

# PAC1 ANÁLISIS DE DATOS ÓMICOS

SERGI PÉREZ

2024-11-06

## Contents

1.INTRODUCCIÓN . . . . .	1
2. MATERIALES Y MÉTODOS . . . . .	1
3. RESULTADOS . . . . .	3
4.DISCUSIÓN . . . . .	4
5. APENDICE . . . . .	4

## 1.INTRODUCCIÓN

En las últimas décadas, la cirugía bariátrica se ha establecido como una intervención eficaz no solo para la pérdida de peso, sino también para el control metabólico y la prevención de la diabetes tipo 2. Las múltiples enfermedades metabólicas asociadas con la obesidad, como la diabetes, a menudo mejoran tras la cirugía, incluso antes de que se produzca una pérdida de peso significativa, aunque las razones detrás de esta mejora aún no están claras.

En la literatura se conocen pocos estudios sobre las consecuencias metabólicas postoperatorias y su relación con la variabilidad en la mejora metabólica. En este estudio, está claro que la ómica que nos ocupa es la metabolómica, la cual es altamente valorada dado a que ayuda en la investigación biomédica a entender mejor las enfermedades metabólicas, y en este caso concreto las relacionadas con la obesidad. En este caso los datos nos ayudan a evaluar el impacto de las intervenciones para la pérdida de peso.

En el estudio se pretende:

- 1) Identificar firmas metabólicas asociadas a la condición basal de los pacientes (metabólicamente no saludables vs saludables)
- 2) Caracterizar metabotipos de respuesta a la cirugía bariátrica basados en la evolución metabólica.
- 3) Correlacionar estos cambios con mejoras metabólicas evaluadas a través de parámetros antropométricos y clínicos.

En el estudio, la condición basal de los pacientes se disipa después de la cirugía. Se han definido dos metabotipos de repuesta independientemente del género, edad o cantidad de pérdida de peso, pero si dependientes de otros factores como resistencia a insulina, colesterol y niveles de ácido úrico.

## 2. MATERIALES Y MÉTODOS

**2.1- Obtención de datos** Los datos brutos se obtienen del repositorio metaboData, repositorio público que incluye varios datasets de diferentes estudios relacionados con la metabolómica, los cuales vienen resumidos en el archivo “Data\_Catalog.xlsx”.

**2.2- Preparación de los datos** Para la preparación de datos para el estudio, se ha generado un contenedor del tipo “Summarized Experiment”, donde se pretende recoger los datos contenidos en el archivo csv “Data\_Values\_S013”, que contiene toda la información clínica i metabólica de 39 pacientes en 5 puntos temporales (T0, T2, T4 Y T5).

Para usar estos datos, de clona el repositorio a nivel local usando el comando system(“git clone https://github.com/nutrimetabolomics/Metabotyping2018.git”) . A continuación obtengo el listado de ficheros usando el comando “list.files”. Con ello ya cargo mis datos en mi Rmd, y los guardo en “datos”.

“datos” va a ser la matriz de datos que usaré para generar el contenedor Summarized Experiment, antes de crear la matriz modifíco los datos quitando en primer lugar la primera columna ya que es repetida y no aporta valor. A continuación, asignamos las 5 primeras columnas de “datos”, a “metabolomic\_data”, que contendrá la información de las filas de nuestro Summarized Experiment. Estas 5 columnas, las quitamos de “datos”.

Por otro lado, añadimos la información relativa a las columnas de nuestro dataframe “datos”, usando el mismo comando anterior cargando el csv “DataInfo\_S013.csv”. A estos valores que nombro “metadatos\_muestras” le quito las primeras 5 columnas para que la dimensión de “datos” y “metadatos\_muestras” sea compatible.

Una vez tenemos los datos preparados podemos generar el contenedor Summarized Experiment que nombraremos “se”. Creamos la matriz de datos y definimos todas las partes necesarias para componer el contenedor. En el tendremos la matriz con todos los datos, así como la información relativa a las filas (cada una de las 39 muestras del estudio) y las columnas (con información relativa a las características de cada una de estas muestras). Por último se añaden metadatos, que son información complementaria a los datos como información relativa al estudio del cual provienen los datos (citación del mismo, doi, etc.). En el repositorio de github además se añade información relativa a todos los metabolitos que se monitorizan en el estudio, se añade en el csv “AAInformation\_S013”.

El objetivo en esta práctica es el poder comparar la variación de los valores de los distintos metabolitos y condiciones de estudio en los 39 pacientes que participan en el estudio. Dividiendo los datos en función del tiempo en el que se midieron podemos observar la tendencia de los mismos y corroborar los efectos que tiene la cirugía bariátrica en estos pacientes.

**2.3- Separación de matriz principal** Para organizar lo datos según su “T”, se divide la matriz “datos” usando el comando “grepl”, en el cual separamos la matriz según tengan T0, T2, T4 O T5. Con estas matrices podremos generar gráficos de densidades de los metabolitos en cada unidad de tiempo.

**2.4-Exploración inicial del “se”** Nos permiten tener un vistazo rápido del contenido de nuestro Summarized Experiment, sus metadatos y dimensiones:

```
se
```

```
## class: SummarizedExperiment
## dim: 39 690
## metadata(2): DOI Citation
## assays(1): counts
## rownames: NULL
## rowData names(5): SUBJECTS SURGERY AGE GENDER Group
## colnames(690): MEDDM_T0 MEDCOL_T0 ... SM.C24.0_T5 SM.C24.1_T5
## colData names(4): X VarName varTpe Description
```

```
summary(se)
```

```
## [1] "SummarizedExperiment object of length 39 with 5 metadata columns"
```

```
dim(se) #dimensiones del dataset
```

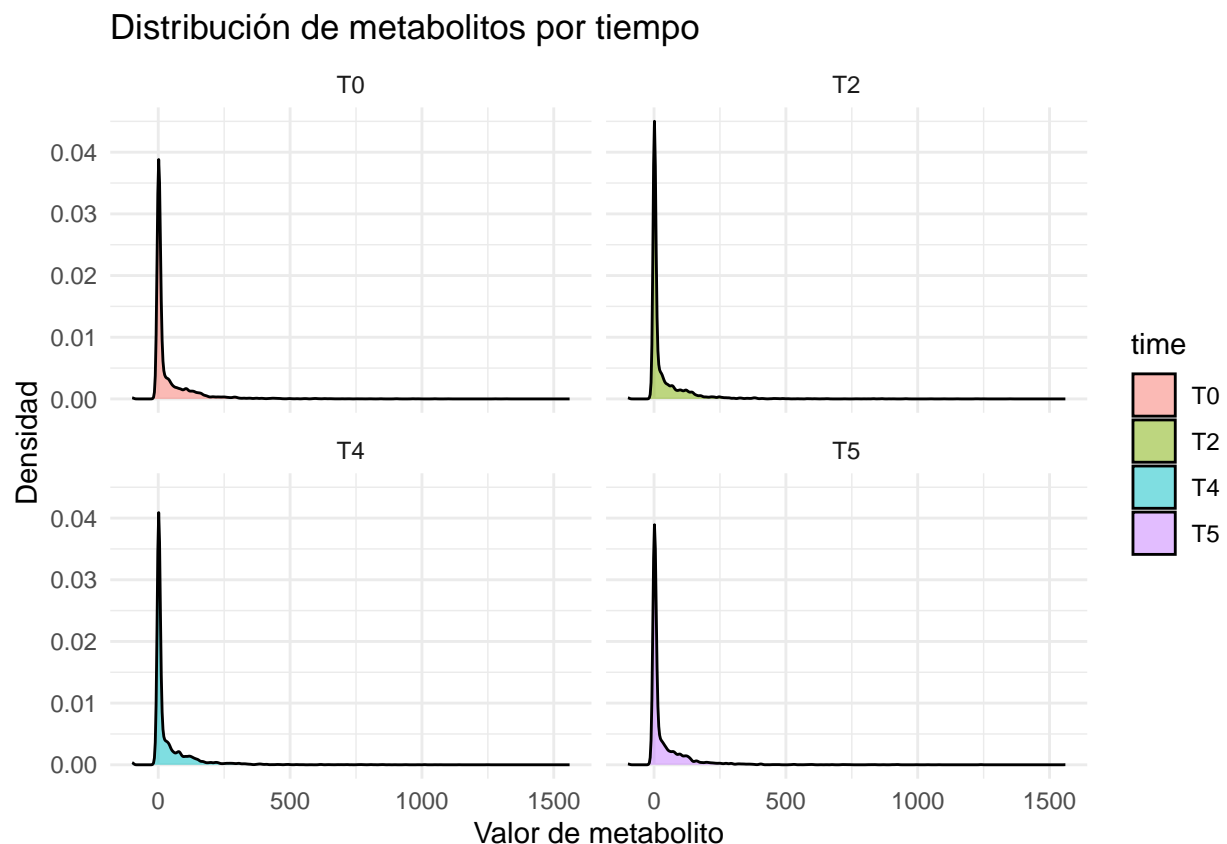
```
## [1] 39 690
```

```
head(colnames(se)) #nombres de los primeros metabolitos/características
```

```
## [1] "MEDDM_T0" "MEDCOL_T0" "MEDINF_T0" "MEDHTA_T0" "GLU_T0" "INS_T0"
```

### 3. RESULTADOS

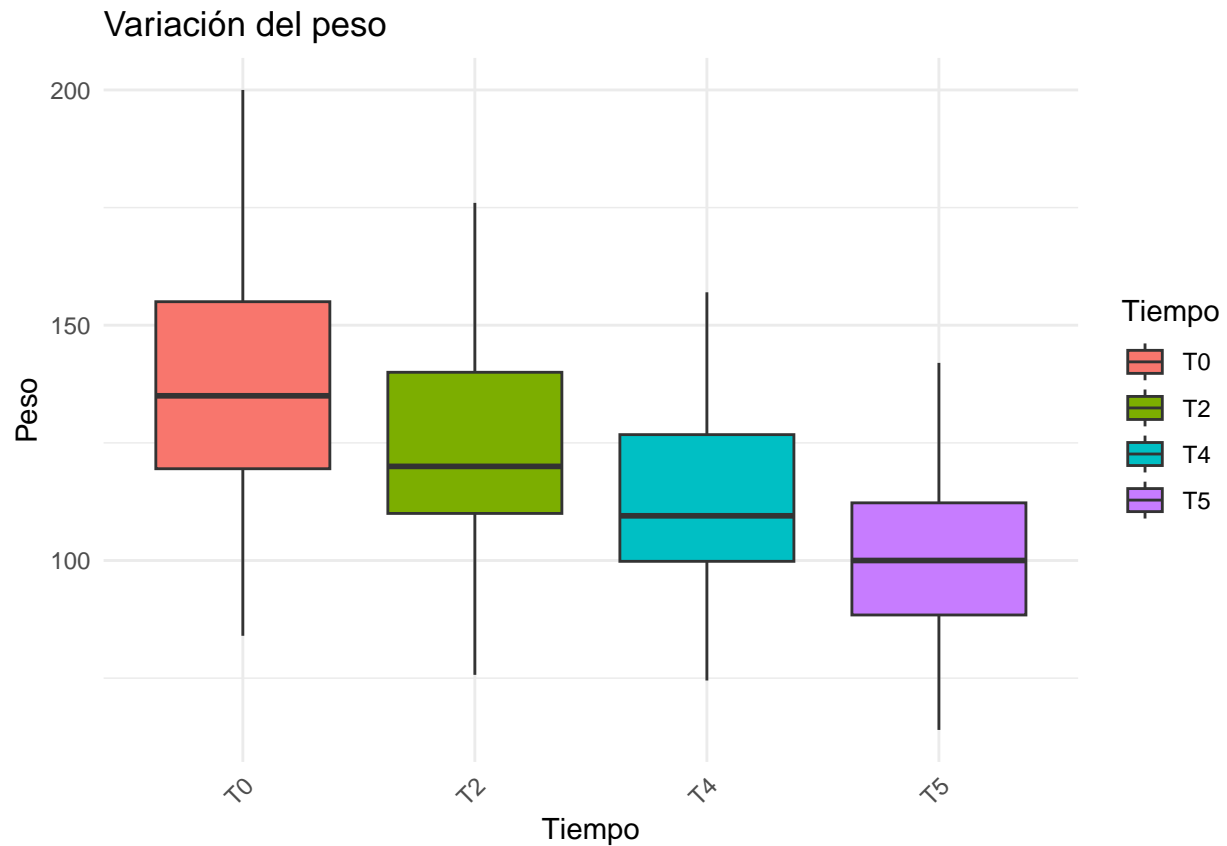
**GRÁFICO DENSIDAD** En primer lugar agrupamos los datos de la matriz de datos inicial. Con estos datos, necesitamos filtrar por columnas según la unidad de tiempo en la que se tomaron los datos de esto metabolitos o características del estudio. Cuando disponemos de las columnas aisladas, podemos usar la librería “ggplot2” para graficar nuestros datos. Para ello generamos un dataframe con nuestros “datos filtrados” y a continuación generamos los 4 gráficos que se corresponden a los 4 tiempos donde se han medido estos metabolitos.



Los gráficos no muestran muchísima variabilidad en los valores de los metabolitos, se aprecia un mayor crecimiento en el valor de estos en el tiempo 5 (T5) en comparación con el resto pero no supone una diferencia muy evidente. Este gráfico nos aporta poco valor en la interpretación y objetivos de este estudio.

**GRÁFICO PESO** Para graficar estos datos, en este caso extraemos de nuestra matriz de datos 4 columnas para conocer la variación del peso en los pacientes en los 4 tiempos que se miden en el estudio. Como en

el ejercicio anterior, creamos un dataframe con nuestros datos seleccionados anteriormente y en este caso dibujamos un boxplot con nuestros datos:



Como era de esperar los pesos disminuyen a medida que el tiempo post-cirugía va pasando. Las diferencias parecen bastante significativas ya que si comparamos las medias partimos de una media cercana a 150 en T0 que consigue disminuir a T5, a casi 100.

#### 4.DISCUSIÓN

Dado que la obtención del contenedor “se” ha supuesto una dificultad para la correcta obtención de datos, no he podido realizar una exploración de los mismos como me hubiese gustado. Aun así he podido navegar por mis datos y hacer breves estudios extrayendo datos del contenedor creado. También añadir información extra de los mismos (metadatos) y graficar algunos ejemplos de cómo se encuentran los datos y su variación en las distintas mediciones en el tiempo.

Considero que el repositorio y en concreto el dataframe elegido tiene mucho más potencial para explorar y que es interesante su revisión a futuro.

#### 5. APENDICE

**REPOSITORIO GITHUB** <https://github.com/SergiP99/Perez-Justribo-Sergi-PEC1.git>

```
system("git clone https://github.com/nutrimetabolomics/Metabotyping2018.git")
```

## OBTENCIÓN DE DATOS

```
## [1] 128
```

```
list.files("Metabotyping2018/Datasets")
```

```
## [1] "AAInformation_S006.csv" "DataInfo_S013.csv"      "DataValues_S013.csv"
```

```
datos <- read.csv("Metabotyping2018/Datasets/DataValues_S013.csv", header = TRUE, stringsAsFactors = FALSE)
```

```
datos <- datos[, -1]
```

```
matabolomic_data<-datos[,1:5]
```

```
datos <- datos[, -c(1:5)]
```

```
metadatos_muestras <- read.csv("Metabotyping2018/Datasets/DataInfo_S013.csv", header = TRUE, stringsAsFactors = FALSE)
```

```
metadatos_muestras <- metadatos_muestras[ -(1:5), ]
```

```
se <- SummarizedExperiment(  
  assays = list(counts = as.matrix(datos)),      # Matriz de datos  
  colData = metadatos_muestras,  
  rowData = matabolomic_data  
)
```

```
metadata(se)$DOI <- "https://doi.org/10.1371/journal.pone.0198214"
```

```
metadata(se)$Citation <- "Palau-Rodriguez, M., Tulipani, S., Marco-Ramell, A., Miñarro, A., Jáuregui, O"
```

```
se
```

```
## class: SummarizedExperiment  
## dim: 39 690  
## metadata(2): DOI Citation  
## assays(1): counts  
## rownames: NULL  
## rowData names(5): SUBJECTS SURGERY AGE GENDER Group  
## colnames(690): MEDDM_T0 MEDCOL_T0 ... SM.C24.0_T5 SM.C24.1_T5  
## colData names(4): X VarName varTpe Description
```

```
summary(se)
```

```
## [1] "SummarizedExperiment object of length 39 with 5 metadata columns"
```

```
# Dividir la matriz de datos en 4 sub-matrices
```

```
matriz_T0 <- assay(se)[, grepl("_T0$", colnames(assay(se)))] # Filtra las columnas que terminan en _T0
```

```
matriz_T2 <- assay(se)[, grepl("_T2$", colnames(assay(se)))] # Filtra las columnas que terminan en _T2
```

```
matriz_T4 <- assay(se)[, grepl("_T4$", colnames(assay(se)))] # Filtra las columnas que terminan en _T4
```

```
matriz_T5 <- assay(se)[, grepl("_T5$", colnames(assay(se)))] # Filtra las columnas que terminan en _T5
```

```

raw_data <- assay(se, "counts")

# filtrar las columnas correspondientes a los tiempos T0, T2, T4, T5
t0_data <- raw_data[, grep("_T0", colnames(raw_data))]
t2_data <- raw_data[, grep("_T2", colnames(raw_data))]
t4_data <- raw_data[, grep("_T4", colnames(raw_data))]
t5_data <- raw_data[, grep("_T5", colnames(raw_data))]

t2_data <- t2_data[, !grepl("lysoPC.a.C14.0_T2", colnames(t2_data))]

library(ggplot2)

data_for_density <- data.frame(
  value = c(as.vector(t0_data), as.vector(t2_data), as.vector(t4_data), as.vector(t5_data)),
  time = rep(c("T0", "T2", "T4", "T5"), each = nrow(t0_data) * ncol(t0_data))
)

ggplot(data_for_density, aes(x = value, fill = time)) +
  geom_density(alpha = 0.5) +
  facet_wrap(~time) +
  theme_minimal() +
  labs(title = "Distribución de metabolitos por tiempo ",
       x = "Valor de metabolito", y = "Densidad")

```

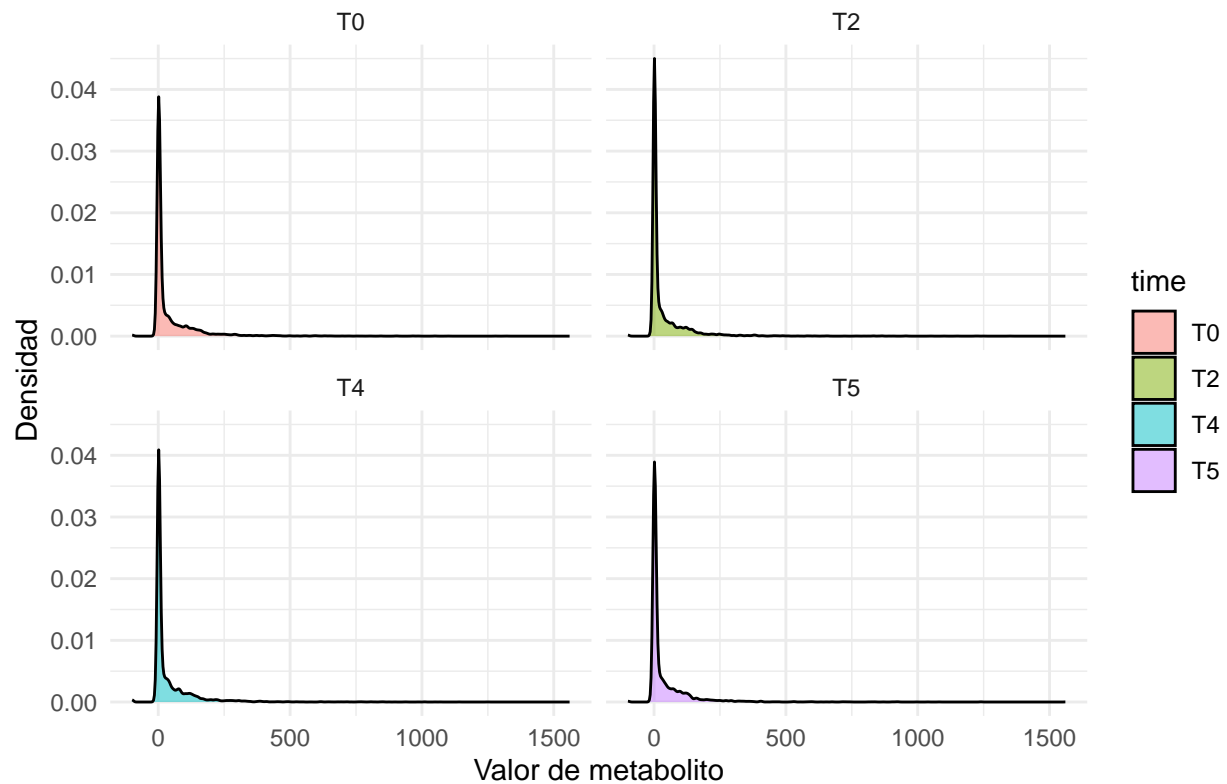
## CREACIÓN GRÁFICOS

```

## Warning: Removed 3350 rows containing non-finite outside the scale range
## ('stat_density()').

```

## Distribución de metabolitos por tiempo



```
raw_data1 <- as.data.frame(raw_data)

peso_t0 <- raw_data1$PESO_T0
peso_t2 <- raw_data1$PESO_T2
peso_t4 <- raw_data1$PESO_T4
peso_t5 <- raw_data1$PESO_T5

peso_data <- data.frame(
  Peso = c(peso_t0, peso_t2, peso_t4, peso_t5),
  Tiempo = rep(c("T0", "T2", "T4", "T5"), each = nrow(raw_data))
)

ggplot(peso_data, aes(x = Tiempo, y = Peso, fill = Tiempo)) +
  geom_boxplot() +
  theme_minimal() +
  labs(title = "Variación del peso", x = "Tiempo", y = "Peso") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

```
## Warning: Removed 10 rows containing non-finite outside the scale range
## ('stat_boxplot()').
```

