  
library(jsonlite)  
json\_data <-jsonlite::fromJSON("ST000298\_AN000476.json", flatten = TRUE)  
# Cerquem i resolem valors nuls #  
null\_values <- sapply(json\_data, function(x) any(is.null(x)))  
na\_values <- sapply(json\_data, function(x) any(is.na(x)))  
# Print results  
print("Valors nuls en:")

## [1] "Valors nuls en:"

print(names(json\_data)[null\_values])

## character(0)

### Creació colData (info i factor mostres) ###  
# Extraem la info de json\_data  
if (is.data.frame(json\_data$SUBJECT\_SAMPLE\_FACTORS)) {  
 sample\_df <- json\_data$SUBJECT\_SAMPLE\_FACTORS  
} else {  
 sample\_info <- as.character(unlist(data$SUBJECT\_SAMPLE\_FACTORS))  
 sample\_df <- read.table(text = sample\_info, sep = "\t", header = FALSE, stringsAsFactors = FALSE)  
}  
colnames(sample\_df) <- c("Subject", "SampleID", "Factors")  
  
# Associem els factors a columnes i creem colData#  
  
# Crear un DataFrame de factores  
factor\_df <- data.frame(Factor = sample\_df$Factors)  
  
# Combinar con el DataFrame original  
colData <- cbind(sample\_df[, "SampleID", drop = FALSE], factor\_df)  
  
# Asignar nombres a las filas  
row.names(colData) <- colData$SampleID  
  
# Ver el resultado  
head(colData)

## SampleID Factor  
## S00017314 S00017314 Psoriasis involved  
## S00017315 S00017315 Psoriasis involved  
## S00017316 S00017316 Psoriasis involved  
## S00017317 S00017317 Psoriasis uninvolved  
## S00017318 S00017318 Psoriasis uninvolved  
## S00017319 S00017319 Psoriasis uninvolved

### Extreuracció de dades metabòliques (fe) ###  
# L'utilitzarem per a fer els anàlisi #  
  
metabolite\_data<-json\_data$MS\_METABOLITE\_DATA$Data  
metabolite\_matrix <- as.data.frame(metabolite\_data)  
rownames(metabolite\_matrix) <- metabolite\_matrix$Metabolite  
metabolite\_matrix <- metabolite\_matrix[, -1] #   
metabolite\_matrix <- t(metabolite\_matrix)  
metabolite\_matrix <- as.matrix(sapply(metabolite\_matrix, as.numeric))  
head(metabolite\_matrix)

## [,1]  
## 0.570 0.57  
## 0.610 0.61  
## 0 0.00  
## 0.31 0.31  
## 2.68 2.68  
## 0.26 0.26

metabolite\_data <- json\_data$MS\_METABOLITE\_DATA$Data  
metabolite\_matrix <- as.data.frame(metabolite\_data)  
rownames(metabolite\_matrix) <- metabolite\_matrix$Metabolite  
metabolite\_matrix <- metabolite\_matrix[, -1] # Eliminar la columna redundant  
metabolite\_matrix[metabolite\_matrix == ""] <- NA  
metabolite\_matrix <- as.matrix(sapply(metabolite\_matrix, as.numeric))  
head(metabolite\_matrix)

## S00017317 S00017315 S00017322 S00017316 S00017318 S00017319 S00017320  
## [1,] 0.57 0.61 0.00 0.31 2.68 0.26 0.93  
## [2,] 1.76 0.48 3.74 0.97 3.05 1.29 5.26  
## [3,] 0.18 6.83 29.85 111.35 109.30 53.91 337.19  
## [4,] 0.60 4.62 4.86 3.82 7.71 8.01 17.09  
## [5,] 0.57 NA 0.85 NA NA 0.82 NA  
## [6,] 6.10 3.30 23.00 24.10 20.10 21.70 130.80  
## S00017321 S00017314  
## [1,] 3.33 1.43  
## [2,] 9.88 2.05  
## [3,] 2126.34 66.98  
## [4,] 105.27 4.02  
## [5,] NA 0.63  
## [6,] 915.40 26.60

1. CREAR OBJECTE ‘SummarizedExperiment’

# Eliminar la fila 5, que conté la info del metabolit amb dades faltants  
  
metabolite\_matrix <- metabolite\_matrix[-5, ]  
metabolite\_matrix <-t(metabolite\_matrix)  
colData <- colData[!rownames(colData) %in% "Corticosterone", ]  
  
# Comprovem que coincideixen les files  
  
nrow(metabolite\_matrix)

## [1] 9

nrow(colData)

## [1] 9

# Generm el SummarizedExperiment   
  
library(POMA)  
poma\_obj <- PomaCreateObject(  
 metadata = colData,   
 features = metabolite\_matrix   
)  
  
# Verifiquem  
  
poma\_obj

## class: SummarizedExperiment   
## dim: 8 9   
## metadata(0):  
## assays(1): ''  
## rownames(8): V1 V2 ... V7 V8  
## rowData names(0):  
## colnames(9): S00017314 S00017315 ... S00017321 S00017322  
## colData names(1): Factor

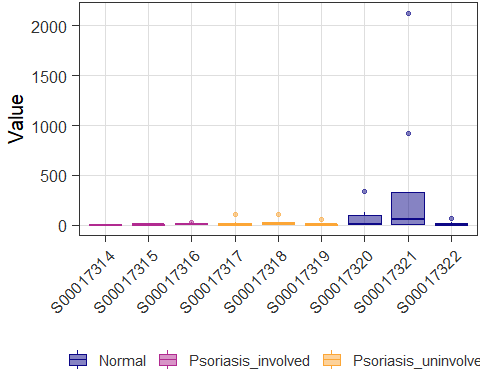
1. AFEGIM METADADES GENERALS DEL PROJECTE

# Afegim les metadades a l'objecte SummarizedExperiment  
for (key in names(json\_data$METABOLOMICS)) {  
 metadata(poma\_obj)$METABOLOMICS[[key]] <- json\_data$METABOLOMICS[[key]]  
}  
for (key in names(json\_data$PROJECT)) {  
 metadata(poma\_obj)$PROJECT[[key]] <- json\_data$PROJECT[[key]]  
}  
  
for (key in names(json\_data$STUDY)) {  
 metadata(poma\_obj)$STUDY[[key]] <- json\_data$STUDY[[key]]  
}  
  
for (key in names(json\_data$SUBJECT)) {  
 metadata(poma\_obj)$SUBJECT[[key]] <- json\_data$SUBJECT[[key]]  
}  
  
for (key in names(json\_data$SUBJECT\_SAMPLE\_FACTORS)) {  
 metadata(poma\_obj)$SUBJECT\_SAMPLE\_FACTORS[[key]] <- json\_data$SUBJECT\_SAMPLE\_FACTORS[[key]]  
}  
  
for (key in names(json\_data$COLLECTION)) {  
 metadata(poma\_obj)$COLLECTION[[key]] <- json\_data$COLLECTION[[key]]  
}  
  
for (key in names(json\_data$TREATMENT)) {  
 metadata(poma\_obj)$TREATMENT[[key]] <- json\_data$TREATMENT[[key]]  
}  
  
for (key in names(json\_data$SAMPLEPREP)) {  
 metadata(poma\_obj)$SAMPLEPREP[[key]] <- json\_data$SAMPLEPREP[[key]]  
}  
  
for (key in names(json\_data$CHROMATOGRAPHY)) {  
 metadata(poma\_obj)$CHROMATOGRAPHY[[key]] <- json\_data$CHROMATOGRAPHY[[key]]  
}  
  
for (key in names(json\_data$ANALYSIS)) {  
 metadata(poma\_obj)$ANALYSIS[[key]] <- json\_data$ANALYSIS[[key]]  
}  
  
for (key in names(json\_data$MS)) {  
 metadata(poma\_obj)$MS[[key]] <- json\_data$MS[[key]]  
}  
  
metadata(poma\_obj)

## $METABOLOMICS  
## $METABOLOMICS$STUDY\_ID  
## [1] "ST000298"  
##   
## $METABOLOMICS$ANALYSIS\_ID  
## [1] "AN000476"  
##   
## $METABOLOMICS$VERSION  
## [1] "1"  
##   
## $METABOLOMICS$CREATED\_ON  
## [1] "December 28, 2015, 11:42 am"  
##   
##   
## $PROJECT  
## $PROJECT$PROJECT\_TITLE  
## [1] "Metabolomics of Psoriasis"  
##   
## $PROJECT$PROJECT\_SUMMARY  
## [1] "Metabolomics of Psoriasis"  
##   
## $PROJECT$INSTITUTE  
## [1] "University of Michigan"  
##   
## $PROJECT$DEPARTMENT  
## [1] "Dermatology"  
##   
## $PROJECT$LABORATORY  
## [1] "Gudjonsson Lab"  
##   
## $PROJECT$LAST\_NAME  
## [1] "Gudjonsson"  
##   
## $PROJECT$FIRST\_NAME  
## [1] "Johann"  
##   
## $PROJECT$ADDRESS  
## [1] "Ann Arbor, MI"  
##   
## $PROJECT$EMAIL  
## [1] "johanng@med.umich.edu"  
##   
## $PROJECT$PHONE  
## [1] "734-615-4508"  
##   
##   
## $STUDY  
## $STUDY$STUDY\_TITLE  
## [1] "Analysis of steroid metabolites in psoriasis."  
##   
## $STUDY$STUDY\_SUMMARY  
## [1] "Analysis of steroid metabolites in tissue of patients with different psoriasis status."  
##   
## $STUDY$INSTITUTE  
## [1] "University of Michigan"  
##   
## $STUDY$DEPARTMENT  
## [1] "Biomedical Research Core Facilities"  
##   
## $STUDY$LABORATORY  
## [1] "Metabolomics core"  
##   
## $STUDY$LAST\_NAME  
## [1] "Kachman"  
##   
## $STUDY$FIRST\_NAME  
## [1] "Maureen"  
##   
## $STUDY$ADDRESS  
## [1] "6300 Brehm Tower, 1000 Wall Street, Ann Arbor, MI 48105-5714"  
##   
## $STUDY$EMAIL  
## [1] "mkachman@med.umich.edu"  
##   
## $STUDY$PHONE  
## [1] "(734) 232-8175"  
##   
## $STUDY$NUM\_GROUPS  
## [1] "3"  
##   
## $STUDY$TOTAL\_SUBJECTS  
## [1] "9"  
##   
##   
## $SUBJECT  
## $SUBJECT$SUBJECT\_TYPE  
## [1] "Human"  
##   
## $SUBJECT$SUBJECT\_SPECIES  
## [1] "Homo sapiens"  
##   
## $SUBJECT$TAXONOMY\_ID  
## [1] "9606"  
##   
##   
## $SUBJECT\_SAMPLE\_FACTORS  
## $SUBJECT\_SAMPLE\_FACTORS$`Subject ID`  
## [1] "SU0011723" "SU0011723" "SU0011723" "SU0011723" "SU0011723" "SU0011723"  
## [7] "SU0011723" "SU0011723" "SU0011723"  
##   
## $SUBJECT\_SAMPLE\_FACTORS$`Sample ID`  
## [1] "S00017314" "S00017315" "S00017316" "S00017317" "S00017318" "S00017319"  
## [7] "S00017320" "S00017321" "S00017322"  
##   
## $SUBJECT\_SAMPLE\_FACTORS$`Factors.Psoriasis Status`  
## [1] "Psoriasis involved" "Psoriasis involved" "Psoriasis involved"   
## [4] "Psoriasis uninvolved" "Psoriasis uninvolved" "Psoriasis uninvolved"  
## [7] "Normal" "Normal" "Normal"   
##   
##   
## $COLLECTION  
## $COLLECTION$COLLECTION\_SUMMARY  
## [1] "-"  
##   
## $COLLECTION$SAMPLE\_TYPE  
## [1] "Biopsy"  
##   
##   
## $TREATMENT  
## $TREATMENT$TREATMENT\_SUMMARY  
## [1] "-"  
##   
##   
## $SAMPLEPREP  
## $SAMPLEPREP$SAMPLEPREP\_SUMMARY  
## [1] "-"  
##   
## $SAMPLEPREP$SAMPLEPREP\_PROTOCOL\_FILENAME  
## [1] "Steroids-20150305.docx"  
##   
##   
## $CHROMATOGRAPHY  
## $CHROMATOGRAPHY$CHROMATOGRAPHY\_TYPE  
## [1] "Reversed phase"  
##   
## $CHROMATOGRAPHY$INSTRUMENT\_NAME  
## [1] "Agilent 1290"  
##   
## $CHROMATOGRAPHY$COLUMN\_NAME  
## [1] "Restek 2.1mm x 50mm 1.9um Pinnacle DB Biphenyl"  
##   
## $CHROMATOGRAPHY$METHODS\_FILENAME  
## [1] "Steroids-20150305.docx"  
##   
##   
## $ANALYSIS  
## $ANALYSIS$ANALYSIS\_TYPE  
## [1] "MS"  
##   
## $ANALYSIS$ANALYSIS\_PROTOCOL\_FILE  
## [1] "Steroids-20150305.docx"  
##   
##   
## $MS  
## $MS$MS\_COMMENTS  
## [1] "-"  
##   
## $MS$INSTRUMENT\_TYPE  
## [1] "Triple quadrupole"  
##   
## $MS$MS\_TYPE  
## [1] "ESI"  
##   
## $MS$ION\_MODE  
## [1] "UNSPECIFIED"  
##   
## $MS$INSTRUMENT\_NAME  
## [1] "Agilent 6410A QQQ Agilent QQQ 6410A"

1. ANÀLISI EXPLORATORI

# MISSING VALUES #  
  
#Ja els hem netejat durant la preparació del SummarizedExperiment. Hem eliminat el metabolit 'Corticosterone' que no presentava dades per a totes les mostres.  
  
# CAL NORMALITZAR LES DADES? #  
  
  
# Dades en cru #  
p1<-PomaBoxplots(  
 poma\_obj,   
 x= "samples",  
 feature\_name = NULL,)  
  
print(p1)



# Dades normalitzades  
normalized<- poma\_obj %>%   
 PomaNorm(method="log\_pareto")  
p2<-PomaBoxplots(  
 normalized,  
 x= "samples",  
 feature\_name = NULL,)  
  
print(p2)

