

---

# PEC 1 - ANÁLISIS DE DATOS ÓMICOS (M0-157)

## CONTENIDO

Abstract .....	1
Objetivos.....	1
Métodos .....	2
Resultados .....	2
Discusión .....	5
Conclusión.....	5
Referencias .....	6

## ABSTRACT

El análisis ómico generalmente nos proporciona muchos datos de un número no muy alto de muestras. Aún así, el análisis es complejo. Podemos utilizar herramientas informáticas tanto para facilitar la visualización de estos datos crudos para encontrar diferencias significativas entre ellos, como para hacer análisis complejos que nos permitan estudiar varias variables de manera simultánea. Para todo ello, una tarea esencial es preparar la base de datos con la que vamos a trabajar y la herramienta que propone utilizar esta actividad es la creación de objeto a través de SumarizedExperiment. Una vez conseguido esto, se estudiará la visualización de estos datos con gráficos como el de heatmap y se realizará un Análisis de Componentes Principales (PCA).

También es importante utilizar las herramientas disponibles hoy en día para el control de versiones y trabajos en red. La web github contiene bases de datos disponibles para su uso. También es una herramienta útil ya que proporciona la posibilidad de tener una copia del proyecto principal guardada mientras se trabaja en la mejora de algunos de los puntos de tal proyecto. Estas mejoras se pueden hacer de manera tanto colectiva como individual, ya que se añadirán a la versión final de una manera controlada.

## OBJETIVOS

El objetivo de esta actividad es familiarizarse con el uso de la programación orientada a objetos (POB) en este caso a través de SumarizedExperiment.

Además, se trabajará en las distintas formas de visualización de estos datos con el objetivo de encontrar diferencias entre los grupos que nos ofrezcan una explicación biológica plausible.

Por último, todo el proyecto se realizará en un repositorio de github para que comprendamos e implantemos su uso en nuestros proyectos.

## MÉTODOS

### **Datos:**

He seleccionado el dataset human\_cachexia del repositorio github (<https://github.com/nutrimetabolomics/metaboData/tree/main/Datasets/2024-Cachexia>). Se trata de una base de datos de 77 observaciones de 65 variables distintas que hacen referencia a 77 muestras de orina de 47 pacientes con caquexia y 30 pacientes control.

### **Repositorio github:**

<https://github.com/luciafdez/FERNANDEZ-DELGADO-LUCIA-PEC1>

Hemos creado (con mkdir) tanto el repositorio principal como las subcarpetas:

- data: los datos de human\_cachexia especificados más arriba
- R: el código de R utilizado
- objetos: los objetos que crearemos

Para añadir los cambios al repositorio iremos utilizando los comandos git add, git commit -m (para especificar cada cambio) y git push.

## RESULTADOS

La diferencia entre SummarizedExperiment y ExpressionSet es que SummarizedExperiment es más nuevo y flexible en la información de las filas. ExpressionSet es del paquete Biobase y SummarizedExperiment del paquete con el mismo nombre.

Usamos unos códigos básicos para visualizar el dataset: head, names, length...

Creamos el objeto que se solicita en el ejercicio:

```
#Pasamos los datos a una matriz desde la tercera columna
#Las dos primeras columnas son identificadores
counts <- as.matrix(human_cachexia[,3:ncol(human_cachexia)])
colData <- DataFrame(
  PatientID = human_cachexia$`Patient ID`,
  MuscleLoss = human_cachexia$`Muscle loss`
)

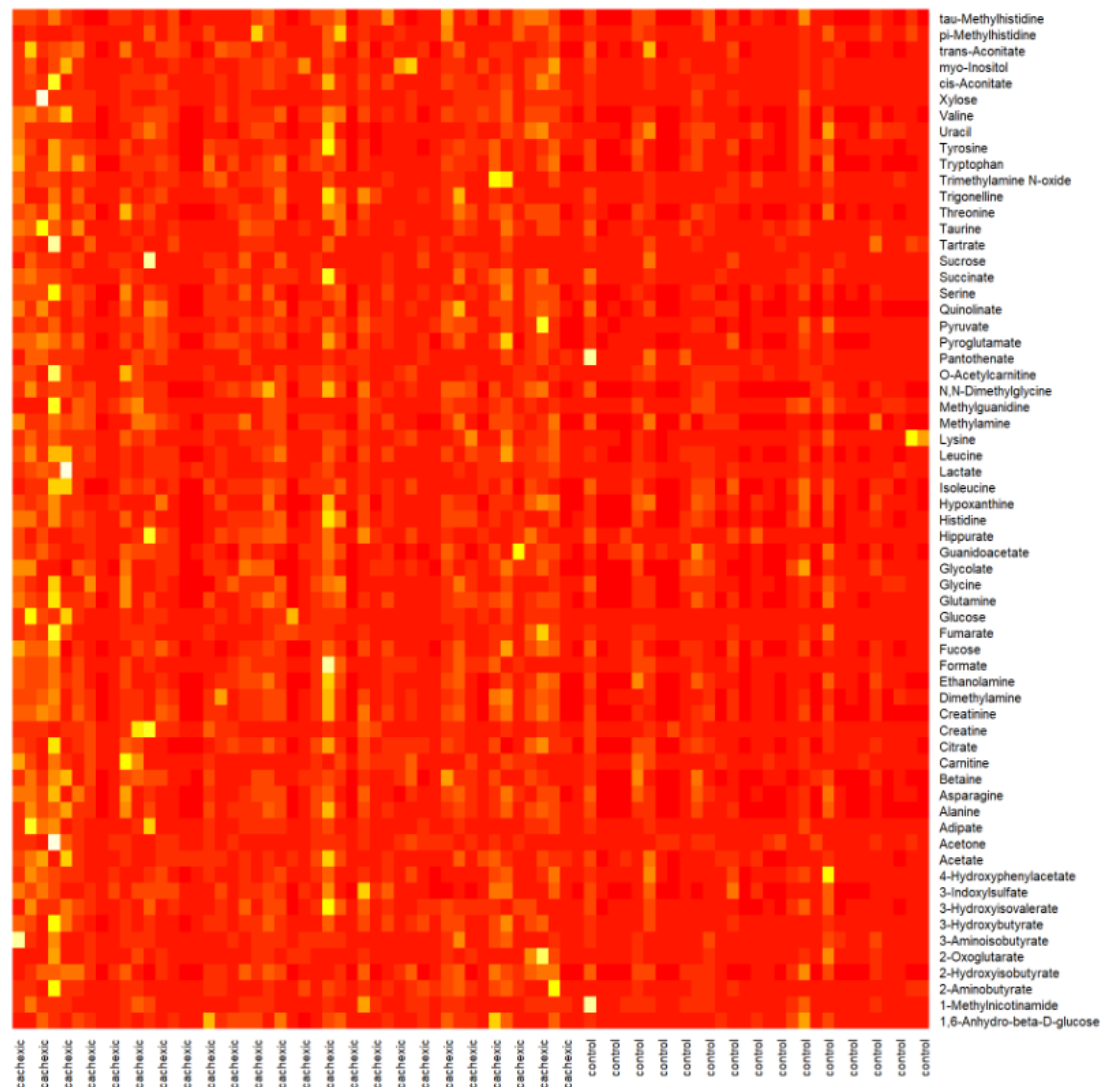
#Trasponemos la matriz ya que las features tienen que estar en filas y
las samples en columns
counts_t <- t(counts)

#Las features serán los distintos metabolitos
rownames(counts_t) <- colnames(counts)
rowData <- DataFrame(metabolite = rownames(counts_t))
```

```
#Creamos el SummarizedExperiment
se <- SummarizedExperiment(assays=list(counts=counts_t),
                           rowData = rowData, colData=colData)

#Guardamos el resultado
save(se, file = "objetos/summarized_experiment.Rda")
```

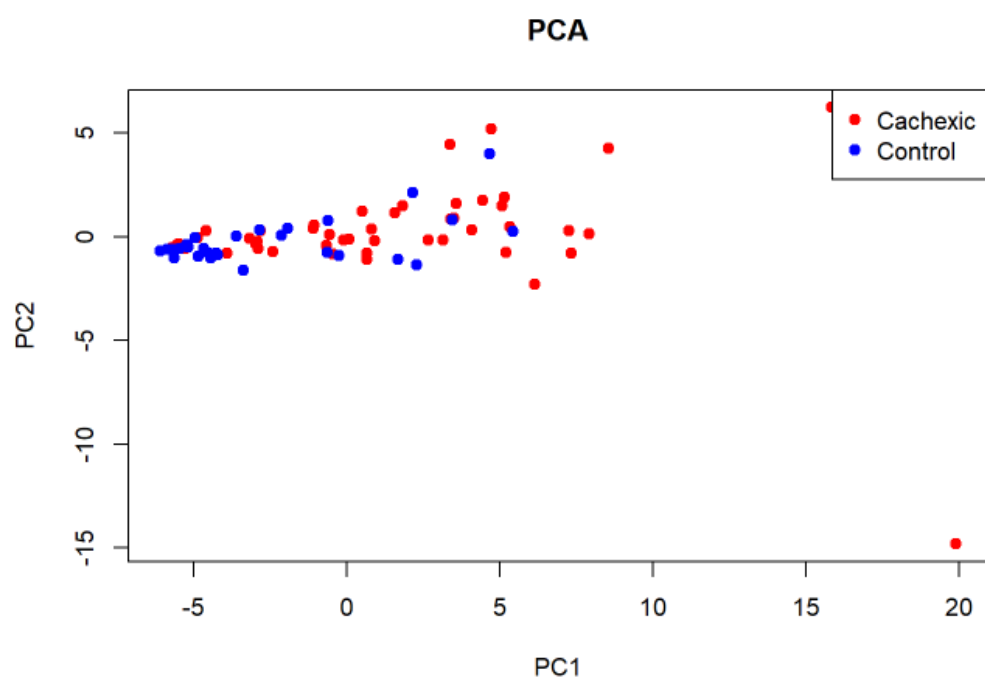
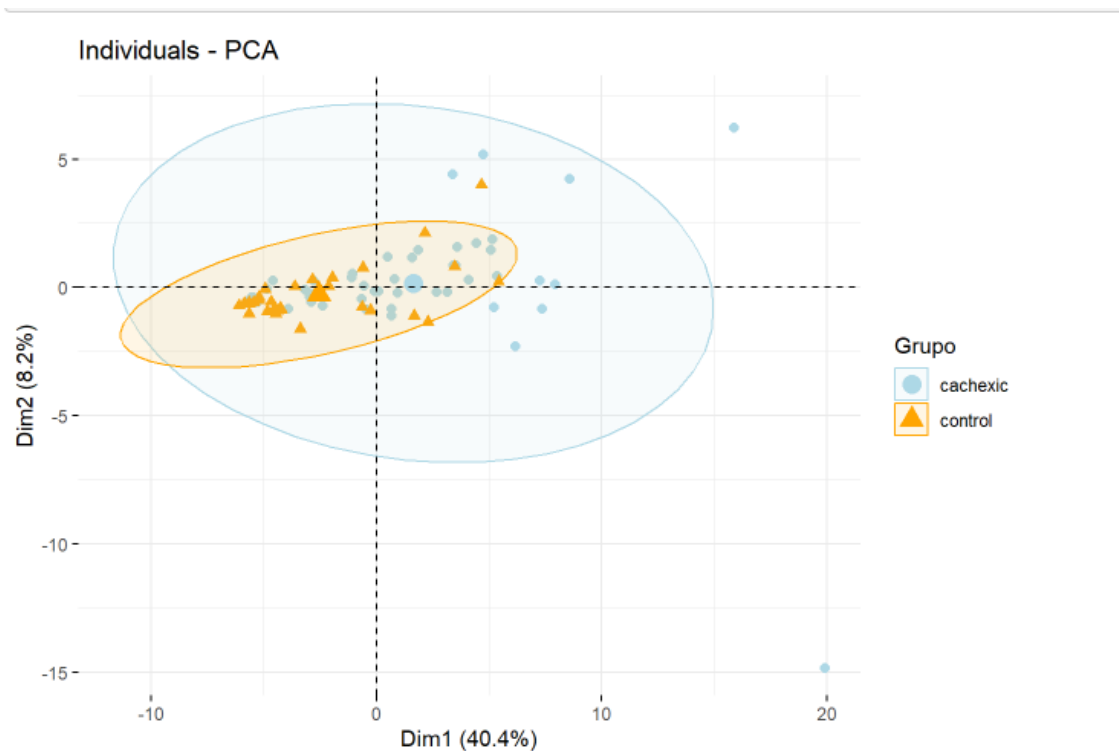
Para la visualización de los datos comenzamos con un heatmap:



Los controles se encuentran a la derecha y los pacientes a la izquierda, pero no parece haber demasiada diferencia entre ellos.

Intentamos realizar un PCA:

PEC 1 - Análisis de datos ómicos (M0-157) - Lucía Fernández Delgado



Vemos aquí también que ambos grupos solapan casi en su totalidad.

No parece haber una excreción distinta en orina en los pacientes seleccionados comparados con los controles. Es cierto que seleccionan pacientes con caquexia por diferentes motivos, comentan que son pacientes con alguna de enfermedad de base como el cáncer, pero no es la misma enfermedad en todos ellos. Puede ser que estas diferencias entre las muestras clasificadas como "pacientes" haga que las diferencias entre ellas sean tan notorias que no se puedan extraer conclusiones al compararlas con un control. Cada enfermedad tendrá unas características e implicará la toma de distintos fármacos, muchos de ellos posiblemente con excreción renal, por lo

que puede haber demasiada variabilidad intra-grupo. Otra posibilidad claro está es que la hipótesis inicial sea incorrecta.

## DISCUSIÓN

El uso de github para el control de versiones permite implementar pequeñas partes de nuestro código. Además, permitiría trabajar en grupo con más gente. Aunque este no era el objetivo principal del proyecto, los conocimientos adquiridos en la actividad permitirán poder utilizarlo en un futuro.

Por otro lado, podemos apreciar las ventajas que ofrece el uso de la POB adaptada los datos ómicos. Una de las tareas más complicadas a la hora de analizar datos es preparar el dataframe: escoger las variables adecuadas, unir los datos numéricos con el metadata, etc. El uso de SummarizedExperiment permite que esta tarea sea sencilla ya que está preparada para este tipo de datos: se espera un número alto de variables para cada muestra y cierto metadata asociado a la muestra. Esto nos facilita por un lado los cálculos sobre nuestras variables, ya que se maneja por separado los datos numéricos de la metadata y también la visualización, ya que es bastante sencillo poder dividir los grupos en base a las características del metadata.

## CONCLUSIÓN

Aunque el análisis de datos ómicos siempre incluye el reto de trabajar con una gran cantidad de datos, los datos obtenidos son similares: no en cuanto al contenido sino en la forma de expresarse. Esto, sumando a los grandes avances de los últimos años, ha favorecido la creación de herramientas diseñadas para este tipo de datos que facilitan su análisis.

## REFERENCIAS

<https://bioconductor.org/packages/release/bioc/vignettes/SummarizedExperiment/inst/doc/SummarizedExperiment.html#constructing-a-summarizedexperiment>

[https://rpubs.com/cristina\\_gil/pca](https://rpubs.com/cristina_gil/pca)

<https://rpubs.com/b25Dev/creacion-mapa-calor-en-r-bryan-cortez>

<http://bioconductor.jp/packages/3.16/bioc/vignettes/POMA/inst/doc/POMA-demo.html>