PAC1 Anàlisi de dades òmiques

Arantxa Vázquez

[illegible]

1. Introducción

[illegible]

2. Objetivos

[illegible]

3. Materiales y métodos

[illegible]

[illegible]

4. Resultados

[illegible]

```
[illegible]
## $ 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 ...
```

A continuación hemos llevado a cabo la extracción de todos los datos y metadatos en un contenedor SummarizedExperiment. Para poder hacer este paso hemos tenido que extraer la matriz de abundancias de las variables, luego adaptar las longitudes y posteriormente hemos podido crear el Dataframe con los metadatos para finalmente guardarlo en un contenedor.

```r
# Cargamos la biblioteca SummarizedExperiment
library(SummarizedExperiment)

# Extraemos la matriz de abundancias, omitiendo las primeras dos columnas (Patient.ID y Muscle.loss)
counts_matrix <- as.matrix(data[, -c(1, 2)])

# Asignamos los nombres de fila utilizando el identificador de paciente
rownames(counts_matrix) <- data$Patient.ID

# Aseguramos que los nombres de fila sean únicos y válidos
rownames(counts_matrix) <- make.names(data$Patient.ID, unique = TRUE)

# Creamos un vector de condiciones que refleje el estado de Muscle.loss para las muestras usando la lon
gitud de counts_matrix para asegurar que coinciden
conditions <- factor(data$Muscle.loss[1:ncol(counts_matrix)])

# Creamos el DataFrame de metadatos
colData <- DataFrame(sample = colnames(counts_matrix), condition = conditions)

# Verificamos la estructura de colData
str(colData)
```

```
## Formal class 'DFrame' [package "S4Vectors"] with 6 slots
##   ..@ rownames       : NULL
##   ..@ nrows          : int 63
##   ..@ elementType    : chr "ANY"
##   ..@ elementMetadata: NULL
##   ..@ metadata       : list()
##   ..@ listData       :List of 2
##   .. ..$ sample   : chr [1:63] "X1.6.Anhydro.beta.D.glucose" "X1.Methylnicotinamide" "X2.Aminobutyra
te" "X2.Hydroxyisobutyrate" ...
##   .. ..$ condition: Factor w/ 2 levels "cachexic","control": 1 1 1 1 1 1 1 1 1 1 ...
```

```r
# Creamos el objeto SummarizedExperiment con los datos de abundancias y los metadatos
se <- SummarizedExperiment(assays = list(counts = counts_matrix), colData = colData)

# Imprimimos el resumen del objeto para comprobar que se haya creado correctamente
print(se)
```

```
## class: SummarizedExperiment
## dim: 77 63
## metadata(0):
## assays(1): counts
## rownames(77): PIF_178 PIF_087 ... NETL_003_V1 NETL_003_V2
## rowData names(0):
## colnames(63): X1.6.Anhydro.beta.D.glucose X1.Methylnicotinamide ...
##   pi.Methylhistidine tau.Methylhistidine
## colData names(2): sample condition
```

Una vez obtenido nuestro objeto 'se' procedemos al análisis exploratorio:

1. Observamos las dimensiones del objeto y los nombres de columnas y filas

```r
# Inspeccionamos las dimensiones
dim(se)
```

```
## [1] 77 63
```

```r
# Observamos los nombres de las filas (muestras) y columnas (metabolitos)
rownames(se)
```

```
##  [1] "PIF_178"     "PIF_087"     "PIF_090"     "NETL_005_V1" "PIF_115"
##  [6] "PIF_110"     "NETL_019_V1" "NETCR_014_V1" "NETCR_014_V2" "PIF_154"
## [11] "NETL_022_V1" "NETL_022_V2" "NETL_008_V1" "PIF_146"     "PIF_119"
## [16] "PIF_099"     "PIF_162"     "PIF_160"     "PIF_113"     "PIF_143"
## [21] "NETCR_007_V1" "NETCR_007_V2" "PIF_137"    "PIF_100"     "NETL_004_V1"
## [26] "PIF_094"     "PIF_132"     "PIF_163"     "NETCR_003_V1" "NETL_028_V1"
## [31] "NETL_028_V2" "NETCR_013_V1" "NETL_020_V1" "NETL_020_V2" "PIF_192"
## [36] "NETCR_012_V1" "NETCR_012_V2" "PIF_089"    "NETCR_002_V1" "PIF_179"
## [41] "PIF_114"     "NETCR_006_V1" "PIF_141"     "NETCR_025_V1" "NETCR_025_V2"
## [46] "NETCR_016_V1" "PIF_116"     "PIF_191"     "PIF_164"     "NETL_013_V1"
## [51] "PIF_188"     "PIF_195"     "NETCR_015_V1" "PIF_102"     "NETL_010_V1"
## [56] "NETL_010_V2" "NETL_001_V1" "NETCR_015_V2" "NETCR_005_V1" "PIF_111"
## [61] "PIF_171"     "NETCR_008_V1" "NETCR_008_V2" "NETL_017_V1" "NETL_017_V2"
## [66] "NETL_002_V1" "NETL_002_V2" "PIF_190"     "NETCR_009_V1" "NETCR_009_V2"
## [71] "NETL_007_V1" "PIF_112"     "NETCR_019_V2" "NETL_012_V1" "NETL_012_V2"
## [76] "NETL_003_V1" "NETL_003_V2"
```

```r
colnames(se)
```

```
##  [1] "X1.6.Anhydro.beta.D.glucose" "X1.Methylnicotinamide"
##  [3] "X2.Aminobutyrate"            "X2.Hydroxyisobutyrate"
##  [5] "X2.Oxoglutarate"             "X3.Aminoisobutyrate"
##  [7] "X3.Hydroxybutyrate"          "X3.Hydroxyisovalerate"
##  [9] "X3.Indoxylsulfate"           "X4.Hydroxyphenylacetate"
## [11] "Acetate"                     "Acetone"
## [13] "Adipate"                     "Alanine"
## [15] "Asparagine"                  "Betaine"
## [17] "Carnitine"                   "Citrate"
## [19] "Creatine"                    "Creatinine"
## [21] "Dimethylamine"               "Ethanolamine"
## [23] "Formate"                     "Fucose"
## [25] "Fumarate"                    "Glucose"
## [27] "Glutamine"                   "Glycine"
## [29] "Glycolate"                   "Guanidoacetate"
## [31] "Hippurate"                   "Histidine"
## [33] "Hypoxanthine"                "Isoleucine"
## [35] "Lactate"                     "Leucine"
## [37] "Lysine"                      "Methylamine"
## [39] "Methylguanidine"             "N.N.Dimethylglycine"
## [41] "O.Acetylcarnitine"           "Pantothenate"
## [43] "Pyroglutamate"               "Pyruvate"
## [45] "Quinolinate"                 "Serine"
## [47] "Succinate"                   "Sucrose"
## [49] "Tartrate"                    "Taurine"
## [51] "Threonine"                   "Trigonelline"
## [53] "Trimethylamine.N.oxide"      "Tryptophan"
## [55] "Tyrosine"                    "Uracil"
## [57] "Valine"                      "Xylose"
## [59] "cis.Aconitate"               "myo.Inositol"
## [61] "trans.Aconitate"             "pi.Methylhistidine"
## [63] "tau.Methylhistidine"
```

2. REalizamos un resumen estadístico para tener una visión general de los datos

```r
# Resumen estadístico de los datos
summary(assay(se))
```

```
##  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
```

3. Observamos la estructura de las condiciones y las muestras

```r
# Ver la estructura
str(colData(se))
```

```
## Formal class 'DFrame' [package "S4Vectors"] with 6 slots
##   ..@ rownames       : chr [1:63] "X1.6.Anhydro.beta.D.glucose" "X1.Methylnicotinamide" "X2.Aminobut
yrate" "X2.Hydroxyisobutyrate" ...
##   ..@ nrows          : int 63
##   ..@ elementType    : chr "ANY"
##   ..@ elementMetadata: NULL
##   ..@ metadata       : list()
##   ..@ listData       :List of 2
##   .. ..$ sample   : chr [1:63] "X1.6.Anhydro.beta.D.glucose" "X1.Methylnicotinamide" "X2.Aminobutyra
te" "X2.Hydroxyisobutyrate" ...
##   .. ..$ condition: Factor w/ 2 levels "cachexic","control": 1 1 1 1 1 1 1 1 1 1 ...
```

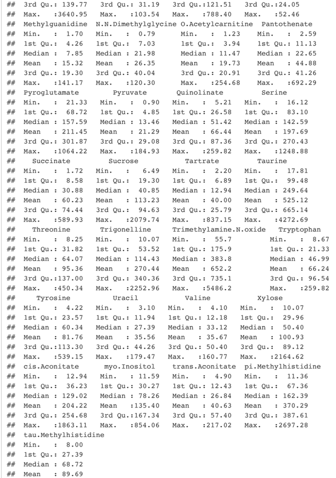
4. Realizamos una tabla de frecuencias de las condiciones

```r
# Tabla de frecuencias para las condiciones
table(colData(se)$condition)
```

```
##
## cachexic  control
##       47       16
```

5. Realizamos una matriz de correlación de las variables

```r
# Importamos la librería
library(corrplot)

# Creamos e imprimimos la matriz
cor_matrix <- cor(t(assay(se)), use = "complete.obs")
corrplot(cor_matrix, method = "circle", tl.col = "black", tl.cex = 0.4)  # Ajustar el tamaño de la fuen
te
```

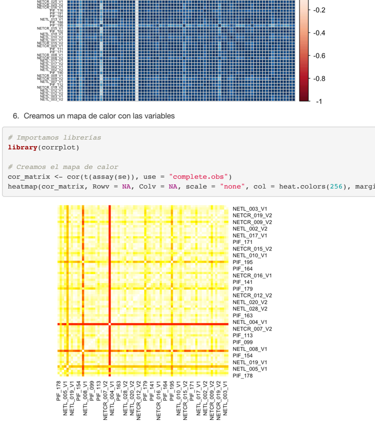


6. Creamos un mapa de calor con las variables

```r
# Importamos librerías
library(corrplot)

# Creamos el mapa de calor
cor_matrix <- cor(t(assay(se)), use = "complete.obs")
heatmap(cor_matrix, Rowv = NA, Colv = NA, scale = "none", col = heat.colors(256), margins = c(10, 10))
```

7. Por último, realizamos un análisis de componentes principales para averiguar qué variables tienen más poder predictivo en nuestro conjunto de datos

```r
# Estandarizaoms los datos
standardized_data <- scale(t(assay(se)))

# Realizamos el PCA
pca_result <- prcomp(standardized_data)

# Imprimimos el gráfico del PCA
plot(pca_result$x[, c(2, 3)], col = as.factor(colData(se)$condition),
     pch = 19, xlab = "PC2", ylab = "PC3", main = "PCA de Metabolitos (PC2 vs PC3")
```



5. Discusión

Hemos realizado un análisis exploratorio de nuestros datos en el cual hemos decidido optar por análisis de correlación puesto que teníamos muchas variables medidas y el número de la muestra era reducido. Tenemos un grupo control muy pequeño y la muestra no es del todo representativa pero sí podemos observar algunos patrones como correlaciones fuertes entre algunos metabolitos según muestran las matrices de correlación. En el análisis de componentes principales hemos visto que hay 1 componente que captura la mayor variabilidad en los datos, eso quiere decir que hay pocos metabolitos que están marcando diferencias grandes entre el grupo control y el grupo con caquexia, así que esos metabolitos podrían tener un fuerte poder predictivo de la enfermedad y eso podría ser útil para decidir tratamiento.

En este trabajo hemos practivado el uso de Bioconductor y contenedores para poder agrupar una gran cantidad de datos y hemos visto lo fácil que resulta el manejo de los datos ómicos con estas herramientas.

6. Referencias

- Instituto Nacional del Cáncer. Caquexia [Internet]. [Actualizado el 20 de marzo de 2020; consultado el 6 de noviembre de 2024]. Disponible en: https://www.cancer.gov/espanol/cancer/tratamiento/investigacion/caquexia.

7. Repositorio gitHub

https://github.com/varantxa/Vazquez-Albacete-Arantxa-PEC1.git

```r
# Guardamos el objeto SummarizedExperiment en un archivo .Rda
save(se, file = "se_metabolitos.Rda")
```