

Informe PEC1. Análisis de datos ómicos (M0-157). Primera prueba de evaluación continua. Marta Torres Valcárcel

2025-04-02

Table of Contents

1. Abstract o resumen.	1
2. Objetivos.	2
3. Métodos:	2
4. Resultados.	3
5. Discusión.	9
6. Conclusiones.	10
7. Referencias.	10

1. Abstract o resumen.

La Cachexia es un síndrome metabólico asociado a la pérdida de masa muscular y a cambios en el metabolismo energético. En este estudio, se analizaron datos de pacientes con Cachexia y pacientes Control para identificar diferencias significativas en los perfiles metabólicos.

Se emplearon herramientas bioinformáticas como Bioconductor y el paquete SummarizedExperiment para el procesamiento de los datos. A nivel estadístico, se realizó la prueba de Mann-Whitney-Wilcoxon, que reveló diferencias significativas en la mayoría de los metabolitos analizados. Además, se aplicaron técnicas de clustering evidenciando patrones en los metabolitos, aunque sin una clara separación entre pacientes sanos y enfermos.

Los resultados sugieren la existencia de alteraciones metabólicas significativas en la Cachexia, destacando ciertos metabolitos como posibles biomarcadores. Sin embargo, la falta de separación clara en los pacientes indica la necesidad de estudios adicionales para comprender mejor la variabilidad metabólica de la enfermedad.

2. Objetivos.

El objetivo general de este estudio trata del estudio metabólico de pacientes con cachexia frente a un grupo control.

Objetivos Específicos

1. Evaluación de la normalidad de los datos mediante pruebas estadísticas para determinar si se sigue una distribución normal.
2. Comparación de la distribución de la población enferma y control.
3. Representación de la población según paciente y según metabolito mediante dendograma.

3. Métodos:

Para la realización del trabajo lo primero que se hizo fue crear una cuenta de GITHUB con el fin de crear un repositorio. Se conectó el proyecto presente, elaborado en RStudio, con dicho repositorio.

A partir de aquí se clonaron del repositorio

<https://github.com/nutrimetabolomics/metaboData.git> la carpeta que contenía los dataset ofrecidos en el enunciado. Se decide elegir los de “2024-Cachexia” ya que es una enfermedad que no conocía y por las información proporcionada por los datos. En esta información se confirmaba que las muestras no estaban apareadas y que había claramente dos grupos. Todos los datos además eran numéricos y no había valores NA

El archivo tratado es un fichero de tipo .csv en el que se incluyen 65 columnas con 77 pacientes, de los cuales se tiene su ID y se tienen clasificados como control y enfermos. Junto a esto, el resto de columnas se tratan de los valores que adquieren cada paciente en 63 metabolitos diferentes y diversos.

Dichos metabolitos son los siguientes: 1,6-Anhidro-beta-D-glucosa, 1-Metilnicotinamida, 2-Aminobutirato, 2-Hidroxibutirato, 2-Oxoglutarato, 3-Aminoisobutirato, 3-Hidroxibutirato, 3-Hidroxiiisovalerato, 3-Indoxilsulfato, 4-Hidroxifenilacetato, Acetato, Acetona, Adipato, Alanina, Asparagina, Betaína, Carnitina, Citrato, Creatina, Creatinina, Dimetilamina, Etanolamina, Formiato, Fucosa, Fumarato, Glucosa, Glutamina, Glicina, Glicolato, Guanidoacetato, Hipurato, Histidina, Hipoxantina, Isoleucina, Lactato, Leucina, Lisina, Metilamina, Metilguanidina, N,N-Dimetilglicina, O-Acetilcarnitina, Pantotenato, Pirrolglutamato, Piruvato, Quinolonato, Serina, Succinato, Sacarosa, Tartrato, Taurina, Treonina, Trigonelina, Óxido de N-Trimetilamina, Triptófano, Tirosina, Uracilo, Valina, Xilosa, cis-Aconitato, Mio-Inositol, trans-Aconitato, pi-Metilhistidina, tau-Metilhistidina

Las herramientas bioinformáticas y bioestadísticas utilizadas fueron diversas en el procesamiento, análisis y visualización de los datos metabólicos del presente estudio. Se empleó el paquete Bioconductor, concretamente una extensión de ExpressionSet, denominada SummarizedExperiment, a partir de la cual se obtuvo un objeto binario en formato Rda. SummarizedExperiment es una estructura de datos diseñada para manejar datos ómicos.

A partir de aquí se plantea la pregunta sobre la normalidad de los datos de manera que se realiza la prueba de Shapiro-Wilk para evaluar la normalidad de los metabolitos, seguida de pruebas no paramétricas como Mann-Whitney-Wilcoxon para comparar los grupos de estudio. Para la exploración de patrones en los datos, se aplicaron técnicas de clustering jerárquico, utilizando la distancia euclidiana agrupando los metabolitos o los pacientes según su perfil metabólico.

Las visualizaciones fueron generadas con el paquete ggplot2, que permitió representar la distribución de los metabolitos a través de diagramas de caja. Además, se utilizó el paquete reshape2 para transformar los datos a formato largo y facilitar su manipulación gráfica.

4. Resultados.

Dado que no siguen una distribución normal, y teniendo en cuenta que son independientes, se realiza una prueba de Mann-Whitney-Wilcoxon para determinar si hay diferencias significativas en el nivel de metabolitos entre los grupos Control y Cachexia.

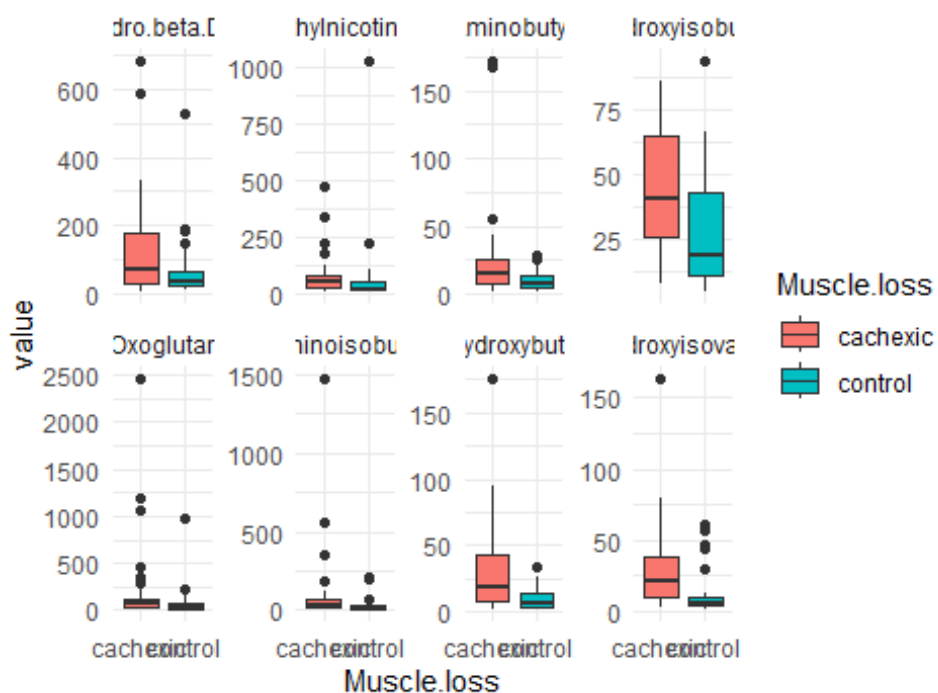
Se obtuvo que la mayoría de los metabolitos presentaban diferencias significativas entre ambos grupos, exceptuándose los siguientes casos: Isoleucina, Pantotenato, Tartrato, Hipoxantina, Aminoisobutirato Metilguanidina, Uracilo, Acetona

Las excepciones mencionadas obtuvieron p valores por encima de 0.05 por lo que no existen evidencias de que existan diferencias significativas.

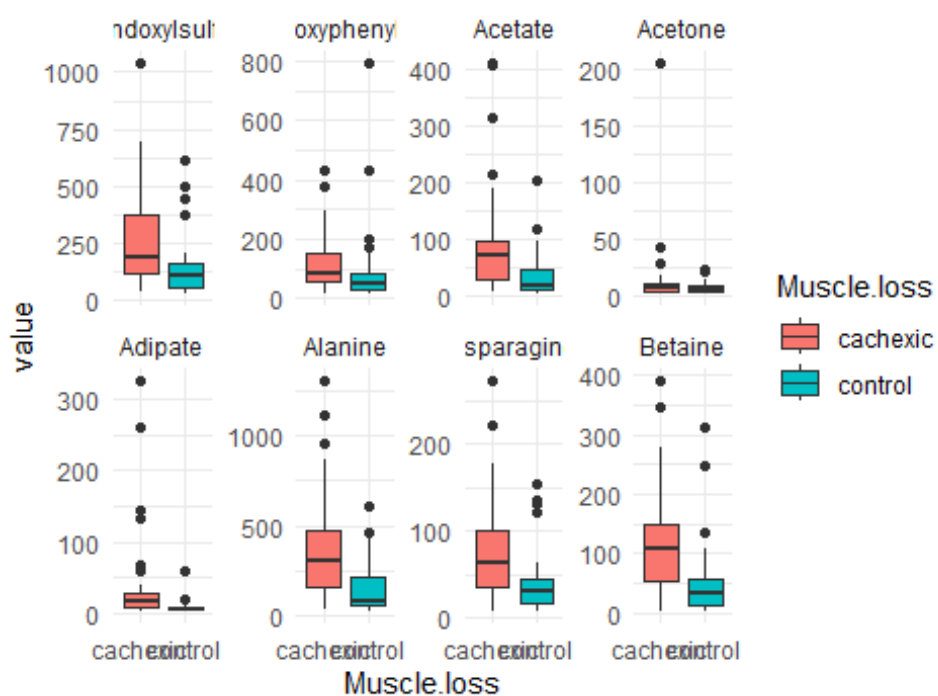
Se realizó una representación gráfica por diagrama de cajas en la que se comparan los niveles de cada metabolito según sea paciente control o enfermo. En la mayoría de los casos, el grupo de enfermos obtuvo valores más altos en cada metabolito que en el caso de los pacientes control, lo cual sugiere que puede haber una cierta relación entre la concentración de dichos metabolitos y la Cachexia, provocando dicha enfermedad una alteración metabólica.

Por ejemplificar, algunos de los valores que se encuentran más elevados son: 1,6-Anhidro-β-D-glucosa, 2-Hidroxibutirato, Alanina, Asparagina o Betaína

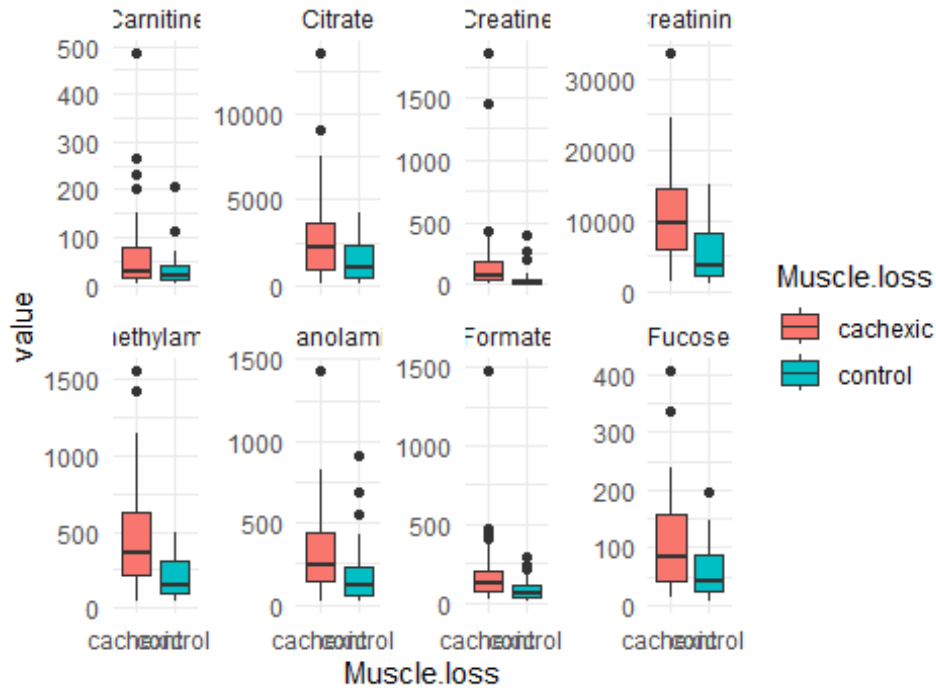
Comparación de metabolitos entre Control y Cachexi



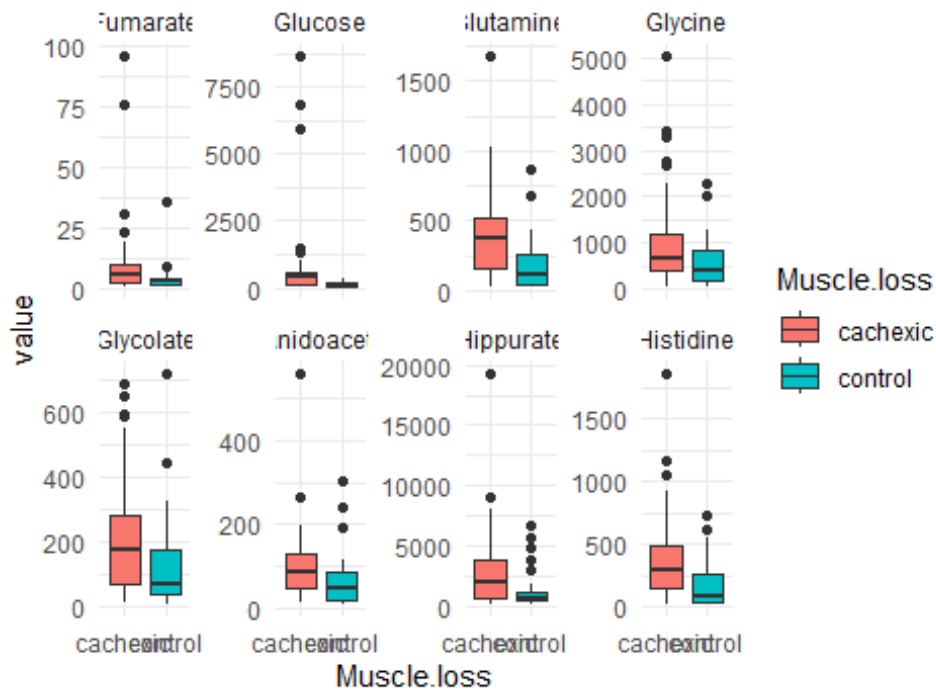
Comparación de metabolitos entre Control y Cachexi



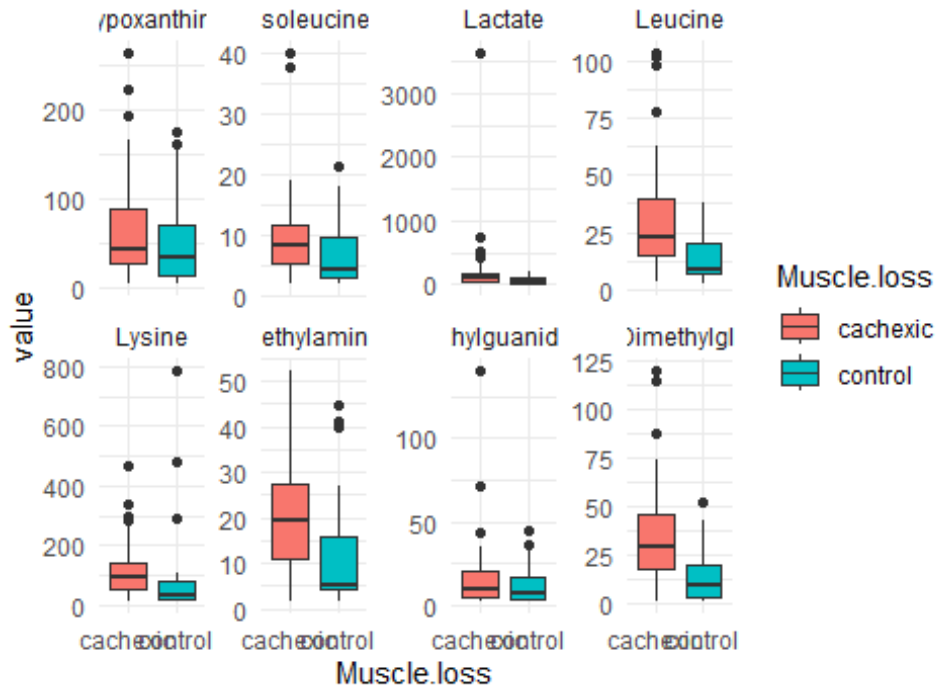
Comparación de metabolitos entre Control y Cachexi



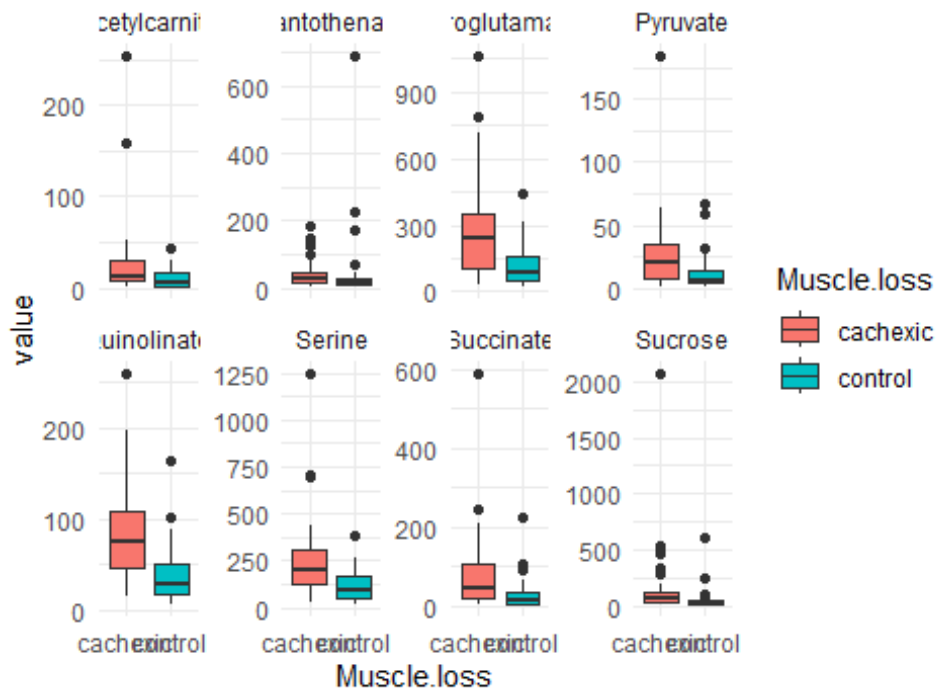
Comparación de metabolitos entre Control y Cachexia



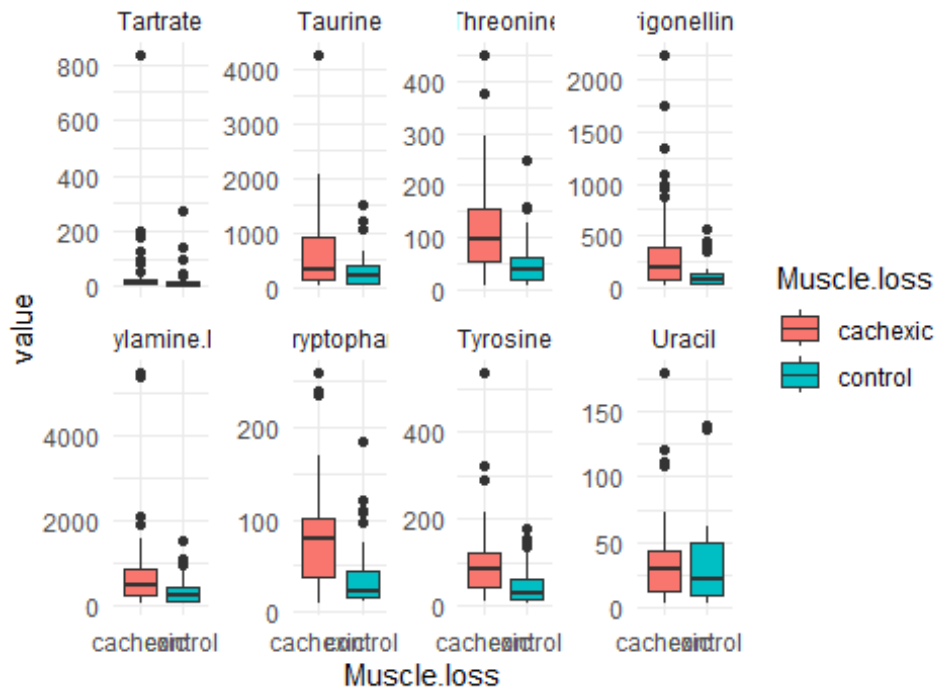
Comparación de metabolitos entre Control y Cachexia



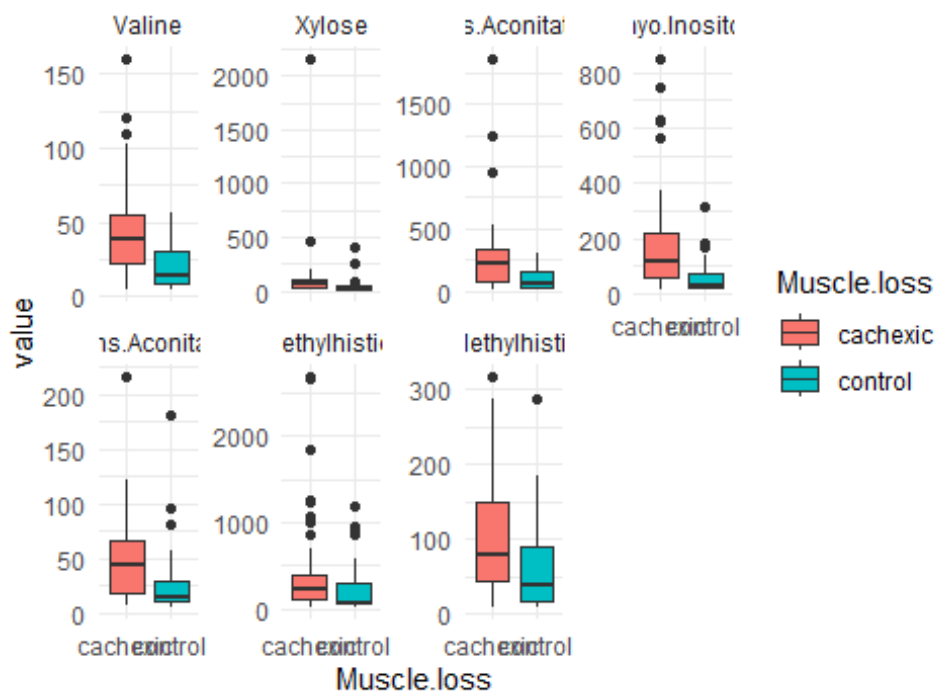
Comparación de metabolitos entre Control y Cachexia



Comparación de metabolitos entre Control y Cachexi

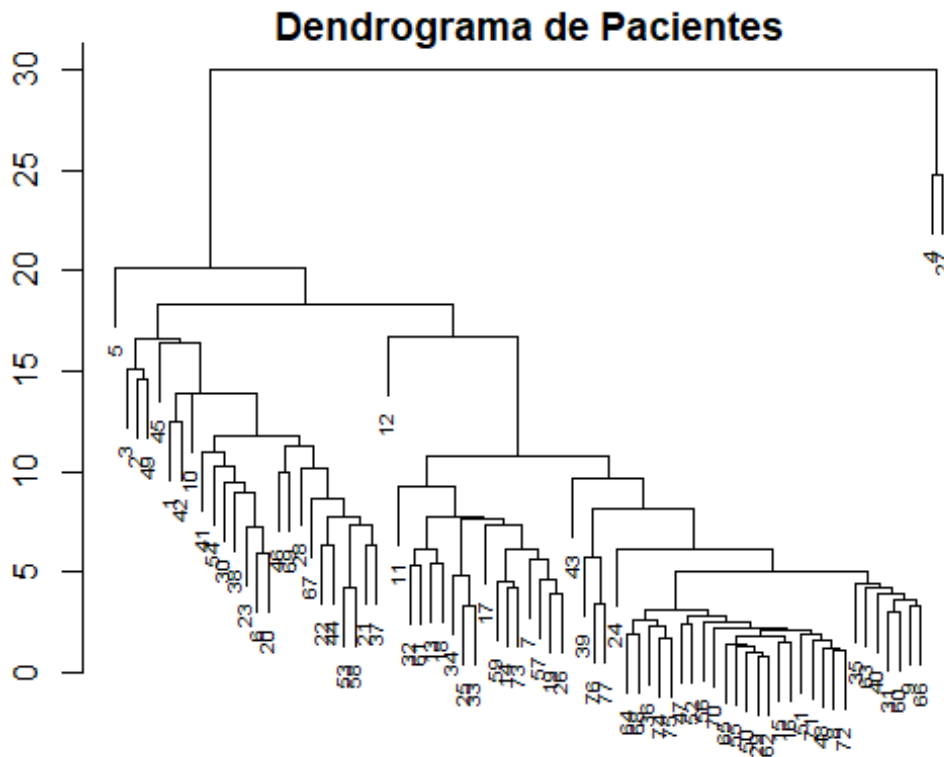


Comparación de metabolitos entre Control y Cachexia



Por último, se decide hacer un cluster para definir grupos de metabolitos con patrones similares, concretamente se consideran dos enfoques:

- Agrupación de pacientes según perfiles metabólicos: identificación de posibles subgrupos de pacientes con características similares.
- Agrupación de metabolitos según su comportamiento en el organismo del paciente.



Los resultados obtenidos del dendrograma de pacientes fueron variados. Aquellos pacientes unidos en los valores más bajos son los que guardan más similitudes entre ellos.

Caso contrario sucede para los pacientes 4 y 7, los cuales se unen en valores más altos lo que sugiere perfiles metabólicos diferentes.

Además, dado que el clustering no separó de forma clara el grupo de pacientes control y enfermo, se sugiere que los metabolitos que se han analizado no son los mejores para diferenciar entre individuos sanos y enfermos.

Por otro lado, los resultados de clustering de metabolitos identifican algunos compuestos que podrían ser posibles marcadores como es el caso de la creatina, el citrato o el Hipurato. Estos, muestran un comportamiento diferenciado en contraste con el resto que parecen agruparse de manera más homogénea, sugiriendo que presentan funciones metabólicas similares.

[illegible]

El estudio tratado ha permitido identificar las diferencias metabólicas entre pacientes con Cachexia y Control. Mediante una serie de análisis estadísticos y diferentes representaciones se ha podido evidenciar que ciertos metabolitos presentan diferencias significativas según el paciente sea sano o enfermo.

Además, aunque los análisis estadísticos mostraron diferencias significativas en los niveles de muchos metabolitos, el clustering de pacientes no reveló una separación clara entre individuos sanos y enfermos. Esto sugiere que, si bien hay alteraciones metabólicas en la Cachexia, el perfil metabólico por sí solo podría no ser suficiente para clasificar a los pacientes de manera precisa. La variabilidad interindividual, factores ambientales o incluso la influencia de otras condiciones médicas podrían estar afectando los resultados.

Finalmente, en futuros estudios, sería beneficioso integrar datos como transcriptómicos o proteómicos, para obtener una visión más completa del proceso patológico. Asimismo, la implementación de modelos predictivos basados en inteligencia artificial podría mejorar la clasificación de pacientes y la identificación de nuevos biomarcadores.

6. Conclusiones.

Los análisis estadísticos sugieren que existe una alteración metabólica significativa en los pacientes que presentan Cachexia, ya que se observa el aumento de niveles de ciertos metabolitos.

Dicha conclusión se sostiene con el clustering de metabolitos realizado, permitiendo identificar posibles biomarcadores clave que podrían ser útiles en la investigación de la enfermedad.

En contraposición, el clustering de pacientes refleja que no hay una división clara entre ambos grupos por lo que se requieren estudios adicionales para validar estos hallazgos.

7. Referencias.

<https://github.com/mtorresvalc/Torres-Valcarcel-Marta-PEC1.git>