

PAC 1 Anàlisi de dades òmiques

<https://github.com/mesquerrap/ESQUERRA-Pares-Marina-PAC1>

Marina Esquerrà a Parés

2024-10-29

Taula de Continguts

Abstract	1
Objectius	1
Materials i Mètodes	2
Resultats	3
Creació del SummarizedExperiment	3
Exploració multivariant de dades	6
Discussió, limitacions i conclusions de l'estudi	13

Abstract

Aquest estudi es basa en l'anàlisi del *dataset* “2024-Cachexia” extret d'un repositori de github. Per a l'explotació s'ha utilitzat eines com Bioconductor i tècniques d'exploració multivariant de dades. L'exercici comença explicant com s'ha adaptat el document.csv a l'estructura requerida per SummarizedExperiment, a continuació, i a través d'exemples, s'explica algun dels avantatges de treballar amb aquest contenidor i per últim en centra en l'exploració multivariant de les dades. En aquesta exploració i després de l'anàlisi de components principals, s'observa que els memetabòlits que més influeixen en la diferenciació de grup són la creatinine, la glutamine, l'ethanolamine, l'asparagine, l'acetone i el tartrate, tots ells relacionats en el cicle de construcció, manteniment i deteriorament dels teixits musculars o adiposos. A l'apartat de material i mètodes, també s'explica com s'ha transformat el contenidor de dades i les metadades en format binari i per últim com s'ha creat el repositori de *github*.

Objectius

- Endinsar-se en l'ús de plataformes utilitzades per a l'anàlisi de dades òmiques com GitHub i Bioconductor, a través de contenidors de dades òmiques com expressionSets.
- Iniciar-se en les anàlisis multivariants de dades òmiques.
- Arribar a respondre algun tipus de pregunta biològica gràcies als resultats extrets de l'anàlisi multivariant, o si més no tenir una idea general del que mostren les dades.

Materials i Mètodes

L'enunciat de la PAC proporcionava un enllaç a un repositori de dades github “<https://github.com/nutrimetabolomics/metaboData/>”, des d'on es va accedir a llistat de *datasets* per escollir-ne un. Després de fer una visió general de les opcions, es va escollir el *dataset* “2024-Cachexia”, el qual conté un document “description.md” amb informació general de les dades i un document “human_cachexia.csv” que contenia les dades.

El programa informàtic utilitzar tant per tractar les dades com per redactar l'informe, va ser R, concretament l'eina R Markdown. Un cop descarregades les dades del repositori, es van importar al R Markdown. A continuació seguint el tutorial, proporcionat a l'apartat de recursos de l'anunciat de la PAC i amb informació general després d'una cerca a Google i amb la funció `help()`, es va crear el contenidor del tipus `SummarizedExperiment`, una extensió d'`ExpressionSet`.

Un cop creat el contenidor, seguint els exemples proporcionats per “Caso de estudio de introducción a Bioconductor” de l'aula virtual (Recursos d'aprenentatge 1.1), es van fer algunes comprovacions i es van donar exemples de com n'és d'útil treballar amb un `SummarizedExperiment` per tractar dades òmiques.

Per a l'anàlisi estadístic de les dades, es van agafat com a referència els documents “Casos y Ejemplos Resueltos de Análisis Multivariante” i “Resúmenes de Algebra Lineal para el análisis multivariante” també de l'aula virtual (Recursos d'aprenentatge 1.3)

La redacció i l'estructura de l'informe es va fer seguint el document “Directrius sobre el format de la PEC” proporcionat a l'aula virtual.

Per guardar l'objecte contenidor amb les dades i les metadades en format binari es va utilitzar la funció “`save(SummarizedExperiment_cachexia, file = "SummarizedExperiment_cachexia.RDA")`”, especificant que el format fos `.RDA`, el qual per defecte ho guarda en format binari. Un cop s'executa aquest codi, es genera un document, `SummarizedExperiment_cachexia.RDA`, al directori de treball.

Per a la creació del repositori de dades a *github*, es va creat un nou repositori públic amb el nom “ESQUERRA-Pares-Marina-PAC1”. A continuació es van afegir els documents requerits; informe en pdf, l'objecte contenidor amb les dades i les metadades en format binari de R (arxiu amb extensió `.Rda`), el codi R per a l'exploració de les dades en format text i les metadades sobre el dataset en un arxiu markdown. El link d'aquest repositori és:

<https://github.com/mesquerrap/ESQUERRA-Pares-Marina-PAC1>

Resultats

Creació del SummarizedExperiment

Per tal de no allargar-se i fer l'informe més llegible, en aquest apartat es descriu com s'ha creat el contenidor, però no es visualitzaran algunes de les dades (head()) a causa de la seva extensió. Tot i això, per tal d'assegurar que les dades s'estaven important i transformant com es demanava, durant la programació sí que s'han utilitzat aquesta eina.

Carreguem les llibreries necessaries a realitzar la PAC

```
#if (!requireNamespace("BiocManager", quietly = TRUE)) {  
#  install.packages("BiocManager")}  
#BiocManager::install("SummarizedExperiment")  
library(SummarizedExperiment)  
library(readr)  
library(ggplot2)  
library(gridExtra)
```

Després de descarregar les dades del repositori, es carreguen a R.

```
cachexia <- read.csv("C:/Users/mesquerra/Desktop/UOC/Analisi_de_dades_omiques/PAC1/2024-Cachexia/human_cachexia.csv")  
# Comprovem que s'han importat correctament.  
# head(cachexia)  
  
# Summary donarà una idea de les dades i la seva organització  
# summary(cachexia)
```

“human_cachexia.csv”, Conté dades d'un total de 77 pacients i presenta informació sobre variables com la pèrdua muscular i diverses concentracions de metabòlits. Les concentracions dels metabòlits mostren una variabilitat considerable la qual analitzarem més endavant.

Ara que les dades .csv ja estan carregades, procedim a crear el contenidor SummarizedExperiment, el qual està format pels elements; assay, rownames, colData i metadata.

Creació de assay: Per a crear la matriu assay cal que les dades estiguin distribuïdes:

- Columnes: cada una de les columnes ha de correspondre a una pacient.
- Files: cada una de les files ha de correspondre a un metabòlit.

```
# Seleccionem totes les files però només les columnes a partir de la tercera  
# (els metabòlits) de cachexia.  
assay <- as.matrix(cachexia[, 3:ncol(cachexia)])  
  
# Per tal que les mostres estiguin a les columnes i els metabòlits en files, és  
# necessari transposar la matriu.  
assay <- t(assay)  
  
# Visualitzem que s'ha fet la transformació de dades correctament  
# head(assay)
```

Creació del DataFrame rowData: Conté el nom dels metabòlits que s'han analitzat.

```

# Els noms dels metabòlits comencen a partir de la tercera columna de la taula
# cachexia.
metabolit_names <- colnames(cachexia)[3:ncol(cachexia)]

# Creem el DataFrame rowData amb els noms dels metabòlits com a files
rowData <- DataFrame(Metabolite = metabolit_names)
# Guardem el nom del metabòlit com a nom de la fila.
rownames(rowData) <- rowData$Metabolite
# Eliminem la columna "Metabolite" perquè ja tenim la informació a les files.
rowData$Metabolite <- NULL
rowData

```

```
## DataFrame with 63 rows and 0 columns
```

Com podem observar a “l'imprimir” el resultat de rowData, la resposta és DataFrame with 63 rows and 0 columns. Aquest element està pensat per incloure informació addicional o característiques d'interès a la fila corresponent. En el nostre cas només tenim el nom del metabòlits (guardats com a nom de les files), però no tenim informació addicional de cap d'ells, per això “0 columns”.

A l'estar treballant amb dades metabolòmiques a diferència de quan es treballa amb dades de seqüenciació no disposem de GRangesList ni GRanges.

Creació del DataFrame colData: Aquest element emmagatzema qualsevol columna amb dades descriptives per a cada mostra. En aquest cas de les dades que proporcionava l'enunciat, hem utilitzat el “Patient.ID” com a nom de cada una de les files i la columna “Muscle.loss” definida com a columna del colData, la qual suggereix l'existència dos grups “cachexic/control” en les dades.

```

# Creem un DataFrame a través de les columnes "Patient ID" i "Muscle loss".
colData <- DataFrame(
  Patient.ID = cachexia$Patient.ID,
  Muscle.loss = cachexia$Muscle.loss)

# Guardem el ID com a nom de la fila.
rownames(colData) <- colData$Patient.ID
# Eliminem la columna "Patient.ID" perquè ja tenim la informació a les files.
colData$Patient.ID <- NULL
head(colData)

```

```

## DataFrame with 6 rows and 1 column
##           Muscle.loss
##           <character>
## PIF_178      cachexic
## PIF_087      cachexic
## PIF_090      cachexic
## NETL_005_V1  cachexic
## PIF_115      cachexic
## PIF_110      cachexic

```

Creació de l'element metadata: es caracteritza per ser un llistat senzill d'informació de l'estudi o qualsevol metadata de l'experiment que es vulgui desar, com per exemple fórmules... En aquest cas, s'ha afegit la informació que proporcionava el document “description.md” del repositori de dades.

```

# Creem un DataFrame per a metadata
metadata <- DataFrame(
  Description = c("This is the famous cachexia dataset used in several MetaboAnalyst tutorials",

```

```

        "Available from: https://rest.xialab.ca/api/download/metaboanalyst/human_cachexia.csv",
        "Successfully passed sanity check!",
        "Samples are not paired.",
        "2 groups were detected in samples.",
        "All data values are numeric.",
        "A total of 0 (0%) missing values were detected.))

# Comprovació el DataFrame creat
print(metadata)

```

```

## DataFrame with 7 rows and 1 column
##           Description
##           <character>
## 1 This is the famous c..
## 2 Available from: http..
## 3 Successfully passed ..
## 4 Samples are not pair..
## 5 2 groups were detect..
## 6 All data values are ..
## 7 A total of 0 (0%) mi..

```

Un cop definits tots els elements del SummarizedExperiment ja es pot crear el contenidor.

```

SummarizedExperiment_cachexia <- SummarizedExperiment(
  assays = list(counts = assay),      # Dades d'assay (metabòlits)
  colData = colData,                 # Metadades de les columnes (pacients i grup)
  rowData = rowData,                 # Nom de les files (metabòlits)
  metadata = list(metadata = metadata)) # Informació general

print(SummarizedExperiment_cachexia)

```

```

## class: SummarizedExperiment
## dim: 63 77
## metadata(1): metadata
## assays(1): counts
## rownames(63): X1.6.Anhydro.beta.D.glucose X1.Methylnicotinamide ...
##      pi.Methylhistidine tau.Methylhistidine
## rowData names(0):
## colnames(77): PIF_178 PIF_087 ... NETL_003_V1 NETL_003_V2
## colData names(1): Muscle.loss

```

```

# Per guardar el contenidor en format binari utilitzem la funció save
save(SummarizedExperiment_cachexia, file = "SummarizedExperiment_cachexia.RDA")

```

Per aquest estudi, treballarem amb la classe *SummarizedExperiment* i no amb la classe *RangedSummarizedExperiment*, ja que aquesta última està pensada per analitzar intervals d'interès genòmic a través de *rowRanges()*, el qual no és el nostre cas d'estudi.

Un cop definit el contenidor, ja està llest per navegar dins les dades i cercar casos més específics. A continuació es mostra de forma breu algunes de les accions que es poden realitzar. Com per exemple a través de *colData->Muscle.loss*, seleccionar aquells pacients que formen part del grup “cachexic”.

```
SummarizedExperiment_cachexia[, SummarizedExperiment_cachexia$Muscle.loss == "cachexic"]
```

```
## class: SummarizedExperiment
## dim: 63 47
## metadata(1): metadata
## assays(1): counts
## rownames(63): X1.6.Anhydro.beta.D.glucose X1.Methylnicotinamide ...
##   pi.Methylhistidine tau.Methylhistidine
## rowData names(0):
## colnames(47): PIF_178 PIF_087 ... NETCR_016_V1 PIF_116
## colData names(1): Muscle.loss
```

Seleccionar una sola fila (metabòlit). Es mostrarà quin és el valor registrat de la seva anàlisi per cada pacient.

```
# Obten el primer assaig i extreu la fila 7
assay(SummarizedExperiment_cachexia, 1)[7, ]
```

Ara també podem crear subconjunts que ens ajudin a analitzar les dades:

```
# Subconjunt només amb els pacients amb cachexia.
se_cachexia <- SummarizedExperiment_cachexia[, SummarizedExperiment_cachexia$Muscle.loss == "cachexic"]
cat("Subconjunt de pacients amb caquèxia:\n")
se_cachexia

# Subconjunt només amb els pacients control.
se_control <- SummarizedExperiment_cachexia[, SummarizedExperiment_cachexia$Muscle.loss == "control"]
cat("\nSubconjunt de pacients control:\n")
se_control
```

Exploració multivariant de dades

Per explorar les dades de cachexia, calcularem una matriu de distàncies que ens permetrà quantificar la semblança per parelles o grups.

```
# Cal transposar la matriu assay perquè es necessiten les mostres en files
distance_matrix <- dist(t(assay), method = "euclidean")
```

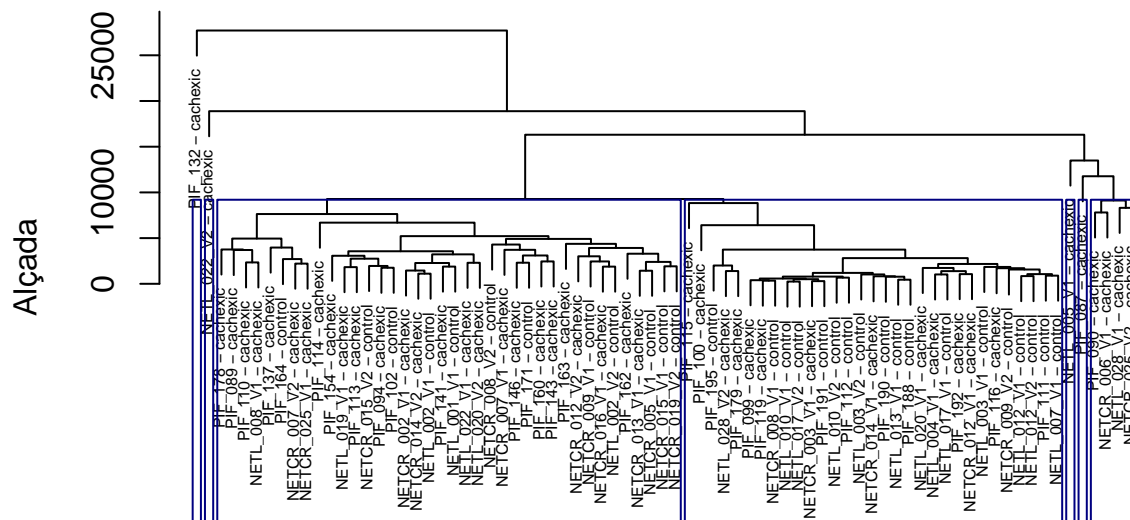
Cluster jeràrquic

Un cop tenim la matriu de distància, podem utilitzar-la per a tècniques de clustering, com per exemple un gràfic de Cluster jeràrquic, la qual ens ajudarà a identificar mostres que són semblants entre elles en funció de la seva similitud.

```
clustering <- hclust(distance_matrix, method = "average")
# Per tal de veure el nom i el grup al qual pertanyen les mostres, creem
# etiquetes costumitzats que continguin el ID i el grup.
custom_labels <- paste(rownames(colData(SummarizedExperiment_cachexia)),
                        colData(SummarizedExperiment_cachexia)$Muscle.loss,
                        sep = " - ")
```

```
# Per explicar gràfic amb més facilitat tallem el dendrograma en 7 grups.
groups <- cutree(clustering, k = 7)
# Visualitzem el dendrograma
plot(clustering, main = "Dendrogram ", xlab = "Patients.ID", ylab = "Alçada",
     cex = 0.5, labels = custom_labels)
rect.hclust(clustering, k = 7, border = "navy")
```

Dendrogram



Patients.ID
hclust (*, "average")

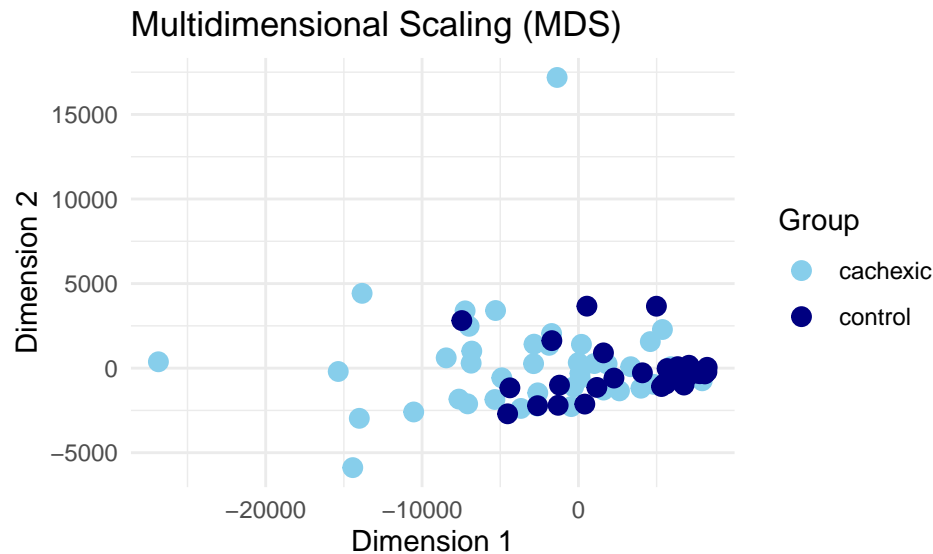
Com podem observar, en el gràfic anterior, es poden veure dos grups a l'extrem l'esquerra i tres a l'extrem de la dreta amb mostres bastant diferenciades, totes elles del grup cachexic. Si ens centrem en la part central del dendrograma, es diferencien dues grans branques. L'extrem esquerre conté la majoria de les mostres del grup cachexic, mentre que el dret agrupa la majoria del grup control. Aquesta observació ens fa pensar existeixen característiques de les mostres cachexics que són distintives.

Multidimensional Scaling (MDS)

L'objectiu de l'MDS és representar gràficament les distàncies/similituds entre les mostres en un espai de dues dimensions a través de la matriu de distàncies. Al gràfic cada punt representa una mostra, i la distància entre els punts reflecteix la semblança entre les mostres.

```
# Calcular el MDS amb dues dimensions
mds <- cmdscale(distance_matrix, k = 2)
# Crear un dataframe
groups <- colData(SummarizedExperiment_cachexia)$Muscle.loss
mds_df <- data.frame(Dim1 = mds[, 1], Dim2 = mds[, 2],
                    Group = as.factor(groups),
                    Label = rownames(colData(SummarizedExperiment_cachexia)))
```

```
ggplot(mds_df, aes(x = Dim1, y = Dim2, color = Group, label = Label)) +
  geom_point(size = 3) +
  labs(title = "Multidimensional Scaling (MDS)", x = "Dimension 1", y = "Dimension 2") +
  theme_minimal() +
  scale_color_manual(values = c("cachexic" = "skyblue", "control" = "navy"))
```



Com podem observar els punts blaus referents al grup control es troben més centralitzats a l'extrem inferior dret, en canvi, les mostres del grup cachexic presenten major dispersió a l'espai. Si ho analitzem des del punt de vista biològic, això té sentit, ja que, demostra que les persones que pateixen cachexia presenten alguna mena d'alteració (a través dels metabòlits) que es podria utilitzar com a predictor de cachexia.

Anàlisi de Components Principals (PCA)

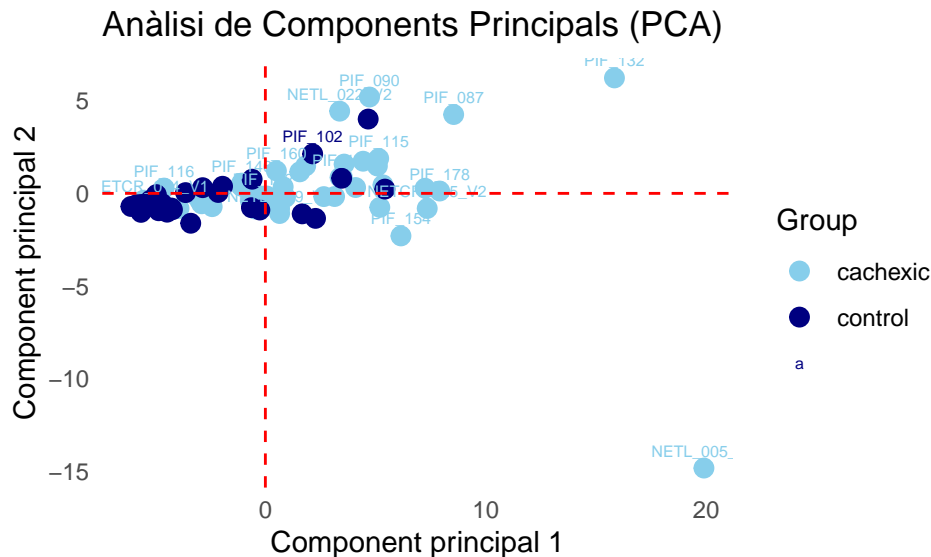
L'objectiu de l'anàlisi de component principal, és estudiar com varien dues o més variables de forma conjunta i posteriorment identificar les relacions lineals entre elles. Aquest estudi es basa en la matriu de covariància, la qual està formada per les covariàncies entre cada parell de variables del conjunt de dades.

```
# Transposem la matriu per tal que les mostres siguin files
pca_results <- prcomp(t(assay(SummarizedExperiment_cachexia)), scale = TRUE)
# Creem un data frame amb els resultats de la PCA
pca_data <- as.data.frame(pca_results$x[, 1:2])
colnames(pca_data) <- c("PC1", "PC2")
# Afegim la variable del grup
pca_data$Group <- colData(SummarizedExperiment_cachexia)$Muscle.loss
# Afegim els noms de les mostres des de colData
pca_data$SampleID <- rownames(colData(SummarizedExperiment_cachexia))
# Calculem les mitjanes per a PC1 i PC2
mean_pc1 <- mean(pca_data$PC1)
mean_pc2 <- mean(pca_data$PC2)

# Creem el gràfic amb línies de mitjana
ggplot(pca_data, aes(x = PC1, y = PC2, color = Group)) +
  geom_point(size = 3) +
  geom_text(aes(label = SampleID), vjust = -1, hjust = 0.5, size = 2,
    check_overlap = TRUE) + # Afegir noms de les mostres
  labs(title = "Anàlisi de Components Principals (PCA)",
```



```
x = "Component principal 1", y = "Component principal 2") +
theme_minimal() +
scale_color_manual(values = c("cachexic" = "skyblue", "control" = "navy")) +
geom_hline(yintercept = mean_pc2, linetype = "dashed", color = "red") +
geom_vline(xintercept = mean_pc1, linetype = "dashed", color = "red") +
theme(panel.grid.major = element_blank(),
      panel.grid.minor = element_blank())
```



A primer cop d'ull en aquest gràfic també podem veure que les mostres del grup control tendeixen a concentrar-se a l'espai superior esquerre, la qual cosa ens indica que les metabòlits dels individus control tendeixen a presentar un patró més homogeni, ja que segurament tenen un perfil metabòlic similar, amb poca variabilitat entre ells. Des del punt de vista biològic és habitual trobar aquest patró en els grups sans, on els metabòlits tendeixen a seguir patrons més previsibles.

En canvi, les mostres del grup cachexic, igual que en el gràfic anterior presenten una major dispersió, existeix una variabilitat més gran en el perfil metabòlic dels individus amb cachexia. Els perfils metabolòmics no segueixen els patrons biològics previsibles.

Com ja hem comentat anteriorment els components principals estan compostos per un conjunt de metabòlits, a continuació intentarem esbrinar quins són aquests metabòlits més influents en cada un dels components. Això ho farem a través de l'estudi de càrregues.

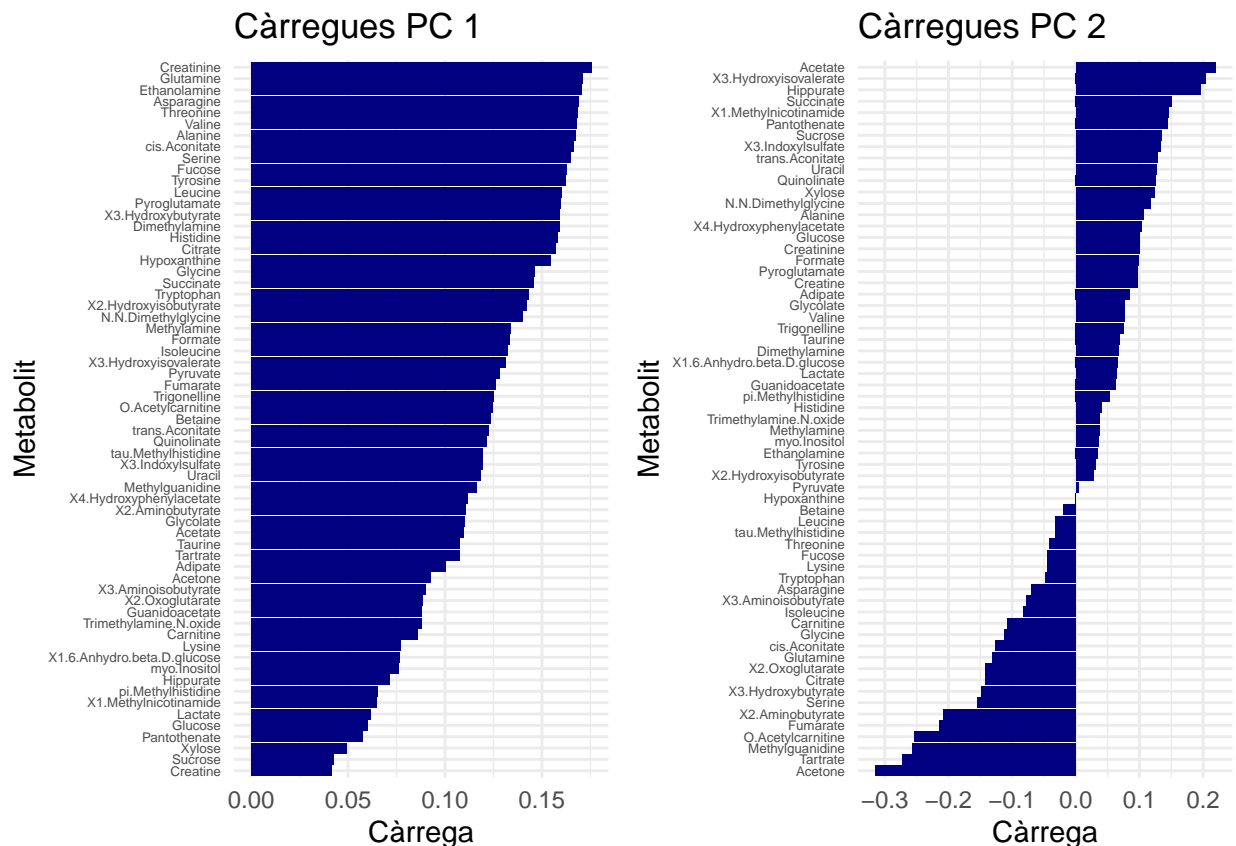
```
# Obtenim les càrregues per a PC1 i PC2
loadings_pc1 <- pca_results$rotation[, 1]
loadings_pc2 <- pca_results$rotation[, 2]
# Creem un dataframe per a les càrregues de cada component
loadings_pc1_df <- data.frame(Metabolit = rownames(pca_results$rotation), Càrrega = loadings_pc1)
loadings_pc2_df <- data.frame(Metabolit = rownames(pca_results$rotation), Càrrega = loadings_pc2)
# Classifiquem les càrregues de PC1 i PC2 de major a menor influència
loadings_pc1_df <- loadings_pc1_df[order(abs(loadings_pc1_df$Càrrega), decreasing = TRUE), ]
loadings_pc2_df <- loadings_pc2_df[order(abs(loadings_pc2_df$Càrrega), decreasing = TRUE), ]

# Visualitzem les càrregues de PC1
PCA_1 <- ggplot(loadings_pc1_df, aes(x = reorder(Metabolit, Càrrega), y = Càrrega)) +
  geom_bar(stat = "identity", fill = "navy") +
  coord_flip() +
  labs(title = "Càrregues PC 1",
```

```

x = "Metabolit", y = "Càrrega") +
theme_minimal() +
theme(axis.text.y = element_text(size = 5)) # Ajusta el valor de size segons sigui necessari
# Visualitzem les càrregues de PC2
PCA_2<-ggplot(loadings_pc2_df, aes(x = reorder(Metabolit, Càrrega), y = Càrrega)) +
geom_bar(stat = "identity", fill = "navy") +
coord_flip() +
labs(title = "Càrregues PC 2",
x = "Metabolit", y = "Càrrega") +
theme_minimal() +
theme(axis.text.y = element_text(size = 5)) # Ajusta el valor de size segons sigui necessari
grid.arrange(PCA_1, PCA_2, ncol = 2)

```



Primer component principal (PC1): és qui té major pes i explica la màxima variància de les dades. Els metabòlits amb càrregues més altes i que per tant major contribueixen a les diferències entre els grups (control i cachexia) són la Creatinine, la Glutamine, el Ethanolamine i l'Asparagine. *Normalment, s'agafa com a llindar un valor absolut de 0.5-0.7, però a la nostra gràfica veiem una gran quantitat de metabòlits que superen aquest número, per facilitar l'estudi, només agafarem els quatre primers.*

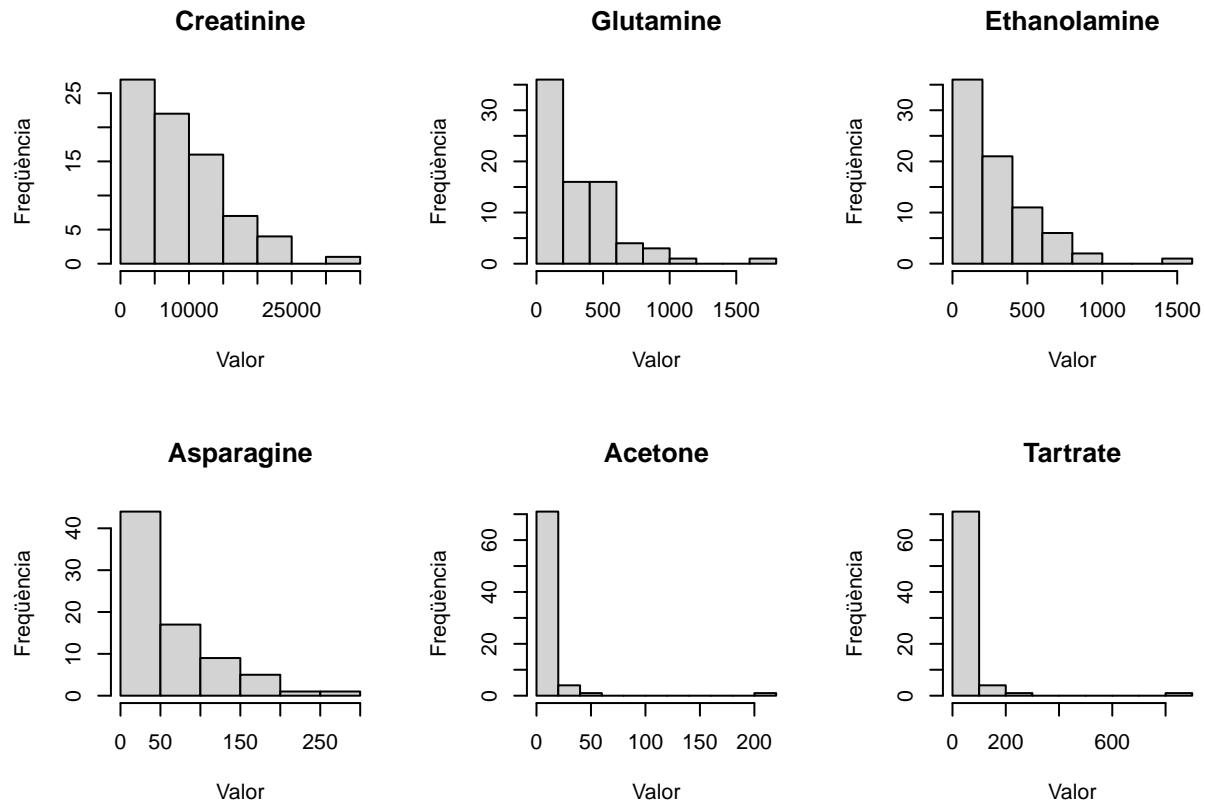
PC2 (Component Principal 2): explica la segona màxima variància, pot revelar diferències no captades per PC1. En aquest cas el valor de les càrregues és molt inferior, a més aquelles amb el número més gran tenen càrrega negativa i es tracta dels metabòlits, Acetone i Tartrate.

Gràfiques de freqüències

Per tal de simplificar l'estudi a partir d'ara ens centrarem a estudiar amb aquests sis metabòlits.

```
# Definim els metabòlits d'interès
selected_metabolites <- c("Creatinine", "Glutamine", "Ethanolamine",
                          "Asparagine", "Acetone", "Tartrate")

# Configurem la figura per mostrar múltiples gràfics
figura_2 <- par(mfrow=c(2, 3))
# Bucle per crear un histograma per a cada metabolit
for (metabolit in selected_metabolites) {
  hist(assay[metabolit, ], main = metabolit, xlab = "Valor", ylab = "Freqüència")
}
```



```
# Restaurem la configuració gràfica original
par(figura_2)
```

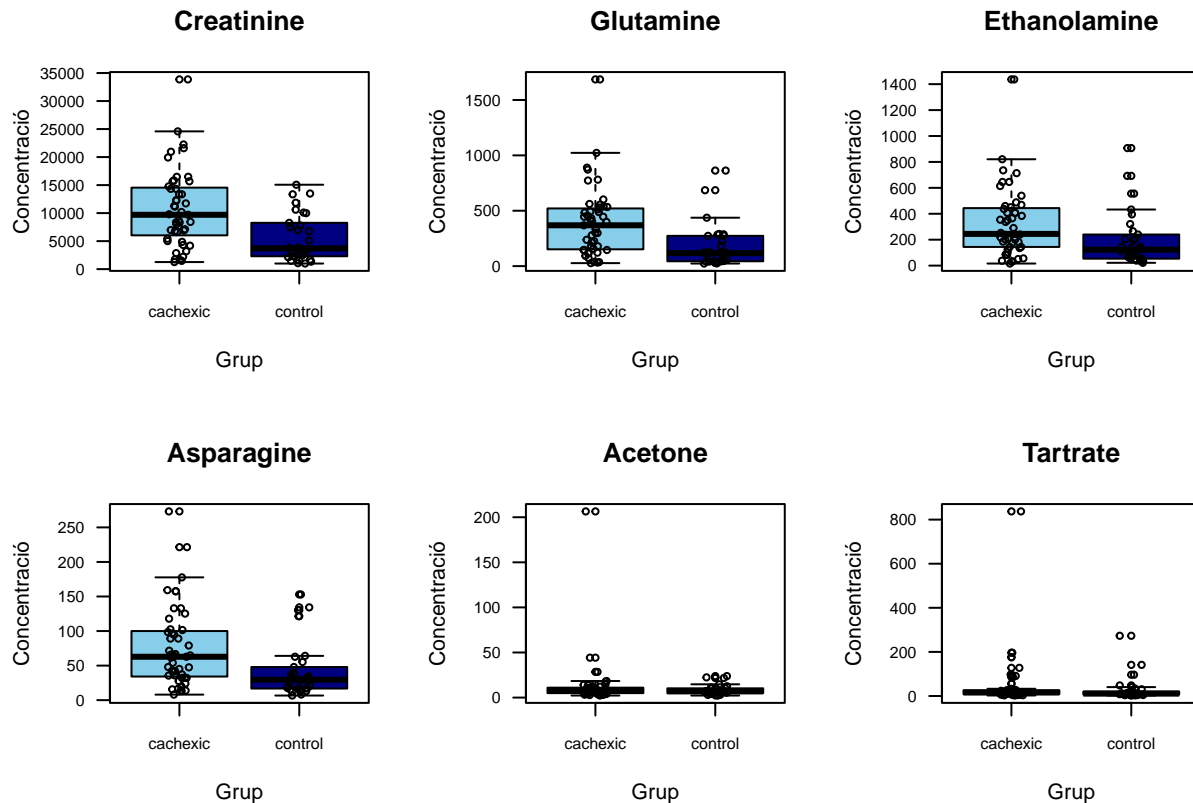
Si ens fixem en aquestes gràfiques de freqüència, podem observar que gairebé en tots els casos les mostres es concentren en unes determinades categories i després sembla que hi ha mostres molt allunyades. Aquestes outliers desconexem si són errors de mesures. Donat que nosaltres pretenem tenir una idea general sobre l'estudi per ara no eliminarem els seus registres, però si es fes un estudi més avançat caldria analitzar-los per valorar la seva eliminació.

Gràfica de concentracions per grups A continuació a través de les gràfiques boxplot, representarem la mitjana per cada grup d'estudi.

```
# Configurem la figura per mostrar múltiples gràfics
par(mfrow = c(2, 3))
# Definim colors per a cada grup
```

```
group_colors <- ifelse(colData(SummarizedExperiment_cachexia)$Muscle.loss ==
                      "cachexic", "skyblue", "navy")
# Creem un boxplot per cada metabòlit
for (metabolit in selected_metabolites) {
  # Boxplot per al metabòlit
  boxplot(assay(SummarizedExperiment_cachexia)[metabolit, ] ~
          colData(SummarizedExperiment_cachexia)$Muscle.loss,
          col = c("skyblue", "navy"),
          main = metabolit, xlab = "Grup", ylab = "Concentració",
          las = 1, cex.axis = 0.7, cex.main = 1.2)
  # Afegim punts per cada mostra
  stripchart(assay(SummarizedExperiment_cachexia)
             [metabolit, ] ~ colData(SummarizedExperiment_cachexia)$Muscle.loss,
             vertical = TRUE, method = "jitter", pch = 1, col = "black",
             add = TRUE)}

```



```
par(mfrow = c(1, 1))
```

A primer cop d'un, és evident que la presència dels outliers fan que els gràfics no quedin del tot polits, tot i això, la diferència entre les mitjanes de les concentracions és notòria.

Si ens centrem en els metabòlits que formaven part del primer component (Creatinine, Glutamine, Ethanolamine i Asparagine), podem observar que en tots els casos la concentració del metabolit estudiat és substancialment major en el grup cachexic. Pel que fa als metabòlits del component 2 (Acetone i Tartrate) la diferència no és tan significativa.

Discussió, limitacions i conclusions de l'estudi

Després de l'exploració de les dades "2024-Cachexia", podem dir que existeixin diferències entre els perfils metabolòmics dels pacients amb cachexia i els pacients del grup control.

La cachexia és una síndrome que es caracteritza per la pèrdua significativa de massa muscular, la qual cosa implica un debilitament general del cos. Sovint la pateixen pacients crònics de càncer, insuficiència cardíaca o malalties pulmonars obstructives cròniques. Cal remarcar que aquesta pèrdua de massa no es deu a una disminució de la ingesta d'aliments, sinó que implica alteracions metabòliques que resulten en un estat inflamatori, provocant que el cos descompongui músculs i greix.

Si ens centrem en el resultat de les càrregues de l'anàlisi de cocomponents principals, podem veure que els memetabolites més influeixen en la diferenciació de grup són la creatinine, la glutamine, l'ethanolamine, l'asparagine, l'acetone i el tartrate. Per entendre els resultats, primer cal entendre la implicació/relació d'aquests metabolites en els teixits de massa muscular.

- Creatinine: és un subproducte del cicle de la creatina, la qual s'utilitza com a indicador de massa muscular, ja que s'utilitza com a font d'energia de curta durada per accions d'alta intensitat.
- Glutamine: és un aminoàcid amb un paper molt important per al manteniment de la massa muscular, essencial per a la recuperació i la construcció de teixit muscular.
- Ethanolamine: és un metabolit altament implicat en la formació de fosfolípids, essencials per a la formació i la integritat de les membranes cel·lulars.
- Asparagine: és un aminoàcid, essencial per a la formació de proteïnes, el manteniment i construcció del teixit muscular.
- Acetone: compost que es genera durant la degradació de greixos.
- Tartrate: metabolit relacionat amb el metabolisme energètic i la recuperació muscular.

En tots els casos com hem pogut veure a l'apartat de resultats, que la concentració és major en el grup de pacients cachectics en comparació al grup control.

Les elevacions en aquests metabolites, i també el fet que estiguin elevades de forma conjunta, ens dona una idea de que en els pacients cachectics existeixen una sèrie de respostes fisiològiques al deteriorament funcional i l'estrès metabòlic que acompanya aquesta síndrome.

Totes les alteracions reflecteixen un desajust en el cicle de creació i destrucció del teixit muscular i adipós, això es transforma en una resposta compensatòria com una adaptació a la malaltia. Tot això ens demostra la batalla constant del pacient per mantenir la massa muscular i gestionar les necessitats energètiques del seu cos.

Cal tenir en compte que l'exploració multivariant de les dades basada només en PCA i clustering jeràrquic, conté algunes limitacions; La pèrdua d'informació, degut a la reducció de la dimensionalitat en la que es basa el PCA, elimina part de la variabilitat. L'assumpció lineal, l'anàlisi PCA assumeix relacions entre variables, en cas de no ser-ho pot ser una limitació. Resultat sensible, tal com hem comentat tant el PCA com el clustering jeràrquic, hem vist outliers, que poden distorsionar la interpretació i formació dels clústers. Per últim, la subjectivitat en la selecció de clústers, no existeix una única manera de determinar el nombre òptim de clústers.

Així doncs, caldria realitzar un anàlisi multivariant amb més profunditat donat que la cachexia, implica un alteració global del perfil metabolòmic. Estudiar l'impacte de més metabolites i la seva implicació a les vies de deteriorament de teixit muscular i adipós per acabar de definir quins són els biomarcadors i els seus valors que indiquen cachexia.

Directori github <https://github.com/mesquerrap/ESQUERRA-Pares-Marina-PAC1>