

PEC1. Análisis de datos ómicos

Mireia García López

2025-03-27

Tabla de Contenidos

1. Abstract
2. Objetivos
3. Métodos
 - 3.1 Selección y Justificación del Dataset
 - 3.2 Creación del Objeto SummarizedExperiment
 - 3.3 Análisis Exploratorio de los Datos
4. Resultados
 - 4.1 Creación del objeto SummarizedExperiment
 - 4.2 Análisis exploratorio
5. Discusión
6. Conclusiones
7. Referencias
8. Anexos

1. Abstract

Este estudio tiene como objetivo integrar y analizar datos de metabolómica relacionados con la salud cardiovascular en el envejecimiento, específicamente en las asociaciones entre las subclases de lipoproteínas. Se seleccionó el dataset ST003791 del National Metabolomics Data Repository, el cual se centra en humanos y contiene información sobre metabolitos y metadatos de muestras. El proceso comenzó con la creación de un objeto de clase SummarizedExperiment, utilizando la matriz de expresión de metabolitos y los metadatos correspondientes a las muestras y características experimentales. Posteriormente, se realizó un análisis exploratorio de los datos, verificando su estructura, calidad y distribución. Los resultados muestran una gran variabilidad en la expresión de los metabolitos, con la presencia de valores atípicos que podrían reflejar estados patológicos, y se destaca la heterogeneidad en la distribución de variables como edad y sexo, lo que podría influir en las conclusiones del estudio. Aunque el análisis exploratorio ha proporcionado una visión detallada de la calidad de los datos, se recomienda realizar análisis más específicos para investigar las relaciones entre los metabolitos y la salud cardiovascular.

2. Objetivos

Los objetivos principales de este trabajo son:

- Seleccionar y justificar un dataset de metabolómica adecuado para el análisis.
- Crear un objeto de clase SummarizedExperiment integrando la matriz de expresión y los metadatos.
- Realizar un análisis exploratorio de los datos para evaluar su calidad y estructura.

- Interpretar los resultados obtenidos en el contexto biológico del estudio, que es el estudio de las asociaciones entre las subclases de lipoproteínas y la salud cardiovascular en el envejecimiento.

3. Métodos

La actividad se ha realizado utilizando la versión 4.4.1 de Rstudio.

3.1 Selección y Justificación del Dataset

Para la realización de este ejercicio el dataset de metabolómica seleccionado es de metabolomicsWorkbench. Pertenece al National Metabolomics Data Repository (NMDR), este te permite buscar los estudios mostrandote el resumen. He aplicado un filtro a la búsqueda para restringirla a estudios en seres humanos. Me he decantado por el estudio ST003791 que estudia las asociaciones entre las subclases de lipoproteínas y la salud cardiovascular en el envejecimiento. He descargado el dataset en formato JSON. Las razones por las que he elegido el dataset de este estudio es por ser un tema dentro de mi área de conocimiento y por presentar una estructura organizada y por la riqueza de la información contenida, lo que será beneficioso para su integración en SummarizedExperiment.

3.2 Creación del Objeto SummarizedExperiment

La creación del objeto SummarizedExperiment se ha realizado en varias etapas. En primer lugar, se extrajo la matriz de expresión desde la sección NMR_METABOLITE_DATA\$Data del archivo JSON. La primera columna contenía los nombres de los metabolitos, que se establecieron como nombres de fila y a continuación se eliminó la columna asegurando que los datos quedaran en formato numérico.

Se prepararon los metadatos de las muestras a partir de dos componentes del JSON: SUBJECT_SAMPLE_FACTORS\$Additional sample data, que contenía información como ID de muestra, edad y sexo, y SUBJECT_SAMPLE_FACTORS\$Factors, que incluía factores experimentales. Ambos fueron combinados en un solo objeto: colData. También se generó rowData, una tabla con los nombres de los metabolitos como filas.

En este punto se presentó el problema de la discrepancia entre el número de muestras en expr_matrix y en total_sample_metadata, ya que la segunda contenía más registros. Para solucionar esto, se filtraron los metadatos de muestra manteniendo únicamente aquellas que coincidían con la matriz de expresión. Luego, se ordenaron en el mismo orden que las columnas de expr_matrix para asegurar que los datos estaban en el orden correcto.

Por último, se creó el objeto SummarizedExperiment con la función SummarizedExperiment(), integrando la matriz de expresión como assays, los metadatos de muestra como colData, los metadatos de metabolitos como rowData, y los metadatos generales del estudio en metadata.

El resultado fue un objeto de clase SummarizedExperiment con las siguientes características: Dimensiones: 112 filas (metabolitos) y 112 columnas (muestras). Metadatos de muestras (colData): 18 columnas, que incluyen detalles sobre las muestras. Metadatos de características

(rowData): 1 columna con el nombre del metabolito. Metadatos del estudio: Información sobre el estudio y sus publicaciones. Esto confirmó la correcta integración de los datos.

3.3 Análisis Exploratorio de los Datos

El análisis exploratorio de los datos se ha llevado a cabo en varias etapas:

Análisis de la estructura. Se verificó la estructura de los datos utilizando varias funciones. Se inspeccionaron las dimensiones de la matriz de expresión con la función `dim(assays(se)$counts)`, las de los metadatos con `(dim(colData(se)))` y las características de los metabolitos (`dim(rowData(se))`). También se inspeccionaron las primeras filas de los metadatos de las muestras y los metabolitos utilizando las funciones `head(colData(se))` y `head(rowData(se))`.

Verificación de la matrix de expresión: se comprobó si los datos de la matriz de expresión eran numéricos para poder calcular medidas para la estadística descriptiva del dataset. Resultó que la matrix no era de tipo numérico, se solucionó el problema y se verificó que la matrix de expresión finalmente era de tipo numérico.

Análisis de Estadísticas Descriptivas: Se calcularon medidas resumen como la media, mediana, desviación estándar y rangos intercuartílicos de cada metabolito.

Inspección de los Metadatos: Se inspeccionaron los metadatos y se verificó la distribución de algunas variables en los metadatos.

Observación de los datos: se visualizó un diagrama de densidad para representar la distribución de los valores de expresión.

En el **Anexo 1** se encuentra el código completo. Ahí se puede ver exactamente cómo se ha desarrollado cada uno de los pasos mencionados.

4. Resultados

El análisis de los datos metabolómicos permite evaluar la estructura, calidad, consistencia y características principales de la información procesada. A continuación, se presentan los resultados obtenidos en cada una de las etapas descritas en la metodología. En el apartado de resultados no se encuentra el código completo, únicamente las partes necesarias para ilustrar la parte del enunciado a la que se está contestando. Para ver el código completo y su compilación consultar el **Anexo 1**.

4.1 Creación del Objeto SummarizedExperiment

El objeto SummarizedExperiment se crea utilizando el dataset cargado en R como `json_data` del cual se extrae la matriz de expresión y los metadatos .

```
# Extraer la matriz de expresión (assay)
expr_data <- json_data$NMR_METABOLITE_DATA$Data
rownames(expr_data) <- expr_data$Metabolite
expr_data <- expr_data[, -1] # Eliminar la columna de nombres de metabolitos
```

```

# Convertir a matriz numérica
expr_matrix <- as.matrix(expr_data)

# Extraer metadatos de muestras (colData)
sample_metadata <- json_data$SUBJECT_SAMPLE_FACTORS$`Additional sample
data`
sample_factors <- json_data$SUBJECT_SAMPLE_FACTORS$Factors

# Unir ambos metadatos de muestras
total_sample_metadata <- cbind(sample_metadata, sample_factors)
rownames(total_sample_metadata) <- total_sample_metadata$SampleID #
Asegurar IDs

# Extraer metadatos de características (rowData)
row_data <- data.frame(Metabolite = rownames(expr_matrix))
rownames(row_data) <- row_data$Metabolite

# Extraer metadatos generales del estudio (metadata)
study_metadata <- list(
  STUDY = json_data$STUDY,
  PROJECT = json_data$PROJECT,
  PUBLICATIONS = json_data$PUBLICATIONS
)

```

Se utiliza la función `SummarizedExperiment()` para crear el objeto tal y como se indica en el código a continuación. Los parámetros pasados fueron:

- assays: Una lista que contenía la matriz de expresión de los metabolitos, la cual fue etiquetada como counts.
- colData: Los metadatos de las muestras (total_sample_metadata).
- rowData: Los metadatos relacionados con las características de las filas, en este caso, los metabolitos (row_data).
- metadata: Los metadatos generales del estudio.

```

# Construir el objeto SummarizedExperiment
se <- SummarizedExperiment(
  assays = list(counts = expr_matrix),
  colData = total_sample_metadata,
  rowData = row_data,
  metadata = study_metadata
)
se

## class: SummarizedExperiment
## dim: 112 112
## metadata(3): STUDY PROJECT PUBLICATIONS
## assays(1): counts
## rownames(112): TPTG TPCH ... H3A2 H4A2
## rowData names(1): Metabolite
## colnames(112): 1 2 ... 111 112
## colData names(18): RAW_FILE_NAME(NMR Exp No.) Age ... Diabetes_Med
Sex

```

El objeto SummarizedExperiment resultante cuenta con unas dimensiones de 112 filas (metabolitos) y 112 columnas (muestras), metadatos de muestras (colData) de 18 columnas que incluyen detalles sobre las muestras, metadatos de características (rowData) con 1 columna con el nombre del metabolito y metadatos del estudio con información sobre el estudio y sus publicaciones.

Llegados a este punto hacemos un pequeño análisis para responder a la pregunta formulada en el enunciado sobre las principales diferencias entre la clase SummarizedExperiment y la clase ExpressionSet. La clase SummarizedExperiment y la clase ExpressionSet son estructuras de datos ampliamente utilizadas en Bioconductor para el manejo de datos ómicos, pero presentan diferencias importantes. ExpressionSet fue diseñado para datos de microarrays, organizando la matriz de expresión en exprs(), mientras que la información de las muestras y de las características se almacena en phenoData() y featureData(), respectivamente. Debido a su estructura específico para microarrays es menos adecuada para otros tipos de datos ómicos. En cambio, SummarizedExperiment es una versión más flexible y generalizada pensada para distintos tipos de experimentos, como RNA-seq, proteómica y metabolómica. Una de sus principales ventajas es que permite almacenar múltiples matrices de datos dentro del mismo objeto (assays()). Además, los metadatos de las muestras (colData()) y de las características (rowData()) se gestionan de forma más clara y estructurada.

4.2 Análisis exploratorio

En este apartado mostraremos el código utilizado pero no la compilación para no sobrepasar la adecuada extensión del informe. Lo primero que se ha realizado es análisis de estructura del objeto:

```
# Verificar la estructura general del objeto SummarizedExperiment
show(se)

# Verificar las dimensiones de la matriz de expresión (assays), los
# metadatos de las muestras (colData) y las características de las filas
# (rowData)
dim(assays(se)$counts) # Dimensiones de la matriz de expresión
dim(colData(se))       # Dimensiones de los metadatos de las muestras
dim(rowData(se))       # Dimensiones de los metadatos de las
# características
# Verificar las primeras filas de los metadatos de las muestras y
# características
head(colData(se))      # Primeras filas de los metadatos de las
# muestras
head(rowData(se))      # Primeras filas de los metadatos de las
# características
```

La inspección inicial muestra que el objeto tiene 112 filas y 112 columnas, lo que coincide con el número esperado de metabolitos y muestras. Además, se identifican tres elementos en la metadata, una matriz de expresión en assays y 18 variables descriptivas en colData. La matriz de expresión tiene dimensiones de 112x112, confirmando la correspondencia entre metabolitos y muestras, mientras que colData también presenta 112 filas y 18 columnas, lo que valida la carga correcta de los metadatos.

Al explorar los primeros registros de colData, se encontraron variables como edad, peso, altura y presión arterial. Por otro lado, rowData, que contiene los nombres de los metabolitos, está bien estructurada con 112 filas y 1 columna donde se almacena el nombre de los metabolitos. La coincidencia entre los nombres de las filas de la matriz de expresión y los de rowData confirma que los datos están alineados correctamente y se puede proseguir con el análisis.

El siguiente paso es el **análisis de la matriz de expresión**:

```
# Estadísticas descriptivas de la matriz de expresión
summary(assays(se)$counts) # Resumen de estadísticas

apply(assays(se)$counts, 1, sd) # Desviación estándar por metabolito

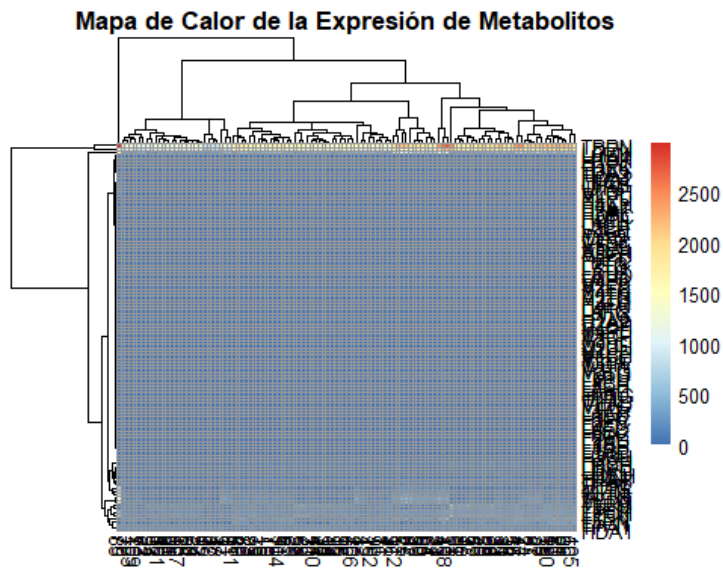
average_per_analyte <- apply(assays(se)$counts, 1, mean) # Media por
analito (por fila)
average_per_analyte # Esto muestra la media de cada metabolito a
través de las muestras
```

En este punto observamos que la distribución de los datos muestra una amplia variabilidad entre las muestras, con valores mínimos cercanos a cero y máximos que superan los 2000, lo que indica una asimetría y la presencia de valores atípicos altos, ya que las medias y medianas son significativamente menores que los máximos. Esto podría reflejar la activación diferencial de genes relacionados con procesos inflamatorios, metabolismo lipídico o estrés oxidativo, factores clave en la salud cardiovascular que dan lugar en distinta medida a los diferentes metabolitos.

Además, hay una gran dispersión de valores, con algunas muestras presentando medias altas (como 86.441 y 91.39) y otras más bajas (como 32.656 y 37.885). La alta variabilidad entre muestras podría indicar la presencia de distintos perfiles de expresión asociados con diferentes estados metabólicos o de envejecimiento cardiovascular en la población estudiada. Al comparar las muestras se observan valores mínimos de 0 en varias muestras mientras que, algunas muestran valores máximos excepcionalmente altos. Esto podría deberse a individuos en un estado especialmente patológico ya que los valores que presenta la gran mayoría de la población suelen ser lo normal, es decir, lo que entendemos por valores de referencia. Sin embargo aquellos valores muy extremos que presentan muy pocos individuos suelen estar asociados a patologías.

Esta misma distribución de los datos se puede ver en el mapa de calor:

```
# Crear un mapa de calor de la matriz de expresión
pheatmap(assays(se)$counts, cluster_rows = TRUE, cluster_cols = TRUE,
          main = "Mapa de Calor de la Expresión de Metabolitos")
```



También se analizan los metadatos en muestras y metabolitos:

#Análisis de Los metadatos de Las muestras

Verificar Los nombres de Las muestras
`colnames(assays(se)$counts)`

Inspeccionar Los metadatos de Las muestras
`head(colData(se))`

Verificar La distribución de algunas variables en Los metadatos de Las muestras

`table(colData(se)$Age)` *# Distribución de La edad*

`table(colData(se)$Sex)` *# Distribución por sexo*

La distribución de las variables en los metadatos de las muestras muestra una amplia diversidad en cuanto a la edad, con un rango que abarca desde los 21 hasta los 78 años, lo que sugiere una muestra representativa de diferentes grupos etarios, aunque con una baja frecuencia por cada valor específico de edad. Esta distribución refleja una muestra continua y heterogénea con una posible falta de equilibrio entre los distintos grupos de edades ya que hay una representación baja de algunas edades intermedias. En cuanto al sexo, se observa un ligero predominio de mujeres (60) sobre hombres (52), lo que refleja una ligera desproporción, aunque no demasiado significativa.

#Análisis de Los metadatos de Las muestras

Verificar Los primeros metabolitos y sus metadatos
`head(rowData(se))`

Verificar La distribución de Los nombres de Los metabolitos
`table(rowData(se)$Metabolite)`

En cuanto al análisis de los metadatos de los metabolitos, la distribución de metabolitos muestra una lista extensa de compuestos, cada uno con una única observación en los datos. La

lista incluye diversas clases de metabolitos relacionados con las subclases de lipoproteínas y marcadores de salud cardiovascular.

5. Discusión

En este estudio, hemos utilizado el objeto de clase SummarizedExperiment para integrar y analizar un dataset de metabolómica relacionado con la salud cardiovascular en el envejecimiento, específicamente en lo que respecta a las subclases de lipoproteínas. Este enfoque ha demostrado ser eficaz para manejar grandes volúmenes de datos complejos, organizando tanto la expresión de metabolitos como los metadatos de las muestras y las características experimentales. Sin embargo, es importante reconocer algunas limitaciones inherentes al estudio y reflexionar sobre las implicaciones biológicas de estas limitaciones.

Una de las principales limitaciones del estudio radica en la naturaleza del dataset utilizado. Aunque el conjunto de datos proviene de un repositorio reconocido y se centra en una población humana, la distribución de algunas variables clave, como la edad y el sexo, muestra una muestra relativamente heterogénea, pero con ciertas desproporciones. Por ejemplo, hay una baja frecuencia de ciertos grupos de edad, lo que podría generar un sesgo en la interpretación de los resultados. Esto podría ser relevante en las subclases de proteínas y los marcadores utilizados para medir el riesgo cardiovascular que varíen con la edad o el sexo, especialmente la edad. Sería conveniente realizar un estudio en una población mayor con una representación equilibrada de todos los grupos de edad.

Además, aunque el análisis exploratorio de los datos revela una gran variabilidad en la expresión de los metabolitos, algunos de los valores atípicos encontrados podrían no ser completamente representativos de la población general y podrían estar vinculados a estados patológicos propios del paciente que afecten a la salud cardiovascular de forma secundaria. Es muy importante tener esto en cuenta a la hora de intentar correlacionar la expresión de metabolitos con la salud cardiovascular en el envejecimiento, ya que las condiciones patológicas preexistentes o coexistentes pueden alterar significativamente los perfiles metabólicos.

Desde una perspectiva metodológica, utilizar el objeto SummarizedExperiment para gestionar los datos ha sido una opción acertada por su diseño flexible y adecuado para manejar múltiples tipos de datos ómicos. Sin embargo, el trabajo realizado se ha centrado principalmente en la organización de los datos y la realización de análisis exploratorios, que es la fase inicial de una investigación, pero para llegar realmente a extraer conclusiones útiles convendría hacer análisis más específicos, por ejemplo, centrándonos en metabolitos concretos.

En cuanto al contexto biológico del estudio, los hallazgos preliminares sugieren que las subclases de lipoproteínas pueden estar fuertemente relacionadas con la salud cardiovascular, especialmente en el envejecimiento. Sin embargo, la alta variabilidad entre los metabolitos y las muestras también refleja la complejidad de los mecanismos biológicos que las regulan.

6. Conclusiones

La elección del dataset de metabolómica del estudio ST003791, que explora las asociaciones entre las subclases de lipoproteínas y la salud cardiovascular en el envejecimiento, se ha mostrado apropiada ya que ha permitido el desarrollo de la actividad.

La creación del objeto SummarizedExperiment ha permitido integrar de manera efectiva la matriz de expresión de metabolitos con los metadatos asociados a las muestras y las características experimentales. Esta integración ha facilitado un análisis exploratorio organizado y completo, asegurando la correcta alineación de los datos para su posterior análisis.

El análisis exploratorio ha permitido evaluar la calidad y estructura de los datos, revelando una gran variabilidad en los metabolitos y mostrando la presencia de valores atípicos que podrían estar asociados a condiciones patológicas. Sin embargo, hay limitaciones como la desproporción en la distribución de las variables edad y sexo que podría influir en los resultados del estudio.

Los resultados obtenidos sugieren que las subclases de lipoproteínas podrían estar relacionadas con la salud cardiovascular en el envejecimiento, aunque la alta variabilidad observada en los datos resalta la complejidad de los mecanismos biológicos involucrados.

Anexo 1

```
# Instalar todos Los paquetes necesarios
options(repos = c(CRAN = "https://cran.rediris.es/")) # Definir el
CRAN
install.packages("readxl", lib = "C:/Users/mireg/AppData/Local/R/win-
library/4.4")
## package 'readxl' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
## C:\Users\mireg\AppData\Local\Temp\RtmpOGT72T\downloaded_packages
# Instalar Los paquetes necesarios
install.packages("BiocManager")
## Installing package into 'C:/Users/mireg/AppData/Local/R/win-
library/4.4'
## (as 'lib' is unspecified)
## package 'BiocManager' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
## C:\Users\mireg\AppData\Local\Temp\RtmpOGT72T\downloaded_packages
BiocManager::install("goProfiles",
                      dep=TRUE)
## 'getOption("repos")' replaces Bioconductor standard repositories,
see
## 'help("repositories", package = "BiocManager")' for details.
## Replacement repositories:
## CRAN: https://cran.rediris.es/
## Bioconductor version 3.20 (BiocManager 1.30.25), R 4.4.1 (2024-06-
14 ucrt)
## Warning: package(s) not installed when version(s) same as or
greater than current; use
## `force = TRUE` to re-install: 'goProfiles'
## Installation paths not writeable, unable to update packages
## path: C:/Program Files/R/R-4.4.1/library
## packages:
## boot, class, cluster, foreign, KernSmooth, MASS, Matrix, nlme,
nnet, rpart,
## spatial, survival
## Old packages: 'abind', 'bit', 'bit64', 'broom', 'bslib', 'chron',
'classInt',
## 'cli', 'colorspace', 'commonmark', 'compareGroups', 'cpp11',
'crayon',
## 'curl', 'data.table', 'digest', 'evaluate', 'faraway',
'flextable',
## 'fontawesome', 'fs', 'gdtools', 'gert', 'ggstats', 'git2r',
'glue', 'gtable',
## 'httr2', 'jsonlite', 'knitr', 'labelled', 'later', 'lme4',
'lubridate',
## 'MatrixModels', 'metR', 'mice', 'mime', 'minqa', 'mvtnorm',
'nloptr',
## 'officer', 'openssl', 'pbkrtest', 'pillar', 'pkgbuild',
'processx',
## 'promises', 'ps', 'purrr', 'quantreg', 'R6', 'Rcpp',
'RcppArmadillo',
```

```

## 'RcppEigen', 'RCurl', 'rlang', 'rmarkdown', 'Rmpfr',
'rstudioapi',
## 'sessioninfo', 'sf', 'shiny', 'sp', 'SparseM', 'stringi',
'SuppDists',
## 'systemfonts', 'testthat', 'textshaping', 'tzdb', 'units',
'usethis',
## 'waldo', 'withr', 'writexl', 'xfun', 'XML', 'xml2', 'yaml',
'zip', 'zoo'
BiocManager::install("SummarizedExperiment")
## 'getOption("repos")' replaces Bioconductor standard repositories,
see
## 'help("repositories", package = "BiocManager")' for details.
## Replacement repositories:
## CRAN: https://cran.rediris.es/
## Bioconductor version 3.20 (BiocManager 1.30.25), R 4.4.1 (2024-06-
14 ucrt)
## Warning: package(s) not installed when version(s) same as or
greater than current; use
## `force = TRUE` to re-install: 'SummarizedExperiment'
## Installation paths not writeable, unable to update packages
## path: C:/Program Files/R/R-4.4.1/library
## packages:
## boot, class, cluster, foreign, KernSmooth, MASS, Matrix, nlme,
nnet, rpart,
## spatial, survival
## Old packages: 'abind', 'bit', 'bit64', 'broom', 'bslib', 'chron',
'classInt',
## 'cli', 'colorspace', 'commonmark', 'compareGroups', 'cpp11',
'crayon',
## 'curl', 'data.table', 'digest', 'evaluate', 'faraway',
'flextable',
## 'fontawesome', 'fs', 'gdtools', 'gert', 'ggstats', 'git2r',
'glue', 'gtable',
## 'httr2', 'jsonlite', 'knitr', 'labelled', 'later', 'lme4',
'lubridate',
## 'MatrixModels', 'metR', 'mice', 'mime', 'minqa', 'mvtnorm',
'nloptr',
## 'officer', 'openssl', 'pbkrtest', 'pillar', 'pkgbuild',
'processx',
## 'promises', 'ps', 'purrr', 'quantreg', 'R6', 'Rcpp',
'RcppArmadillo',
## 'RcppEigen', 'RCurl', 'rlang', 'rmarkdown', 'Rmpfr',
'rstudioapi',
## 'sessioninfo', 'sf', 'shiny', 'sp', 'SparseM', 'stringi',
'SuppDists',
## 'systemfonts', 'testthat', 'textshaping', 'tzdb', 'units',
'usethis',
## 'waldo', 'withr', 'writexl', 'xfun', 'XML', 'xml2', 'yaml',
'zip', 'zoo'
install.packages("jsonlite", dependencies = TRUE)
## Installing package into 'C:/Users/mireg/AppData/Local/R/win-
library/4.4'
## (as 'lib' is unspecified)
## package 'jsonlite' successfully unpacked and MD5 sums checked

```

```
## Warning: cannot remove prior installation of package 'jsonlite'
## Warning in file.copy(savedcopy, lib, recursive = TRUE): problema al
copiar
## C:\Users\mireg\AppData\Local\R\win-
library\4.4\00LOCK\jsonlite\libs\x64\jsonlite.dll
## a
## C:\Users\mireg\AppData\Local\R\win-
library\4.4\jsonlite\libs\x64\jsonlite.dll:
## Permission denied
## Warning: restored 'jsonlite'
##
## The downloaded binary packages are in
## C:\Users\mireg\AppData\Local\Temp\RtmpOGT72T\downloaded_packages
install.packages("git2r")
## Installing package into 'C:/Users/mireg/AppData/Local/R/win-
library/4.4'
## (as 'lib' is unspecified)
## package 'git2r' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
## C:\Users\mireg\AppData\Local\Temp\RtmpOGT72T\downloaded_packages
```

1. **Seleccionad y descargad un dataset de metabolómica, que podéis obtener de metabolomicsWorkbench o de este repositorio de GitHub.**

```
# Cargar La Librerías
library(jsonlite)
library(SummarizedExperiment)
## Cargando paquete requerido: MatrixGenerics
## Warning: package 'MatrixGenerics' was built under R version 4.4.2
## Cargando paquete requerido: matrixStats
## Warning: package 'matrixStats' was built under R version 4.4.3
##
## Adjuntando el paquete: 'MatrixGenerics'
## The following objects are masked from 'package:matrixStats':
##
##      colAlls, colAnyNAs, colAnys, colAvgsPerRowSet, colCollapse,
##      colCounts, colCummaxs, colCummins, colCumprods, colCumsums,
##      colDiffs, colIQRDiffs, colIQRs, colLogSumExps, colMadDiffs,
##      colMads, colMaxs, colMeans2, colMedians, colMins,
colOrderStats,
##      colProds, colQuantiles, colRanges, colRanks, colSdDiffs,
colSds,
##      colSums2, colTabulates, colVarDiffs, colVars, colWeightedMads,
##      colWeightedMeans, colWeightedMedians, colWeightedSds,
##      colWeightedVars, rowAlls, rowAnyNAs, rowAnys, rowAvgsPerColSet,
##      rowCollapse, rowCounts, rowCummaxs, rowCummins, rowCumprods,
##      rowCumsums, rowDiffs, rowIQRDiffs, rowIQRs, rowLogSumExps,
##      rowMadDiffs, rowMads, rowMaxs, rowMeans2, rowMedians, rowMins,
##      rowOrderStats, rowProds, rowQuantiles, rowRanges, rowRanks,
##      rowSdDiffs, rowSds, rowSums2, rowTabulates, rowVarDiffs,
rowVars,
##      rowWeightedMads, rowWeightedMeans, rowWeightedMedians,
##      rowWeightedSds, rowWeightedVars
```

```

## Cargando paquete requerido: GenomicRanges
## Cargando paquete requerido: stats4
## Cargando paquete requerido: BiocGenerics
##
## Adjuntando el paquete: 'BiocGenerics'
## The following objects are masked from 'package:stats':
##
##     IQR, mad, sd, var, xtabs
## The following objects are masked from 'package:base':
##
##     anyDuplicated, aperm, append, as.data.frame, basename, cbind,
##     colnames, dirname, do.call, duplicated, eval, evalq, Filter,
##     Find,
##     get, grep, grepl, intersect, is.unsorted, lapply, Map, mapply,
##     match, mget, order, paste, pmax, pmax.int, pmin, pmin.int,
##     Position, rank, rbind, Reduce, rownames, sapply, saveRDS,
##     setdiff,
##     table, tapply, union, unique, unsplit, which.max, which.min
## Cargando paquete requerido: S4Vectors
##
## Adjuntando el paquete: 'S4Vectors'
## The following object is masked from 'package:utils':
##
##     findMatches
## The following objects are masked from 'package:base':
##
##     expand.grid, I, unname
## Cargando paquete requerido: IRanges
## Warning: package 'IRanges' was built under R version 4.4.2
##
## Adjuntando el paquete: 'IRanges'
## The following object is masked from 'package:grDevices':
##
##     windows
## Cargando paquete requerido: GenomeInfoDb
## Warning: package 'GenomeInfoDb' was built under R version 4.4.2
## Cargando paquete requerido: Biobase
## Welcome to Bioconductor
##
##     Vignettes contain introductory material; view with
##     'browseVignettes()'. To cite Bioconductor, see
##     'citation("Biobase")', and for packages 'citation("pkgname")'.
##
## Adjuntando el paquete: 'Biobase'
## The following object is masked from 'package:MatrixGenerics':
##
##     rowMedians
## The following objects are masked from 'package:matrixStats':
##
##     anyMissing, rowMedians
library(git2r)
## Warning: package 'git2r' was built under R version 4.4.3
##
## Adjuntando el paquete: 'git2r'

```

```

## The following objects are masked from
'package:SummarizedExperiment':
##
##      as.data.frame, merge
## The following objects are masked from 'package:Biobase':
##
##      content, notes
## The following objects are masked from 'package:GenomicRanges':
##
##      as.data.frame, merge
## The following objects are masked from 'package:GenomeInfoDb':
##
##      as.data.frame, merge
## The following objects are masked from 'package:IRanges':
##
##      as.data.frame, diff, merge
## The following objects are masked from 'package:S4Vectors':
##
##      as.data.frame, head, merge
## The following object is masked from 'package:BiocGenerics':
##
##      as.data.frame
# Leer el archivo JSON
json_data <- fromJSON("C:\\Users\\mireg\\Desktop\\Master
Bioinformatica y Bioestadística\\Análisis de datos
ómicos\\PAC1\\ST003791_AN006232.json")

# Estudiar el json_data
str(json_data)
## List of 11
## $ METABOLOMICS WORKBENCH:List of 5
## ..$ STUDY_ID : chr "ST003791"
## ..$ ANALYSIS_ID: chr "AN006232"
## ..$ PROJECT_ID : chr "PR002367"
## ..$ VERSION : chr "1"
## ..$ CREATED_ON : chr "March 12, 2025, 6:25 pm"
## $ PROJECT :List of 10
## ..$ PROJECT_TITLE : chr "Associations between lipoprotein
subclasses and cardiovascular health in ageing"
## ..$ PROJECT_SUMMARY: chr "Abnormal lipoprotein profiles have been
frequently described in association with atherogenic cardiovascular
dis"| __truncated__
## ..$ INSTITUTE : chr "Nanyang Technological University"
## ..$ DEPARTMENT : chr "Lee Kong Chian School of Medicine"
## ..$ LABORATORY : chr "Singapore Phenome Centre"
## ..$ LAST_NAME : chr "Yu King Hing"
## ..$ FIRST_NAME : chr "Nathaphon"
## ..$ ADDRESS : chr "59 Nanyang Drive Singapore"
## ..$ EMAIL : chr "joel.yukh@ntu.edu.sg"
## ..$ PHONE : chr "83199405"
## $ STUDY :List of 11
## ..$ STUDY_TITLE : chr "Associations between lipoprotein
subclasses and cardiovascular health in ageing"
## ..$ STUDY_SUMMARY : chr "Abnormal lipoprotein profiles have been

```

```

frequently described in association with atherogenic cardiovascular
dis"| __truncated__
## ..$ INSTITUTE      : chr "Nanyang Technological University"
## ..$ DEPARTMENT     : chr "Lee Kong Chian School of Medicine"
## ..$ LABORATORY     : chr "Singapore Phenome Centre"
## ..$ LAST_NAME      : chr "Yu King Hing"
## ..$ FIRST_NAME     : chr "Nathaphon"
## ..$ ADDRESS        : chr "59 Nanyang Drive"
## ..$ EMAIL          : chr "joel.yukh@ntu.edu.sg"
## ..$ PHONE          : chr "83199405"
## ..$ TOTAL_SUBJECTS: chr "474"
## $ SUBJECT          :List of 5
## ..$ SUBJECT_TYPE   : chr "Human"
## ..$ SUBJECT_SPECIES: chr "Homo sapiens"
## ..$ TAXONOMY_ID    : chr "9606"
## ..$ AGE_OR_AGE_RANGE: chr "21-84"
## ..$ GENDER         : chr "Male and female"
## $ SUBJECT_SAMPLE_FACTORS:'data.frame': 474 obs. of 4 variables:
## ..$ Subject ID      : chr [1:474] "-" "-" "-" "-" ...
## ..$ Sample ID       : chr [1:474] "1" "2" "3" "4" ...
## ..$ Factors         : 'data.frame': 474 obs. of 9
variables:
## .. ..$ Sample source : chr [1:474] "Blood" "Blood" "Blood"
"Blood" ...
## .. ..$ E/A < 0.8     : chr [1:474] "0" "0" "1" "0" ...
## .. ..$ Hypertension   : chr [1:474] "0" "0" "1" "1" ...
## .. ..$ Dyslipidemia   : chr [1:474] "0" "0" "1" "0" ...
## .. ..$ Diabetes mellitus: chr [1:474] "0" "0" "0" "0" ...
## .. ..$ Hypertension_Med : chr [1:474] "0" "0" "1" "1" ...
## .. ..$ Cholesterol_Med : chr [1:474] "0" "0" "1" "0" ...
## .. ..$ Diabetes_Med    : chr [1:474] "0" "0" "0" "0" ...
## .. ..$ Sex            : chr [1:474] "F" "M" "M" "F" ...
## ..$ Additional sample data:'data.frame': 474 obs. of 9
variables:
## .. ..$ RAW_FILE_NAME(NMR Exp No.) : chr [1:474]
"Box1.zip/100/fid" "Box1.zip/110/fid" "Box1.zip/120/fid"
"Box1.zip/130/fid" ...
## .. ..$ Age           : chr [1:474]
"25.2247" "74.8986" "78.4548" "78.4438" ...
## .. ..$ Weight (kg)    : chr [1:474]
"53.9000" "62.8000" "67.9000" "64.6000" ...
## .. ..$ Height (cm)    : chr [1:474]
"164.0000" "164.0000" "167.0000" "158.0000" ...
## .. ..$ Pulse         : chr [1:474]
"80" "88" "89" "64" ...
## .. ..$ WaistCircumference (cm) : chr [1:474]
"76.0" "83.0" "92.0" "84.0" ...
## .. ..$ SBP           : chr [1:474]
"110" "146" "168" "158" ...
## .. ..$ DBP           : chr [1:474]
"67" "77" "65" "67" ...
## .. ..$ Smoking (Never 0, Current 1, Past 2): chr [1:474] "0"
"0" "2" "0" ...
## $ COLLECTION        :List of 3

```

```

## ..$ COLLECTION_SUMMARY: chr "The subjects were recruited from the
Cardiac Ageing Study (CAS) 34, a prospective study initiated in 2014
exami"| __truncated__
## ..$ SAMPLE_TYPE : chr "Blood (serum)"
## ..$ STORAGE_CONDITIONS: chr "-80°C"
## $ TREATMENT :List of 1
## ..$ TREATMENT_SUMMARY: chr "Samples were not treated. Factors (0
= negative, 1 = positive): If E/A < 0.8 = 1, the subject's
echocardiogr"| __truncated__
## $ SAMPLEPREP :List of 2
## ..$ SAMPLEPREP_SUMMARY : chr "350 µL of each blood
serum sample were mixed with 350 µL of phosphate buffer for plasma
(Bruker Biospin, German"| __truncated__
## ..$ PROCESSING_STORAGE_CONDITIONS: chr "-80°C"
## $ ANALYSIS :List of 1
## ..$ LABORATORY_NAME: chr "Singapore Phenome Centre"
## $ NM :List of 5
## ..$ INSTRUMENT_NAME : chr "Bruker 600 MHz AVANCE III"
## ..$ INSTRUMENT_TYPE : chr "FT-NMR"
## ..$ NMR_EXPERIMENT_TYPE : chr "1D-1H"
## ..$ SPECTROMETER_FREQUENCY: chr "600 MHz"
## ..$ TEMPERATURE : chr "4"
## $ NMR_METABOLITE_DATA :List of 3
## ..$ Units : chr "mg/dL"
## ..$ Data :'data.frame': 112 obs. of 113 variables:
## .. ..$ Metabolite: chr [1:112] "TPTG" "TPCH" "LDCH" "HDCH" ...
## .. ..$ 1 : chr [1:112] "129.76" "255.19" "154.07" "68.03"
...
## .. ..$ 2 : chr [1:112] "285.98" "268.05" "130.39" "75.47"
...
## .. ..$ 3 : chr [1:112] "307.48" "163.48" "61.89" "49.13"
...
## .. ..$ 4 : chr [1:112] "213.37" "185.11" "86.35" "61.18"
...
## .. ..$ 5 : chr [1:112] "418.47" "246.79" "108.49" "49.66"
...
## .. ..$ 6 : chr [1:112] "67.37" "148.92" "74.52" "62.35"
...
## .. ..$ 7 : chr [1:112] "170.28" "177.55" "74.96" "75.57"
...
## .. ..$ 8 : chr [1:112] "188.77" "201.9" "102.31" "66.64"
...
## .. ..$ 9 : chr [1:112] "135.09" "178.7" "99.02" "57.86"
...
## .. ..$ 10 : chr [1:112] "172.07" "234.51" "129.4" "70.12"
...
## .. ..$ 11 : chr [1:112] "108.5" "181.02" "100.51" "59.32"
...
## .. ..$ 12 : chr [1:112] "199.17" "167.31" "80.88" "46.97"
...
## .. ..$ 13 : chr [1:112] "205.64" "200.81" "99.18" "56.99"
...
## .. ..$ 14 : chr [1:112] "84.02" "207.35" "124.52" "64.15"
...

```



```

## .. ..$ 15      : chr [1:112] "105.41" "224.18" "136.82" "64.34"
...
## .. ..$ 16      : chr [1:112] "113.59" "299.56" "188.76" "64.01"
...
## .. ..$ 17      : chr [1:112] "300.49" "234.47" "42.55" "63.8"
...
## .. ..$ 18      : chr [1:112] "198.93" "186.02" "97.25" "53.68"
...
## .. ..$ 19      : chr [1:112] "90.18" "113.92" "50.17" "48.22"
...
## .. ..$ 20      : chr [1:112] "378.22" "231.64" "89.29" "67.32"
...
## .. ..$ 21      : chr [1:112] "162.98" "300.76" "177.17" "77.33"
...
## .. ..$ 22      : chr [1:112] "73.95" "174.39" "93.6" "64.1" ...
## .. ..$ 23      : chr [1:112] "68.5" "353.83" "230.37" "85.71"
...
## .. ..$ 24      : chr [1:112] "57.82" "186.83" "106.11" "68.79"
...
## .. ..$ 25      : chr [1:112] "472.63" "273.24" "103.47" "62.91"
...
## .. ..$ 26      : chr [1:112] "93.17" "200.58" "100.6" "81.21"
...
## .. ..$ 27      : chr [1:112] "294.05" "257.66" "128.46" "47.66"
...
## .. ..$ 28      : chr [1:112] "84.35" "288.69" "160.71" "110.38"
...
## .. ..$ 29      : chr [1:112] "82.49" "222.98" "119.28" "80.77"
...
## .. ..$ 30      : chr [1:112] "58.05" "217.91" "123.03" "85.03"
...
## .. ..$ 31      : chr [1:112] "120.09" "264.82" "152.23" "70.29"
...
## .. ..$ 32      : chr [1:112] "149.35" "251.49" "142.35" "80.7"
...
## .. ..$ 33      : chr [1:112] "418.89" "279.23" "127.65" "54.62"
...
## .. ..$ 34      : chr [1:112] "180.4" "282.34" "141.65" "92.72"
...
## .. ..$ 35      : chr [1:112] "282.81" "269.39" "142.21" "62.5"
...
## .. ..$ 36      : chr [1:112] "92.33" "251.95" "153.68" "74.86"
...
## .. ..$ 37      : chr [1:112] "57.79" "144.99" "61.28" "70.06"
...
## .. ..$ 38      : chr [1:112] "71.29" "228.95" "121" "89.76" ...
## .. ..$ 39      : chr [1:112] "192.23" "287.68" "152.41" "78.93"
...
## .. ..$ 40      : chr [1:112] "103.3" "271.22" "158.62" "74.83"
...
## .. ..$ 41      : chr [1:112] "154.24" "212.48" "111.79" "68.28"
...
## .. ..$ 42      : chr [1:112] "300.85" "224.54" "104.83" "47.93"
...

```

##\$ 43	: chr [1:112]	"350.23"	"315.92"	"158.77"	"64.55"
...						
##\$ 44	: chr [1:112]	"37.87"	"220.82"	"122.56"	"98.14"
...						
##\$ 45	: chr [1:112]	"70.33"	"215.71"	"115.77"	"85.42"
...						
##\$ 46	: chr [1:112]	"63.27"	"187.63"	"94.44"	"72.47"
...						
##\$ 47	: chr [1:112]	"201.99"	"306.56"	"175.04"	"61.21"
...						
##\$ 48	: chr [1:112]	"45.29"	"154.13"	"75.34"	"75.92"
...						
##\$ 49	: chr [1:112]	"165.61"	"236.94"	"127.17"	"58.27"
...						
##\$ 50	: chr [1:112]	"161.5"	"234.67"	"129.6"	"62.64"
...						
##\$ 51	: chr [1:112]	"126.77"	"188.85"	"111.62"	"46.12"
...						
##\$ 52	: chr [1:112]	"59.02"	"269.85"	"164.81"	"87.61"
...						
##\$ 53	: chr [1:112]	"79.1"	"241.88"	"136.11"	"80.47"
...						
##\$ 54	: chr [1:112]	"140.85"	"264.34"	"152.33"	"71.3"
...						
##\$ 55	: chr [1:112]	"89.13"	"269.88"	"163.48"	"76.35"
...						
##\$ 56	: chr [1:112]	"101.34"	"223.3"	"116.72"	"84.91"
...						
##\$ 57	: chr [1:112]	"337.58"	"219.6"	"106.83"	"48.54"
...						
##\$ 58	: chr [1:112]	"97.24"	"213.87"	"107.13"	"81.34"
...						
##\$ 59	: chr [1:112]	"191.19"	"298.66"	"170.42"	"79.33"
...						
##\$ 60	: chr [1:112]	"318.28"	"239.7"	"106.4"	"51.87"
...						
##\$ 61	: chr [1:112]	"72.08"	"197.02"	"99.36"	"80.19"
...						
##\$ 62	: chr [1:112]	"68.23"	"188.76"	"120.3"	"51.11"
...						
##\$ 63	: chr [1:112]	"63.32"	"187.98"	"87.66"	"85.07"
...						
##\$ 64	: chr [1:112]	"288.94"	"228.05"	"93.11"	"64.96"
...						
##\$ 65	: chr [1:112]	"174.41"	"204.12"	"103.78"	"55.79"
...						
##\$ 66	: chr [1:112]	"164.55"	"286.51"	"169.22"	"74.25"
...						
##\$ 67	: chr [1:112]	"107.71"	"249.32"	"144.75"	"71.04"
...						
##\$ 68	: chr [1:112]	"85.72"	"314.82"	"184.72"	"102.03"
...						
##\$ 69	: chr [1:112]	"62.05"	"229.09"	"131.5"	"80.4"
...						

##\$ 70	: chr [1:112]	"167.01"	"185.01"	"89.91"	"57"	...
##\$ 71	: chr [1:112]	"312.82"	"221.39"	"103.6"	"58.36"	
...							
##\$ 72	: chr [1:112]	"133.55"	"229.64"	"106.05"	"94.46"	
...							
##\$ 73	: chr [1:112]	"169.7"	"260.59"	"119.68"	"98.69"	
...							
##\$ 74	: chr [1:112]	"244.24"	"250.86"	"127.55"	"60.16"	
...							
##\$ 75	: chr [1:112]	"139.51"	"296.89"	"169.71"	"92.01"	
...							
##\$ 76	: chr [1:112]	"136.33"	"210.02"	"111.94"	"57.88"	
...							
##\$ 77	: chr [1:112]	"131.33"	"205.12"	"92.43"	"78.43"	
...							
##\$ 78	: chr [1:112]	"203.28"	"218.29"	"115.62"	"71.67"	
...							
##\$ 79	: chr [1:112]	"227.36"	"205.69"	"100.25"	"49.1"	
...							
##\$ 80	: chr [1:112]	"224.29"	"297.62"	"167.22"	"71.08"	
...							
##\$ 81	: chr [1:112]	"228.2"	"160.89"	"55.59"	"55.87"	
...							
##\$ 82	: chr [1:112]	"431.98"	"266.62"	"106.13"	"71.56"	
...							
##\$ 83	: chr [1:112]	"1217.79"	"403.74"	"28.76"	"89.22"	
...							
##\$ 84	: chr [1:112]	"88.84"	"205.31"	"106.13"	"77.05"	
...							
##\$ 85	: chr [1:112]	"194.37"	"212.72"	"118.3"	"50.53"	
...							
##\$ 86	: chr [1:112]	"104.55"	"225.39"	"125.93"	"65.85"	
...							
##\$ 87	: chr [1:112]	"276.9"	"257.39"	"133.05"	"67.84"	
...							
##\$ 88	: chr [1:112]	"81.88"	"212.73"	"103.73"	"85.04"	
...							
##\$ 89	: chr [1:112]	"82.17"	"183.77"	"91.47"	"75.05"	
...							
##\$ 90	: chr [1:112]	"88.99"	"272.3"	"157.58"	"84.02"	
...							
##\$ 91	: chr [1:112]	"98.96"	"259.36"	"150.14"	"86.5"	
...							
##\$ 92	: chr [1:112]	"75.78"	"199.44"	"108.5"	"77.12"	
...							
##\$ 93	: chr [1:112]	"154.08"	"240.22"	"107.49"	"70.38"	
...							
##\$ 94	: chr [1:112]	"81.13"	"180.51"	"91.9"	"63.32"	
...							
##\$ 95	: chr [1:112]	"347.06"	"227.53"	"89.81"	"73.3"	
...							
##\$ 96	: chr [1:112]	"322.17"	"236.08"	"116.62"	"54.59"	
...							
##\$ 97	: chr [1:112]	"78.68"	"165.14"	"93.59"	"57.49"	

```

...
## .. ..$ 98 : chr [1:112] "59.99" "204.88" "115.62" "82.43"
...
## .. .. [list output truncated]
## ..$ Metabolites:'data.frame': 112 obs. of 2 variables:
## .. ..$ Metabolite: chr [1:112] "TPTG" "TPCH" "LDCH" "HDCH" ...
## .. ..$ Full Name : chr [1:112] "Triglyceride" "Cholesterol" "LDL-
Cholesterol" "HDL-Cholesterol" ...
names(json_data)
## [1] "METABOLOMICS WORKBENCH" "PROJECT" "STUDY"

## [4] "SUBJECT" "SUBJECT_SAMPLE_FACTORS"
"COLLECTION"
## [7] "TREATMENT" "SAMPLEPREP" "ANALYSIS"

## [10] "NM" "NMR_METABOLITE_DATA"

```

2 . Cread un objeto de clase SummarizedExperiment que contenga los datos y los metadatos (información acerca del dataset, sus filas y columnas). La clase SummarizedExperiment es una extensión de ExpressionSet, utilizada por muchas aplicaciones y bases de datos (como es el caso de metabolomicsWorkbench). ¿Cuáles son sus principales diferencias con la clase ExpressionSet?

```

# Extraer la matriz de expresión (assay)
expr_data <- json_data$NMR_METABOLITE_DATA$Data
rownames(expr_data) <- expr_data$Metabolite
expr_data <- expr_data[, -1] # Eliminar la columna de nombres de
metabolitos

# Convertir a matriz numérica
expr_matrix <- as.matrix(expr_data)

# Extraer metadatos de muestras (colData)
sample_metadata <- json_data$SUBJECT_SAMPLE_FACTORS$`Additional sample
data`
sample_factors <- json_data$SUBJECT_SAMPLE_FACTORS$Factors

# Unir ambos metadatos de muestras
total_sample_metadata <- cbind(sample_metadata, sample_factors)
rownames(total_sample_metadata) <- total_sample_metadata$SampleID #
Asegurar IDs

# Extraer metadatos de características (rowData)
row_data <- data.frame(Metabolite = rownames(expr_matrix))
rownames(row_data) <- row_data$Metabolite

# Extraer metadatos generales del estudio (metadata)
study_metadata <- list(
  STUDY = json_data$STUDY,
  PROJECT = json_data$PROJECT,
  PUBLICATIONS = json_data$PUBLICATIONS
)

```

```

# Revisar dimensiones ya que hay un error
dim(expr_matrix)
## [1] 112 112
dim(row_data)
## [1] 112 1
dim(total_sample_metadata)
## [1] 474 18
# Comprobamos que las dimensiones de total_sample_data no coinciden
con las de expr_matrix

# Filtrar total_sample_metadata para que coincida con las muestras de
expr_matrix
total_sample_metadata <-
total_sample_metadata[rownames(total_sample_metadata) %in%
colnames(expr_matrix), ] #Genera un vector lógico que indica qué filas
de total_sample_metadata tienen un nombre de muestra que está presente
en expr_matrix y selecciona los datos con true.

# Verificar nueva dimensión
dim(total_sample_metadata)
## [1] 112 18
# Ordenar total_sample_metadata para que coincida con
colnames(expr_matrix)
total_sample_metadata <- total_sample_metadata[colnames(expr_matrix),
]

# Construir el objeto SummarizedExperiment
se <- SummarizedExperiment(
  assays = list(counts = expr_matrix),
  colData = total_sample_metadata,
  rowData = row_data,
  metadata = study_metadata
)
se
## class: SummarizedExperiment
## dim: 112 112
## metadata(3): STUDY PROJECT PUBLICATIONS
## assays(1): counts
## rownames(112): TPTG TPCH ... H3A2 H4A2
## rowData names(1): Metabolite
## colnames(112): 1 2 ... 111 112
## colData names(18): RAW_FILE_NAME(NMR Exp No.) Age ... Diabetes_Med
Sex

```

3. Llevad a cabo un análisis exploratorio que os proporcione una visión general del dataset en la línea de lo que hemos visto en las actividades de este reto.

Análisis de la estructura:

```

# Verificar la estructura general del objeto SummarizedExperiment
show(se)
## class: SummarizedExperiment
## dim: 112 112

```

```
## metadata(3): STUDY PROJECT PUBLICATIONS
## assays(1): counts
## rownames(112): TPTG TPCB ... H3A2 H4A2
## rowData names(1): Metabolite
## colnames(112): 1 2 ... 111 112
## colData names(18): RAW_FILE_NAME(NMR Exp No.) Age ... Diabetes_Med
Sex
# Verificar Las dimensiones de La matriz de expresión (assays), Los
metadatos de Las muestras (colData) y Las características de Las filas
(rowData)
dim(assays(se)$counts) # Dimensiones de La matriz de expresión
## [1] 112 112
dim(colData(se))      # Dimensiones de Los metadatos de Las muestras
## [1] 112 18
dim(rowData(se))      # Dimensiones de Los metadatos de Las
características
## [1] 112 1
# Verificar Las primeras filas de Los metadatos de Las muestras y
características
head(colData(se))     # Primeras filas de Los metadatos de Las
muestras
## DataFrame with 6 rows and 18 columns
## RAW_FILE_NAME(NMR Exp No.) Age Weight (kg) Height (cm)
Pulse
##          <character> <character> <character> <character>
<character>
## 1          Box1.zip/100/fid      25.2247      53.9000      164.0000
80
## 2          Box1.zip/110/fid      74.8986      62.8000      164.0000
88
## 3          Box1.zip/120/fid      78.4548      67.9000      167.0000
89
## 4          Box1.zip/130/fid      78.4438      64.6000      158.0000
64
## 5          Box1.zip/140/fid      75.6082      66.6000      158.0000
73
## 6          Box1.zip/150/fid      63.8740      73.2000      178.0000
78
## WaistCircumference (cm)      SBP      DBP
##          <character> <character> <character>
## 1          76.0      110      67
## 2          83.0      146      77
## 3          92.0      168      65
## 4          84.0      158      67
## 5          92.0      164      85
## 6          89.0      139      83
## Smoking (Never 0, Current 1, Past 2) Sample source E/A <
0.8
##          <character> <character>
<character>
## 1          0      Blood
0
## 2          0      Blood
0
```

```

## 3      2      Blood
1
## 4      0      Blood
0
## 5      0      Blood
0
## 6      N/A     Blood
0
## Hypertension Dyslipidemia Diabetes mellitus Hypertension_Med
Cholesterol_Med
## <character> <character> <character> <character>
<character>
## 1      0      0      0      0
0
## 2      0      0      0      0
0
## 3      1      1      0      1
1
## 4      1      0      0      1
0
## 5      0      0      0      0
0
## 6      0      1      0      0
1
## Diabetes_Med Sex
## <character> <character>
## 1      0      F
## 2      0      M
## 3      0      M
## 4      0      F
## 5      0      M
## 6      0      M
head(rowData(se))      # Primeras filas de los metadatos de las
características
## DataFrame with 6 rows and 1 column
## Metabolite
## <character>
## TPTG      TPTG
## TPCH      TPCH
## LDCH      LDCH
## HDCH      HDCH
## TPA1      TPA1
## TPA2      TPA2

```

Análisis de la matrix de expresión:

```

# Verificar si la matriz es numérica antes de la conversión
is.numeric(assays(se)$counts) # Debería devolver FALSE si no es
numérica
## [1] FALSE
class(assays(se)$counts)
## [1] "matrix" "array"
# Guardar los nombres de filas y columnas
row_names <- rownames(assays(se)$counts)

```

```
col_names <- colnames(assays(se)$counts)

# Convertir la matriz a numérico, sin modificar los nombres
assays(se, withDimnames = FALSE)$counts <- apply(assays(se)$counts, 2,
function(x) as.numeric(as.character(x)))

# Restaurar los nombres de filas y columnas para que coincidan con los
originales
rownames(assays(se)$counts) <- row_names
colnames(assays(se)$counts) <- col_names

# Verificar la conversión
is.numeric(assays(se)$counts)
## [1] TRUE
# Estadísticas descriptivas de la matriz de expresión
summary(assays(se)$counts) # Resumen de estadísticas
##           1           2           3           4

## Min.      : 0.56    Min.      : 0.620    Min.      : 0.510    Min.      :
0.50
## 1st Qu.:  4.37    1st Qu.:  6.178    1st Qu.:  3.570    1st Qu.:
4.55
## Median : 12.71    Median : 15.395    Median :  9.345    Median :
11.03
## Mean   : 68.62    Mean   : 77.377    Mean   : 51.504    Mean   :
56.20
## 3rd Qu.: 31.95    3rd Qu.: 36.462    3rd Qu.: 28.302    3rd Qu.:
28.59
## Max.    :1839.69   Max.    :2027.120   Max.    :1291.680   Max.
:1410.51
##           5           6           7           8

## Min.      : 0.000    Min.      : 0.000    Min.      : 0.000    Min.      :
0.200
## 1st Qu.:  4.617    1st Qu.:  2.538    1st Qu.:  3.953    1st Qu.:
4.645
## Median : 12.185    Median :  7.260    Median :  9.465    Median :
11.915
## Mean   : 81.751    Mean   : 38.429    Mean   : 48.041    Mean   :
60.760
## 3rd Qu.: 32.763    3rd Qu.: 22.992    3rd Qu.: 29.675    3rd Qu.:
26.727
## Max.    :2196.530   Max.    :950.440    Max.    :1110.720   Max.
:1558.120
##           9          10          11          12

## Min.      : 0.140    Min.      : 0.230    Min.      : 0.000    Min.      :
0.570
## 1st Qu.:  4.325    1st Qu.:  4.117    1st Qu.:  3.833    1st Qu.:
3.917
## Median :  9.785    Median : 11.240    Median :  9.715    Median :
9.920
## Mean   : 54.241    Mean   : 63.551    Mean   : 53.295    Mean   :
52.752
```


## 3rd Qu.: 22.988 27.275	3rd Qu.: 29.977	3rd Qu.: 26.808	3rd Qu.: 27.275
## Max. :1424.130 :1372.350	Max. :1601.400	Max. :1328.820	Max. :1372.350
## 13	14	15	16
## Min. : 0.500 0.590	Min. : 0.190	Min. : 0.000	Min. : 0.590
## 1st Qu.: 4.582 5.383	1st Qu.: 2.808	1st Qu.: 2.998	1st Qu.: 5.383
## Median : 11.800 14.720	Median : 9.575	Median : 11.055	Median : 14.720
## Mean : 58.395 86.441	Mean : 55.989	Mean : 60.267	Mean : 86.441
## 3rd Qu.: 28.340 31.360	3rd Qu.: 26.635	3rd Qu.: 27.223	3rd Qu.: 31.360
## Max. :1500.140 :2440.820	Max. :1457.970	Max. :1542.880	Max. :2440.820
## 17	18	19	20
## Min. : 0.000 0.510	Min. : 0.130	Min. : 0.200	Min. : 0.510
## 1st Qu.: 4.492 4.745	1st Qu.: 4.855	1st Qu.: 2.510	1st Qu.: 4.745
## Median : 13.855 13.710	Median : 10.080	Median : 6.045	Median : 13.710
## Mean : 55.906 64.846	Mean : 59.026	Mean : 32.656	Mean : 64.846
## 3rd Qu.: 31.922 35.547	3rd Qu.: 27.120	3rd Qu.: 17.990	3rd Qu.: 35.547
## Max. :1402.510 :1573.560	Max. :1521.190	Max. :800.900	Max. :1573.560
## 21	22	23	24
## Min. : 0.630 0.060	Min. : 0.370	Min. : 0.370	Min. : 0.060
## 1st Qu.: 4.575 2.567	1st Qu.: 2.405	1st Qu.: 4.122	1st Qu.: 2.567
## Median : 15.390 7.530	Median : 7.530	Median : 12.955	Median : 7.530
## Mean : 81.502 50.120	Mean : 44.237	Mean : 90.432	Mean : 50.120
## 3rd Qu.: 32.818 25.258	3rd Qu.: 26.968	3rd Qu.: 37.850	3rd Qu.: 25.258
## Max. :2191.990 :1303.160	Max. :1073.750	Max. :2517.110	Max. :1303.160
## 25	26	27	28
## Min. : 0.000 0.000	Min. : 0.150	Min. : 0.380	Min. : 0.000
## 1st Qu.: 5.607 4.152	1st Qu.: 3.425	1st Qu.: 6.245	1st Qu.: 4.152
## Median : 14.695 12.475	Median : 9.950	Median : 12.495	Median : 12.475

## Mean : 85.362	Mean : 51.222	Mean : 79.360	Mean : 74.787
## 3rd Qu.: 45.830	3rd Qu.: 32.465	3rd Qu.: 35.597	3rd Qu.: 34.358
## Max. :2172.840	Max. :1229.630	Max. :2159.110	Max. :1885.370
## 29	30	31	32
## Min. : 0.050	Min. : 0.000	Min. : 0.180	Min. : 0.000
## 1st Qu.: 3.485	1st Qu.: 2.728	1st Qu.: 4.168	1st Qu.: 4.077
## Median : 8.930	Median : 9.150	Median : 12.035	Median : 13.165
## Mean : 57.170	Mean : 55.485	Mean : 71.225	Mean : 67.902
## 3rd Qu.: 35.595	3rd Qu.: 30.922	3rd Qu.: 31.050	3rd Qu.: 34.240
## Max. :1430.550	Max. :1379.480	Max. :1926.290	Max. :1728.730
## 33	34	35	36
## Min. : 0.70	Min. : 0.56	Min. : 0.530	Min. : 0.070
## 1st Qu.: 5.86	1st Qu.: 5.90	1st Qu.: 6.275	1st Qu.: 3.857
## Median : 14.37	Median : 14.10	Median : 15.830	Median : 13.465
## Mean : 91.39	Mean : 77.49	Mean : 80.353	Mean : 69.949
## 3rd Qu.: 43.59	3rd Qu.: 40.07	3rd Qu.: 36.847	3rd Qu.: 29.215
## Max. :2433.08	Max. :2065.00	Max. :2110.560	Max. :1842.000
## 37	38	39	40
## Min. : 0.000	Min. : 0.000	Min. : 0.560	Min. : 0.540
## 1st Qu.: 2.525	1st Qu.: 2.917	1st Qu.: 5.643	1st Qu.: 4.838
## Median : 5.430	Median : 9.535	Median : 14.345	Median : 13.750
## Mean : 37.885	Mean : 56.440	Mean : 80.204	Mean : 75.483
## 3rd Qu.: 23.715	3rd Qu.: 34.620	3rd Qu.: 37.475	3rd Qu.: 32.170
## Max. :893.150	Max. :1318.230	Max. :2092.990	Max. :2047.000
## 41	42	43	44
## Min. : 0.390	Min. : 0.460	Min. : 0.640	Min. : 0.000
## 1st Qu.: 3.987	1st Qu.: 4.907	1st Qu.: 7.168	1st Qu.: 2.123

## Median : 11.335 8.740	Median : 11.045	Median : 16.995	Median :
## Mean : 55.966 53.501	Mean : 67.939	Mean : 99.208	Mean :
## 3rd Qu.: 30.800 29.477	3rd Qu.: 35.312	3rd Qu.: 44.560	3rd Qu.:
## Max. :1369.070 :1282.730	Max. :1777.840	Max. :2679.600	Max.
## 45	46	47	48
## Min. : 0.000 0.00	Min. : 0.270	Min. : 0.730	Min. :
## 1st Qu.: 3.215 1.96	1st Qu.: 3.280	1st Qu.: 6.973	1st Qu.:
## Median : 8.265 6.51	Median : 8.315	Median : 15.035	Median :
## Mean : 56.475 39.12	Mean : 47.561	Mean : 89.933	Mean :
## 3rd Qu.: 30.918 27.10	3rd Qu.: 29.270	3rd Qu.: 37.005	3rd Qu.:
## Max. :1412.620 :890.93	Max. :1163.840	Max. :2500.900	Max.
## 49	50	51	52
## Min. : 0.49 0.00	Min. : 0.38	Min. : 0.050	Min. :
## 1st Qu.: 5.22 2.96	1st Qu.: 4.70	1st Qu.: 3.730	1st Qu.:
## Median : 11.92 9.80	Median : 11.21	Median : 9.035	Median :
## Mean : 67.01 67.68	Mean : 74.66	Mean : 58.278	Mean :
## 3rd Qu.: 31.19 35.83	3rd Qu.: 30.83	3rd Qu.: 24.415	3rd Qu.:
## Max. :1768.49 :1738.65	Max. :2052.04	Max. :1564.680	Max.
## 53	54	55	56
## Min. : 0.300 0.080	Min. : 0.470	Min. : 0.240	Min. :
## 1st Qu.: 3.382 4.395	1st Qu.: 5.223	1st Qu.: 4.128	1st Qu.:
## Median : 10.140 10.400	Median : 14.125	Median : 13.260	Median :
## Mean : 62.234 61.099	Mean : 73.616	Mean : 72.605	Mean :
## 3rd Qu.: 35.847 34.385	3rd Qu.: 32.312	3rd Qu.: 32.227	3rd Qu.:
## Max. :1567.080 :1513.090	Max. :1931.640	Max. :1933.220	Max.
## 57	58	59	60
## Min. : 0.000 0.210	Min. : 0.380	Min. : 0.440	Min. :

## 1st Qu.: 5.393	1st Qu.: 3.547	1st Qu.: 5.572	1st Qu.: 5.095
## Median : 11.575	Median : 9.705	Median : 13.960	Median : 11.635
## Mean : 71.058	Mean : 53.535	Mean : 83.790	Mean : 71.067
## 3rd Qu.: 34.605	3rd Qu.: 32.620	3rd Qu.: 35.320	3rd Qu.: 36.500
## Max. :1835.320	Max. :1317.840	Max. :2212.070	Max. :1820.100
## 61	62	63	64
## Min. : 0.200	Min. : 0.17	Min. : 0.000	Min. : 0.500
## 1st Qu.: 2.785	1st Qu.: 2.69	1st Qu.: 3.337	1st Qu.: 4.805
## Median : 8.925	Median : 9.32	Median : 7.600	Median : 14.405
## Mean : 50.160	Mean : 54.86	Mean : 45.904	Mean : 66.136
## 3rd Qu.: 29.218	3rd Qu.: 22.93	3rd Qu.: 28.745	3rd Qu.: 36.625
## Max. :1177.000	Max. :1498.17	Max. :1070.360	Max. :1660.360
## 65	66	67	68
## Min. : 0.530	Min. : 0.650	Min. : 0.170	Min. : 0.170
## 1st Qu.: 4.707	1st Qu.: 5.258	1st Qu.: 4.695	1st Qu.: 3.877
## Median : 12.290	Median : 14.270	Median : 12.915	Median : 13.730
## Mean : 61.051	Mean : 81.795	Mean : 67.021	Mean : 77.373
## 3rd Qu.: 28.953	3rd Qu.: 32.615	3rd Qu.: 29.455	3rd Qu.: 41.477
## Max. :1556.700	Max. :2187.020	Max. :1777.060	Max. :1950.470
## 69	70	71	72
## Min. : 0.240	Min. : 0.030	Min. : 0.01	Min. : 0.000
## 1st Qu.: 2.958	1st Qu.: 3.908	1st Qu.: 5.01	1st Qu.: 2.705
## Median : 9.230	Median : 10.370	Median : 12.15	Median : 9.610
## Mean : 59.703	Mean : 51.160	Mean : 65.66	Mean : 56.876
## 3rd Qu.: 31.475	3rd Qu.: 26.523	3rd Qu.: 33.40	3rd Qu.: 33.458
## Max. :1520.950	Max. :1242.680	Max. :1678.85	Max. :1361.270
## 73	74	75	76

## Min. : 0.380	Min. : 0.620	Min. : 0.580	Min. :
## 1st Qu.: 4.438	1st Qu.: 5.982	1st Qu.: 6.128	1st Qu.:
## Median : 11.435	Median : 14.030	Median : 15.445	Median :
## Mean : 64.962	Mean : 75.297	Mean : 82.951	Mean :
## 3rd Qu.: 42.890	3rd Qu.: 36.028	3rd Qu.: 41.065	3rd Qu.:
## Max. :1551.600	Max. :2011.420	Max. :2202.430	Max. :
## 77	78	79	80
## Min. : 0.38	Min. : 0.530	Min. : 0.330	Min. :
## 1st Qu.: 4.85	1st Qu.: 4.942	1st Qu.: 5.268	1st Qu.:
## Median : 10.82	Median : 12.495	Median : 11.985	Median :
## Mean : 52.15	Mean : 65.200	Mean : 61.393	Mean :
## 3rd Qu.: 31.03	3rd Qu.: 31.090	3rd Qu.: 28.840	3rd Qu.:
## Max. :1236.15	Max. :1660.480	Max. :1572.410	Max. :
## 81	82	83	
## Min. : 0.000	Min. : 0.120	Min. : 0.0000	Min. :
## 1st Qu.: 2.705	1st Qu.: 5.125	1st Qu.: 0.7025	1st Qu.:
## Median : 8.595	Median : 14.420	Median : 17.8300	Median :
## Mean : 47.108	Mean : 79.045	Mean : 132.6531	Mean :
## 3rd Qu.: 26.325	3rd Qu.: 43.380	3rd Qu.: 67.9550	3rd Qu.:
## Max. :1134.380	Max. :1975.510	Max. :2981.0500	Max. :
## 85	86	87	88
## Min. : 0.130	Min. : 0.35	Min. : 0.63	Min. :
## 1st Qu.: 5.178	1st Qu.: 4.81	1st Qu.: 6.09	1st Qu.:
## Median : 11.380	Median : 10.02	Median : 15.02	Median :
## Mean : 65.303	Mean : 65.51	Mean : 78.09	Mean :
## 3rd Qu.: 28.898	3rd Qu.: 30.39	3rd Qu.: 36.90	3rd Qu.:
## Max. :1769.230	Max. :1770.00	Max. :2047.31	Max. :

##	89	90	91	92
##	Min. : 0.060 0.010	Min. : 0.180	Min. : 0.090	Min. :
##	1st Qu.: 3.130 2.825	1st Qu.: 3.717	1st Qu.: 3.765	1st Qu.:
##	Median : 8.095 9.895	Median : 12.440	Median : 13.090	Median :
##	Mean : 48.940 51.720	Mean : 67.569	Mean : 68.331	Mean :
##	3rd Qu.: 26.665 28.635	3rd Qu.: 36.605	3rd Qu.: 33.240	3rd Qu.:
##	Max. :1183.360 :1274.870	Max. :1744.220	Max. :1705.350	Max.
##	93	94	95	96
##	Min. : 0.520 0.280	Min. : 0.340	Min. : 0.19	Min. :
##	1st Qu.: 5.513 5.438	1st Qu.: 3.507	1st Qu.: 4.79	1st Qu.:
##	Median : 12.060 12.255	Median : 8.850	Median : 14.23	Median :
##	Mean : 65.415 75.954	Mean : 46.437	Mean : 69.40	Mean :
##	3rd Qu.: 33.775 39.355	3rd Qu.: 25.957	3rd Qu.: 43.03	3rd Qu.:
##	Max. :1736.260 :1989.950	Max. :1172.580	Max. :1708.39	Max.
##	97	98	99	100
##	Min. : 0.330 0.340	Min. : 0.000	Min. : 0.000	Min. :
##	1st Qu.: 3.522 4.575	1st Qu.: 2.562	1st Qu.: 3.075	1st Qu.:
##	Median : 9.360 14.825	Median : 7.755	Median : 9.390	Median :
##	Mean : 50.319 80.927	Mean : 53.364	Mean : 55.423	Mean :
##	3rd Qu.: 25.640 32.028	3rd Qu.: 30.457	3rd Qu.: 32.170	3rd Qu.:
##	Max. :1283.440 :2197.250	Max. :1320.350	Max. :1348.940	Max.
##	101	102	103	104
##	Min. : 0.000 0.340	Min. : 0.420	Min. : 0.100	Min. :
##	1st Qu.: 1.698 5.255	1st Qu.: 3.797	1st Qu.: 3.725	1st Qu.:
##	Median : 8.065 12.480	Median : 9.825	Median : 8.865	Median :
##	Mean : 47.399 66.528	Mean : 55.705	Mean : 48.948	Mean :
##	3rd Qu.: 25.040 33.935	3rd Qu.: 30.628	3rd Qu.: 26.988	3rd Qu.:

```

## Max. :1112.320 Max. :1366.780 Max. :1196.310 Max.
:1694.490
## 105 106 107 108

## Min. : 0.270 Min. : 0.00 Min. : 0.110 Min. :
0.49
## 1st Qu.: 4.327 1st Qu.: 2.52 1st Qu.: 2.965 1st Qu.:
6.74
## Median : 12.030 Median : 9.28 Median : 8.350 Median :
15.71
## Mean : 73.904 Mean : 54.31 Mean : 50.969 Mean :
99.87
## 3rd Qu.: 28.233 3rd Qu.: 27.02 3rd Qu.: 31.360 3rd Qu.:
48.42
## Max. :2025.460 Max. :1440.72 Max. :1252.870 Max.
:2726.75
## 109 110 111 112

## Min. : 0.040 Min. : 0.360 Min. : 0.65 Min. :
0.03
## 1st Qu.: 4.585 1st Qu.: 3.935 1st Qu.: 4.89 1st Qu.:
5.21
## Median : 10.565 Median : 9.225 Median : 12.13 Median :
13.51
## Mean : 51.899 Mean : 57.288 Mean : 66.98 Mean :
78.78
## 3rd Qu.: 32.938 3rd Qu.: 36.240 3rd Qu.: 27.73 3rd Qu.:
37.13
## Max. :1228.170 Max. :1354.080 Max. :1781.09 Max.
:2126.05
apply(assays(se)$counts, 1, sd) # Desviación estándar por metabolito
## TPTG TPCB LDCH HDCH TPA1
TPA2
## 141.3947436 45.8442246 32.7329093 13.7170362 21.7394052
5.1599368
## TPAB LDHD ABA1 TBPN VLPN
IDPN
## 23.5499304 0.4949865 0.1480981 428.2082754 122.2184789
77.3568137
## LDPN L1PN L2PN L3PN L4PN
L5PN
## 340.3343097 83.4266293 88.9126629 79.2706724 77.5609899
92.1143242
## L6PN VLTG IDTG LDTG HDTG
VLCH
## 206.8854210 103.0793375 20.1202220 6.5338877 4.9339714
19.7256312
## IDCH VLFC IDFC LDFC HDFC
VLPL
## 12.0101111 9.1371918 3.4458223 9.2260652 4.1200176
19.4580633
## IDPL LDPL HDPL HDA1 HDA2
VLAB
## 5.9110025 16.5379880 17.7117818 24.1406077 5.7186639

```

```

6.7217680
##          IDAB          LDAB          V1TG          V2TG          V3TG
V4TG
##    4.2544317  18.7170899  77.7678017   9.2502474   8.0645467
5.6624162
##          V5TG          V1CH          V2CH          V3CH          V4CH
V5CH
##    1.0208385  13.8450466   2.6749968   3.1750689   2.9012337
0.5497136
##          V1FC          V2FC          V3FC          V4FC          V5FC
V1PL
##    4.3652176   1.7457586   1.9853780   1.7740049   1.3863939
9.8640860
##          V2PL          V3PL          V4PL          V5PL          L1TG
L2TG
##    2.2957928   2.9647948   2.3913670   0.6694992   3.2754570
0.8462064
##          L3TG          L4TG          L5TG          L6TG          L1CH
L2CH
##    0.6785750   0.9033815   1.2055145   2.0505181   9.2788996
9.8209679
##          L3CH          L4CH          L5CH          L6CH          L1FC
L2FC
##    8.3987686   7.0411761   7.3307482  11.8944097   2.6004348
2.8049943
##          L3FC          L4FC          L5FC          L6FC          L1PL
L2PL
##    2.3236915   1.7974982   1.6675710   1.9122954   4.6250740
5.0152030
##          L3PL          L4PL          L5PL          L6PL          L1AB
L2AB
##    4.2824896   3.6195618   3.6062658   5.8371607   4.5885314
4.8899077
##          L3AB          L4AB          L5AB          L6AB          H1TG
H2TG
##    4.3595893   4.2656522   5.0663426  11.3783591   2.9175129
0.9613341
##          H3TG          H4TG          H1CH          H2CH          H3CH
H4CH
##    1.0501215   1.2124991  10.3151975   2.8927404   2.0481556
4.5756730
##          H1FC          H2FC          H3FC          H4FC          H1PL
H2PL
##    2.3203692   0.7483269   0.7356488   1.1636426  12.1849931
4.2628864
##          H3PL          H4PL          H1A1          H2A1          H3A1
H4A1
##    3.4502193   5.2609174  17.5232825   5.1986078   5.7634948
13.5279364
##          H1A2          H2A2          H3A2          H4A2
##    2.0029987   1.9766994   1.9061099   3.6354732
average_per_analyte <- apply(assays(se)$counts, 1, mean) # Media por
analito (por fila)

```


average_per_analyte # Esto muestra la media de cada metabolito a través de las muestras

##	TPTG	TPCH	LDCH	HDCH	TPA1
TPA2					
## 171.6844643	230.1240179	119.0497321	70.4203571	171.4629464	
38.1174107					
##	TPAB	LDHD	ABA1	TBPN	VLPN
IDPN					
## 90.9807143	1.7302679	0.5369643	1654.2713393	166.3433036	
116.0906250					
##	LDPN	L1PN	L2PN	L3PN	L4PN
L5PN					
## 1353.3644643	231.2171429	191.6153571	176.8106250	155.9041964	
213.0021429					
##	L6PN	VLTG	IDTG	LDTG	HDTG
VLCH					
## 422.7819643	115.8727679	18.9650000	18.9619643	11.7853571	
24.6983929					
##	IDCH	VLFC	IDFC	LDFC	HDFC
VLPL					
## 16.5586607	11.6059821	4.7581250	34.6433929	16.0193750	
25.8539286					
##	IDPL	LDPL	HDPL	HDA1	HDA2
VLAB					
## 10.3333929	64.7145536	91.1580357	171.4297321	38.3012500	
9.1481250					
##	IDAB	LDAB	V1TG	V2TG	V3TG
V4TG					
## 6.3842857	74.4310714	67.8867857	14.9811607	12.9531250	
9.9804464					
##	V5TG	V1CH	V2CH	V3CH	V4CH
V5CH					
## 2.7242857	11.6245536	3.5778571	4.2458929	4.9720536	
0.7142857					
##	V1FC	V2FC	V3FC	V4FC	V5FC
V1PL					
## 4.3249107	1.7413393	2.1530357	2.3985714	0.7569643	
9.6710714					
##	V2PL	V3PL	V4PL	V5PL	L1TG
L2TG					
## 3.8188393	4.6828571	4.5937500	1.3067857	5.6032143	
2.4075893					
##	L3TG	L4TG	L5TG	L6TG	L1CH
L2CH					
## 2.1083929	1.5440179	2.2837500	4.6241071	24.1394643	
19.8725000					
##	L3CH	L4CH	L5CH	L6CH	L1FC
L2FC					
## 16.8866071	13.6479464	16.7165179	27.9702679	7.5473214	
6.8772321					
##	L3FC	L4FC	L5FC	L6FC	L1PL
L2PL					
## 5.8668750	4.7953571	4.9008036	7.1549107	13.4253571	
10.6850893					

##	L3PL	L4PL	L5PL	L6PL	L1AB
	L2AB				
##	9.2008036	7.4500893	8.9643750	15.6487500	12.7161607
	10.5383929				
##	L3AB	L4AB	L5AB	L6AB	H1TG
	H2TG				
##	9.7240179	8.5745536	11.7148214	23.2516071	4.7591964
	2.1166071				
##	H3TG	H4TG	H1CH	H2CH	H3CH
	H4CH				
##	2.3420536	3.1299107	26.4902679	11.4107143	12.3613393
	19.2645536				
##	H1FC	H2FC	H3FC	H4FC	H1PL
	H2PL				
##	6.7743750	2.7917857	2.8228571	3.8841964	30.7636607
	16.9415179				
##	H3PL	H4PL	H1A1	H2A1	H3A1
	H4A1				
##	19.0404464	25.9058036	40.3289286	22.6289286	31.7095536
	73.4633036				
##	H1A2	H2A2	H3A2	H4A2	
##	4.1672321	5.2052679	8.3866071	19.5838393	

Análisis de los metadatos en muestras:

Verificar Los nombres de Las muestras

`colnames(assays(se)$counts)`

```
## [1] "1" "2" "3" "4" "5" "6" "7" "8" "9" "10"
"11" "12"
## [13] "13" "14" "15" "16" "17" "18" "19" "20" "21" "22"
"23" "24"
## [25] "25" "26" "27" "28" "29" "30" "31" "32" "33" "34"
"35" "36"
## [37] "37" "38" "39" "40" "41" "42" "43" "44" "45" "46"
"47" "48"
## [49] "49" "50" "51" "52" "53" "54" "55" "56" "57" "58"
"59" "60"
## [61] "61" "62" "63" "64" "65" "66" "67" "68" "69" "70"
"71" "72"
## [73] "73" "74" "75" "76" "77" "78" "79" "80" "81" "82"
"83" "84"
## [85] "85" "86" "87" "88" "89" "90" "91" "92" "93" "94"
"95" "96"
## [97] "97" "98" "99" "100" "101" "102" "103" "104" "105" "106"
"107" "108"
## [109] "109" "110" "111" "112"
```

Inspeccionar Los metadatos de Las muestras

`head(colData(se))`

DataFrame with 6 rows and 18 columns

```
## RAW_FILE_NAME(NMR Exp No.) Age Weight (kg) Height (cm)
Pulse
```

```
## <character> <character> <character> <character>
<character>
```

```
## 1 Box1.zip/100/fid 25.2247 53.9000 164.0000
```

	80				
## 2	Box1.zip/110/fid	74.8986	62.8000	164.0000	
	88				
## 3	Box1.zip/120/fid	78.4548	67.9000	167.0000	
	89				
## 4	Box1.zip/130/fid	78.4438	64.6000	158.0000	
	64				
## 5	Box1.zip/140/fid	75.6082	66.6000	158.0000	
	73				
## 6	Box1.zip/150/fid	63.8740	73.2000	178.0000	
	78				
##	WaistCircumference (cm)	SBP	DBP		
##	<character>	<character>	<character>		
## 1	76.0	110	67		
## 2	83.0	146	77		
## 3	92.0	168	65		
## 4	84.0	158	67		
## 5	92.0	164	85		
## 6	89.0	139	83		
##	Smoking (Never 0, Current 1, Past 2)	Sample source	E/A	<	
##		<character>	<character>		
<character>					
## 1		0	Blood		
0					
## 2		0	Blood		
0					
## 3		2	Blood		
1					
## 4		0	Blood		
0					
## 5		0	Blood		
0					
## 6		N/A	Blood		
0					
##	Hypertension	Dyslipidemia	Diabetes mellitus	Hypertension_Med	
	Cholesterol_Med				
##	<character>	<character>	<character>	<character>	
<character>					
## 1	0	0	0	0	
0					
## 2	0	0	0	0	
0					
## 3	1	1	0	1	
1					
## 4	1	0	0	1	
0					
## 5	0	0	0	0	
0					
## 6	0	1	0	0	
1					
##	Diabetes_Med	Sex			
##	<character>	<character>			
## 1	0	F			

```

## 2      0      M
## 3      0      M
## 4      0      F
## 5      0      M
## 6      0      M
# Verificar la distribución de algunas variables en los metadatos de
# las muestras
table(colData(se)$Age) # Distribución de la edad
##
## 21.2959 23.2027 24.9370 25.2247 25.6274 25.6329 25.6493 26.0795
27.3753 27.9589
##      1      1      1      1      1      1      1      1
1      1
## 28.7370 28.7781 28.9589 29.9096 30.0301 32.7452 34.9863 35.9068
38.0411 39.0932
##      1      1      1      1      1      1      1      1
1      1
## 39.3151 39.8411 39.9562 39.9781 41.2822 41.3370 42.6329 42.7890
43.1260 43.2137
##      1      1      1      1      1      1      1      1
1      1
## 43.2575 43.4658 43.5808 44.4247 45.1014 45.2301 46.1096 46.2849
47.2164 47.2548
##      1      1      1      1      1      1      1      1
1      1
## 47.3233 47.9452 48.9945 49.0740 50.0329 51.0466 51.3205 51.4301
51.8685 51.8877
##      1      1      1      1      1      1      1      1
1      1
## 52.2137 52.5562 52.9452 53.4493 53.5671 53.8000 53.8630 54.0301
54.5753 55.5452
##      1      1      1      1      1      1      1      1
1      1
## 55.7726 56.2219 56.4986 56.8986 57.8767 58.1863 58.4685 60.1890
60.5014 60.6959
##      1      1      1      1      1      1      1      1
1      2
## 61.2795 61.3041 61.5205 61.7781 61.8164 62.3068 62.4027 62.6795
62.7041 63.2575
##      1      1      2      1      1      1      1      1
1      1
## 63.5699 63.8438 63.8740 64.1370 64.9315 64.9945 65.3726 65.7753
65.8301 67.0603
##      1      1      1      1      1      1      1      1
1      1
## 67.9205 68.1288 68.2247 68.4767 68.6110 69.0575 69.2082 69.8411
70.3753 70.5973
##      1      1      1      1      1      1      1      1
1      1
## 70.6603 73.0712 73.3726 73.5452 74.8986 75.6082 77.0438 77.5836
78.4438 78.4548
##      1      1      1      1      1      1      1      1
1      1
table(colData(se)$Sex) # Distribución por sexo

```

```
##
## F M
## 60 52
```

Análisis de los metadatos de los metabolitos:

```
# Verificar los primeros metabolitos y sus metadatos
head(rowData(se))
## DataFrame with 6 rows and 1 column
##      Metabolite
##      <character>
## TPTG          TPTG
## TPCH          TPCH
## LDCH          LDCH
## HDCH          HDCH
## TPA1          TPA1
## TPA2          TPA2
# Verificar la distribución de los nombres de los metabolitos
table(rowData(se)$Metabolite)
##
## ABA1 H1A1 H1A2 H1CH H1FC H1PL H1TG H2A1 H2A2 H2CH H2FC H2PL H2TG
H3A1 H3A2 H3CH
##      1      1      1      1      1      1      1      1      1      1      1      1      1
1      1      1
## H3FC H3PL H3TG H4A1 H4A2 H4CH H4FC H4PL H4TG HDA1 HDA2 HDCH HDCH
HDPL HDTG IDAB
##      1      1      1      1      1      1      1      1      1      1      1      1      1
1      1      1
## IDCH IDFC IDPL IDPN IDTG L1AB L1CH L1FC L1PL L1PN L1TG L2AB L2CH
L2FC L2PL L2PN
##      1      1      1      1      1      1      1      1      1      1      1      1      1
1      1      1
## L2TG L3AB L3CH L3FC L3PL L3PN L3TG L4AB L4CH L4FC L4PL L4PN L4TG
L5AB L5CH L5FC
##      1      1      1      1      1      1      1      1      1      1      1      1      1
1      1      1
## L5PL L5PN L5TG L6AB L6CH L6FC L6PL L6PN L6TG LDAB LDCH LDCH LDCH
LDPL LDPN LDTG
##      1      1      1      1      1      1      1      1      1      1      1      1      1
1      1      1
## TBPB TPA1 TPA2 TPAB TPCH TPTG V1CH V1FC V1PL V1TG V2CH V2FC V2PL
V2TG V3CH V3FC
##      1      1      1      1      1      1      1      1      1      1      1      1      1
1      1      1
## V3PL V3TG V4CH V4FC V4PL V4TG V5CH V5FC V5PL V5TG VLAB VLCH VLFC
VLPL VLPN VLTG
##      1      1      1      1      1      1      1      1      1      1      1      1      1
1      1      1
```

Observar los datos:

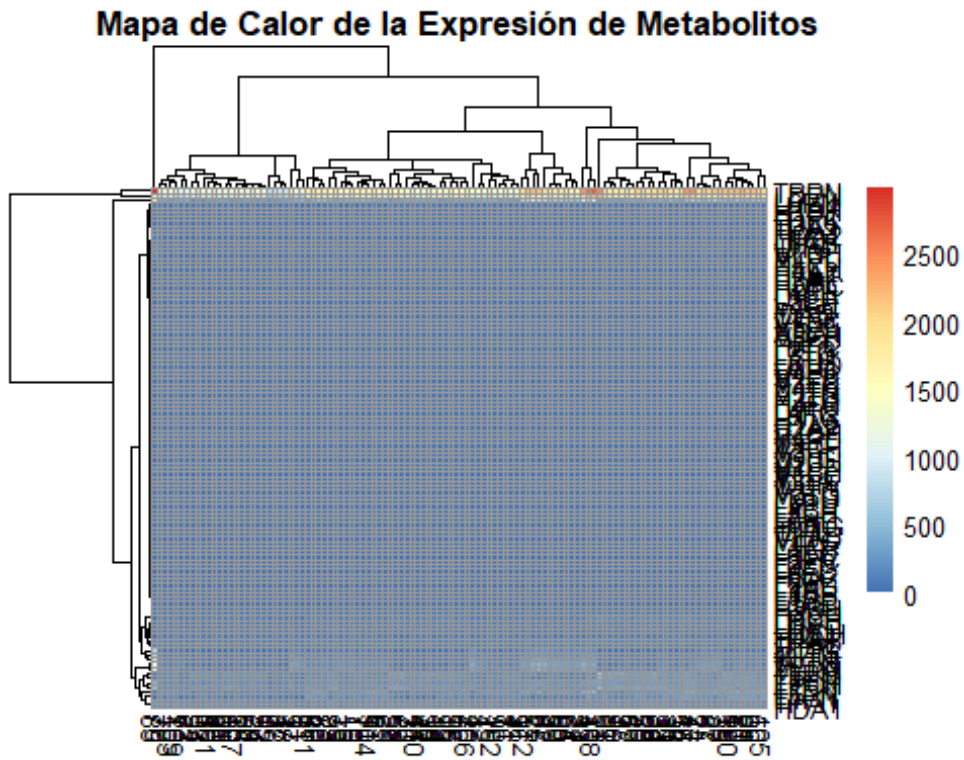
```
# Instalar y cargar la librería pheatmap si no está instalada
if (!require(pheatmap)) install.packages("pheatmap")
## Cargando paquete requerido: pheatmap
## Warning: package 'pheatmap' was built under R version 4.4.3
```

```
library(heatmap)

# Crear un mapa de calor de la matriz de expresión
heatmap(assays(se)$counts, cluster_rows = TRUE, cluster_cols = TRUE,
        main = "Mapa de Calor de la Expresión de Metabolitos")
```

```
# Crear un mapa de calor de la matriz de expresión
pheatmap(assays(se)$counts, cluster_rows = TRUE, cluster_cols = TRUE,
          main = "Mapa de Calor de la Expresión de Metabolitos")
```

```
pheatmap(assays(se)$counts, cluster_rows = TRUE, cluster_cols = TRUE,
  main = "Mapa de Calor de la Expresión de Metabolitos")
```



```
# Guardar el objeto de clase SummarizedExperiment
save(se, file = "metabolite_data.Rda")
write.csv(assays(se)$counts, "expression_data.csv", row.names = TRUE)
write.csv(colData(se), "samples_metadata.csv", row.names = TRUE)
write.csv(rowData(se), "metabolites_metadata.csv", row.names = TRUE)
```

```
save(se, file = "metabolite_data.Rda")
write.csv(assays(se)$counts, "expression_data.csv", row.names = TRUE)
write.csv(colData(se), "samples_metadata.csv", row.names = TRUE)
write.csv(rowData(se), "metabolites_metadata.csv", row.names = TRUE)
```

```
write.csv(assays(se)$counts, "expression_data.csv", row.names = TRUE)
write.csv(colData(se), "samples_metadata.csv", row.names = TRUE)
write.csv(rowData(se), "metabolites_metadata.csv", row.names = TRUE)
```

```
write.csv(colData(se), "samples_metadata.csv", row.names = TRUE)
write.csv(rowData(se), "metabolites_metadata.csv", row.names = TRUE)
```

```
write.csv(rowData(se), "metabolites_metadata.csv", row.names = TRUE)
```