

PEC 1 – ANÁLISIS DE DATOS ÓMICOS

Tabla de contenidos

1. Abstract.....	1
2. Objetivos.....	1
3. Metodología.....	2
5. Resultados.....	3
Creación de SummarizedExperiment.....	3
Análisis exploratorio.....	4
6. Discusión de resultados	8
7. Conclusiones.....	9
8. Referencias	10
9. Anexos.....	11

1. Abstract

La caquexia es un síndrome complejo que se caracteriza por una importante pérdida de masa muscular asociado también a alteraciones metabólicas. Es muy frecuente que se dé en enfermedades crónicas, como el cáncer. [2]

En esta práctica se analizan los datos metabolómicos de pacientes con caquexia vs pacientes control para identificar qué metabolitos están asociados a la pérdida de masa muscular. Los datos se han extraído del dataset human_cachexia.csv, proporcionado en los materiales de la asignatura

Mediante el paquete SummarizedExperiment en R, se estructuran los datos y se realiza un análisis exploratorio. Este incluye un análisis estadístico, pruebas de hipótesis y visualización gráfica.

2. Objetivos

El objetivo general del trabajo es identificar metabolitos diferenciales asociados a la caquexia mediante el análisis de datos metabolómicos, empleando herramientas bioinformáticas en R.

Para ello, se ha realizado lo siguiente:

- Preparación y estructuración de la información, creando el objeto SummarizedExperiment para un mejor manejo de datos y metadatos.

- Análisis exploratorio, en el que se describen las características generales de los datos y se da una idea de las diferencias entre el grupo caquexia y el grupo control.
- Identificación de metabolitos relevantes, mediante pruebas estadísticas (t-test) y corrección por el método Bonferroni.
- Interpretación biológica, relacionando los metabolitos identificados con rutas metabólicas importantes en la caquexia.

3. Metodología

La fuente de datos es el dataset human_cachexia.csv proporcionado en los materiales de la práctica, descargado de GitHub. Contiene 77 muestras, de las cuales 47 son pacientes con caquexia y 30 pertenecen al grupo control. Hay 65 variables, dos de ellas categóricas, Patient_ID y Muscle_loss (caquexia/control) y las 63 restantes son datos metabolómicos de concentraciones (valores numéricos continuos).

Los pasos que se han seguido se detallan a continuación:

1. Pre-procesamiento de datos: se han corregido los nombres de las columnas, así como sustituido los caracteres que R no es capaz de interpretar. También se han convertido ciertos valores a numéricos.
2. Normalización o escalado de los metabolitos, para ser capaces de comparar metabolitos en distintas unidades.
3. Estructuración de datos con SummarizedExperiment de bioconductor. Facilita la integración de datos y metadatos y, en un entorno de trabajo real, nos aporta información compatible con otras herramientas bioinformáticas. Los componentes de SummarizedExperiment son:
 - Assays: matriz de intensidades de los metabolitos.
 - colData: metadatos de muestras.
 - rowData: nombres de metabolitos.
4. Análisis exploratorio:
 - Estadística descriptiva (mediante summary). Permite observar medidas de tendencia central e identificar sesgos de distribución para cada metabolito.
 - Visualización mediante gráficos de barras y de tipo boxplot.
5. Análisis estadístico: se aplica t-test como prueba de hipótesis, para comparar las medias entre los grupos. Después, el método Bonferroni de corrección múltiple permite controlar los falsos positivos y, con todo ello, se seleccionan los tres metabolitos más relevantes del estudio, con un umbral de p-valor ajustado de $< 0,05$ (se han elegido solamente tres con el fin de reducir el informe a la vez que se explica el proceso de trabajo correctamente).

La herramienta utilizada ha sido el lenguaje R, empleando los paquetes clave readr, SummarizedExperiment, ggplot2 y dplyr.

5. Resultados

Creación de SummarizedExperiment

```
# Cargar la biblioteca necesaria

# Leer el archivo CSV
data <- read_csv("human_cachexia.csv", locale = locale(decimal_mark =
","))

# Corregir nombres de columnas
colnames(data) <- gsub(",", "_", colnames(data)) # Reemplazar comas por
guiones bajos
colnames(data) <- gsub(" ", "_", colnames(data)) # Reemplazar espacios
por guiones bajos
colnames(data) <- gsub("-", "_", colnames(data)) # Reemplazar los
guiones por guiones bajos

# Separar la matriz de datos (intensidades de los metabolitos)
data_matrix <- as.matrix(data[, 3:ncol(data)]) # Selecciona las columnas
numéricas

# Metadatos de las muestras
col_data <- data[, 1:2] # Selecciona "Patient ID" y "Muscle Loss"

# Metadatos de los metabolitos (nombres de los metabolitos)
row_data <- data.frame(metabolite = colnames(data)[3:ncol(data)])

# Transponer la matriz de datos
data_matrix <- t(data_matrix)

# Asignar nombres de filas (metabolitos)
rownames(data_matrix) <- row_data$metabolite

# Asignar nombres de columnas (muestras)
colnames(data_matrix) <- col_data$Patient_ID

# Instalar y cargar SummarizedExperiment

# Crear el objeto SummarizedExperiment
se <- SummarizedExperiment(assays = list(counts = data_matrix),
                           colData = col_data,
                           rowData = row_data)

# Ver el objeto
se

## 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(1): metabolite
## colnames(77): PIF_178 PIF_087 ... NETL_003_V1 NETL_003_V2
## colData names(2): Patient_ID Muscle_loss

# Guardar en formato binario
save(se, file = "summarizedexperimenta.Rda", compress = TRUE)
```

Análisis exploratorio

Resumen estadístico de las variables (en ANEXO 1):

- El dataset contiene una gran variabilidad en las concentraciones de los metabolitos, muchos de ellos con distribuciones sesgadas y rangos muy amplios.
- Las distribuciones de muchas variables están sesgadas, lo que sugiere que la media no es buen representante de la tendencia.

Este resumen estadístico es un punto de partida para comenzar un análisis más detallado.

```
# Cargar librerías necesarias

# Excluir columnas no numéricas (Patient_ID y Muscle_Loss)
numeric_data <- data[, -c(1, 2)]

# Algunas columnas están en formato chr
problem_columns <- c("3_Hydroxyisovalerate", "Fumarate",
"N_N_Dimethylglycine", "Pyruvate")

# Limpiar valores no numéricos
for (col in problem_columns) {
  numeric_data[[col]] <- gsub(",", "", numeric_data[[col]]) # Eliminar
comas
  numeric_data[[col]] <- gsub(" ", "", numeric_data[[col]]) # Eliminar
espacios
  numeric_data[[col]] <- gsub("[^0-9.]", "", numeric_data[[col]]) #
Eliminar todo excepto números y puntos
}

# Convertir las columnas a numéricas
for (col in problem_columns) {
  numeric_data[[col]] <- as.numeric(numeric_data[[col]])
}

# Escalar los datos (normalización)
scaled_data <- scale(numeric_data)

# Dividir los datos en dos grupos: cachexic y control
cachexic_data <- scaled_data[data$Muscle_loss == "cachexic", ]
control_data <- scaled_data[data$Muscle_loss == "control", ]
```

```
# Comparar Las medias de cada metabolito entre grupos
mean_diff <- colMeans(cachexic_data) - colMeans(control_data)

# Ver Los metabolitos con mayores diferencias
sort(abs(mean_diff), decreasing = TRUE)

##           N_N_Dimethylglycine           Glutamine
##           0.85147426           0.83911893
##           Formate           Valine
##           0.82346704           0.80088366
```

Se han representado los cuatro metabolitos más relevantes, de los cuales se escogen los tres primeros para el estudio. El resto se encuentra en el ANEXO 2.

Estos valores representan la diferencia absoluta de entre las medias de cada metabolito para el grupo cachexic y el grupo control. A mayor diferencia, mayor relevancia podría tener el metabolito para distinguir entre los dos grupos.

Los metabolitos con más diferencia son el N-N-dimethylglycine, Glutamine y Formate.

```
# Análisis estadístico

# Función para aplicar t-test a cada metabolito
perform_t_test <- function(metabolite) {
  t.test(metabolite ~ data$Muscle_loss)
}

# Aplicar t-test a todos Los metabolitos
t_test_results <- apply(numeric_data, 2, perform_t_test)

# Extraer p-valores
p_values <- sapply(t_test_results, function(x) x$p.value)

# Ajustar p-valores para múltiples comparaciones (corrección de Bonferroni)
adjusted_p_values <- p.adjust(p_values, method = "bonferroni")

# Crear un dataframe con Los resultados
results <- data.frame(
  Metabolite = colnames(numeric_data),
  Mean_Difference = mean_diff,
  P_Value = p_values,
  Adjusted_P_Value = adjusted_p_values
)

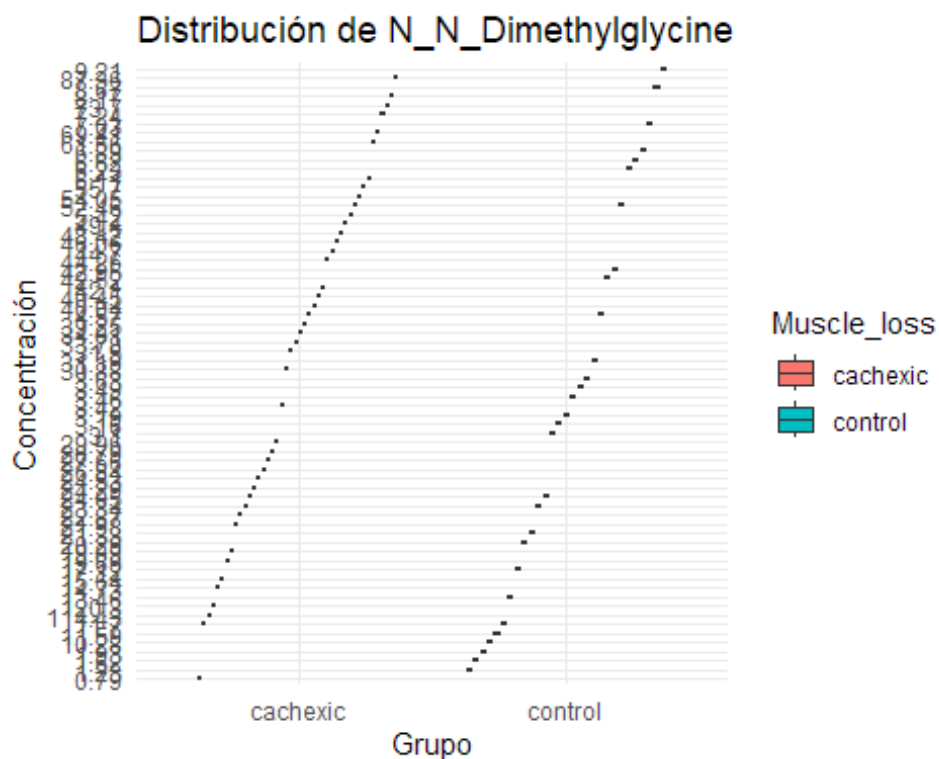
# Ordenar por p-valor ajustado
results <- results[order(results$Adjusted_P_Value), ]
```

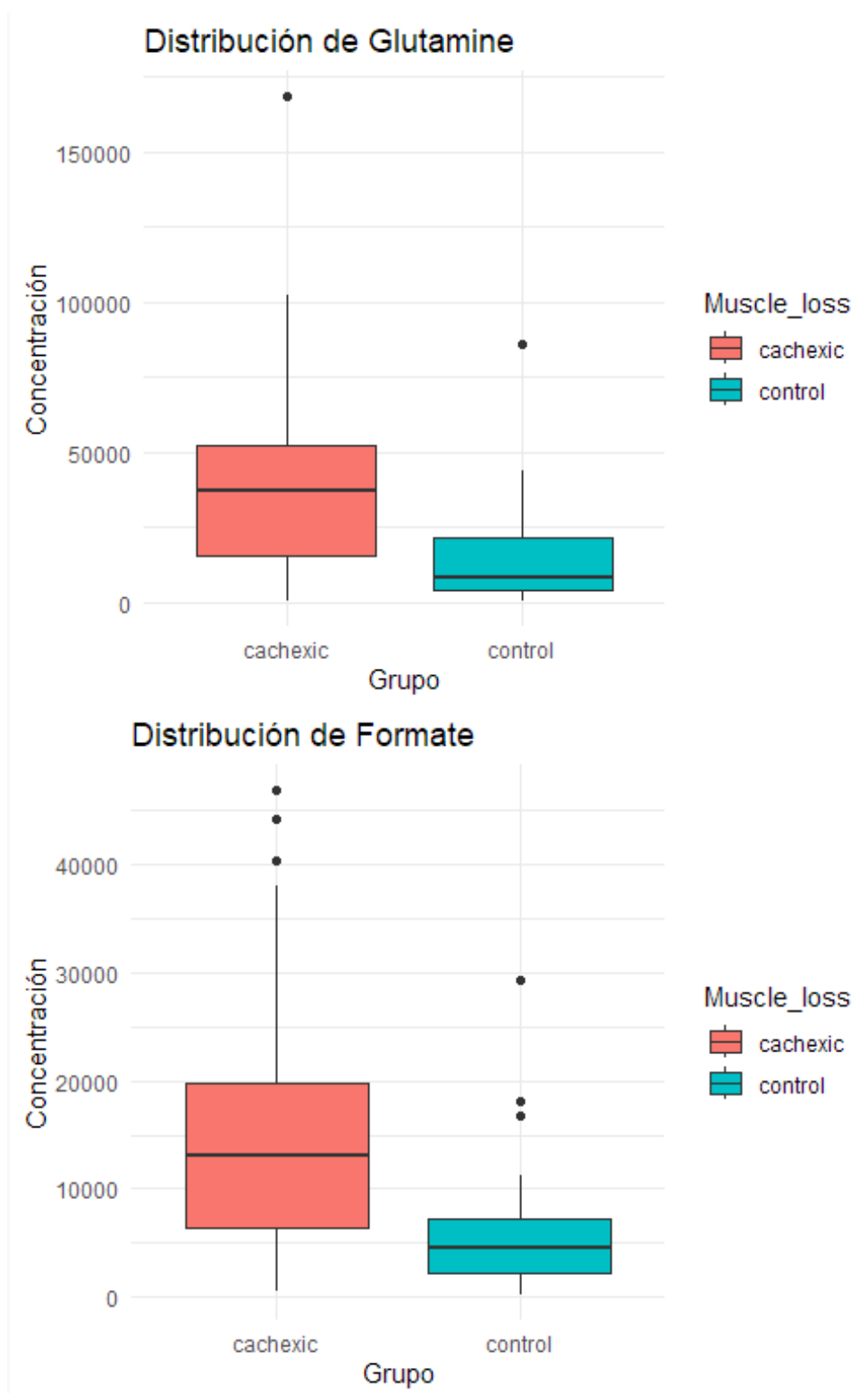
```
# Ver los resultados (resto en ANEXO 3)
print(results)

##                               Metabolite Mean_Difference
## N_N_Dimethylglycine          N_N_Dimethylglycine    0.85147426
## Glutamine                    Glutamine             0.83911893
## Formate                      Formate                0.82346704

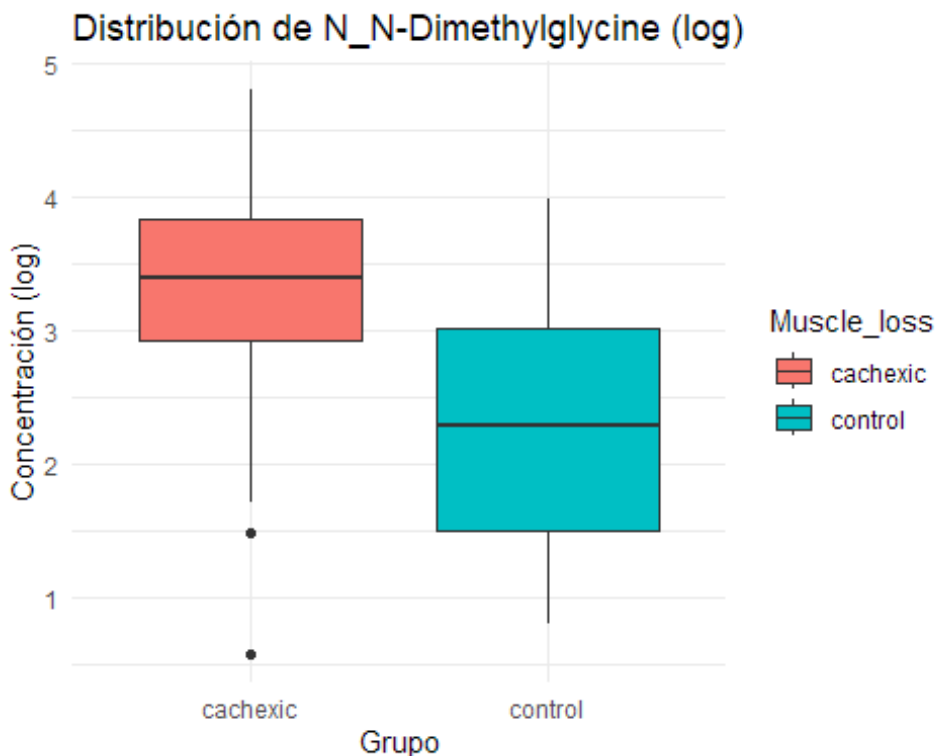
# Seleccionar los 3 metabolitos más significativos
top_metabolites <- results$Metabolite[1:3]

# Crear un gráfico de cajas (boxplot) para cada metabolito significativo
library(ggplot2)
for (metabolite in top_metabolites) {
  print(
    ggplot(data, aes(x = Muscle_loss, y = .data[[metabolite]], fill =
Muscle_loss)) +
    geom_boxplot() +
    labs(title = paste("Distribución de", metabolite),
         x = "Grupo",
         y = "Concentración") +
    theme_minimal()
  )
}
```





NOTA: Transformación logarítmica en ANEXO 4.



6. Discusión de resultados

El análisis realizado identifica tres metabolitos sustantivamente diferenciales entre pacientes con caquexia y pacientes del grupo control: N-N-Dimetilglicina, Glutamina o Formato.

- La N-N-dimetilglicina es la parte activa de la vitamina B15 y forma parte del metabolismo energético. Participa en ciertas reacciones enzimáticas implicadas en rutas de síntesis de vitaminas, hormonas, neurotransmisores y otras biomoléculas del organismo. Este metabolito actúa como “intensificador metabólico”, ayudando a mejorar funciones fisiológicas a nivel del sistema inmune, cardiovascular y función muscular. Este último papel es el que sobre todo se ve implicado en pacientes con caquexia. [1]
- La glutamina es un aminoácido que en personas sanas es no esencial, pero en situaciones de estrés, como el caso que se estudia, pasa a ser esencial. Es el aminoácido más abundante en el organismo, tanto en el plasma como en los tejidos, especialmente a nivel muscular. Además, participa en el sistema inmune y ejerce funciones como preservador del glutatión y otras moléculas antioxidantes. Como se comentaba previamente, el músculo esquelético alberga un gran depósito de glutamina, lo cual es coherente con que sea un metabolito a estudiar en pacientes con caquexia. [3] [4]
- El formato (o formiato) es un metabolito intermedio del metabolismo del ácido fórmico, el ácido carbónico más sencillo ya que solo tiene un carbono. Se trata

de un compuesto esencial, pues participa en la síntesis de purinas y en la metilación. Es producido por dos vías, a nivel mitocondrial y a nivel intestinal, en rutas dependientes del folato y la vitamina B12. Sin embargo, su acumulación excesiva puede llegar a ser perjudicial, pues genera estrés oxidativo por radicales libres. Su acumulación puede conllevar alteraciones mitocondriales y a déficit de vitamina B12 y folato, en enfermedades como el cáncer. Así, se observa su relación con la caquexia, lo cual ocupa este trabajo. [5]

Por otro lado, se deben discutir las diferencias entre SummarizedExperiment, el objeto empleado en el análisis de la práctica, versus ExpressionSet. [6] [7]

- SummarizedExperiment (SE) maneja múltiples datos y matrices (assays), mientras que ExpressionSet (ES) está más orientado a datos de expresión génica con una sola matriz de expresión.
- También hay diferencias en la estructura, el almacenamiento de metadatos de muestras, características y experimentales. En SE emplea colData, rowDara y metadata, respectivamente, y en ES empleamos phenoData, featureData y experimentData.
- SE es más moderno y compatible con herramientas actuales de Bioconductor.
- ES está siendo reemplazado por SE porque se emplea en muchas aplicaciones y bases de datos modernas, como metabolomicsWorkbench. Está siendo referente para el manejo de datos de alto rendimiento.

Sin embargo, el estudio presenta importantes limitaciones, que se exponen a continuación.

- Los resultados se han obtenido empleando estadística fiable pero, dada la extensión reducida del trabajo, es posible que ciertos metabolitos se hayan quedado fuera del estudio, pues se ha escogido profundizar en los tres primeros por simplificación.
- El tamaño de la muestra es reducido (77 pacientes), lo cual podría limitar también la generalización de los resultados.
- Algunos metabolitos presentan distribuciones sesgadas, concretamente N-N-dimetilglicina, por lo que se ha optado por una transformación logarítmica para un mejor análisis.
- El empleo de la corrección de Bonferroni puede descartar falsos negativos, pues aunque es un método riguroso, puede resultar conservador.

7. Conclusiones

- Los metabolitos identificados están vinculados a rutas metabólicas importantes en la caquexia, tales como el estrés oxidativo, la degradación muscular y la síntesis de proteínas. Esto respalda el hecho de que la caquexia

produce alteraciones importantes a nivel sistémico, en el metabolismo energético y proteico.

- Estos metabolitos podrían emplearse como biomarcadores potenciales en el diagnóstico en fases iniciales de la caquexia, así como en monitorización de pacientes con cáncer u otras enfermedades crónicas.
- Las limitaciones previamente explicadas hacen necesario un estudio más exhaustivo, con cohortes más grandes y que integren datos de otras ómicas para una mejor comprensión del síndrome.

8. Referencias

[1] Kendall, R. (1995). La dimetilglicina (DMG), un normalizador fisiológico con acción inmunoestimulante. *Natura Medicatrix: Revista médica para el estudio y difusión de las medicinas alternativas*, (41), 28-30.

[2] Gómez, R., & Fernández, M. (2017). Suplementos nutricionales como modificadores de morbilidad en pacientes con cáncer. *Archivos Venezolanos de Farmacología y Terapéutica*, 36(4), 169-180.
https://ve.scielo.org/scielo.php?script=sci_arttext&pid=S0004-06222017000400169

[3] Castejón, E., Lambruschini, N., Meavilla, S. M. y Catalán, N. (2010). Manejo farmacológico en el síndrome anorexia-caquexia. Sección de Gastroenterología, Hepatología y Nutrición Infantil, Hospital Sant Joan de Déu, Esplugues de Llobregat, Barcelona, España

[4] López Pérez, A., Perales Pascual, J., Escolano Pueyo, Á., López Pérez, M., & Serrano Vicente, C. (2022). Revisión narrativa del papel de la glutamina en la prevención y el tratamiento de diferentes patologías. *Revista de Nutrición Clínica y Metabolismo*, 5(4), 51-64. <https://doi.org/10.35454/rncm.v5n4.434>

[5] Pietzke, M., Meiser, J., & Vazquez, A. (2020). Formate metabolism in health and disease. *Molecular Metabolism*, 33, 23-37.
<https://doi.org/10.1016/j.molmet.2019.05.012>

[6] Morgan, M., Obenchain, V., Hester, J., & Pagès, H. (2023). SummarizedExperiment: SummarizedExperiment container [Software]. Bioconductor.
<https://bioconductor.org/packages/release/bioc/html/SummarizedExperiment.html>

[7] Gentleman, R., Carey, V., Huber, W., Irizarry, R., & Dudoit, S. (2023). Biobase: ExpressionSet class [Software]. RDocumentation.
<https://www.rdocumentation.org/packages/Biobase/versions/2.32.0/topics/ExpressionSet>

[8] Repositorio github: <https://www.github.com/tamcasti/Castillo-Blanco-Tamara-PEC1>

9. Anexos

ANEXO 1 – RESUMEN ESTADÍSTICO DE LAS VARIABLES

Resumen estadístico de Las variables

summary(data)

```
## Patient_ID      Muscle_loss      1_6_Anhydro_beta_D_glucose
## Length:77      Length:77      Min.   : 158
## Class :character Class :character 1st Qu.: 2357
## Mode  :character Mode  :character Median : 3847
##                                     Mean  : 9088
##                                     3rd Qu.:12774
##                                     Max.   :58993
## 1_Methylnicotinamide 2_Aminobutyrate 2_Hydroxyisobutyrate
2_Oxoglutarate
## Min.   : 108      Min.   : 5      Min.   : 93      Min.   :
85
## 1st Qu.: 1255      1st Qu.: 466      1st Qu.:1503      1st Qu.:
2049
## Median : 3587      Median : 894      Median :3214      Median :
4699
## Mean   : 6925      Mean    : 1715      Mean    :3586      Mean    :
13061
## 3rd Qu.: 7011      3rd Qu.: 1836      3rd Qu.:5194      3rd Qu.:
9276
## Max.   :103277      Max.    :17243      Max.    :9369      Max.
:246513
## 3_Aminoisobutyrate 3_Hydroxybutyrate 3_Hydroxyisovalerate
3_Indoxylsulfate
## Min.   : 93      Min.   : 17      Length:77      Min.   :
137
## 1st Qu.: 841      1st Qu.: 537      Class :character 1st Qu.:
5027
## Median : 2133      Median : 1125      Mode  :character Median :
12647
## Mean   : 5706      Mean    : 2070      Mean    :
19131
## 3rd Qu.: 5626      3rd Qu.: 2908      3rd Qu.:
24715
## Max.   :56116      Max.    :17591      Max.
:104315
## 4_Hydroxyphenylacetate Acetate      Acetone      Adipate
## Min.   : 315      Min.   : 57      Min.   : 3      Min.   : 8
## 1st Qu.: 2908      1st Qu.: 1307      1st Qu.: 401      1st Qu.: 481
## Median : 5974      Median : 3734      Median : 682      Median : 903
## Mean   :10238      Mean    : 6490      Mean    : 1049      Mean    : 2376
## 3rd Qu.:13295      3rd Qu.: 8649      3rd Qu.: 1028      3rd Qu.: 1854
## Max.   :79632      Max.    :41158      Max.    :20644      Max.    :32701
## Alanine      Asparagine      Betaine      Carnitine
```

```

## Min. : 1678 Min. : 8 Min. : 137 Min. : 117
## 1st Qu.: 6156 1st Qu.: 1746 1st Qu.: 2176 1st Qu.: 841
## Median : 16402 Median : 3734 Median : 5405 Median : 1969
## Mean : 24313 Mean : 5952 Mean : 7909 Mean : 4579
## 3rd Qu.: 35781 3rd Qu.: 8912 3rd Qu.:10995 3rd Qu.: 4252
## Max. :131291 Max. :27314 Max. :39151 Max. :48785
## Citrate Creatine Creatinine Dimethylamine
## Min. : 5027 Min. : 8 Min. : 14803 Min. : 447
## 1st Qu.: 42411 1st Qu.: 1294 1st Qu.: 248991 1st Qu.: 8912
## Median : 166903 Median : 4126 Median : 697439 Median : 21486
## Mean : 206064 Mean : 12526 Mean : 807068 Mean : 28988
## 3rd Qu.: 283557 3rd Qu.: 11792 3rd Qu.:1184901 3rd Qu.: 37993
## Max. :1362961 Max. :186311 Max. :3386035 Max. :142226
## Ethanolamine Formate Fucose Fumarate
## Min. : 557 Min. : 207 Min. : 57 Length:77
## 1st Qu.: 8227 1st Qu.: 3623 1st Qu.: 2579 Class :character
## Median : 20034 Median : 8393 Median : 6095 Mode :character
## Mean : 25912 Mean :11689 Mean : 8469
## 3rd Qu.: 38375 3rd Qu.:16077 3rd Qu.:11792
## Max. :143655 Max. :46872 Max. :40748
## Glucose Glutamine Glycine Glycolate
## Min. : 628 Min. : 366 Min. : 3303 Min. : 207
## 1st Qu.: 7826 1st Qu.: 7519 1st Qu.: 23746 1st Qu.: 4653
## Median : 17768 Median : 21486 Median : 49275 Median :11217
## Mean : 53894 Mean : 29525 Mean : 84104 Mean :17798
## 3rd Qu.: 39941 3rd Qu.: 44586 3rd Qu.: 99227 3rd Qu.:25724
## Max. :869062 Max. :168581 Max. :506445 Max. :72054
## Guanidoacetate Hippurate Histidine Hypoxanthine
## Min. : 137 Min. : 1133 Min. : 285 Min. : 93
## 1st Qu.: 1764 1st Qu.: 45825 1st Qu.: 5352 1st Qu.: 1218
## Median : 5298 Median : 108572 Median : 14841 Median : 3378
## Mean : 7737 Mean : 218126 Mean : 26886 Mean : 5755
## 3rd Qu.:10354 3rd Qu.: 286407 3rd Qu.: 38375 3rd Qu.: 8393
## Max. :56116 Max. :1934134 Max. :186311 Max. :26507
## Isoleucine Lactate Leucine Lysine
Methylamine
## Min. : 5 Min. : 137 Min. : 71 Min. : 137 Min.
: 5
## 1st Qu.: 269 1st Qu.: 2176 1st Qu.: 494 1st Qu.: 1781 1st
Qu.: 397
## Median : 553 Median : 6156 Median : 1415 Median : 6218
Median :1091
## Mean : 760 Mean : 14816 Mean : 2056 Mean : 8955 Mean
:1554
## 3rd Qu.:1049 3rd Qu.: 13295 3rd Qu.: 2553 3rd Qu.:11558 3rd
Qu.:2405
## Max. :4004 Max. :364095 Max. :10354 Max. :47819 Max.
:5246
## Methylguanidine N_N_Dimethylglycine O_Acetylcarnitine Pantothenate
## Min. : 5 Length:77 Min. : 16 Min. : 31

```

```

## 1st Qu.: 253   Class :character   1st Qu.: 386   1st Qu.: 717
## Median : 642   Mode  :character   Median : 1059   Median : 2009
## Mean : 1358                                     Mean : 1844   Mean : 4102
## 3rd Qu.: 1729                                   3rd Qu.: 1854   3rd Qu.: 3246
## Max. :14117                                     Max. :25468   Max. :69229
## Pyroglutamate      Pyruvate      Quinolinate      Serine
## Min. : 447   Length:77   Min. : 108   Min. : 421
## 1st Qu.: 5668   Class :character   1st Qu.: 2198   1st Qu.: 5194
## Median : 14259   Mode  :character   Median : 4699   Median : 13838
## Mean : 18607                                     Mean : 6143   Mean : 18912
## 3rd Qu.: 27866                                   3rd Qu.: 8563   3rd Qu.: 27043
## Max. :106422                                     Max. :25982   Max. :124888
## Succinate      Sucrose      Tartrate      Taurine
## Min. : 19   Min. : 193   Min. : 22   Min. : 137
## 1st Qu.: 739   1st Qu.: 1729   1st Qu.: 649   1st Qu.: 8563
## Median : 2822   Median : 3965   Median : 1281   Median : 24715
## Mean : 5948   Mean : 10545   Mean : 3886   Mean : 50029
## 3rd Qu.: 7444   3rd Qu.: 8393   3rd Qu.: 2478   3rd Qu.: 64548
## Max. :58993   Max. :207974   Max. :83715   Max. :427269
## Threonine      Trigonelline   Trimethylamine_N_oxide   Tryptophan
## Min. : 137   Min. : 499   Min. : 557   Min. : 123
## 1st Qu.: 1854   1st Qu.: 3809   1st Qu.: 14841   1st Qu.: 1746
## Median : 6034   Median : 8912   Median : 32376   Median : 4004
## Mean : 8809   Mean : 25369   Mean : 59960   Mean : 5851
## 3rd Qu.:13564   3rd Qu.: 32054   3rd Qu.: 65852   3rd Qu.: 9092
## Max. :45034   Max. :225296   Max. :548625   Max. :25982
## Tyrosine      Uracil      Valine      Xylose
## Min. : 137   Min. : 31   Min. : 41   Min. : 143
## 1st Qu.: 1503   1st Qu.: 761   1st Qu.: 1059   1st Qu.: 2176
## Median : 4295   Median : 2009   Median : 3088   Median : 4252
## Mean : 7329   Mean : 3064   Mean : 3303   Mean : 9549
## 3rd Qu.: 9849   3rd Qu.: 3847   3rd Qu.: 4699   3rd Qu.: 8736
## Max. :53915   Max. :17947   Max. :16077   Max. :216462
## cis_Aconitate   myo_Inositol   trans_Aconitate   pi_Methylhistidine
## Min. : 193   Min. : 137   Min. : 49   Min. : 494
## 1st Qu.: 3214   1st Qu.: 2357   1st Qu.: 825   1st Qu.: 6736
## Median : 10354   Median : 6472   Median : 2453   Median : 16239
## Mean : 19777   Mean :12349   Mean : 3808   Mean : 36885
## 3rd Qu.: 25468   3rd Qu.:16077   3rd Qu.: 5626   3rd Qu.: 38761
## Max. :186311   Max. :85406   Max. :21702   Max. :269728
## tau_Methylhistidine
## Min. : 8
## 1st Qu.: 1746
## Median : 6095
## Mean : 8234
## 3rd Qu.:12774
## Max. :31735

```

ANEXO 2 – METABOLITOS CON MAYORES DIFERENCIAS EN LA MEDIA

Ver los metabolitos con mayores diferencias

`sort(abs(mean_diff), decreasing = TRUE)`

##	N_N_Dimethylglycine	Glutamine
##	0.85147426	0.83911893
##	Formate	Valine
##	0.82346704	0.80088366
##	3_Indoxylsulfate	Quinolate
##	0.79402292	0.75880853
##	Dimethylamine	3_Hydroxybutyrate
##	0.74180940	0.71006672
##	Serine	Creatinine
##	0.70756779	0.69110489
##	Leucine	Methylamine
##	0.68144090	0.67491308
##	cis_Aconitate	Pyroglutamate
##	0.65939765	0.65804870
##	Tryptophan	Histidine
##	0.65769005	0.65053915
##	myo_Inositol	Fucose
##	0.64727945	0.63314825
##	Threonine	2_Hydroxyisobutyrate
##	0.63226405	0.62124436
##	Acetate	Asparagine
##	0.62084662	0.61801074
##	3_Hydroxyisovalerate	Alanine
##	0.61307064	0.60817196
##	Glycine	Trigonelline
##	0.60260818	0.58462753
##	Succinate	Taurine
##	0.57899126	0.56965704
##	Betaine	Pyruvate
##	0.55847863	0.55058423
##	Citrate	Hippurate
##	0.51864416	0.50278456
##	2_Aminobutyrate	Adipate
##	0.50109019	0.49708797
##	tau_Methylhistidine	Glucose
##	0.49005339	0.47982544
##	Lysine	trans_Aconitate
##	0.47614229	0.46107956
##	Creatine	Fumarate
##	0.45338650	0.44030461
##	Trimethylamine_N_oxide	Tyrosine
##	0.43179105	0.41685982
##	Ethanolamine	2_Oxoglutarate
##	0.41243314	0.38676587
##	Carnitine	Guanidoacetate

```
##          0.38481377          0.37973978
##      O_Acetylcarnitine 1_6_Anhydro_beta_D_glucose
##          0.37842575          0.37611640
##          Glycolate          Lactate
##          0.37574215          0.34795757
##      pi_Methylhistidine      3_Aminoisobutyrate
##          0.34645558          0.33560097
##          Sucrose          Methylguanidine
##          0.32305597          0.31261023
##          Xylose          Hypoxanthine
##          0.30827270          0.25062877
##          Acetone          Pantothenate
##          0.22361169          0.19130343
##      4_Hydroxyphenylacetate      Tartrate
##          0.17858279          0.17814394
##          Isoleucine          Uracil
##          0.14410105          0.05340809
##      1_Methylnicotinamide
##          0.02978337
```

ANEXO 3 – RESULTADO DEL ANÁLISIS ESTADÍSTICO

Ver Los resultados

```
print(results)
```

```
##          Metabolite Mean_Difference
## N_N_Dimethylglycine      N_N_Dimethylglycine      0.85147426
## Glutamine          Glutamine      0.83911893
## Formate          Formate      0.82346704
## Valine          Valine      0.80088366
## 3_Indoxylsulfate      3_Indoxylsulfate      0.79402292
## Dimethylamine      Dimethylamine      0.74180940
## 3_Hydroxybutyrate      3_Hydroxybutyrate      0.71006672
## Quinolinate      Quinolinate      0.75880853
## Serine          Serine      0.70756779
## Leucine          Leucine      0.68144090
## cis_Aconitate      cis_Aconitate      0.65939765
## Creatinine          Creatinine      0.69110489
## myo_Inositol      myo_Inositol      0.64727945
## Pyroglutamate      Pyroglutamate      0.65804870
## Histidine          Histidine      0.65053915
## Methylamine          Methylamine      0.67491308
## Fucose          Fucose      0.63314825
## Tryptophan          Tryptophan      0.65769005
## Acetate          Acetate      0.62084662
## Threonine          Threonine      0.63226405
## Trigonelline      Trigonelline      0.58462753
## Glycine          Glycine      0.60260818
## Alanine          Alanine      0.60817196
## 3_Hydroxyisovalerate      3_Hydroxyisovalerate      0.61307064
## Asparagine          Asparagine      0.61801074
```

## Taurine	Taurine	0.56965704
## Succinate	Succinate	0.57899126
## 2_Hydroxyisobutyrate	2_Hydroxyisobutyrate	0.62124436
## Pyruvate	Pyruvate	0.55058423
## 2_Aminobutyrate	2_Aminobutyrate	0.50109019
## Adipate	Adipate	0.49708797
## Glucose	Glucose	0.47982544
## Betaine	Betaine	0.55847863
## Citrate	Citrate	0.51864416
## Hippurate	Hippurate	0.50278456
## 1_6_Anhydro_beta_D_glucose	1_6_Anhydro_beta_D_glucose	0.37611640
## 1_Methylnicotinamide	1_Methylnicotinamide	-0.02978337
## 2_Oxoglutarate	2_Oxoglutarate	0.38676587
## 3_Aminoisobutyrate	3_Aminoisobutyrate	0.33560097
## 4_Hydroxyphenylacetate	4_Hydroxyphenylacetate	0.17858279
## Acetone	Acetone	0.22361169
## Carnitine	Carnitine	0.38481377
## Creatine	Creatine	0.45338650
## Ethanolamine	Ethanolamine	0.41243314
## Fumarate	Fumarate	0.44030461
## Glycolate	Glycolate	0.37574215
## Guanidoacetate	Guanidoacetate	0.37973978
## Hypoxanthine	Hypoxanthine	0.25062877
## Isoleucine	Isoleucine	0.14410105
## Lactate	Lactate	0.34795757
## Lysine	Lysine	0.47614229
## Methylguanidine	Methylguanidine	0.31261023
## O_Acetylcarnitine	O_Acetylcarnitine	0.37842575
## Pantothenate	Pantothenate	-0.19130343
## Sucrose	Sucrose	0.32305597
## Tartrate	Tartrate	0.17814394
## Trimethylamine_N_oxide	Trimethylamine_N_oxide	0.43179105
## Tyrosine	Tyrosine	0.41685982
## Uracil	Uracil	0.05340809
## Xylose	Xylose	0.30827270
## trans_Aconitate	trans_Aconitate	0.46107956
## pi_Methylhistidine	pi_Methylhistidine	0.34645558
## tau_Methylhistidine	tau_Methylhistidine	0.49005339
##	P_Value Adjusted_P_Value	
## N_N_Dimethylglycine	2.114015e-05	0.001331829
## Glutamine	3.569296e-05	0.002248656
## Formate	4.223756e-05	0.002660966
## Valine	7.009518e-05	0.004415997
## 3_Indoxylsulfate	7.429254e-05	0.004680430
## Dimethylamine	1.966380e-04	0.012388195
## 3_Hydroxybutyrate	2.406200e-04	0.015159059
## Quinolinolate	2.519296e-04	0.015871567
## Serine	3.850957e-04	0.024261026
## Leucine	5.160085e-04	0.032508535
## cis_Aconitate	6.703090e-04	0.042229469

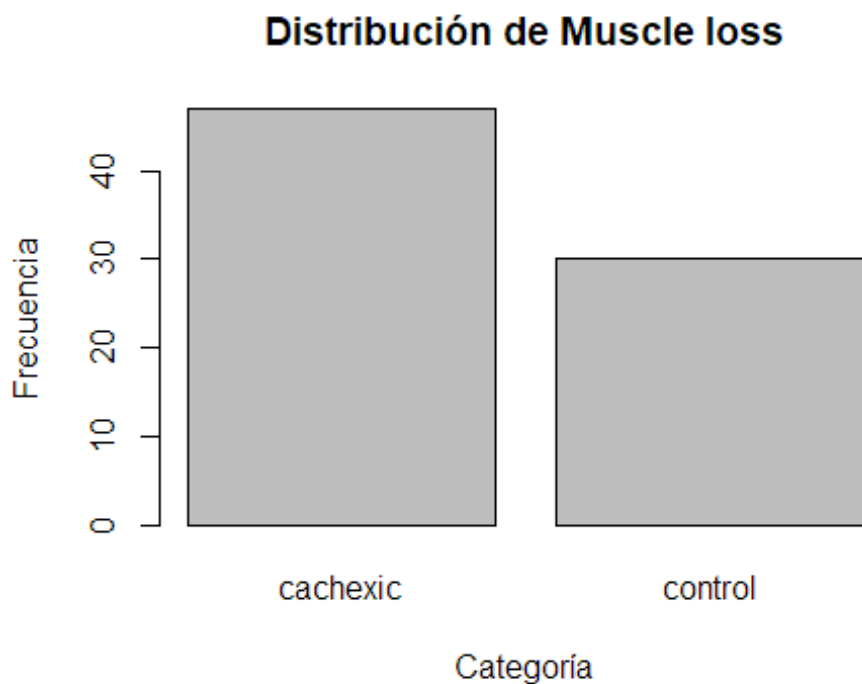
## Creatinine	8.357591e-04	0.052652822
## myo_Inositol	1.048377e-03	0.066047773
## Pyroglutamate	1.219365e-03	0.076819983
## Histidine	1.288378e-03	0.081167843
## Methylamine	1.525107e-03	0.096081767
## Fucose	1.756860e-03	0.110682153
## Tryptophan	2.103360e-03	0.132511692
## Acetate	2.107729e-03	0.132786919
## Threonine	2.281433e-03	0.143730306
## Trigonelline	2.531383e-03	0.159477119
## Glycine	2.616481e-03	0.164838299
## Alanine	2.984559e-03	0.188027248
## 3_Hydroxyisovalerate	3.458034e-03	0.217856150
## Asparagine	3.843171e-03	0.242119786
## Taurine	4.030135e-03	0.253898533
## Succinate	4.046232e-03	0.254912639
## 2_Hydroxyisobutyrate	6.741298e-03	0.424701805
## Pyruvate	7.357362e-03	0.463513809
## 2_Aminobutyrate	9.283377e-03	0.584852774
## Adipate	9.481537e-03	0.597336829
## Glucose	1.155836e-02	0.728176905
## Betaine	1.172988e-02	0.738982448
## Citrate	1.173970e-02	0.739601053
## Hippurate	1.430867e-02	0.901446033
## 1_6_Anhydro_beta_D_glucose	9.784289e-02	1.000000000
## 1_Methylnicotinamide	9.134834e-01	1.000000000
## 2_Oxoglutarate	4.280939e-02	1.000000000
## 3_Aminoisobutyrate	1.010100e-01	1.000000000
## 4_Hydroxyphenylacetate	4.929102e-01	1.000000000
## Acetone	2.455619e-01	1.000000000
## Carnitine	5.854324e-02	1.000000000
## Creatine	1.942113e-02	1.000000000
## Ethanolamine	6.480049e-02	1.000000000
## Fumarate	2.620524e-02	1.000000000
## Glycolate	1.025792e-01	1.000000000
## Guanidoacetate	8.718822e-02	1.000000000
## Hypoxanthine	2.720059e-01	1.000000000
## Isoleucine	5.086887e-01	1.000000000
## Lactate	6.712148e-02	1.000000000
## Lysine	3.737421e-02	1.000000000
## Methylguanidine	1.271945e-01	1.000000000
## O_Acetylcarnitine	5.039199e-02	1.000000000
## Pantothenate	5.036630e-01	1.000000000
## Sucrose	1.033021e-01	1.000000000
## Tartrate	3.795001e-01	1.000000000
## Trimethylamine_N_oxide	2.766311e-02	1.000000000
## Tyrosine	4.281529e-02	1.000000000
## Uracil	8.208692e-01	1.000000000
## Xylose	1.123071e-01	1.000000000
## trans_Aconitate	4.206952e-02	1.000000000

```
## pi_Methylhistidine      9.799334e-02      1.0000000000
## tau_Methylhistidine     2.695402e-02      1.0000000000

# Frecuencia de La variable categórica
table(data$Muscle_loss)

##
## cachexic    control
##          47         30

# Gráfico de barras para La variable categórica
barplot(table(data$Muscle_loss), main = "Distribución de Muscle loss",
xlab = "Categoría", ylab = "Frecuencia")
```



ANEXO 4 - TRANSFORMACIÓN LOGARÍTMICA

```
# Dada la diferencia de valores en N_N-Dimethylglycine se procede a una
transformación Logarítmica
# Convertir a numérico
data$N_N_Dimethylglycine <- as.numeric(data$N_N_Dimethylglycine)

# Transformación Logarítmica
data$N_N_Dimethylglycine_log <- log(data$N_N_Dimethylglycine + 1) # +1
para evitar Log(0)

# Gráfico de cajas
```

```
ggplot(data, aes(x = Muscle_loss, y = N_N_Dimethylglycine_log, fill =  
Muscle_loss)) +  
  geom_boxplot() +  
  labs(title = "Distribución de N_N-Dimethylglycine (log)",  
        x = "Grupo",  
        y = "Concentración (log)") +  
  theme_minimal()
```