

# Análisis de datos ómicos (M0-157) Primera prueba de evaluación continua

Pablo Gómez Rodríguez

2024-04-01

## Contents

Selección del dataset de metabolómica . . . . .	1
Contenedor tipo SummarizedExperiment . . . . .	1
Exploración del dataset . . . . .	3
Creación del repositorio en Github . . . . .	7

### Selección del dataset de metabolómica

Los datos para esta actividad se han escogido de una lista ubicada en un repositorio en Github procedente de **nutrimetabolomics**. De la carpeta de *Datasets* escogemos la de *Cachexia* por ser de las más recientes. Para ello empleamos los datos en formato original para poder importarlos directamente desde la página, acceso que se encuentra visible en el botón de **Raw**.

```
library(readr)
met_cachexia <- read_csv("https://raw.githubusercontent.com/nutrimetabolomics/metaboData/refs/heads/main/0157_Cachexia.csv")

## Rows: 77 Columns: 65
## -- Column specification -----
## Delimiter: ","
## chr (2): Patient ID, Muscle loss
## dbl (63): 1,6-Anhydro-beta-D-glucose, 1-Methylnicotinamide, 2-Aminobutyrate,...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

### Contenedor tipo SummarizedExperiment

SumarizedExperiment tiene la ventaja de poder manejar diferentes resultados experimentales siempre que tengan la misma dimensión, además de coordinar el metadata y los datos del ensayo. Esto es útil pues permite mantenerlos en sincronía y efectuar cambios en ambos al mismo tiempo de querer eliminar una muestra o cambiarla si se diera el caso, lo que lo hace más flexible que ExpressionSet.

```
# Cargamos los paquetes necesarios
library(BiocManager)
library(SummarizedExperiment)

## Cargando paquete requerido: MatrixGenerics
## Cargando paquete requerido: matrixStats
## Warning: package 'matrixStats' was built under R version 4.4.2
```

```

##
## Adjuntando el paquete: 'MatrixGenerics'
## The following objects are masked from 'package:matrixStats':
##
##   colAlls, colAnyNAs, colAnys, colAvgPerRowSet, 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, rowAvgPerColSet,
##   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
# Creamos los objetos de metadatos
row_data <- Dataframe(row.names = rownames(met_cachexia))
col_data <- Dataframe(col.names = colnames(met_cachexia))
# Creamos el contenedor de tipo SummarizedExperiment
se <- SummarizedExperiment(assays = list(counts = as.matrix(met_cachexia)),
                           rowData = row_data,
                           colData = col_data)
```

## Exploración del dataset

```
# Obtenemos información sobre las dimensiones del dataset
dim(se)

## [1] 77 65
dimnames(se)

## [[1]]
## [1] "1" "2" "3" "4" "5" "6" "7" "8" "9" "10" "11" "12" "13" "14" "15"
## [16] "16" "17" "18" "19" "20" "21" "22" "23" "24" "25" "26" "27" "28" "29" "30"
## [31] "31" "32" "33" "34" "35" "36" "37" "38" "39" "40" "41" "42" "43" "44" "45"
## [46] "46" "47" "48" "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" "74" "75"
## [76] "76" "77"
##
## [[2]]
## [1] "Patient ID" "Muscle loss"
## [3] "1,6-Anhydro-beta-D-glucose" "1-Methylnicotinamide"
```

```
## [5] "2-Aminobutyrate"      "2-Hydroxyisobutyrate"
## [7] "2-Oxoglutarate"      "3-Aminoisobutyrate"
## [9] "3-Hydroxybutyrate"    "3-Hydroxyisovalerate"
## [11] "3-Indoxylsulfate"     "4-Hydroxyphenylacetate"
## [13] "Acetate"              "Acetone"
## [15] "Adipate"              "Alanine"
## [17] "Asparagine"           "Betaine"
## [19] "Carnitine"            "Citrate"
## [21] "Creatine"             "Creatinine"
## [23] "Dimethylamine"        "Ethanolamine"
## [25] "Formate"              "Fucose"
## [27] "Fumarate"             "Glucose"
## [29] "Glutamine"            "Glycine"
## [31] "Glycolate"            "Guanidoacetate"
## [33] "Hippurate"            "Histidine"
## [35] "Hypoxanthine"         "Isoleucine"
## [37] "Lactate"              "Leucine"
## [39] "Lysine"               "Methylamine"
## [41] "Methylguanidine"      "N,N-Dimethylglycine"
## [43] "O-Acetylcarnitine"    "Pantothenate"
## [45] "Pyroglutamate"        "Pyruvate"
## [47] "Quinolate"            "Serine"
## [49] "Succinate"            "Sucrose"
## [51] "Tartrate"             "Taurine"
## [53] "Threonine"            "Trigonelline"
## [55] "Trimethylamine N-oxide" "Tryptophan"
## [57] "Tyrosine"             "Uracil"
## [59] "Valine"               "Xylose"
## [61] "cis-Aconitate"         "myo-Inositol"
## [63] "trans-Aconitate"      "pi-Methylhistidine"
## [65] "tau-Methylhistidine"
```

```
# Obtenemos un resumen de los datos
summary(assay(se))
```

```
## Patient ID      Muscle loss      1,6-Anhydro-beta-D-glucose
## Length:77      Length:77      Length:77
## Class :character Class :character Class :character
## Mode :character Mode :character Mode :character
## 1-Methylnicotinamide 2-Aminobutyrate 2-Hydroxyisobutyrate
## Length:77      Length:77      Length:77
## Class :character Class :character Class :character
## Mode :character Mode :character Mode :character
## 2-Oxoglutarate 3-Aminoisobutyrate 3-Hydroxybutyrate 3-Hydroxyisovalerate
## Length:77      Length:77      Length:77      Length:77
## Class :character Class :character Class :character Class :character
## Mode :character Mode :character Mode :character Mode :character
## 3-Indoxylsulfate 4-Hydroxyphenylacetate Acetate
## Length:77      Length:77      Length:77
## Class :character Class :character Class :character
## Mode :character Mode :character Mode :character
## Acetone      Adipate      Alanine      Asparagine
## Length:77      Length:77      Length:77      Length:77
## Class :character Class :character Class :character Class :character
## Mode :character Mode :character Mode :character Mode :character
```

## Betaine	Carnitine	Citrate	Creatine
## Length:77	Length:77	Length:77	Length:77
## Class :character	Class :character	Class :character	Class :character
## Mode :character	Mode :character	Mode :character	Mode :character
## Creatinine	Dimethylamine	Ethanolamine	Formate
## Length:77	Length:77	Length:77	Length:77
## Class :character	Class :character	Class :character	Class :character
## Mode :character	Mode :character	Mode :character	Mode :character
## Fucose	Fumarate	Glucose	Glutamine
## Length:77	Length:77	Length:77	Length:77
## Class :character	Class :character	Class :character	Class :character
## Mode :character	Mode :character	Mode :character	Mode :character
## Glycine	Glycolate	Guanidoacetate	Hippurate
## Length:77	Length:77	Length:77	Length:77
## Class :character	Class :character	Class :character	Class :character
## Mode :character	Mode :character	Mode :character	Mode :character
## Histidine	Hypoxanthine	Isoleucine	Lactate
## Length:77	Length:77	Length:77	Length:77
## Class :character	Class :character	Class :character	Class :character
## Mode :character	Mode :character	Mode :character	Mode :character
## Leucine	Lysine	Methylamine	Methylguanidine
## Length:77	Length:77	Length:77	Length:77
## Class :character	Class :character	Class :character	Class :character
## Mode :character	Mode :character	Mode :character	Mode :character
## N,N-Dimethylglycine	O-Acetylcarnitine	Pantothenate	Pyroglutamate
## Length:77	Length:77	Length:77	Length:77
## Class :character	Class :character	Class :character	Class :character
## Mode :character	Mode :character	Mode :character	Mode :character
## Pyruvate	Quinolate	Serine	Succinate
## Length:77	Length:77	Length:77	Length:77
## Class :character	Class :character	Class :character	Class :character
## Mode :character	Mode :character	Mode :character	Mode :character
## Sucrose	Tartrate	Taurine	Threonine
## Length:77	Length:77	Length:77	Length:77
## Class :character	Class :character	Class :character	Class :character
## Mode :character	Mode :character	Mode :character	Mode :character
## Trigonelline	Trimethylamine N-oxide	Tryptophan	
## Length:77	Length:77	Length:77	
## Class :character	Class :character	Class :character	
## Mode :character	Mode :character	Mode :character	
## Tyrosine	Uracil	Valine	Xylose
## Length:77	Length:77	Length:77	Length:77
## Class :character	Class :character	Class :character	Class :character
## Mode :character	Mode :character	Mode :character	Mode :character
## cis-Aconitate	myo-Inositol	trans-Aconitate	pi-Methylhistidine
## Length:77	Length:77	Length:77	Length:77
## Class :character	Class :character	Class :character	Class :character
## Mode :character	Mode :character	Mode :character	Mode :character
## tau-Methylhistidine			
## Length:77			
## Class :character			
## Mode :character			

```
# Exploramos los metadatos
head(rowData(se))
```

```
## DataFrame with 6 rows and 0 columns
```

```
head(colData(se))
```

```
## DataFrame with 6 rows and 1 column
```

```
##                               col.names
##                               <character>
## Patient ID                    Patient ID
## Muscle loss                   Muscle loss
## 1,6-Anhydro-beta-D-glucose 1,6-Anhydro-beta-D-g..
## 1-Methylnicotinamide          1-Methylnicotinamide
## 2-Aminobutyrate               2-Aminobutyrate
## 2-Hydroxyisobutyrate          2-Hydroxyisobutyrate
```

Primero obtenemos el número de variables (= columnas - 2, por ser la primera el id del paciente y la segunda si presentaba Cachexia o era control), siendo **63** y el número de sujetos (= filas) siendo **77**. Entre las variables encontramos las biomoléculas que se midieron, estando listadas todas por el nombre de las columnas.

```
library(ggplot2)
```

```
anhy <- ggplot(met_cachexia, aes(x = met_cachexia$'Muscle loss', y = met_cachexia$'1,6-Anhydro-beta-D-g..
```

```
methy1 <- ggplot(met_cachexia, aes(x = met_cachexia$'Muscle loss', y = met_cachexia$'1-Methylnicotinami
```

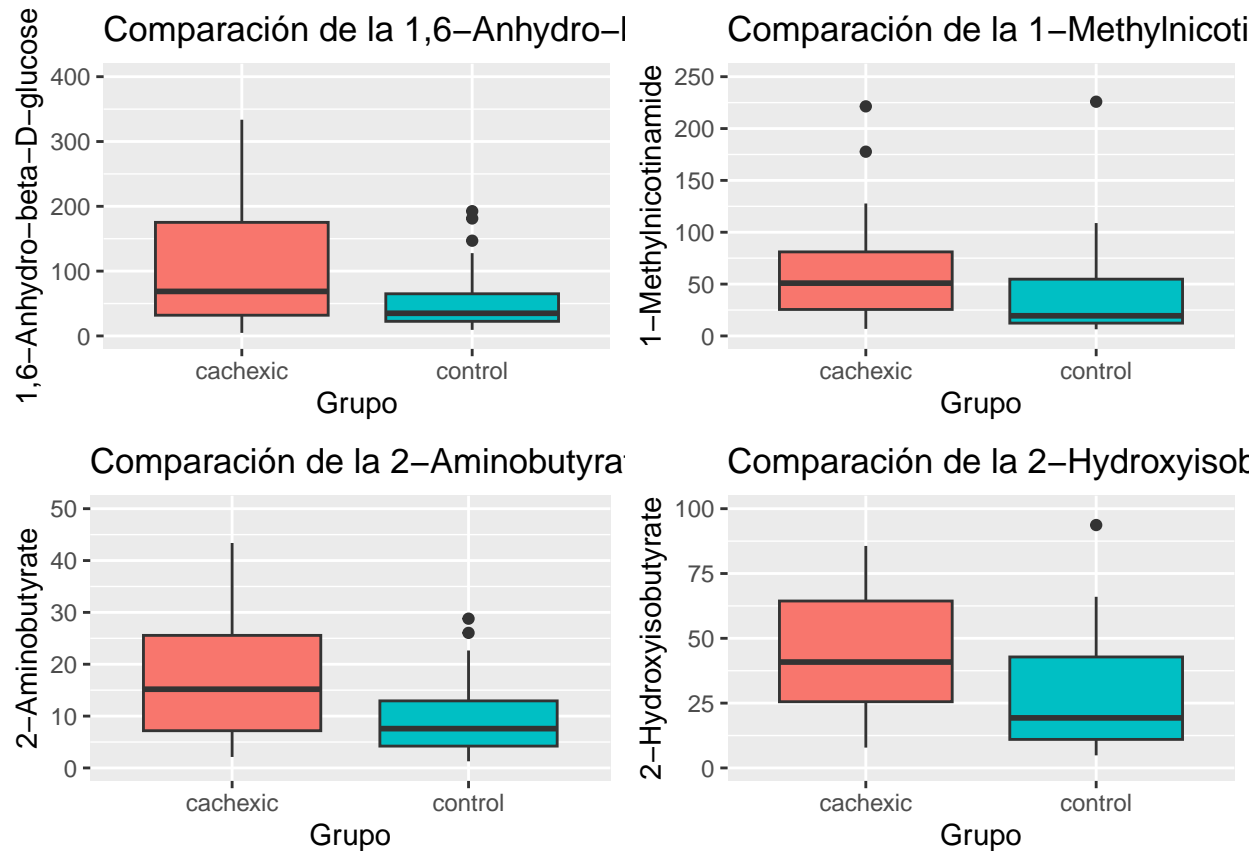
```
amino <- ggplot(met_cachexia, aes(x = met_cachexia$'Muscle loss', y = met_cachexia$'2-Aminobutyrate', f
```

```
hydro <- ggplot(met_cachexia, aes(x = met_cachexia$'Muscle loss', y = met_cachexia$'2-Hydroxyisobutyrat
```

```
library(ggpubr)
```

```
## Warning: package 'ggpubr' was built under R version 4.4.3
```

```
ggarrange(anhy, methy1, amino, hydro, ncol= 2, nrow =2)
```



Creamos unas gráficas de cajas para observar los datos y poder tener una comparativa rápida entre el grupo control y el que padece la caquexia. Para ello empleamos el paquete **ggplot2**, creando un gráfico que emplee una de las biomoléculas analizadas y separando en grupo *control* y *cachexic*. Luego ajustamos la escala del eje y para que sea más visible la diferencia media. Por último unimos el grupo de tablas creadas en una sola imagen con la función **ggarrange** del paquete **ggpubr**. Sería posible hacerlo para las 63 variables pero quedaría masificado y se debería hacer un análisis primero de las que cuentan con diferencias significativas estadísticamente.

### Creación del repositorio en Github

Para esta parte, creamos una nueva cuenta con el correo propio de la universidad (especificando ser estudiantes y solo trabajar nosotros dentro del repositorio). Dentro de la cuenta creamos un nuevo repositorio donde agregar los elementos que son necesarios para la entrega de la prueba:

- Informe
- Objeto contenedor de los datos y metadatos en formato binario (.Rda)
- Datos en formato texto
- Metadatos en un archivo markdown (.Rmd)
  - Archivo README (se marca esta opción a la hora de crear el repositorio)

La dirección del repositorio la encontramos dentro del desplegable de *Code* y es la siguiente: <https://github.com/pblgmzrdrz/PEC1.git>