

# Sanchez-Sierra-David-PEC1

David Sánchez Sierra

2025-04-01

## **Análisis de datos ómicos (M0-157). Prueba de evaluación continua.**

### **Índice**

1. Resumen
2. Objetivos
3. Enfoque y método seguido
  - 3.1. Selección y descarga del dataset
  - 3.2. Creación del objeto SummarizedExperiment
  - 3.3. Diferencias entre SummarizedExperiment y ExpressionSet
  - 3.4. Análisis exploratorio
4. Resultados
5. Discusión
6. Conclusiones
7. Bibliografía

### **1. Resumen**

### **2. Objetivos**

### **3. Enfoque y método seguido**

#### **3.1. Selección y descarga del dataset**

Para esta actividad, se seleccionó el dataset “caquexia” obtenido del repositorio de github adjunto al enunciado. Este conjunto de datos ha sido ampliamente utilizado en diversos tutoriales de MetaboAnalyst, lo que permitió expandir el análisis realizado al disponer de información sobre numerosos enfoques y metodologías establecidas.

Las técnicas metabolómicas se han utilizado para estudiar los cambios metabólicos, incluyendo las variaciones en las concentraciones de metabolitos y las vías metabólicas alteradas en la progresión de la caquexia relacionada con el cáncer (CC), así como para ampliar la comprensión fundamental de la pérdida muscular (Cui et al. 2022). En este dataset, se incluyen datos sobre 77 pacientes oncológicos, clasificados en dos grupos según su estado de pérdida muscular: controles y caquéxicos.

La descarga de los datos se realizó utilizando R, empleando la función `read.csv()` para cargar el archivo en formato tabular:

```
dataset <- read.csv("human_cachexia.csv", header = TRUE, sep = ",", stringsAsFactors = FALSE)
str(dataset, max.level = 0)
```

```
## 'data.frame': 77 obs. of 65 variables:
```

Como puede observarse, el dataset contiene información sobre 65 variables, incluyendo las concentraciones de diversos metabolitos en las muestras de orina asociadas con los dos grupos de pacientes. Además, el conjunto de datos incluye el identificador de cada paciente y los metadatos relacionados con las características de estos, en este caso el tipo categórico (control o caquéticos).

### 3.2. Creación del objeto SummarizedExperiment

Tras cargar los datos, se creó un objeto de clase `SummarizedExperiment` proporcionada por el paquete `Bioconductor`, el cual permite estructurar los datos y metadatos asociados a cada observación, facilitando su manipulación y análisis. Primero se asignaron los valores de `Patient.ID` como nombres de fila para el dataset, asegurando que cada muestra tuviera una identificación única en la matriz. Tras ello, se eliminó la columna `Patient.ID` del conjunto de datos, ya que solo se necesitaba para asignar los nombres de las filas, y no debía formar parte de los datos de expresión.

A continuación, se extrajeron las mediciones de los metabolitos del dataset, excluyendo la columna `Muscle.loss`, que corresponde a la información de los metadatos y no a las concentraciones de los metabolitos. Las concentraciones de metabolitos se almacenaron en una matriz de expresión (`expr_matrix`), la cual fue transpuesta para que las muestras fueran representadas en las columnas, lo que facilita la asociación con los metadatos. En el siguiente paso, se creó el `DataFrame` para los metadatos que debía contener la columna `Muscle.loss` (información sobre el estado de caquexia de las muestras).

El código que implementa este procedimiento es el siguiente:

```
## Warning: package 'MatrixGenerics' was built under R version 4.4.2
```

```
## Warning: package 'matrixStats' was built under R version 4.4.2
```

```
## Warning: package 'IRanges' was built under R version 4.4.2
```

```
## Warning: package 'GenomeInfoDb' was built under R version 4.4.2
```

```
# Asignar ID como nombre de fila
rownames(dataset) <- dataset$Patient.ID
dataset <- dataset[, -1] # Eliminar la columna 'Patient ID'

# Asegurarse de que 'Muscle.loss' es excluida correctamente
expr_matrix <- as.matrix(dataset[, -which(names(dataset) == "Muscle.loss")])

# Transponer la matriz para que las muestras sean columnas
expr_matrix <- t(expr_matrix)

# Crear el dataframe de metadatos (solo 'Muscle.loss')
metadata_df <- DataFrame(Muscle.loss = dataset$Muscle.loss)
rownames(metadata_df) <- rownames(dataset) # Asegurar nombres de fila en metadatos
```

```

# Crear el objeto SummarizedExperiment
se_object <- SummarizedExperiment(
  assays = list(counts = expr_matrix),
  colData = metadata_df
)

# Verificar el objeto creado
se_object

## class: SummarizedExperiment
## dim: 63 77
## metadata(0):
## assays(1): counts
## rownames(63): X1.6.Anhydro.beta.D.glucose X1.Methylnicotinamide ...
##   pi.Methylhistidine tau.Methylhistidine
## rowData names(0):
## colnames(77): PIF_178 PIF_087 ... NETL_003_V1 NETL_003_V2
## colData names(1): Muscle.loss

```

### 3.3. Diferencias entre SummarizedExperiment y ExpressionSet

La clase SummarizedExperiment es una extensión moderna de ExpressionSet, con varias mejoras en cuanto a la estructura de almacenamiento y la interoperabilidad con otras herramientas de Bioconductor. Una de las diferencias más notables es que SummarizedExperiment utiliza el componente assays para manejar las matrices de datos en lugar de la función exprs() que se usaba en ExpressionSet. Además, SummarizedExperiment también incorpora colData y rowData para el almacenamiento de los metadatos relativos a las muestras y las características, respectivamente, sustituyendo a las funciones pData() y fData() que se encontraban en ExpressionSet. Esta modificación simplifica y vuelve más adaptable la gestión de dichos metadatos. Un beneficio adicional significativo de SummarizedExperiment reside en su mejorada interoperabilidad con análisis de datos multi-ómicos, lo que la establece como una herramienta más idónea para la realización de estudios que combinan distintos tipos de datos ómicos, tales como datos genómicos, transcriptómicos y proteómicos (“Convert Seurat Object to Summarised Experiment or ExpressionSet” n.d.).

Estas diferencias hacen que SummarizedExperiment resulte una opción más adecuada para manejar y analizar datos provenientes de diversas fuentes dentro de un mismo entorno de trabajo, otorgándole una mayor flexibilidad en comparación con ExpressionSet.

### 3.4. Análisis exploratorio

Para obtener una visión general del dataset, se realizaron análisis descriptivos como distribución de valores, presencia de datos faltantes y comparación entre grupos experimentales.

Se generaron gráficos para visualizar patrones en los datos y se identificaron posibles outliers o tendencias relevantes.

## 4. Resultados

La caquexia es una enfermedad compleja y multifactorial, caracterizada por la pérdida de peso a través de la pérdida de masa muscular esquelética y tejido adiposo, un desequilibrio en la regulación metabólica y una reducción en la ingesta de alimentos. Las principales causas de este trastorno son factores catabólicos producidos por tumores en la circulación sistémica, así como factores fisiológicos como la activación inflamatoria

desequilibrada, la proteólisis, la autofagia y la lipólisis que pueden ocurrir en cánceres gástrico, pancreático, esofágico, pulmonar, hepático y colorrectal (“Caquexia y cáncer - Efectos secundarios” 2024).

En el contexto oncológico, la caquexia es una de las principales causas de morbilidad y mortalidad, contribuyendo directamente a aproximadamente el 25% de los fallecimientos relacionados con el cáncer. Esta condición no solo deteriora la calidad de vida del paciente, sino que también puede limitar la eficacia de los tratamientos antineoplásicos y aumentar la toxicidad asociada a los mismos.

Las causas de la caquexia en pacientes con cáncer son complejas y multifactoriales. Se cree que los tumores malignos pueden liberar sustancias como el factor de necrosis tumoral alfa (TNF- $\alpha$ ), la interleucina-6 (IL-6) y otras citoquinas inflamatorias que inducen la degradación de proteínas y lípidos, llevando a la pérdida de masa muscular y tejido adiposo. Además, la inflamación sistémica y los cambios metabólicos asociados al cáncer pueden provocar un desequilibrio entre la ingesta y el gasto energético, exacerbando la pérdida de peso y la atrofia muscular.

## Discusión

## Conclusiones

## Bibliografía

- “Caquexia y cáncer - Efectos secundarios.” 2024. {cgvArticle}. <https://www.cancer.gov/espanol/cancer/tratamiento/efectos-sekundarios/caquexia-cancer>.
- “Convert Seurat Object to Summarised Experiment or ExpressionSet.” n.d. <https://support.bioconductor.org/p/9157595/>. Accessed April 1, 2025.
- Cui, Pengfei, Xiaoyi Li, Caihua Huang, Qinxu Li, and Donghai Lin. 2022. “Metabolomics and Its Applications in Cancer Cachexia.” *Frontiers in Molecular Biosciences* 9 (February): 789889. <https://doi.org/10.3389/fmolb.2022.789889>.