

PEC 1 - Análisis de datos ómicos

Nombre: Sergio Moya Millán

Fecha: 27/3/2025

Enlace a Github: <https://github.com/docktoo/Moya-Millan-Sergio-PEC1.git>

Índice

1. Introducción.....	3
2. Objetivo	3
3. Materiales y métodos	3
3.1. Creación del objeto SummarizedExperiment	3
3.2. Análisis exploratorio	5
3.2.1. Pruebas t	5
3.2.2. Análisis de correlaciones	5
3.2.3. Dendrograma y k-means	6
3.2.4. Análisis de Componentes Principales	7
4. Interpretación biológica	8

1. Introducción

El presente análisis se lleva a cabo utilizando un conjunto de datos ampliamente utilizado en estudios de metabolómica relacionado con la caquexia. Esta patología es una condición médica compleja que se caracteriza por la pérdida progresiva de masa muscular y grasa corporal, generalmente asociada con enfermedades crónicas graves como el cáncer, la insuficiencia renal, y enfermedades cardíacas. Esta condición no solo afecta la calidad de vida del paciente, sino que también puede influir negativamente en la respuesta al tratamiento y en el pronóstico de la enfermedad subyacente. El análisis de metabolitos en estos pacientes ha ganado relevancia, ya que permite entender los mecanismos metabólicos involucrados en la caquexia y cómo estos podrían ser modulados para mejorar el manejo clínico de los pacientes.

2. Objetivo

El objetivo principal de este estudio es realizar un análisis comparativo entre las concentraciones de metabolitos en pacientes con caquexia y un grupo control. El análisis se realizará mediante el uso de herramientas estadísticas en R studio y a partir de la creación del objeto SummarizedExperiment. Se realizarán análisis como por ejemplo, las pruebas t, para identificar metabolitos que estén significativamente asociados con la enfermedad. A través de esta batería de análisis estadísticos, se busca proporcionar una visión más clara de las alteraciones metabólicas presentes en los pacientes enfermos y cómo estas pueden diferir de los individuos saludables.

3. Materiales y métodos

3.1. Creación del objeto SummarizedExperiment

Antes de la creación del objeto se hace un análisis de los datos de metabolitos de un conjunto de pacientes, con el objetivo de explorar y procesar

la información para realizar comparaciones entre grupos de pacientes que sufren pérdida muscular. En primer lugar, se cargan y se hace una exploración inicial para ver la estructura y la distribución de la base de datos cargada. Después, se observa que los datos necesitan ser transformados, para ello se utiliza la transformación logarítmica y se escalan para que tengan una media de 0 y una desviación estándar de 1. En la figura 1 se representa mediante boxplot el cambio en la distribución de los datos después de la transformación.

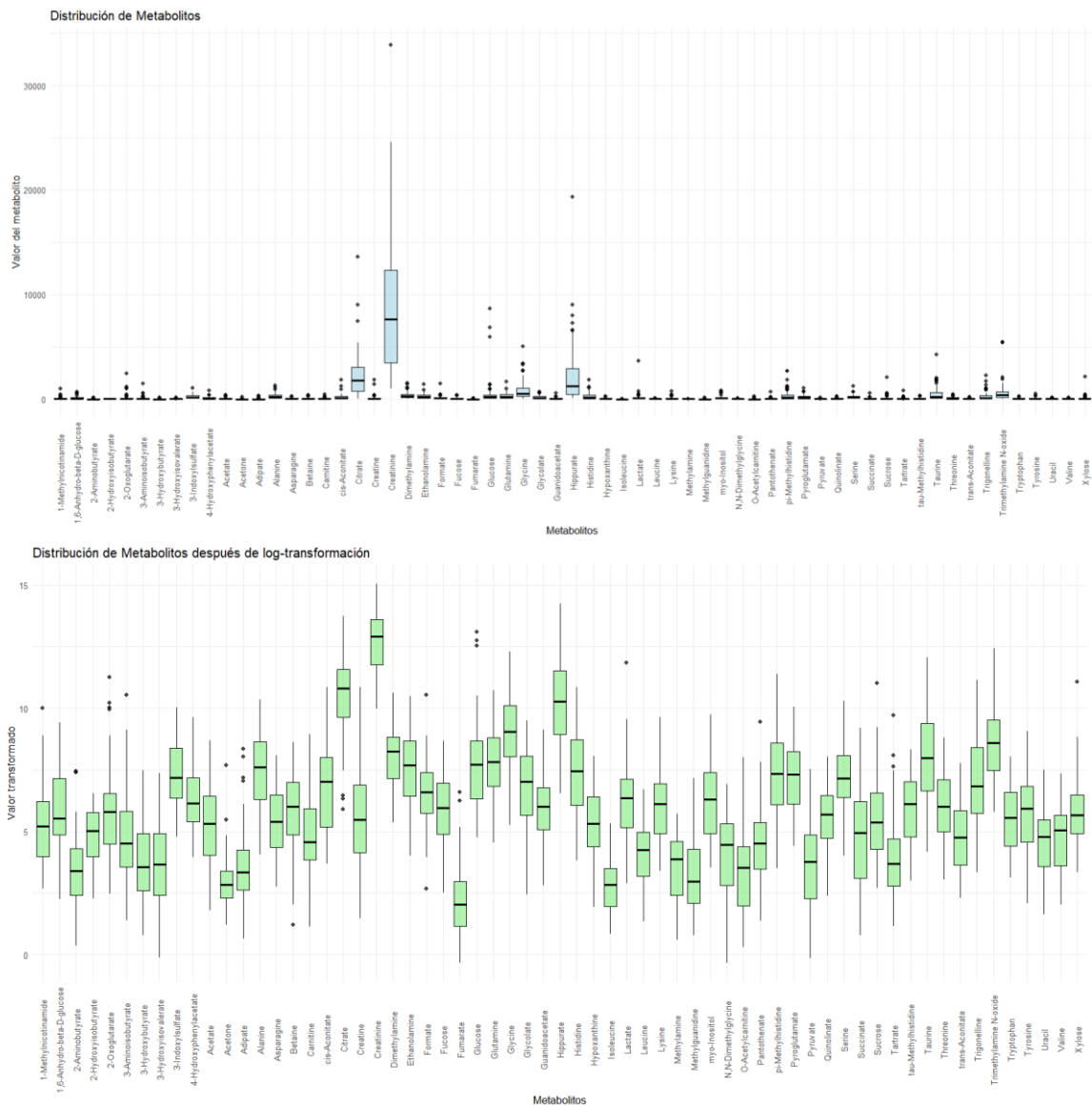


Figura 1: Arriba distribución de los datos crudos; abajo distribución de los datos después de la transformación.

Una vez transformados los datos, se transpone la matriz para que las columnas sean los metabolitos y se crea el objeto SummarizedExperiment. En la figura 2 vemos la estructura del objeto creado, llamado “se”, el cual consta de 63 filas y 77 columnas,

```

class: SummarizedExperiment
dim: 63 77
metadata(0):
assays(1): counts
rownames(63): 1,6-Anhydro-beta-D-glucose 1-Methylnicotinamide ... pi-Methylhistidine tau-Methylhistidine
rowData names(0):
colnames(77): PIF_178 PIF_087 ... NETL_003_V1 NETL_003_V2
colData names(2): PatientID MuscleLoss

```

Figura 2: Resumen del objeto creado.

3.2. Análisis exploratorio

3.2.1. Pruebas t

Se crea un script en R que realiza pruebas t para comparar los niveles de cada metabolito en función de si tienen caquexia o son control. Los valores p obtenidos se ajustan usando el método FDR para controlar el error por comparaciones múltiples. La figura 3 muestra los 10 metabolitos más significativos según el p valor ajustado, en el script adjuntado en Github se puede ver todos los valores. Estos metabolitos muestran diferencias significativas entre los grupos.

Feature	t_statistic	p_value	adjusted_p_value
Glucose	5.091330	2.563849e-06	0.0001615225
Adipate	4.738656	1.014722e-05	0.0002582590
Quinolate	4.816955	1.229805e-05	0.0002582590
Leucine	4.633705	1.770981e-05	0.0002789295
Valine	4.518369	2.671384e-05	0.0002949168
myo-Inositol	4.493841	2.808731e-05	0.0002949168
3-Hydroxyisovalerate	4.438793	4.298485e-05	0.0003489020
N,N-Dimethylglycine	4.397987	4.430502e-05	0.0003489020
3-Hydroxybutyrate	4.259178	5.940527e-05	0.0004158369
Succinate	4.301428	7.055359e-05	0.0004444876

Figura 3: Los 10 metabolitos más significativos según la prueba t.

3.2.2. Análisis de correlaciones

A continuación, se realiza un análisis de correlaciones de Pearson entre los niveles de cada par de metabolitos. Para su representación se ha realizado un mapa de calor (Fig. 4) en el cual en azul representa correlaciones cercanas al 0 y en rojo correlaciones cercanas al 1. Las filas representan metabolitos y están agrupadas jerárquicamente según su similitud en los valores de correlación.

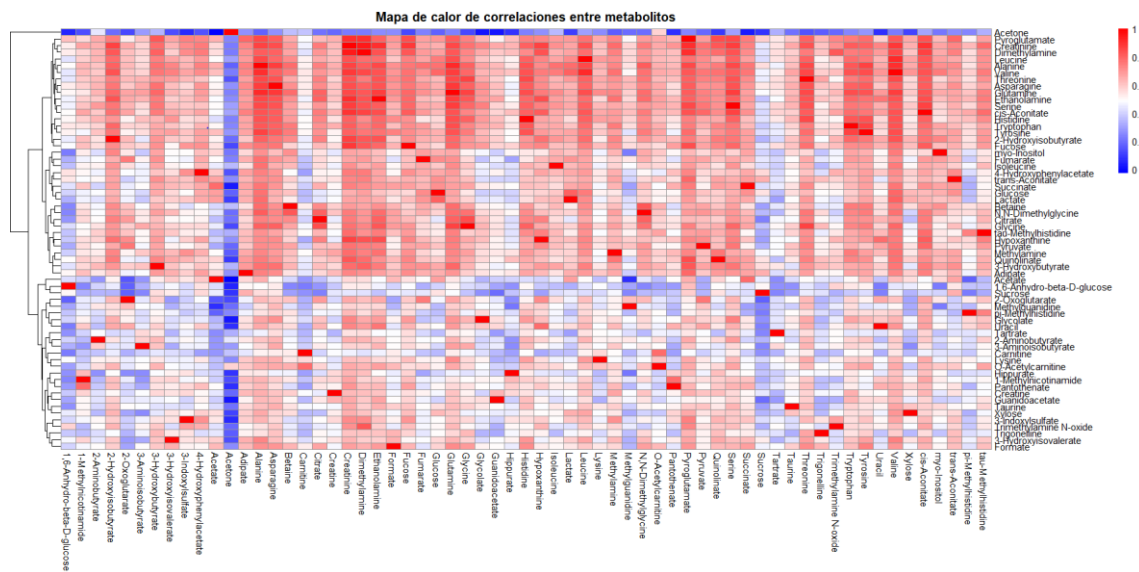


Figura 4: Mapa de calor de correlaciones entre metabolitos.

3.2.3. Dendrograma y k-means

Se procede con un agrupamiento jerárquico (Fig. 5) mediante la distancia euclidiana para medir similitud entre los pacientes, para detectar si se dividen perfectamente entre grupo control y enfermos.

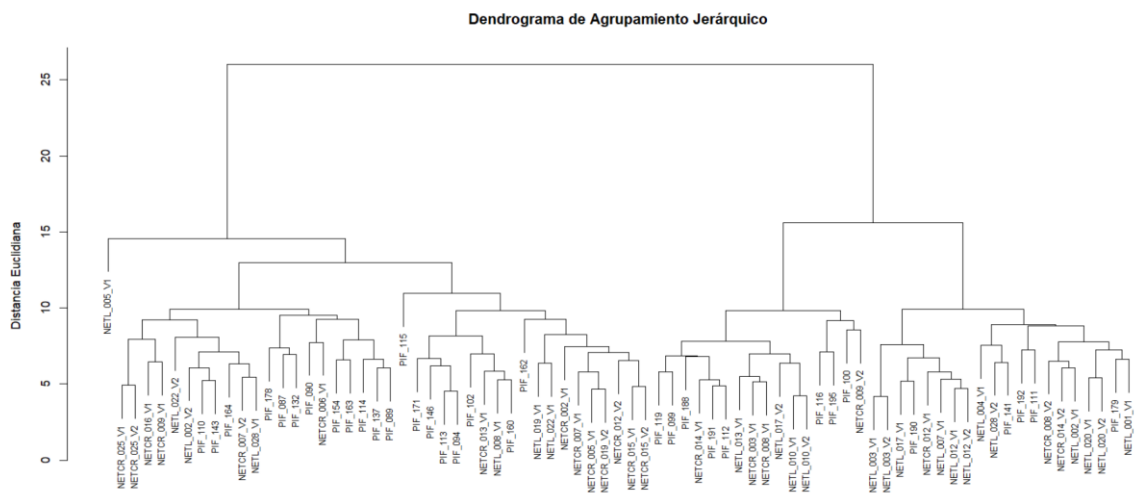


Figura 5: Dendrograma de los datos.

Mirando la figura 5 podemos pensar en dividir en 2 o en 4 grupos, para terminar de decidir, se calcula la suma de los cuadrados dentro de los cluster o WSS para valores de $k=1$ a $K=10$. A continuación se grafica en la figura 6 para identificar el codo o donde la disminución en WSS se estabiliza.

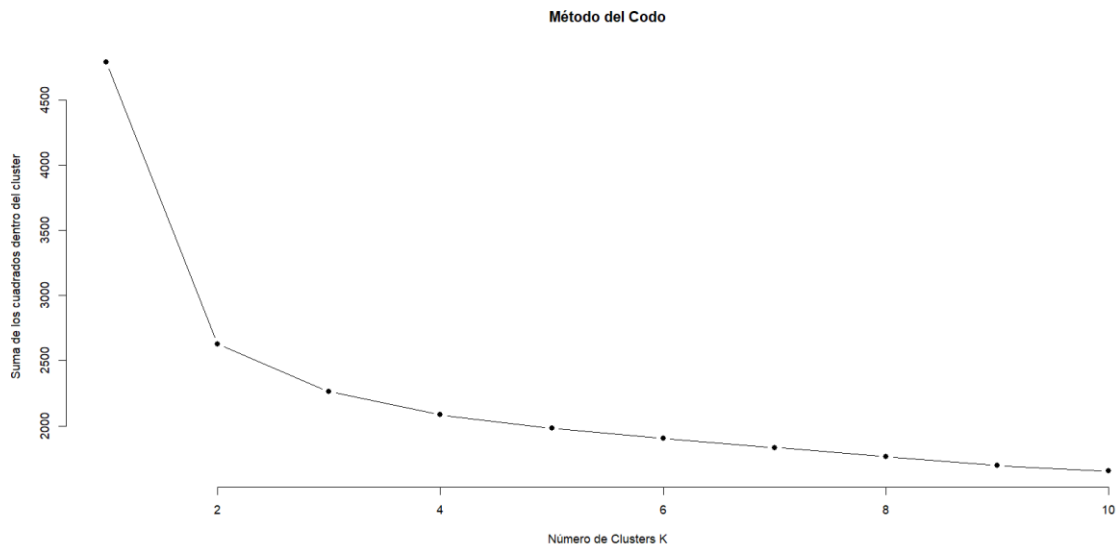


Figura 6: Grafico del numero de cluster y WSS.

3.2.4. Análisis de Componentes Principales

Esta técnica se usa para reducir la dimensionalidad de los datos mientras se conserva la mayor variabilidad posible. En este caso en particular, se aplica para ver si las muestras se agrupan según los clusters obtenidos con k-means. La figura 7 representa los datos en 2 dimensiones de los 4 cluster, vemos que las nubes de puntos de cada cluster quedan separadas entre ellas.



Figura 7: Representación gráfica de ACP.

4. Interpretación biológica

El análisis de los datos metabólicos ha revelado diferencias significativas entre los grupos estudiados. A través de pruebas t, se han identificado varios metabolitos que presentan diferencias relevantes entre las condiciones evaluadas, con valores de p ajustados que indican una alta significancia estadística. Entre los metabolitos más destacados se encuentran la glucosa, el adipato, el quinolinate, la leucina y la valina, entre otros. Esto sugiere que estos compuestos podrían estar implicados en procesos metabólicos claves relacionados con la caquexia.

El uso de técnicas de agrupamiento, como el método del codo, ha permitido determinar que cuatro es el número óptimo de clusters, lo que sugiere que las muestras pueden dividirse en grupos bien diferenciados. El dendrograma de agrupamiento jerárquico refuerza esta idea y muestra que se agrupan de manera clara en función de su condición, lo que podría indicar un perfil metabólico distintivo para cada grupo.

Por otra parte, el análisis de correlaciones entre metabolitos ha permitido visualizar patrones de relación entre ellos a través de un mapa de calor. Este tipo de análisis es útil para detectar metabolitos que presentan variaciones similares y que podrían estar involucrados en rutas metabólicas comunes. La identificación de correlaciones significativas puede proporcionar información sobre interacciones metabólicas que expliquen las diferencias observadas entre los grupos. Los 5 aminoácidos más correlacionados son, Dimethylamina y Creatinina con un valor de 0,945, la Valina con la Leucina 0,914, Glutamina con Alanina 0,912, Piroglutamato y Creatinina 0,905 y el piroglutamato con la dimethylamina con un valor de 0,905.

Por otro lado, el ACP ha permitido reducir la dimensionalidad de los datos y muestra una separación clara entre cluster, por lo tanto confirma que hay diferencias estructurales en los datos metabólicos y que K-Means hizo una buena clasificación.

En conclusión, los resultados obtenidos indican que existen diferencias metabólicas entre los grupos y que ciertos metabolitos pueden desempeñar un papel clave en la caracterización de la condición estudiada. Esto abre la posibilidad de utilizar estos metabolitos como posibles biomarcadores para la identificación y el estudio de la alteración metabólica en futuras investigaciones.