PEC1_Alejandro Gonzalez

2024-11-06

En primer lugar seleccionamos un dataset del repositorio "https://github.com/nutrimetabolomics/ metaboData/". A través de ese enlace entramos en un repositorio de GitHub denominado "metabo-Data" que contiene algunos datasets de metabolómica obtenidos de fuentes públicas. Dentro de la carpeta "Datasets", he decidido trabajar con los datos de "2024-Cachexia". Me descargo el archivo "human_cachexia.csv", pinchando sobre el se visualizan los datos y arriba a la derecha está la opción de "Download raw file".

A contiuación cargamos los datos en R

```r
# Instalamos y cargamos las librerías necesarias
if (!requireNamespace("BiocManager", quietly = TRUE))
  install.packages("BiocManager")
BiocManager::install(c("SummarizedExperiment", "BiocGenerics"))
```

```
## Bioconductor version 3.19 (BiocManager 1.30.25), R 4.4.1 (2024-06-14)
```

```
## Warning: package(s) not installed when version(s) same as or greater than current; use
##    `force = TRUE` to re-install: 'SummarizedExperiment' 'BiocGenerics'
```

```r
# Cargamos las librerías adicionales para análisis y visualización
library(SummarizedExperiment)
```

```
## Loading required package: MatrixGenerics
```

```
## Loading required package: matrixStats
```

```
##
## Attaching package: '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


## Loading required package: GenomicRanges


## Loading required package: stats4


## Loading required package: BiocGenerics


##
## Attaching package: '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, setdiff, table,
##      tapply, union, unique, unsplit, which.max, which.min


## Loading required package: S4Vectors


##
## Attaching package: 'S4Vectors'


## The following object is masked from 'package:utils':
##
##      findMatches


## The following objects are masked from 'package:base':
##
##      expand.grid, I, unname


## Loading required package: IRanges


## Loading required package: GenomeInfoDb


## Loading required package: Biobase


## Welcome to Bioconductor
##
##      Vignettes contain introductory material; view with
##      'browseVignettes()'. To cite Bioconductor, see
##      'citation("Biobase")', and for packages 'citation("pkgname")'.
```

```
##
## Attaching package: 'Biobase'

## The following object is masked from 'package:MatrixGenerics':
##
##     rowMedians

## The following objects are masked from 'package:matrixStats':
##
##     anyMissing, rowMedians
```

```r
library(BiocGenerics)
library(ggplot2)

# Cargamos los datos
dataset <- read.csv("/Users/alejandrogonzalez/Desktop/Alex/Curro/UOC/Master Bioinformatica y Bioestadis
```

El dataset contiene principalmente columnas con parámetros numéricos, pero la primera columna corresponde
a texto con los los identificadores de la condición ("cachexic" y "control"). Procesamos el dataset para separar
los datos de texto y los datos numéricos

```r
# Procesamos el dataset para separar los datos de texto y los datos numéricos
# La primera columna contiene los identificadores de la condición ("cachexic" y "control")
conditions <- dataset[, 1]                # Guardamos la columna de condiciones
expression_data <- as.matrix(dataset[, -1])  # Seleccionamos solo las columnas numéricas
```

Creamos el objeto SummarizedExperiment

```r
# Creamos el objeto SummarizedExperiment
# Primero creamos un dataframe con los metadatos de filas (id de pacientes) y columnas (parámetros de m
row_metadata <- data.frame(
  Condition = conditions,  # Usamos las condiciones como metadatos de los pacientes
  stringsAsFactors = FALSE
)

# Usamos los nombres de las columnas como nombres de los parámetros
col_metadata <- data.frame(
  Parameter = colnames(expression_data),
  Description = "Parámetro metabólico",
  stringsAsFactors = FALSE
)

# Creamos el objeto SummarizedExperiment
se_object <- SummarizedExperiment(
  assays = list(counts = expression_data),
  rowData = row_metadata,
  colData = col_metadata
)
```

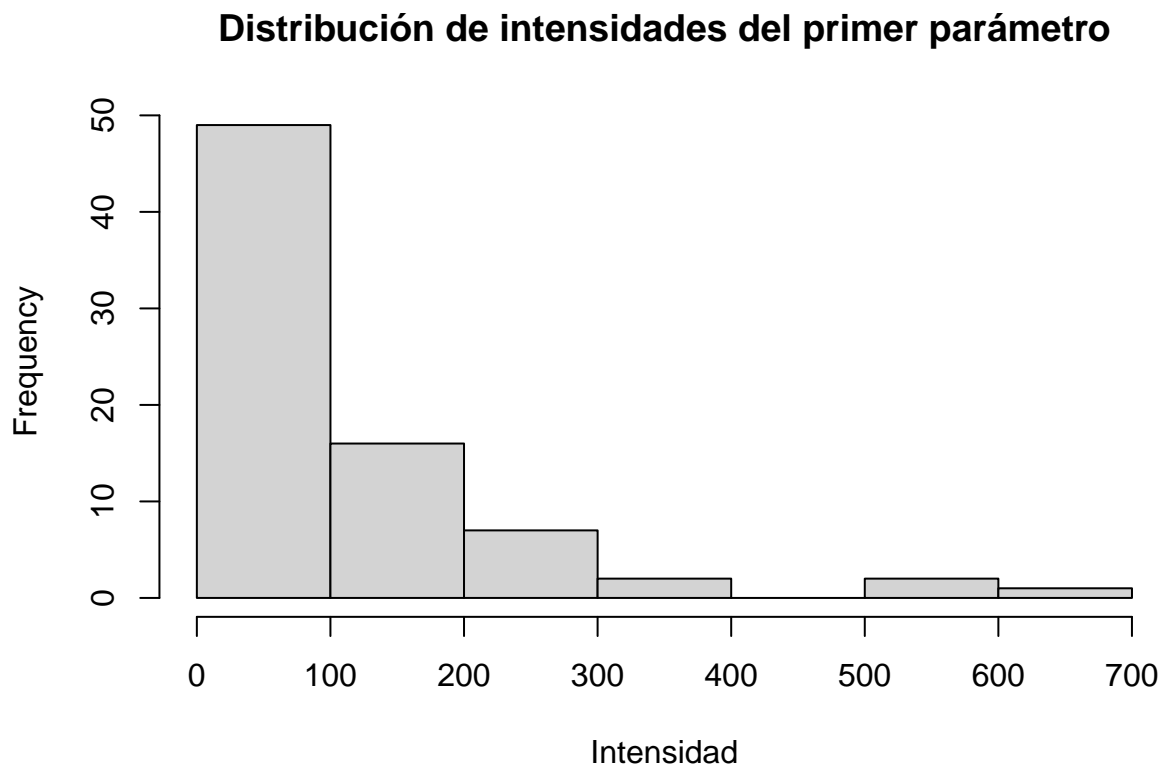Una vez creado el objeto SummarizedExperiment, exploramos el dataset

3

```r
# Realizamos una Visualización básica de los datos, su distribución y resumen estadístico
summary(assay(se_object))
```

```
##  X1.6.Anhydro.beta.D.glucose X1.Methylnicotinamide X2.Aminobutyrate
##  Min.   :  4.71              Min.   :   6.42       Min.   :  1.28
##  1st Qu.: 28.79              1st Qu.:  15.80       1st Qu.:  5.26
##  Median : 45.60              Median :  36.60       Median : 10.49
##  Mean   :105.63              Mean   :  71.57       Mean   : 18.16
##  3rd Qu.:141.17              3rd Qu.:  73.70       3rd Qu.: 19.49
##  Max.   :685.40              Max.   :1032.77       Max.   :172.43
##  X2.Hydroxyisobutyrate X2.Oxoglutarate  X3.Aminoisobutyrate X3.Hydroxybutyrate
##  Min.   : 4.85         Min.   :   5.53  Min.   :   2.61     Min.   :  1.70
##  1st Qu.:15.80         1st Qu.:  22.42  1st Qu.:  11.70     1st Qu.:  5.99
##  Median :32.46         Median :  55.15  Median :  22.65     Median : 11.70
##  Mean   :37.25         Mean   : 145.09  Mean   :  76.76     Mean   : 21.72
##  3rd Qu.:54.60         3rd Qu.:  92.76  3rd Qu.:  56.26     3rd Qu.: 29.96
##  Max.   :93.69         Max.   :2465.13  Max.   :1480.30     Max.   :175.91
##  X3.Hydroxyisovalerate X3.Indoxylsulfate X4.Hydroxyphenylacetate
##  Min.   :  0.92        Min.   :  27.66   Min.   : 15.49
##  1st Qu.:  5.26        1st Qu.:  82.27   1st Qu.: 41.68
##  Median : 12.55        Median : 144.03   Median : 70.11
##  Mean   : 21.65        Mean   : 218.88   Mean   :112.02
##  3rd Qu.: 30.27        3rd Qu.: 333.62   3rd Qu.:145.47
##  Max.   :164.02        Max.   :1043.15   Max.   :796.32
##     Acetate          Acetone          Adipate          Alanine
##  Min.   :  3.49   Min.   :  2.29   Min.   :  1.55   Min.   :  16.78
##  1st Qu.: 16.28   1st Qu.:  4.95   1st Qu.:  6.11   1st Qu.:  78.26
##  Median : 39.65   Median :  7.10   Median : 10.18   Median : 194.42
##  Mean   : 66.14   Mean   : 11.43   Mean   : 24.76   Mean   : 273.56
##  3rd Qu.: 86.49   3rd Qu.: 10.49   3rd Qu.: 19.11   3rd Qu.: 399.41
##  Max.   :411.58   Max.   :206.44   Max.   :327.01   Max.   :1312.91
##    Asparagine        Betaine         Carnitine         Citrate
##  Min.   :  6.69   Min.   :  2.29   Min.   :  2.18   Min.   :   59.74
##  1st Qu.: 20.49   1st Qu.: 28.79   1st Qu.: 14.44   1st Qu.:  788.40
##  Median : 42.10   Median : 64.72   Median : 23.81   Median : 1790.05
##  Mean   : 62.28   Mean   : 90.32   Mean   : 52.09   Mean   : 2235.35
##  3rd Qu.: 89.12   3rd Qu.:127.74   3rd Qu.: 60.95   3rd Qu.: 3071.74
##  Max.   :273.14   Max.   :391.51   Max.   :487.85   Max.   :13629.61
##     Creatine        Creatinine     Dimethylamine     Ethanolamine
##  Min.   :   2.75   Min.   : 1002   Min.   :  41.26   Min.   :  16.12
##  1st Qu.:  17.64   1st Qu.: 3498   1st Qu.: 142.59   1st Qu.:  86.49
##  Median :  44.26   Median : 7631   Median : 304.90   Median : 204.38
##  Mean   : 126.83   Mean   : 8734   Mean   : 358.17   Mean   : 276.26
##  3rd Qu.: 117.92   3rd Qu.:12333   3rd Qu.: 454.86   3rd Qu.: 407.48
##  Max.   :1863.11   Max.   :33860   Max.   :1556.20   Max.   :1436.55
##     Formate          Fucose          Fumarate         Glucose
##  Min.   :   6.42   Min.   :  5.70   Min.   : 0.79   Min.   :  26.84
##  1st Qu.:  53.52   1st Qu.: 29.37   1st Qu.: 2.23   1st Qu.:  80.64
##  Median :  95.58   Median : 61.56   Median : 4.10   Median : 210.61
##  Mean   : 147.40   Mean   : 88.67   Mean   : 8.44   Mean   : 559.85
##  3rd Qu.: 167.34   3rd Qu.:123.97   3rd Qu.: 7.85   3rd Qu.: 407.48
##  Max.   :1480.30   Max.   :407.48   Max.   :96.54   Max.   :8690.62
##     Glutamine         Glycine         Glycolate        Guanidoacetate
```

```
##   Min.   :  23.34   Min.   :  38.09   Min.   :  5.42   Min.    :  7.03
##   1st Qu.: 113.30   1st Qu.: 262.43   1st Qu.: 50.91   1st Qu.: 33.78
##   Median : 225.88   Median : 528.48   Median :130.32   Median : 64.72
##   Mean   : 306.87   Mean   : 880.72   Mean   :187.99   Mean    : 86.37
##   3rd Qu.: 445.86   3rd Qu.:1096.63   3rd Qu.:267.74   3rd Qu.:108.85
##   Max.   :1685.81   Max.   :5064.45   Max.   :720.54   Max.    :561.16
##     Hippurate        Histidine       Hypoxanthine       Isoleucine
##   Min.   :   92.76  Min.   :  14.15  Min.   :  3.78   Min.   : 1.790
##   1st Qu.:  492.75  1st Qu.:  66.69  1st Qu.: 20.70   1st Qu.: 3.900
##   Median : 1224.15  Median : 174.16  Median : 40.04   Median : 7.170
##   Mean   : 2286.84  Mean   : 292.64  Mean   : 61.10   Mean   : 8.709
##   3rd Qu.: 2921.93  3rd Qu.: 419.89  3rd Qu.: 83.93   3rd Qu.:11.250
##   Max.   :19341.34  Max.   :1863.11  Max.   :265.07   Max.   :40.040
##     Lactate          Leucine          Lysine        Methylamine
##   Min.   :   7.32  Min.   :  2.51  Min.   : 10.49  Min.   : 1.51
##   1st Qu.:  35.52  1st Qu.:  9.12  1st Qu.: 30.27  1st Qu.: 5.26
##   Median :  81.45  Median : 19.11  Median : 69.41  Median :14.73
##   Mean   : 158.46  Mean   : 24.36  Mean   :108.79  Mean   :17.38
##   3rd Qu.: 139.77  3rd Qu.: 31.19  3rd Qu.:121.51  3rd Qu.:24.05
##   Max.   :3640.95  Max.   :103.54  Max.   :788.40  Max.   :52.46
##   Methylguanidine  N.N.Dimethylglycine O.Acetylcarnitine  Pantothenate
##   Min.   :  1.70   Min.   :  0.79      Min.   :  1.23    Min.   :  2.59
##   1st Qu.:  4.26   1st Qu.:  7.03      1st Qu.:  3.94    1st Qu.: 11.13
##   Median :  7.85   Median : 21.98      Median : 11.47    Median : 22.65
##   Mean   : 15.32   Mean   : 26.35      Mean   : 19.73    Mean   : 44.88
##   3rd Qu.: 19.30   3rd Qu.: 40.04      3rd Qu.: 20.91    3rd Qu.: 41.26
##   Max.   :141.17   Max.   :120.30      Max.   :254.68    Max.   :692.29
##   Pyroglutamate      Pyruvate       Quinolinate        Serine
##   Min.   :  21.33  Min.   :  0.90  Min.   :  5.21  Min.   :  16.12
##   1st Qu.:  68.72  1st Qu.:  4.85  1st Qu.: 26.58  1st Qu.:  83.10
##   Median : 157.59  Median : 13.46  Median : 51.42  Median : 142.59
##   Mean   : 211.45  Mean   : 21.29  Mean   : 66.44  Mean   : 197.69
##   3rd Qu.: 301.87  3rd Qu.: 29.08  3rd Qu.: 87.36  3rd Qu.: 270.43
##   Max.   :1064.22  Max.   :184.93  Max.   :259.82  Max.   :1248.88
##     Succinate        Sucrose          Tartrate        Taurine
##   Min.   :  1.72   Min.   :   6.49  Min.   :  2.20  Min.   :  17.81
##   1st Qu.:  8.58   1st Qu.:  19.30  1st Qu.:  6.89  1st Qu.:  99.48
##   Median : 30.88   Median :  40.85  Median : 12.94  Median : 249.64
##   Mean   : 60.23   Mean   : 113.23  Mean   : 40.00  Mean   : 525.12
##   3rd Qu.: 74.44   3rd Qu.:  94.63  3rd Qu.: 25.79  3rd Qu.: 665.14
##   Max.   :589.93   Max.   :2079.74  Max.   :837.15  Max.   :4272.69
##     Threonine       Trigonelline    Trimethylamine.N.oxide   Tryptophan
##   Min.   :  8.25   Min.   :  10.07  Min.   :  55.7         Min.   :  8.67
##   1st Qu.: 31.82   1st Qu.:  53.52  1st Qu.: 175.9         1st Qu.: 21.33
##   Median : 64.07   Median : 114.43  Median : 383.8         Median : 46.99
##   Mean   : 95.36   Mean   : 270.44  Mean   : 652.2         Mean   : 66.24
##   3rd Qu.:137.00   3rd Qu.: 340.36  3rd Qu.: 735.1         3rd Qu.: 96.54
##   Max.   :450.34   Max.   :2252.96  Max.   :5486.2         Max.   :259.82
##     Tyrosine         Uracil          Valine          Xylose
##   Min.   :  4.22   Min.   :  3.10  Min.   :  4.10  Min.   :  10.07
##   1st Qu.: 23.57   1st Qu.: 11.94  1st Qu.: 12.18  1st Qu.:  29.96
##   Median : 60.34   Median : 27.39  Median : 33.12  Median :  50.40
##   Mean   : 81.76   Mean   : 35.56  Mean   : 35.67  Mean   : 100.93
##   3rd Qu.:113.30   3rd Qu.: 44.26  3rd Qu.: 50.40  3rd Qu.:  89.12
```

```
## Max.     :539.15    Max.     :179.47    Max.     :160.77    Max.     :2164.62
## cis.Aconitate       myo.Inositol    trans.Aconitate  pi.Methylhistidine
## Min.    :  12.94   Min.   :  11.59   Min.    :  4.90   Min.    :  11.36
## 1st Qu.:  36.23    1st Qu.: 30.27    1st Qu.: 12.43    1st Qu.:  67.36
## Median :  129.02   Median : 78.26    Median : 26.84    Median : 162.39
## Mean   :  204.22   Mean   :135.40    Mean    : 40.63   Mean    : 370.29
## 3rd Qu.:  254.68   3rd Qu.:167.34    3rd Qu.: 57.40    3rd Qu.: 387.61
## Max.   : 1863.11   Max.   :854.06    Max.    :217.02   Max.    :2697.28
## tau.Methylhistidine
## Min.    :  8.00
## 1st Qu.: 27.39
## Median : 68.72
## Mean   : 89.69
## 3rd Qu.:130.32
## Max.   :317.35
```
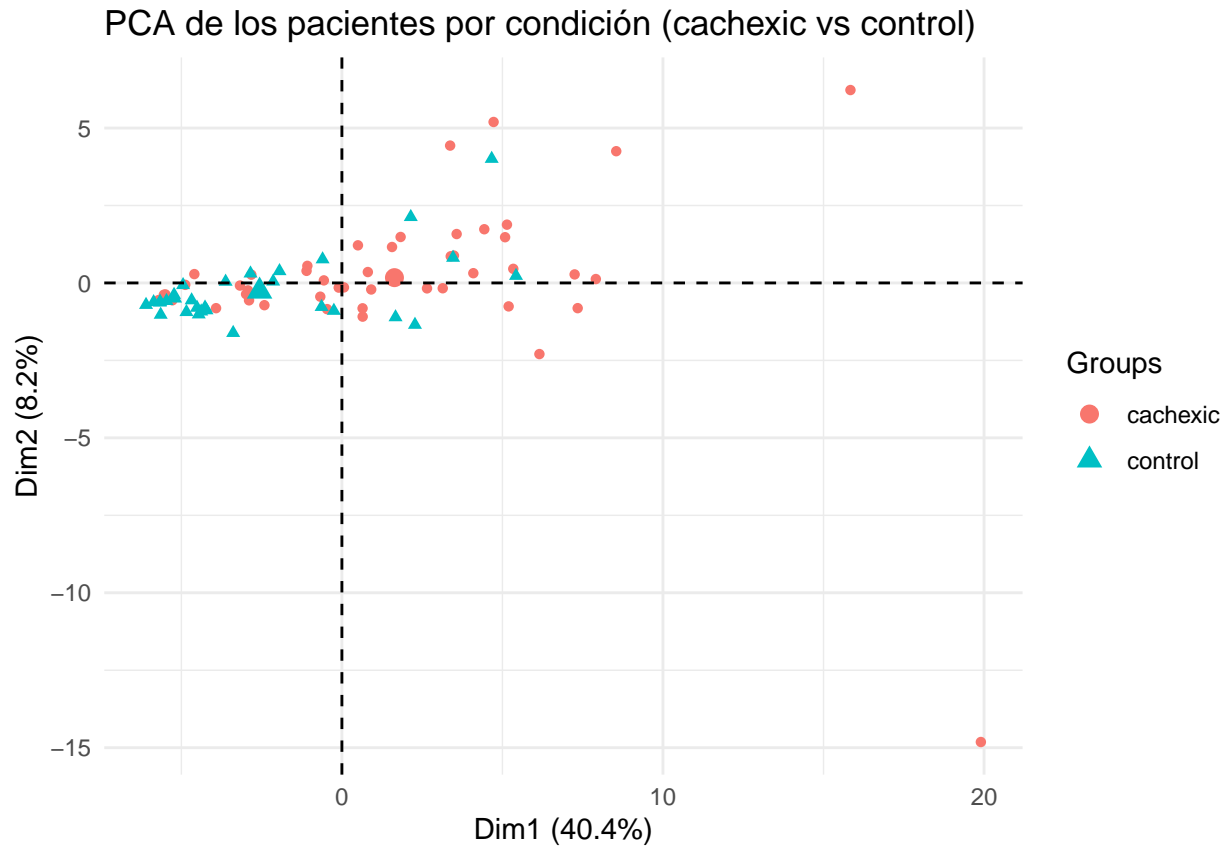
```r
hist(assay(se_object)[,1], main = "Distribución de intensidades del primer parámetro", xlab = "Intensida
```

Distribución de intensidades del primer parámetro


```r
# Realizamos un Análisis de componentes principales (PCA) para la exploración de posibles agrupamientos
if (!requireNamespace("factoextra", quietly = TRUE))
  install.packages("factoextra")
library(factoextra)
```

```
## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa
```

```
# Realizamos PCA y graficamos
pca <- prcomp(assay(se_object), scale. = TRUE)  # Cada fila es un paciente, cada columna es un parámetr
fviz_pca_ind(pca, label = "none", habillage = row_metadata$Condition) +
  labs(title = "PCA de los pacientes por condición (cachexic vs control)") +
  theme_minimal()
```

PCA de los pacientes por condición (cachexic vs control)

Por último guardamos el objeto SummarizedExperiment y los otros elementos para el repositorio de GitHub

```
# Guardamos el objeto SummarizedExperiment
save(se_object, file = "se_object.Rda")

# Guardamos los datos y metadatos en formato texto
write.csv(expression_data, "expression_data.csv", row.names = FALSE)
write.csv(row_metadata, "row_metadata.csv", row.names = FALSE)
write.csv(col_metadata, "col_metadata.csv", row.names = FALSE)
```