# Huertas Torres Sergio PEC1

Sergio Huertas Torres

2024-11-04

# Informe PEC1 Análisis Datos Ómicos

## Tabla de contendido



## 1- Descarga de los datos

Descargamos los datos de human cachexia del repositorio de github y los importamos al proyecto.

```
human_cachexia<-read.csv("human_cachexia.csv")
```

Ejecutamos class para saber de que forma esta organizada la información

```
class(human_cachexia)
```

```
## [1] "data.frame"
```

## 2- Abstract

Human_cachexia de metaboAnalyst se trata de una tabla de concentración de metabolitos en muestras de orina humana de dos grupos (cachexic y control).

La caquexia se trata de un síndrome metabólico complicado y relacionado con una enfermedad subyacente caracterizado por la perdida de masa muscular (con o sin perdida de masa grasa).

El dataframe human_cachexia probablemente sea el resultado de un análisis metabólico comparativo entre pacientes con caquexia y un grupo control (quizá en un intento de encontrar metabolitos en orina humana que puedan servir como un indicador del síndrome).

# 3- Objetivo del estudio

Teniendo en cuenta que no tenemos información del objetivo puesto que carecemos de metadatos que lo describan u otras consideraciones podemos realizar una inferencia del objetivo final del estudio:

¿Hay diferencias significativas para la cantidad de alguno del listado de metabolitos detectados en orina entre los grupos Control y Cachexic?

# 4- Exploración de los datos

Realizamos str para hacernos una idea general de la estructura interna del data frame.

```
str(human_cachexia)
```

```
## '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 ...
##  $ Quinolinate                : 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 ...
```

Observamos que se trata de un dataframe con 65 columnas, 63 de ellas son numéricas y dos de “characters”

# 4.1- SummarizedExperiment object

Para generar el objeto SummarizedExperiment debemos importar la librería de SummarizedExperiment.

```
if (!requireNamespace("BiocManager", quietly = TRUE)) {
    install.packages("BiocManager")
}
BiocManager::install("SummarizedExperiment")
```

```
## Bioconductor version 3.20 (BiocManager 1.30.25), R 4.4.0 (2024-04-24 ucrt)
```

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

```
## Installation paths not writeable, unable to update packages
##   path: C:/Program Files/R/R-4.4.0/library
##   packages:
##     boot, foreign, KernSmooth, MASS, Matrix, nlme, survival
```

```
## Old packages: 'Biobase', 'GenomicRanges', 'httr2', 'IRanges', 'S4Arrays',
##   'S4Vectors', 'xfun', 'XVector', 'zlibbioc'
```

```
library(SummarizedExperiment)
```

```
## Warning: package 'SummarizedExperiment' was built under R version 4.4.1
```

```
## Cargando paquete requerido: MatrixGenerics
```

```
## Warning: package 'MatrixGenerics' was built under R version 4.4.1
```

```
## Cargando paquete requerido: matrixStats
```

```
## Warning: package 'matrixStats' was built under R version 4.4.1
```

```
##
## Adjuntando el paquete: '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
```

```
## Cargando paquete requerido: GenomicRanges
```

```
## Cargando paquete requerido: stats4
```

```
## Cargando paquete requerido: BiocGenerics
```

```
## Warning: package 'BiocGenerics' was built under R version 4.4.1
```

```
##
## 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
```

```
## Warning: package 'S4Vectors' was built under R version 4.4.1
```

```
##
## 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 4.4.1
```

```
##
## 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 4.4.1
```

```
## 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
```

```
## Warning: replacing previous import 'S4Arrays::read_block' by
## 'DelayedArray::read_block' when loading 'SummarizedExperiment'
```

Generamos un data frame solo con los datos numéricos de human_cachexia transpuestos, puesto que para poder generar el object summarizedexperiment las variables están en las filas y las samples en las columnas y en nuestro dataframe están al revés.

Con esto tendremos los datos del assay.

Después generamos el dataframe para col_data y la lista de nombres para colnames.

```
datos_cachexiat <- t(as.matrix(human_cachexia[, -c(1, 2)]))
colnames(datos_cachexiat) <- human_cachexia$Patient.ID
col_data <- DataFrame(Patient.ID = human_cachexia$Patient.ID,
                      Muscle.loss =human_cachexia$Muscle.loss)

se <- SummarizedExperiment(
  assays = list(counts = datos_cachexiat),
  colData = col_data
)
se
```

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

Por último, creamos el objeto en formato binario para añadirlo al repositorio de github:

```
save(se, file = "summarized_experiment.Rda")
```

# 4.2- Análisis estadístico

Para realizar el análisis estadístico que responda a la pregunta biológica planteada en el objetivo del estudio realizaremos un conjunto de estudios.

En los datos proporcionados en el repositorio de github nos indican que se tratan de datos desapareados (diferente proporción de control y cachexic). Lo comprobamos mediante:

```
table(human_cachexia$Muscle.loss)
```

```
##
## cachexic  control
##       47       30
```

Tendremos que tener este dato en cuenta a la hora de realizar una análisis multivariante. Empezamos por crear un data.frame solo con las variables numéricas, sin transponerlas.

```
datos_cachexia<-human_cachexia[,-c(1:2)]
```

Después pasamos la variable Muscle.loss de un character a factor para poder utilizarlo en las funciones. Realizamos un estudio general de la media, la mean y la desviación estándar.

La revisión de estos datos ya nos puede dar una idea general de las diferencias para los metabolitos entre los dos grupos.

```
muscle<-as.factor(human_cachexia$Muscle.loss)
apply(datos_cachexia, MARGIN = 2, FUN = tapply,INDEX=muscle, mean)
```

```
##          X1.6.Anhydro.beta.D.glucose X1.Methylnicotinamide X2.Aminobutyrate
## cachexic                   128.68894              70.56426        23.669149
## control                     69.50533              73.15500         9.528333
##          X2.Hydroxyisobutyrate X2.Oxoglutarate X3.Aminoisobutyrate
## cachexic              43.23766       183.11043            100.2747
## control               27.87100        85.51733             39.9110
##          X3.Hydroxybutyrate X3.Hydroxyisovalerate X3.Indoxylsulfate
## cachexic          29.260638              27.60638          265.1577
## control            9.898667              12.31267          146.3763
##          X4.Hydroxyphenylacetate  Acetate  Acetone   Adipate  Alanine
## cachexic               119.82255 85.63298 13.34638 34.817872 347.5911
## control                 99.79867 35.60467  8.42000  8.993333 157.5840
##          Asparagine   Betaine Carnitine  Citrate  Creatine Creatinine
## cachexic   75.39085 112.25298  64.62213 2720.853 174.91340  10722.140
## control    41.74900  55.97033  32.44367 1474.719  51.50433   5619.175
##          Dimethylamine Ethanolamine   Formate    Fucose Fumarate  Glucose
## cachexic      453.5806     326.7721 187.56447 108.59915 10.92191 827.2189
## control       208.6833     197.1253  84.48333  57.44467  4.55200 140.9580
##          Glutamine   Glycine Glycolate Guanidoacetate Hippurate Histidine
## cachexic  391.4104 1069.3779  219.2696       97.62426   2875.73  364.2323
## control   174.4273  585.1493  138.9837       68.73967   1364.24  180.4723
##          Hypoxanthine Isoleucine   Lactate  Leucine    Lysine Methylamine
## cachexic     67.08702   9.660851 217.63191 31.26170 121.28234    21.21638
## control      51.71433   7.218000  65.74833 13.55667  89.22933    11.36000
##          Methylguanidine N.N.Dimethylglycine O.Acetylcarnitine Pantothenate
## cachexic        17.36468             34.48979          25.56447     39.94404
## control         12.12833             13.59667          10.59800     52.62267
##          Pyroglutamate Pyruvate Quinolinate   Serine Succinate   Sucrose
## cachexic      270.2923 26.86553    83.74723 245.8298  79.62894 150.02447
## control       119.2580 12.56633    39.32400 122.2630  29.83600  55.57967
##          Tartrate  Taurine Threonine Trigonelline Trimethylamine.N.oxide
## cachexic 47.23468 655.7200 118.23319     359.6377               820.3406
## control  28.67600 320.5223  59.51867     130.6870               388.6690
##          Tryptophan Tyrosine   Uracil   Valine    Xylose cis.Aconitate
## cachexic   81.82404 100.7423 37.51362 45.58255 129.28915      276.0255
## control    41.83300  52.0140 32.49333 20.13267  56.50933       91.7240
##          myo.Inositol trans.Aconitate pi.Methylhistidine tau.Methylhistidine
## cachexic    181.83766        48.81404           441.5532           105.66766
## control      62.64133        27.80933           258.6400            64.65033
```

```
apply(datos_cachexia, MARGIN = 2, FUN = tapply,INDEX=muscle, median)
```

```
##          X1.6.Anhydro.beta.D.glucose X1.Methylnicotinamide X2.Aminobutyrate
## cachexic                      68.720               50.910            15.18
## control                       34.985               19.435             7.58
##          X2.Hydroxyisobutyrate X2.Oxoglutarate X3.Aminoisobutyrate
## cachexic                40.850          67.360               29.67
## control                 19.305          31.685               18.41
##          X3.Hydroxybutyrate X3.Hydroxyisovalerate X3.Indoxylsulfate
## cachexic             19.490                21.120            188.67
## control               6.655                 5.625            105.11
##          X4.Hydroxyphenylacetate Acetate Acetone Adipate Alanine Asparagine
## cachexic                   86.49   70.81   7.920  15.030  304.90      62.80
## control                    50.70   17.32   6.925   6.295   78.65      29.52
##          Betaine Carnitine Citrate Creatine Creatinine Dimethylamine
## cachexic  107.77    28.790 2230.54   76.710    9701.15        361.41
## control    32.55    19.205 1043.62   19.515    3697.13        149.91
##          Ethanolamine Formate Fucose Fumarate Glucose Glutamine Glycine
## cachexic       244.69  130.32  82.27    6.050 387.610   368.710 671.830
## control        123.03   61.56  42.59    3.225 103.595   117.405 417.045
##          Glycolate Guanidoacetate Hippurate Histidine Hypoxanthine Isoleucine
## cachexic   172.430          83.93  1919.850   284.290        43.82       8.17
## control     70.605          48.00   602.595    88.375        33.67       4.31
##          Lactate Leucine Lysine Methylamine Methylguanidine N.N.Dimethylglycine
## cachexic 106.700   23.10  96.54       19.30            8.85                28.79
## control   40.885    9.12  35.34        5.34            6.82                 8.94
##          O.Acetylcarnitine Pantothenate Pyroglutamate Pyruvate Quinolinate
## cachexic            13.460       25.790        244.69    21.12       74.44
## control              6.675       14.735         83.95     6.62       27.26
##          Serine Succinate Sucrose Tartrate Taurine Threonine Trigonelline
## cachexic 206.44    45.150  58.560    15.96  336.97     98.49      186.790
## control   98.28    11.805  19.395    10.75  192.21     39.45       72.275
##          Trimethylamine.N.oxide Tryptophan Tyrosine Uracil Valine Xylose
## cachexic                 482.99      79.04   85.630 29.370  38.47  70.11
## control                  257.70      22.69   27.005 21.355  13.47  34.88
##          cis.Aconitate myo.Inositol trans.Aconitate pi.Methylhistidine
## cachexic       232.760      117.920          44.260            232.760
## control         56.305       30.725          14.315             73.335
##          tau.Methylhistidine
## cachexic              79.840
## control               37.185
```

```
apply(datos_cachexia, MARGIN = 2, FUN = tapply,INDEX=muscle, sd)
```

```
##          X1.6.Anhydro.beta.D.glucose X1.Methylnicotinamide X2.Aminobutyrate
## cachexic                   142.43720              85.25987        33.861804
## control                     99.56823             186.96823         7.307203
##          X2.Hydroxyisobutyrate X2.Oxoglutarate X3.Aminoisobutyrate
## cachexic              23.43735        411.9266           238.16424
## control               21.99241        179.7103            57.84174
##          X3.Hydroxybutyrate X3.Hydroxyisovalerate X3.Indoxylsulfate
## cachexic          30.734216              27.43863          211.6573
## control            7.985001              16.99889          146.9752
##          X4.Hydroxyphenylacetate  Acetate   Acetone  Adipate  Alanine
## cachexic                94.05853 90.60022 29.656970 62.18722 281.7285
## control                154.80432 42.93314  5.652959 10.42638 156.1359
##          Asparagine  Betaine Carnitine  Citrate  Creatine Creatinine
## cachexic   57.79106 83.19913  87.11959 2459.721 335.36486   6905.406
## control    40.19878 70.37315  40.45228 1313.056  87.23761   4228.759
##          Dimethylamine Ethanolamine   Formate   Fucose  Fumarate    Glucose
## cachexic      347.0764     264.5155 224.50099 91.12714 17.432383 1727.55510
## control       139.1947     211.0371  71.86832 47.17225  6.391354   99.19766
##          Glutamine   Glycine Glycolate Guanidoacetate Hippurate Histidine
## cachexic  311.5721 1091.7293  186.3896       89.81720  3298.445  350.9996
## control   195.3615  570.1131  161.6588       69.08304  1700.305  199.0066
##          Hypoxanthine Isoleucine   Lactate   Leucine    Lysine Methylamine
## cachexic     60.18018   7.572630 529.95631 24.231769  97.51977    13.78307
## control      52.72308   5.596286  55.81337  9.157233 162.20024    12.03258
##          Methylguanidine N.N.Dimethylglycine O.Acetylcarnitine Pantothenate
## cachexic        23.35677            26.59008          42.32996     41.09855
## control         12.23669            13.44479          10.28234    130.19429
##          Pyroglutamate Pyruvate Quinolinate   Serine Succinate  Sucrose
## cachexic      212.1925 29.62169    53.40643 214.4016  98.57757 315.0315
## control        98.6342 15.70713    33.76190  88.8920  44.99090 113.2921
##           Tartrate  Taurine Threonine Trigonelline Trimethylamine.N.oxide
## cachexic 125.12896 786.4167  95.29957     476.9450              1099.2591
## control   54.87093 371.5426  57.42642     146.2309               369.4847
##          Tryptophan Tyrosine   Uracil   Valine   Xylose cis.Aconitate
## cachexic   59.07686 94.08034 35.81805 32.48843 311.8836     331.28732
## control    42.08062 51.21703 34.05826 15.10151  80.2464      84.89998
##          myo.Inositol trans.Aconitate pi.Methylhistidine tau.Methylhistidine
## cachexic    197.76882        39.95090           612.8804            80.29638
## control      70.45362        35.95093           348.0329            65.91869
```

Después de evaluar de manera superficial los datos de cada metabolito debemos comprobar que los datos sigan una distribución normal (aunque se podría asumir puesto que la n>=30).

Para ello hacemos una función con el Saphiro test que nos devuelva la normalidad o no de cada uno de los grupos para cada uno de los metabolitos.

```
grupo_cachexic <- human_cachexia[human_cachexia$Muscle.loss == "cachexic", ]
grupo_control <- human_cachexia[human_cachexia$Muscle.loss == "control", ]

p_valores_normalidad <- sapply(names(datos_cachexia), function(metabolito) {

  shapiro_cachexic <- shapiro.test(grupo_cachexic[[metabolito]])$p.value
  shapiro_control <- shapiro.test(grupo_control[[metabolito]])$p.value
  c(cachexic = shapiro_cachexic, control = shapiro_control)
})

p_valores_normalidad
```

```
##          X1.6.Anhydro.beta.D.glucose X1.Methylnicotinamide X2.Aminobutyrate
## cachexic               1.999866e-07          1.584013e-09     6.170993e-11
## control                1.613558e-08          1.631778e-10     9.541639e-04
##          X2.Hydroxyisobutyrate X2.Oxoglutarate X3.Aminoisobutyrate
## cachexic           0.039032782    2.303795e-12        1.465323e-12
## control            0.001626644    8.919452e-10        3.120749e-08
##          X3.Hydroxybutyrate X3.Hydroxyisovalerate X3.Indoxylsulfate
## cachexic       1.144397e-07          5.567290e-08      3.522801e-05
## control        1.960501e-04          1.123566e-07      4.252464e-06
##          X4.Hydroxyphenylacetate      Acetate      Acetone      Adipate
## cachexic            3.095323e-05 3.081564e-08 8.188739e-14 1.285528e-11
## control             6.849287e-09 2.386493e-06 7.606672e-05 2.078146e-08
##               Alanine   Asparagine      Betaine    Carnitine      Citrate
## cachexic 5.169099e-05 1.701573e-04 2.707096e-04 2.456377e-09 8.321229e-07
## control  6.674537e-05 4.929412e-06 1.011844e-06 2.278765e-07 2.235414e-03
##              Creatine  Creatinine Dimethylamine Ethanolamine      Formate
## cachexic 5.864870e-12 0.012599185  0.0001238763 3.949457e-05 1.233095e-10
## control  3.317946e-08 0.002615974  0.0071950351 1.183419e-05 5.613942e-04
##                Fucose     Fumarate      Glucose    Glutamine      Glycine
## cachexic 0.0000361081 2.261986e-11 2.862457e-12 3.908030e-05 6.618303e-07
## control  0.0013534871 2.881270e-09 1.288603e-03 3.445112e-06 2.551701e-04
##             Glycolate Guanidoacetate    Hippurate    Histidine Hypoxanthine
## cachexic 1.699577e-04   2.886359e-08 2.314484e-08 9.460144e-07 1.121557e-05
## control  5.164470e-06   1.866486e-05 6.496493e-07 2.765833e-05 9.258419e-05
##            Isoleucine      Lactate      Leucine       Lysine  Methylamine
## cachexic 1.289632e-07 1.580594e-13 7.102276e-06 3.859168e-05 4.354943e-03
## control  1.911635e-04 5.792578e-04 9.180866e-03 3.268827e-09 2.929615e-06
##          Methylguanidine N.N.Dimethylglycine O.Acetylcarnitine Pantothenate
## cachexic    5.426061e-10        0.0002357476      1.698556e-11 2.879223e-07
## control     3.359130e-05        0.0000823241      2.738320e-04 3.790773e-10
##          Pyroglutamate     Pyruvate  Quinolinate       Serine    Succinate
## cachexic  9.497402e-05 7.209991e-09 1.534567e-03 2.913715e-07 4.172836e-09
## control   2.609957e-04 5.114396e-07 5.658773e-05 6.241602e-03 1.812575e-07
##                Sucrose     Tartrate      Taurine    Threonine Trigonelline
## cachexic 1.523150e-12 2.278197e-13 6.743389e-08 9.369478e-05 1.621765e-08
## control  6.125045e-10 3.952077e-09 1.035237e-05 4.538956e-05 5.288891e-06
##          Trimethylamine.N.oxide   Tryptophan     Tyrosine       Uracil
## cachexic           2.559041e-10 1.648129e-04 1.272778e-07 3.187664e-07
## control            4.925928e-05 2.994641e-06 6.269224e-05 7.906170e-06
##                Valine       Xylose cis.Aconitate myo.Inositol trans.Aconitate
## cachexic 0.0003602307 7.416826e-14   3.54981e-09 1.009535e-07    1.745467e-05
## control  0.0007869125 5.004286e-09   1.96704e-04 2.153350e-06    5.209227e-08
##          pi.Methylhistidine tau.Methylhistidine
## cachexic       1.883894e-09        5.052134e-04
## control        1.051396e-06        5.480573e-05
```

Con los resultados <0.05 podemos asumir la normalidad de todos los parámetros.

Después deberemos comprobar si los datos de los metabolitos mantienen su homocedasticidad (homogeneidad de varianzas), factor muy importante cuando se trata de muestras con datos desapareados (diferente n). Para ello realizaremos el test de Barlet.

```
p_valores_varianzas <- sapply(names(datos_cachexia), function(metabolito) {
  bartlett.test(list(grupo_cachexic[[metabolito]], grupo_control[[metabolito]]))$p.value
})

p_valores_varianzas
```

```
## X1.6.Anhydro.beta.D.glucose       X1.Methylnicotinamide
##                4.078556e-02                2.667623e-06
##            X2.Aminobutyrate       X2.Hydroxyisobutyrate
##                2.361929e-13                7.077151e-01
##             X2.Oxoglutarate         X3.Aminoisobutyrate
##                9.226100e-06                5.555439e-12
##          X3.Hydroxybutyrate       X3.Hydroxyisovalerate
##                3.273701e-11                7.067918e-03
##           X3.Indoxylsulfate     X4.Hydroxyphenylacetate
##                3.734419e-02                2.675249e-03
##                     Acetate                     Acetone
##                5.270733e-05                8.108063e-15
##                     Adipate                     Alanine
##                2.368442e-16                1.076292e-03
##                  Asparagine                     Betaine
##                3.820288e-02                3.285443e-01
##                   Carnitine                     Citrate
##                3.463111e-05                5.426541e-04
##                    Creatine                  Creatinine
##                3.380446e-11                5.877088e-03
##               Dimethylamine                Ethanolamine
##                1.429341e-06                1.901399e-01
##                     Formate                      Fucose
##                6.714089e-09                3.037699e-04
##                    Fumarate                     Glucose
##                1.793217e-07                1.520834e-29
##                   Glutamine                     Glycine
##                8.533954e-03                3.589739e-04
##                   Glycolate              Guanidoacetate
##                4.049773e-01                1.294174e-01
##                   Hippurate                    Histidine
##                2.803221e-04                1.611793e-03
##                Hypoxanthine                  Isoleucine
##                4.385478e-01                8.196354e-02
##                     Lactate                     Leucine
##                5.037986e-22                3.646063e-07
##                      Lysine                  Methylamine
##                2.169188e-03                4.265723e-01
##             Methylguanidine         N.N.Dimethylglycine
##                3.816691e-04                1.929635e-04
##           O.Acetylcarnitine                Pantothenate
##                5.582396e-12                1.800147e-11
##               Pyroglutamate                     Pyruvate
##                3.541953e-05                4.788725e-04
##                 Quinolinate                      Serine
##                9.685024e-03                3.015905e-06
##                   Succinate                     Sucrose
##                2.415734e-05                1.134576e-07
##                    Tartrate                     Taurine
##                1.030899e-05                4.955005e-05
##                   Threonine                Trigonelline
##                4.526319e-03                2.287092e-09
##      Trimethylamine.N.oxide                  Tryptophan
##                2.226998e-08                5.205522e-02
##                    Tyrosine                      Uracil
```

```
##                7.789801e-04                7.663420e-01
##                      Valine                      Xylose
##                3.540937e-05                2.539016e-11
##               cis.Aconitate               myo.Inositol
##                2.287431e-11                9.060095e-08
##             trans.Aconitate         pi.Methylhistidine
##                5.357232e-01                1.656837e-03
##         tau.Methylhistidine
##                2.510004e-01
```

Comprobamos si alguno no mantiene la homogeneidad de las varianzas:

```
resultados_no_homocedasticidad <- p_valores_varianzas[p_valores_varianzas > 0.05]
resultados_no_homocedasticidad
```

```
## X2.Hydroxyisobutyrate               Betaine          Ethanolamine
##            0.70771510            0.32854430            0.19013987
##             Glycolate        Guanidoacetate          Hypoxanthine
##            0.40497734            0.12941742            0.43854776
##            Isoleucine           Methylamine            Tryptophan
##            0.08196354            0.42657234            0.05205522
##                Uracil       trans.Aconitate   tau.Methylhistidine
##            0.76634204            0.53572316            0.25100042
```

Como hemos encontrado valores que no conservan la homocedasticidad deberemos realizar una t de Student con la corrección de Welch para buscar las diferencias en cada uno de los grupos e intentaremos observar que metabolitos tienen diferencias significativas.

Como se suele hacer:

H0: No hay diferencias significativas entre las medias de los 2 grupos, por lo que no podemos asumir que las diferencias observadas no sean fruto del azar.

H1: hay diferencias significativas entre las medias de los 2 grupos, por lo que podemos asumir que las diferencias observadas se deben a los factores dentro de los grupos.

```
resultados_ttest <- sapply(names(datos_cachexia), function(metabolito) {
    t.test(grupo_cachexic[[metabolito]], grupo_control[[metabolito]],var.equal=FALSE)$p.value
  })
resultados_ttest
```

```
## X1.6.Anhydro.beta.D.glucose       X1.Methylnicotinamide
##                3.531943e-02                9.435367e-01
##            X2.Aminobutyrate       X2.Hydroxyisobutyrate
##                7.859048e-03                4.893295e-03
##             X2.Oxoglutarate         X3.Aminoisobutyrate
##                1.585788e-01                1.021946e-01
##          X3.Hydroxybutyrate       X3.Hydroxyisovalerate
##                1.334590e-04                3.458034e-03
##           X3.Indoxylsulfate     X4.Hydroxyphenylacetate
##                4.849645e-03                5.272919e-01
##                     Acetate                     Acetone
##                1.740727e-03                2.731685e-01
##                     Adipate                     Alanine
##                7.515595e-03                2.960803e-04
##                  Asparagine                     Betaine
##                3.566603e-03                2.174920e-03
##                   Carnitine                     Citrate
##                3.192506e-02                5.098507e-03
##                    Creatine                  Creatinine
##                1.984702e-02                1.369605e-04
##               Dimethylamine                Ethanolamine
##                5.348990e-05                2.011237e-02
##                     Formate                      Fucose
##                4.909304e-03                1.863466e-03
##                    Fumarate                     Glucose
##                2.620524e-02                9.239445e-03
##                   Glutamine                     Glycine
##                3.391600e-04                1.304117e-02
##                   Glycolate              Guanidoacetate
##                4.940718e-02                1.165822e-01
##                   Hippurate                   Histidine
##                1.015460e-02                4.541154e-03
##                Hypoxanthine                  Isoleucine
##                2.421171e-01                1.087749e-01
##                     Lactate                     Leucine
##                5.733204e-02                2.662822e-05
##                      Lysine                  Methylamine
##                3.347563e-01                1.496975e-03
##              Methylguanidine         N.N.Dimethylglycine
##                2.027750e-01                2.114015e-05
##           O.Acetylcarnitine                Pantothenate
##                2.418993e-02                6.085017e-01
##               Pyroglutamate                     Pyruvate
##                7.289814e-05                7.357362e-03
##                 Quinolinate                      Serine
##                2.708732e-05                8.173888e-04
##                   Succinate                     Sucrose
##                3.676506e-03                6.558505e-02
##                    Tartrate                     Taurine
##                3.758724e-01                1.418748e-02
##                   Threonine                Trigonelline
##                1.183249e-03                3.219917e-03
##      Trimethylamine.N.oxide                  Tryptophan
##                1.586788e-02                8.885986e-04
##                    Tyrosine                       Uracil
```

```
##                4.456642e-03                5.386741e-01
##                      Valine                      Xylose
##                1.574212e-05                1.335317e-01
##               cis.Aconitate               myo.Inositol
##                6.190184e-04                3.611776e-04
##             trans.Aconitate         pi.Methylhistidine
##                1.953314e-02                9.958934e-02
##         tau.Methylhistidine
##                1.710741e-02
```

Aplicamos el test de Welch con var.equal=FALSE y puesto que el resultado de la comparativa de la t de student es bastante extensa con $p.value hacemos que solo nos devuelva el valor del p.valor para simplificar la obtención de los resultados.

Del mismo modo podemos obtener un listado con los metabolitos que presenten diferencias significativas y los que no.

```
resultados_significativos <- resultados_ttest[resultados_ttest < 0.05]
resultados_significativos
```

```
## X1.6.Anhydro.beta.D.glucose           X2.Aminobutyrate
##               3.531943e-02               7.859048e-03
##       X2.Hydroxyisobutyrate         X3.Hydroxybutyrate
##               4.893295e-03               1.334590e-04
##       X3.Hydroxyisovalerate          X3.Indoxylsulfate
##               3.458034e-03               4.849645e-03
##                    Acetate                    Adipate
##               1.740727e-03               7.515595e-03
##                    Alanine                  Asparagine
##               2.960803e-04               3.566603e-03
##                    Betaine                   Carnitine
##               2.174920e-03               3.192506e-02
##                    Citrate                    Creatine
##               5.098507e-03               1.984702e-02
##                  Creatinine               Dimethylamine
##               1.369605e-04               5.348990e-05
##               Ethanolamine                    Formate
##               2.011237e-02               4.909304e-03
##                     Fucose                    Fumarate
##               1.863466e-03               2.620524e-02
##                    Glucose                   Glutamine
##               9.239445e-03               3.391600e-04
##                    Glycine                   Glycolate
##               1.304117e-02               4.940718e-02
##                   Hippurate                   Histidine
##               1.015460e-02               4.541154e-03
##                    Leucine                  Methylamine
##               2.662822e-05               1.496975e-03
##         N.N.Dimethylglycine          O.Acetylcarnitine
##               2.114015e-05               2.418993e-02
##              Pyroglutamate                    Pyruvate
##               7.289814e-05               7.357362e-03
##                Quinolinate                     Serine
##               2.708732e-05               8.173888e-04
##                   Succinate                    Taurine
##               3.676506e-03               1.418748e-02
##                   Threonine                Trigonelline
##               1.183249e-03               3.219917e-03
##      Trimethylamine.N.oxide                  Tryptophan
##               1.586788e-02               8.885986e-04
##                    Tyrosine                     Valine
##               4.456642e-03               1.574212e-05
##              cis.Aconitate                myo.Inositol
##               6.190184e-04               3.611776e-04
##             trans.Aconitate         tau.Methylhistidine
##               1.953314e-02               1.710741e-02
```

```
resultados_no_significativos <- resultados_ttest[resultados_ttest > 0.05]
resultados_no_significativos
```

```
##   X1.Methylnicotinamide         X2.Oxoglutarate     X3.Aminoisobutyrate
##               0.94353674              0.15857875              0.10219457
## X4.Hydroxyphenylacetate                 Acetone         Guanidoacetate
##               0.52729195              0.27316846              0.11658217
##             Hypoxanthine             Isoleucine                 Lactate
##               0.24211709              0.10877489              0.05733204
##                   Lysine         Methylguanidine            Pantothenate
##               0.33475631              0.20277495              0.60850167
##                  Sucrose                Tartrate                   Uracil
##               0.06558505              0.37587240              0.53867410
##                   Xylose      pi.Methylhistidine
##               0.13353169              0.09958934
```

Y generar un listado solo con el nombre de lo metabolitos significativos.

```
metabolitos_significativos <- names(resultados_ttest)[resultados_ttest < 0.05]
metabolitos_significativos
```

```
##  [1] "X1.6.Anhydro.beta.D.glucose" "X2.Aminobutyrate"
##  [3] "X2.Hydroxyisobutyrate"       "X3.Hydroxybutyrate"
##  [5] "X3.Hydroxyisovalerate"       "X3.Indoxylsulfate"
##  [7] "Acetate"                     "Adipate"
##  [9] "Alanine"                     "Asparagine"
## [11] "Betaine"                     "Carnitine"
## [13] "Citrate"                     "Creatine"
## [15] "Creatinine"                  "Dimethylamine"
## [17] "Ethanolamine"                "Formate"
## [19] "Fucose"                      "Fumarate"
## [21] "Glucose"                     "Glutamine"
## [23] "Glycine"                     "Glycolate"
## [25] "Hippurate"                   "Histidine"
## [27] "Leucine"                     "Methylamine"
## [29] "N.N.Dimethylglycine"         "O.Acetylcarnitine"
## [31] "Pyroglutamate"               "Pyruvate"
## [33] "Quinolinate"                 "Serine"
## [35] "Succinate"                   "Taurine"
## [37] "Threonine"                   "Trigonelline"
## [39] "Trimethylamine.N.oxide"      "Tryptophan"
## [41] "Tyrosine"                    "Valine"
## [43] "cis.Aconitate"               "myo.Inositol"
## [45] "trans.Aconitate"             "tau.Methylhistidine"
```

# 5- Resultados

Con este listado podríamos contestar a la pregunta:

¿hay diferencias significativas entre control y cachexic para la cantidad de alguno de los metabolitos presentes en orina?

Si. Hay diferencias significativas.

Si asumimos que la función de este estudio tiene un perfil orientado al diagnostico del síndrome de caquexia se podría utilizar una combinación de ellos para definir un panel de diagnostico que permitiese su utilización en un contexto hospitalario.

La combinación final de metabolitos deberá ser evaluada en función de otras consideraciones como: su detección, su presencia en otras enfermedades o la factibilidad de ser implementado en un laboratorio hospitalario, entre muchas otras.

# 6- URL repositorio Github

https://github.com/SergioHHTT/Huertas-Torres-Sergio-PEC1 (https://github.com/SergioHHTT/Huertas-Torres-Sergio-PEC1)