

# PEC1 - Análisis de datos ómicos

Anna Aguila Griful

1 de April, 2025

## Contents

<b>0. Repositorio GitHub</b>	<b>1</b>
<b>1. Importación de los datos</b>	<b>1</b>
<b>2. Creación de objeto de clase SummarizedExperiment</b>	<b>3</b>
<b>3. Análisis exploratorio</b>	<b>3</b>
3.1. Breve descripción de los datos . . . . .	3
3.2. Estructura de los datos . . . . .	3
3.3. Exploración de los datos numéricos . . . . .	4
<b>4. Interpretación de los resultados</b>	<b>5</b>
4.1. Leucina . . . . .	5
4.2. N.N.Dimetilglicina . . . . .	6
4.3. Quinolinato . . . . .	6
4.4. Valina . . . . .	6
4.5. Conclusiones . . . . .	6
<b>5. Referencias</b>	<b>6</b>

## 0. Repositorio GitHub

Acceso directo al repositorio GitHub con los datos.

## 1. Importación de los datos

Usaremos el dataset **2024-Cachexia** del repositorio de GitHub, que contiene información metabólica de sujetos enfermos (cachexia) y de sujetos sanos (control).

Se ha escogido este dataset debido a la calidad de los datos que contiene. Como se especifica en la descripción de los datos, estos ya ha pasado un control de calidad gracias al cual es posible trabajar directamente con un mínimo pre-procesamiento.

Vemos la estructura de este dataset:

```
## 'data.frame':   77 obs. of  65 variables:
## $ Patient.ID      : chr  "PIF_178" "PIF_087" "PIF_090" "NETL_005_V1" ...
## $ Muscle.loss      : chr  "cachexic" "cachexic" "cachexic" "cachexic" ...
## $ X1.6.Anhydro.beta.D.glucose: num  40.9 62.2 270.4 154.5 22.2 ...
## $ X1.Methylnicotinamide : num  65.4 340.4 64.7 53 73.7 ...
## $ X2.Aminobutyrate  : num  18.7 24.3 12.2 172.4 15.6 ...
```

##	\$ X2.Hydroxyisobutyrate	: num	26.1 41.7 65.4 74.4 83.9 ...
##	\$ X2.Oxoglutarate	: num	71.5 67.4 23.8 1199.9 33.1 ...
##	\$ X3.Aminoisobutyrate	: num	1480.3 116.8 14.3 555.6 29.7 ...
##	\$ X3.Hydroxybutyrate	: num	56.83 43.82 5.64 175.91 76.71 ...
##	\$ X3.Hydroxyisovalerate	: num	10.1 79.8 23.3 25 69.4 ...
##	\$ X3.Indoxylsulfate	: num	567 369 665 412 166 ...
##	\$ X4.Hydroxyphenylacetate	: num	120.3 432.7 292.9 214.9 97.5 ...
##	\$ Acetate	: num	126.5 212.7 314.2 37.3 407.5 ...
##	\$ Acetone	: num	9.49 11.82 4.44 206.44 44.26 ...
##	\$ Adipate	: num	38.1 327 131.6 144 15 ...
##	\$ Alanine	: num	314 871 464 590 1119 ...
##	\$ Asparagine	: num	159.2 157.6 89.1 273.1 42.5 ...
##	\$ Betaine	: num	110 245 117 279 392 ...
##	\$ Carnitine	: num	265.1 120.3 25 200.3 84.8 ...
##	\$ Citrate	: num	3714 2618 863 13630 854 ...
##	\$ Creatine	: num	196.4 212.7 221.4 85.6 105.6 ...
##	\$ Creatinine	: num	16482 15835 24588 20952 6768 ...
##	\$ Dimethylamine	: num	633 608 735 1064 242 ...
##	\$ Ethanolamine	: num	645 488 407 821 365 ...
##	\$ Formate	: num	441 252 250 469 114 ...
##	\$ Fucose	: num	337 198.3 186.8 407.5 26.1 ...
##	\$ Fumarate	: num	7.69 18.92 7.1 96.54 19.69 ...
##	\$ Glucose	: num	395 8691 1353 863 6836 ...
##	\$ Glutamine	: num	871 602 302 1686 433 ...
##	\$ Glycine	: num	2039 1108 620 5064 395 ...
##	\$ Glycolate	: num	685.4 652 141.2 70.8 26.6 ...
##	\$ Guanidoacetate	: num	154 110 183 103 53 ...
##	\$ Hippurate	: num	4582 1737 4316 757 1153 ...
##	\$ Histidine	: num	925 846 284 1043 327 ...
##	\$ Hypoxanthine	: num	97.5 82.3 114.4 223.6 66.7 ...
##	\$ Isoleucine	: num	5.58 8.17 9.3 37.71 40.04 ...
##	\$ Lactate	: num	107 369 750 369 3641 ...
##	\$ Leucine	: num	42.1 77.5 31.5 103.5 101.5 ...
##	\$ Lysine	: num	146.9 284.3 97.5 290 122.7 ...
##	\$ Methylamine	: num	52.5 23.6 18.7 48.9 27.9 ...
##	\$ Methylguanidine	: num	9.97 7.69 4.66 141.17 5.31 ...
##	\$ N.N.Dimethylglycine	: num	23.3 87.4 24.5 40 46.1 ...
##	\$ O.Acetylcarnitine	: num	52.98 50.4 5.58 254.68 45.6 ...
##	\$ Pantothenate	: num	25.8 186.8 145.5 42.5 74.4 ...
##	\$ Pyroglutamate	: num	437 437 713 567 185 ...
##	\$ Pyruvate	: num	21.1 37 29.4 64.1 12.3 ...
##	\$ Quinolate	: num	165.7 73 192.5 86.5 38.1 ...
##	\$ Serine	: num	284 392 296 1249 206 ...
##	\$ Succinate	: num	154.5 244.7 142.6 144 68.7 ...
##	\$ Sucrose	: num	45.1 459.4 160.8 111 75.2 ...
##	\$ Tartrate	: num	97.51 32.79 16.28 837.15 4.53 ...
##	\$ Taurine	: num	1920 1261 4273 1525 469 ...
##	\$ Threonine	: num	184.9 198.3 110 376.1 64.1 ...
##	\$ Trigonelline	: num	943.9 208.5 192.5 992.3 86.5 ...
##	\$ Trimethylamine.N.oxide	: num	2122 639 1153 1451 172 ...
##	\$ Tryptophan	: num	259.8 83.1 82.3 235.1 103.5 ...
##	\$ Tyrosine	: num	290 167.3 60.3 323.8 142.6 ...
##	\$ Uracil	: num	111 47 31.5 30.6 44.3 ...
##	\$ Valine	: num	86.5 110 59.1 102.5 160.8 ...

```
## $ Xylose : num 72.2 192.5 2164.6 125.2 186.8 ...
## $ cis.Aconitate : num 237 334 330 1863 101 ...
## $ myo.Inositol : num 135.6 376.1 86.5 247.2 750 ...
## $ trans.Aconitate : num 51.9 217 58.6 75.9 98.5 ...
## $ pi.Methylhistidine : num 157.6 308 145.5 249.6 84.8 ...
## $ tau.Methylhistidine : num 160.8 130.3 83.9 254.7 79.8 ...
```

De las 65 variables, 63 son de tipo numérica y 2 de tipo carácter. De estas dos últimas, una corresponde al ID del paciente (información irrelevante para el estudio) y la otra define dos grupos de estudio: pacientes con caquexia (*cachexic*) y pacientes sin caquexia (*control*).

## 2. Creación de objeto de clase SummarizedExperiment

Antes que nada, se elimina la primera columna del dataset con el que vamos a trabajar, dado que contiene los IDs de los pacientes, información que no nos interesa. Una vez realizada esta modificación, creamos el objeto de clase `SummarizedExperiment`:

```
sum_exp <- SummarizedExperiment(assays = list(original=data_set))
```

¿Cuáles son sus principales diferencias con la clase `ExpressionSet`?

`SummarizedExperiment` se diferencia de `ExpressionSet` de la siguiente manera:

- Mientras que `ExpressionSet` puede contener únicamente una sola matriz de expresión, `SummarizedExperiment` puede almacenar varias matrices o *assays*.
- `SummarizedExperiment` es más flexible que `ExpressionSet` con el tipo de información con la que puede tratar: `ExpressionSet` está limitado a datos provenientes de microarrays, mientras que `SummarizedExperiment` puede contener información de una variedad más amplia de experimentos, como RNA-Seq o Chip-Seq.

## 3. Análisis exploratorio

### 3.1. Breve descripción de los datos

Medida	Valor
Nombre del archivo	human_cachexia.csv
Numero de observaciones	77
Numero de variables	64
Número de valores NA	0

### 3.2. Estructura de los datos

Vemos qué contiene objeto `SummarizedExperiment` que hemos creado:

```
## class: SummarizedExperiment
## dim: 77 64
## metadata(0):
## assays(1): original
## rownames(77): 1 2 ... 76 77
## rowData names(0):
## colnames(64): Muscle.loss X1.6.Anhydro.beta.D.glucose ...
## pi.Methylhistidine tau.Methylhistidine
## colData names(0):
```

Como se puede observar, el objeto que se ha creado no contiene metadata, y por el momento contiene un solo *assay*. Vamos a crear un segundo *assay* llamado **numeric** que guardaremos en el objeto **sum\_exp** y que contendrá todos los datos en formato numérico para facilitar su exploración.

Para ello:

- Se crea una función propia llamada **to.binary()** que permite convertir un vector a código binario.
- Extraemos los datos de expresión del único assay del dataset y convertimos la variable **Muscle.loss** a código binario usando nuestra función propia. **cachexic** será 1, y **control**, 0.
- Una vez realizada la modificación, volvemos a guardar el dataset en el objeto de clase **SummarizedExperiment**, y al nuevo *assay* que generará le daremos el nombre de **numeric**.

Esta modificación nos permitirá realizar más exploraciones con esta variable.

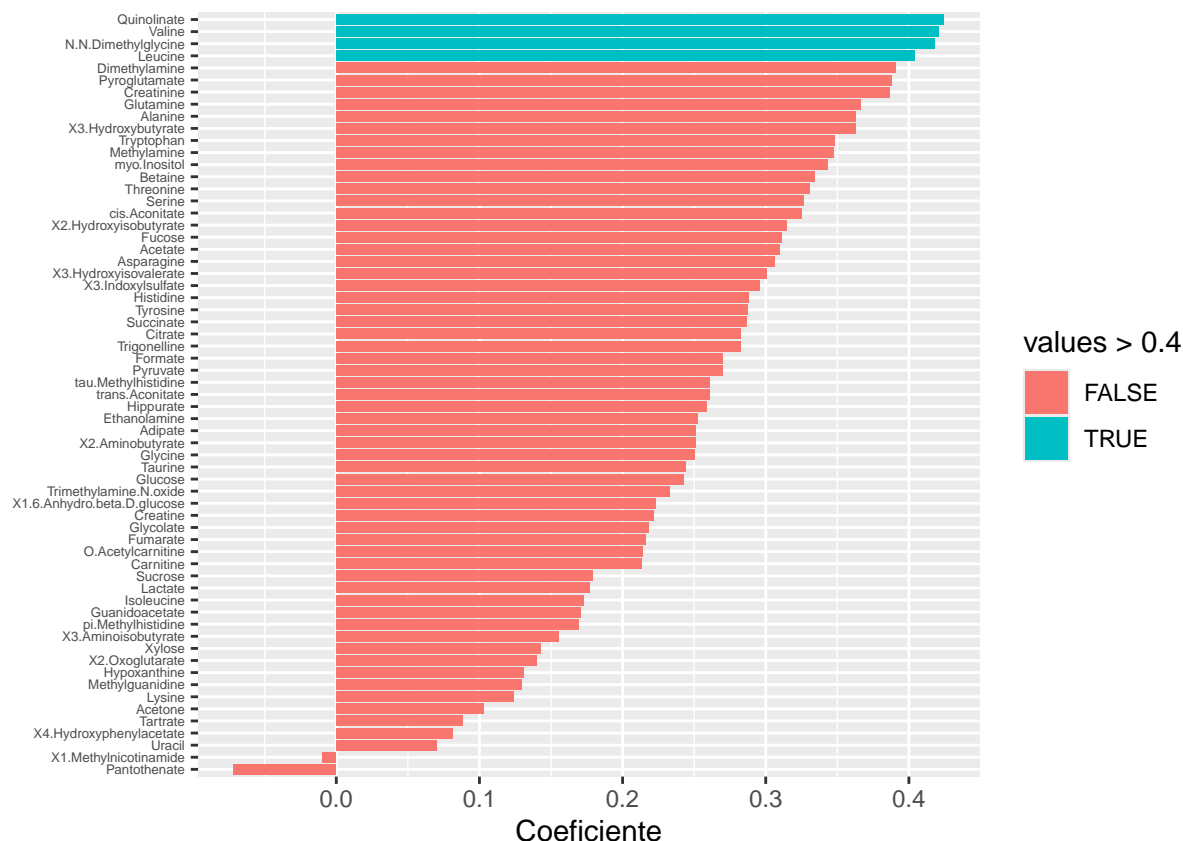
Si observamos el contenido del objeto de clase **SummarizedExperiment** resultante, vemos que ahora contiene dos *assays*:

```
## class: SummarizedExperiment
## dim: 77 64
## metadata(0):
## assays(2): original numeric
## rownames(77): 1 2 ... 76 77
## rowData names(0):
## colnames(64): Muscle.loss X1.6.Anhydro.beta.D.glucose ...
##   pi.Methylhistidine tau.Methylhistidine
## colData names(0):
```

### 3.3. Exploración de los datos numéricos

Realizados estos cambios, es interesante explorar la relación entre la cantidad de cada metabolito y la presencia o no de caquexia, lo cual se consigue obteniendo sus coeficientes de relación.

Dado que este dataset contiene información de 63 metabolitos, para comprender mejor la información que nos proporcionan estos valores, vamos a representarlos en un gráfico de barras usando **ggplot()**. En este gráfico de barras colorearemos de otro color aquellos coeficientes de relación que se encuentren por encima de 0.4:



Como se observa, los metabolitos que presentan un mayor coeficiente de relación con la presencia o ausencia de caquexia son: *Leucine*, *N.N.Dimethylglycine*, *Quinolate*, *Valine*.

## 4. Interpretación de los resultados

Para entender la relación entre estos cuatro metabolitos y la caquexia, necesitamos entender en qué consiste la caquexia [1]:

- Es una afección que consiste en una pérdida de más de un 10% de masa corporal (muscular y grasa) en aquellos pacientes que no están en un proceso activo de perder peso.
- La caquexia ocurre cuándo cambia la manera en que el cuerpo asimila y utiliza los nutrientes, debido a factores como una respuesta inflamatoria o con un aumento del metabolismo.
- Y este cambio suele ser causado por enfermedades en estado avanzado como el cáncer o la SIDA, es muy difícil de revertir y suele ocurrir en el tramo final de la vida del paciente.

La relación entre la presencia de caquexia y la cantidad de los cuatro metabolitos es positiva, es decir, se detecta mayor cantidad de cada metabolito en aquellos pacientes que padecen de caquexia.

### 4.1. Leucina

La leucina es un aminoácido esencial que el cuerpo humano no puede sintetizar por sí mismo. Es decir, es obtenida a través de la dieta. Este aminoácido interviene en el proceso de síntesis de proteína muscular, y por lo tanto tiene un rol relevante en procesos metabólicos relacionados con la caquexia.

Dado que el aumento del consumo de alimentos ricos en leucina o suplementos de leucina ha sido relacionado con una reducción de la pérdida de masa corporal, es posible que el motivo por el que los pacientes con caquexia tienen niveles más altos de este aminoácido sea debido a un aumento del consumo de leucina como tratamiento para su afección. [2,3]

## 4.2. N.N.Dimetilglicina

La dimetilglicina es un compuesto derivado del aminoácido glicina y se utiliza como biomarcador. Su presencia en la orina de un paciente es indicador de la pérdida de masa corporal. Por ello, los pacientes que padecen caquexia tienen un mayor contenido de este compuesto que los pacientes que no tienen esta afección. [4]

## 4.3. Quinolinato

El quinolinato (o ácido quinolínico) es un ácido dicarboxílico unido a una pirimidina y es un producto sintetizado en la ruta de síntesis del triptófano. En altas cantidades tiene un efecto neurotóxico. Por ello, su presencia está relacionada con varias enfermedades neurológicas.

En concreto, este compuesto se produce en grandes cantidades cuándo hay una inflamación, y dado que la presencia de una inflamación es uno de los factores que provocan la caquexia, este es posiblemente la razón por la que se observa una cantidad mayor de este compuesto en pacientes con caquexia. [5]

## 4.4. Valina

La valina es un aminoácido de la misma familia que la leucina que también interviene en el proceso de síntesis de proteínas musculares. Por ello, como con la leucina, se ha observado que la ingestión de una cantidad más elevada de este aminoácido ayuda a reducir la pérdida muscular de pacientes, y por tanto que su mayor presencia en pacientes con caquexia esté relacionada con una dieta diseñada para paliar su afección. [3]

## 4.5. Conclusiones

Aunque el coeficiente de correlación de cada uno de estos componentes no llega a 0.5, puede que este coeficiente sea mayor si se tiene en cuenta el efecto en conjunto de todos estos metabolitos (y de otros que no llegaban a 0.4). Por ello, si se lleva a cabo un estudio estadístico en profundidad en el que se estudiaran estos efectos y correlaciones, se podrán extraer conclusiones más relevantes. Con esta información (aunque tal vez requiera de un *dataset* más extenso) se podría detectar un patrón con técnicas de *machine learning* que permita el diagnóstico de la caquexia disponiendo solamente del perfil metabólico del paciente.

## 5. Referencias

- [1] NIH National Cancer Institute. (n.d.) *Cachexia*. <https://www.cancer.gov/espanol/publicaciones/diccionarios/diccionario-cancer/def/caquexia>
- [2] Beaudry, A. G., & Law, M. L. (2022). Leucine Supplementation in Cancer Cachexia: Mechanisms and a Review of the Pre-Clinical Literature. *Nutrients*, 14(14), 2824. <https://doi.org/10.3390/nu14142824>
- [3] Eley, H. L., Russell, S. T., & Tisdale, M. J. (2007). Effect of branched-chain amino acids on muscle atrophy in cancer cachexia. *The Biochemical journal*, 407(1), 113–120. <https://doi.org/10.1042/BJ20070651>
- [4] Oliveira, M. S., Santo, R. C. E., Silva, J. M. S., Alabarse, P. V. G., Brenol, C. V., Young, S. P., & Xavier, R. M. (2023). Urinary metabolomic biomarker candidates for skeletal muscle wasting in patients with rheumatoid arthritis. *Journal of cachexia, sarcopenia and muscle*, 14(4), 1657–1669. <https://doi.org/10.1002/jcsm.13240>
- [5] Moffett, J. R., Arun, P., Puthillathu, N., Vengilote, R., Ives, J. A., Badawy, A. A., & Namboodiri, A. M. (2020). Quinolate as a Marker for Kynurenine Metabolite Formation and the Unresolved Question of NAD<sup>+</sup> Synthesis During Inflammation and Infection. *Frontiers in immunology*, 11, 31. <https://doi.org/10.3389/fimmu.2020.00031>