

Primera prueba de evaluación continua
ANALISIS DE DATOS ÓMICOS

Ana Chen

2024-10-27

Contents

0. Carga de paquetes	3
1. Introducción	4
2. Selección de base de datos	4
3. Cargamos los datos y los metadatos	4
4. Exploración de los datos	7
5. Repositorio Github	19
6. Referencias	20

0. Carga de paquetes

```
if (!require("BiocManager", quietly = TRUE))  
  install.packages("BiocManager")  
BiocManager::install(version = "3.18")  
BiocManager::install("GEOquery")  
BiocManager::install("SummarizedExperiment")  
if (!require(ggplot2)) install.packages("ggplot2")
```

1. Introducción

El objetivo de la siguiente PEC es familiarizarse con las herramientas y la exploración multivariante de los datos. Básicamente el ejercicio nos pide trabajar con dataset del repositorio de github i del repositorio metabolomicsWorkbench.

Posteriormente, creamos un contenedor del tipo SummarizedExperiment, con los datos y los metadatos por separado. Con esos datos realizaremos una visión general de los datos.

Finalmente la elaboración del informe final.

2. Selección de base de datos

En esta PEC he escogido la base de datos de human_cachexia.csv, en el que consiste en un estudio sobre el estado de pérdida muscular de paciente probablemente con alguna enfermedad crónica (cáncer, infecciones graves...). En la base nos dan las dos primeras columnas, la primera que consiste en el ID del paciente y la segunda que consiste en la categoría que pertenece cada paciente respecto a la pérdida de masa muscular (cachexia/normal). El resto de variables en las columnas, corresponden a variables bioquímicas, metabolitos comunes como biomarcadores específicos.

3. Cargamos los datos y los metadatos

Cargamos los datos de expresión metabolómica del data set seleccionado

```
data <- read.csv("/Users/anachen/Desktop/human_cachexia.csv")
colnames(data)
```

```
## [1] "Patient.ID" "Muscle.loss"
## [3] "X1.6.Anhydro.beta.D.glucose" "X1.Methylnicotinamide"
## [5] "X2.Aminobutyrate" "X2.Hydroxyisobutyrate"
## [7] "X2.Oxoglutarate" "X3.Aminoisobutyrate"
## [9] "X3.Hydroxybutyrate" "X3.Hydroxyisovalerate"
## [11] "X3.Indoxylsulfate" "X4.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"
```

```
head(data)
```

```
## Patient.ID Muscle.loss X1.6.Anhydro.beta.D.glucose X1.Methylnicotinamide
## 1 PIF_178 cachexic 40.85 65.37
## 2 PIF_087 cachexic 62.18 340.36
## 3 PIF_090 cachexic 270.43 64.72
## 4 NETL_005_V1 cachexic 154.47 52.98
## 5 PIF_115 cachexic 22.20 73.70
## 6 PIF_110 cachexic 212.72 31.82
## X2.Aminobutyrate X2.Hydroxyisobutyrate X2.Oxoglutarate X3.Aminoisobutyrate
## 1 18.73 26.05 71.52 1480.30
## 2 24.29 41.68 67.36 116.75
## 3 12.18 65.37 23.81 14.30
## 4 172.43 74.44 1199.91 555.57
## 5 15.64 83.93 33.12 29.67
## 6 18.36 80.64 47.94 17.46
## X3.Hydroxybutyrate X3.Hydroxyisovalerate X3.Indoxylsulfate
## 1 56.83 10.07 566.80
## 2 43.82 79.84 368.71
## 3 5.64 23.34 665.14
## 4 175.91 25.03 411.58
## 5 76.71 69.41 165.67
## 6 31.82 35.16 183.09
## X4.Hydroxyphenylacetate Acetate Acetone Adipate Alanine Asparagine Betaine
## 1 120.30 126.47 9.49 38.09 314.19 159.17 109.95
## 2 432.68 212.72 11.82 327.01 871.31 157.59 244.69
## 3 292.95 314.19 4.44 131.63 464.05 89.12 116.75
## 4 214.86 37.34 206.44 144.03 589.93 273.14 278.66
## 5 97.51 407.48 44.26 15.03 1118.79 42.52 391.51
## 6 132.95 81.45 14.44 25.28 237.46 157.59 66.69
## Carnitine Citrate Creatine Creatinine Dimethylamine Ethanolamine Formate
## 1 265.07 3714.50 196.37 16481.60 632.70 645.48 441.42
## 2 120.30 2617.57 212.72 15835.35 607.89 487.85 252.14
## 3 25.03 862.64 221.41 24587.66 735.10 407.48 249.64
## 4 200.34 13629.61 85.63 20952.22 1064.22 820.57 468.72
## 5 84.77 854.06 105.64 6768.26 242.26 365.04 114.43
## 6 40.04 1958.63 200.34 15677.78 614.00 459.44 314.19
## Fucose Fumarate Glucose Glutamine Glycine Glycolate Guanidoacetate Hippurate
## 1 336.97 7.69 395.44 871.31 2038.56 685.40 154.47 4582.50
## 2 198.34 18.92 8690.62 601.85 1107.65 651.97 109.95 1737.15
## 3 186.79 7.10 1352.89 301.87 620.17 141.17 183.09 4315.64
## 4 407.48 96.54 862.64 1685.81 5064.45 70.81 102.51 757.48
## 5 26.05 19.69 6836.29 432.68 395.44 26.58 52.98 1152.86
## 6 123.97 5.05 512.86 298.87 482.99 428.38 57.97 3568.85
```

##	Histidine	Hypoxanthine	Isoleucine	Lactate	Leucine	Lysine	Methylamine	
## 1	925.19	97.51	5.58	106.70	42.10	146.94	52.46	
## 2	845.56	82.27	8.17	368.71	77.48	284.29	23.57	
## 3	284.29	114.43	9.30	749.95	31.50	97.51	18.73	
## 4	1043.15	223.63	37.71	368.71	103.54	290.03	48.91	
## 5	327.01	66.69	40.04	3640.95	101.49	122.73	27.94	
## 6	459.44	62.80	8.17	113.30	28.79	120.30	36.97	
##	Methylguanidine	N.N.Dimethylglycine	O.Acetylcarnitine	Pantothenate				
## 1	9.97	23.34	52.98	25.79				
## 2	7.69	87.36	50.40	186.79				
## 3	4.66	24.53	5.58	145.47				
## 4	141.17	40.04	254.68	42.52				
## 5	5.31	46.06	45.60	74.44				
## 6	43.38	24.29	13.46	35.52				
##	Pyroglutamate	Pyruvate	Quinolinat	Serine	Succinate	Sucrose	Tartrate	Taurine
## 1	437.03	21.12	165.67	284.29	154.47	45.15	97.51	1919.85
## 2	437.03	36.97	72.97	391.51	244.69	459.44	32.79	1261.43
## 3	713.37	29.37	192.48	295.89	142.59	160.77	16.28	4272.69
## 4	566.80	64.07	86.49	1248.88	144.03	111.05	837.15	1525.38
## 5	184.93	12.30	38.09	206.44	68.72	75.19	4.53	468.72
## 6	432.68	32.79	112.17	387.61	33.45	336.97	24.05	2059.05
##	Threonine	Trigonelline	Trimethylamine.N.oxide	Tryptophan	Tyrosine	Uracil		
## 1	184.93	943.88	2121.76	259.82	290.03	111.05		
## 2	198.34	208.51	639.06	83.10	167.34	46.99		
## 3	109.95	192.48	1152.86	82.27	60.34	31.50		
## 4	376.15	992.27	1450.99	235.10	323.76	30.57		
## 5	64.07	86.49	172.43	103.54	142.59	44.26		
## 6	105.64	862.64	880.07	239.85	127.74	29.67		
##	Valine	Xylose	cis.Aconitate	myo.Inositol	trans.Aconitate	pi.Methylhistidine		
## 1	86.49	72.24	237.46	135.64	51.94	157.59		
## 2	109.95	192.48	333.62	376.15	217.02	307.97		
## 3	59.15	2164.62	330.30	86.49	58.56	145.47		
## 4	102.51	125.21	1863.11	247.15	75.94	249.64		
## 5	160.77	186.79	101.49	749.95	98.49	84.77		
## 6	36.97	89.12	287.15	129.02	121.51	399.41		
##	tau.Methylhistidine							
## 1	160.77							
## 2	130.32							
## 3	83.93							
## 4	254.68							
## 5	79.84							
## 6	68.72							

Separamos como nos especifica en el enunciado las columnas de expresión y los metadatos: Creamos una matriz para los datos de expresión (los metabolitos concretamente), para ello seleccionamos todas las columnas excepto las dos primeras que corresponderán a las Id del paciente y el estado de pérdida muscular). Posteriormente, creamos dos data.frames para los metadatos (será la información del dataset). Una que se llamará row_metadata que incluirá información sobre el ID del paciente y el estado de pérdida muscular. Otra, col_metadata que corresponderá al dataframe de la expresión de metabolitos.

```
# Cargamos la librería
library(SummarizedExperiment)

expression_data <- as.matrix(data[, -c(1:2)])
```

```
row_metadata <- data[,c("Patient.ID", "Muscle.loss")]
col_metadata <- data.frame(Metabolite = colnames(expression_data))
```

Creamos el objeto SummarizedExperiment como nos pide en el enunciado.

```
se <- SummarizedExperiment(
  assays = list(counts = expression_data),
  rowData = row_metadata,
  colData = col_metadata
)

print(se)
```

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

4. Exploración de los datos

Mostramos las dimensiones del expressionset

```
dim(se)
```

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

Nombres de los metadatos de filas y columnas

```
colnames(rowData(se))
```

```
## [1] "Patient.ID" "Muscle.loss"
```

```
colnames(colData(se))
```

```
## [1] "Metabolite"
```

Miramos la estructura del objeto

```
str(se)
```

```
## Formal class 'SummarizedExperiment' [package "SummarizedExperiment"] with 5 slots
## ..@ colData          :Formal class 'DFrame' [package "S4Vectors"] with 6 slots
## .. ..@ rownames      : chr [1:63] "X1.6.Anhydro.beta.D.glucose" "X1.Methylnicotinamide" "X2.Am"
```

```
## ..@ nrow      : int 63
## ..@ elementType : chr "ANY"
## ..@ elementMetadata: NULL
## ..@ metadata    : list()
## ..@ listData     :List of 1
## ..@ $ Metabolite: chr [1:63] "X1.6.Anhydro.beta.D.glucose" "X1.Methylnicotinamide" "X2.Aminobutyrate"
## ..@ assays       :Formal class 'SimpleAssays' [package "SummarizedExperiment"] with 1 slot
## ..@ data:Formal class 'SimpleList' [package "S4Vectors"] with 4 slots
## ..@ listData     :List of 1
## ..@ $ counts: num [1:77, 1:63] 40.9 62.2 270.4 154.5 22.2 ...
## ..@ attr(*, "dimnames")=List of 2
## ..@ $ : NULL
## ..@ $ : chr [1:63] "X1.6.Anhydro.beta.D.glucose" "X1.Methylnicotinamide" "X2.Aminobutyrate"
## ..@ elementType : chr "ANY"
## ..@ elementMetadata: NULL
## ..@ metadata     : list()
## ..@ NAMES        : NULL
## ..@ elementMetadata:Formal class 'DFrame' [package "S4Vectors"] with 6 slots
## ..@ rownames      : NULL
## ..@ nrow          : int 77
## ..@ elementType   : chr "ANY"
## ..@ elementMetadata: NULL
## ..@ metadata      : list()
## ..@ listData      :List of 2
## ..@ $ Patient.ID : chr [1:77] "PIF_178" "PIF_087" "PIF_090" "NETL_005_V1" ...
## ..@ $ Muscle.loss: chr [1:77] "cachexic" "cachexic" "cachexic" "cachexic" ...
## ..@ metadata      : list()
```

Resumen estadístico de los datos de expresión, que nos proporciona un resumen estadístico de las variables de expresión de cada metabolito medido en las muestras obtenidas de los diferentes pacientes del dataset. Podemos observar la distribución y la variabilidad de los datos (valor mínimo, primer cuartil, mediana, media, tercer cuartil y valor máximo). Esta información es fundamental a la hora de realizar cualquier análisis posterior.

```
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
```

Para hacerlo más visual podemos crear histogramas para expresar los valores anteriores.

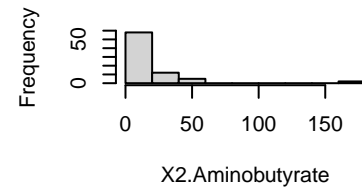
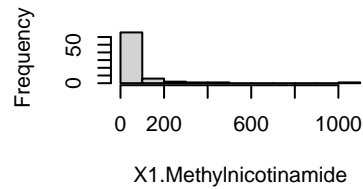
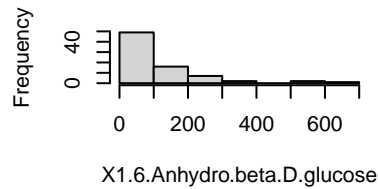
```
library(ggplot2)

numeric_cols <- sapply(data, is.numeric)
```

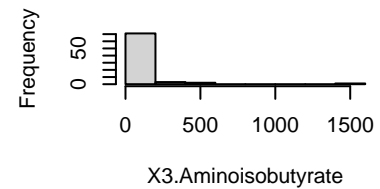
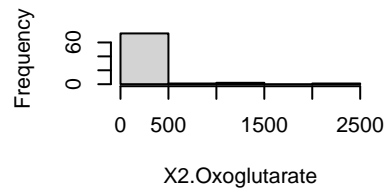
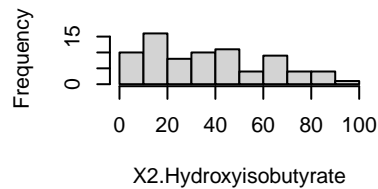
```
data_numeric <- data[, numeric_cols]

par(mfrow = c(3, 3))
for (col in colnames(data_numeric)) {
  hist(data_numeric[[col]], main = paste("Histograma de", col), xlab = col)
}
```

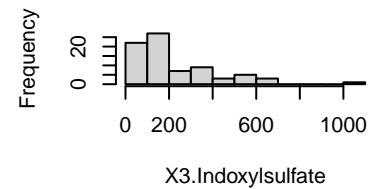
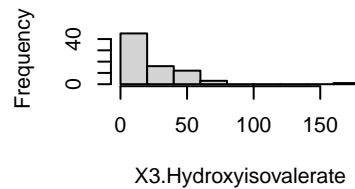
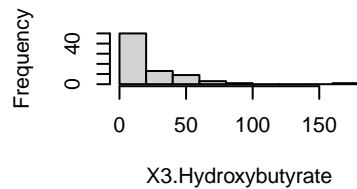
Histograma de X1.6.Anhydro.beta.D. **Histograma de X1.Methylnicotina** **Histograma de X2.Aminobutyra**



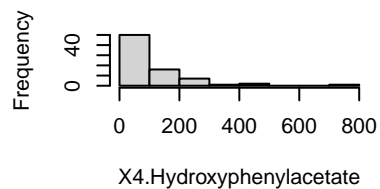
Histograma de X2.Hydroxyisobuty **Histograma de X2.Oxoglutarat** **Histograma de X3.Aminoisobuty**



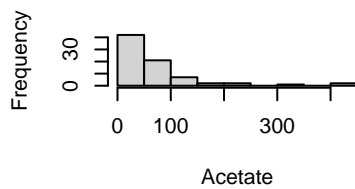
Histograma de X3.Hydroxybutyr **Histograma de X3.Hydroxyisoval** **Histograma de X3.Indoxylsulfa**



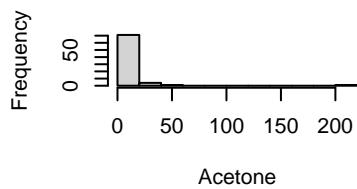
Histograma de X4.Hydroxyphenyla



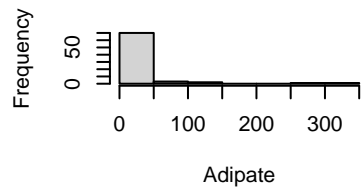
Histograma de Acetate



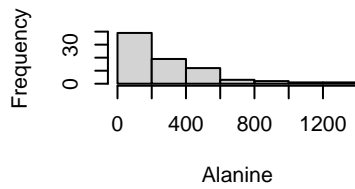
Histograma de Acetone



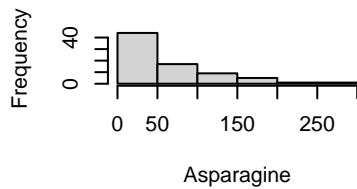
Histograma de Adipate



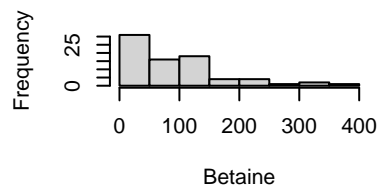
Histograma de Alanine



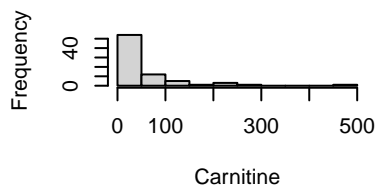
Histograma de Asparagine



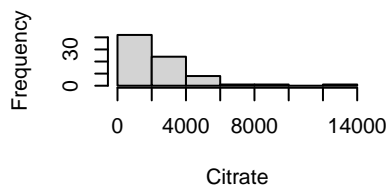
Histograma de Betaine

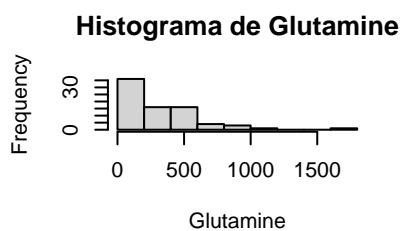
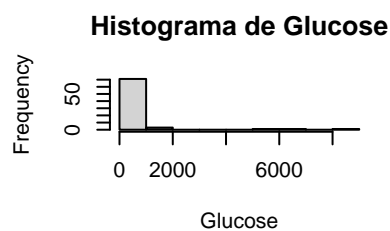
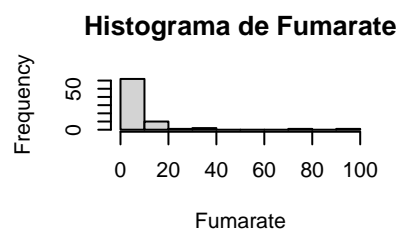
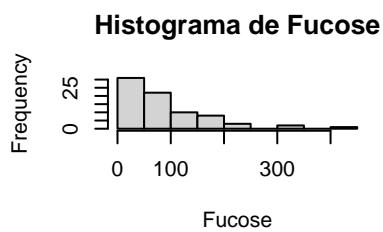
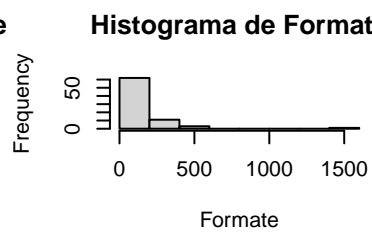
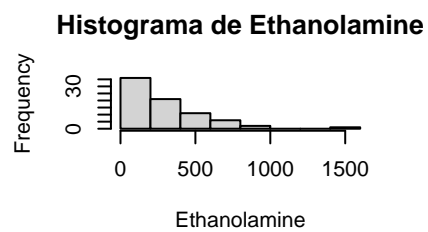
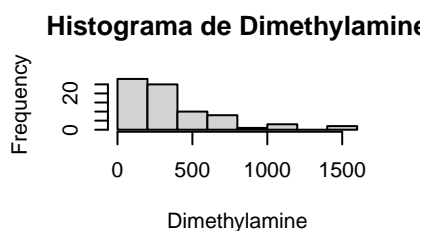
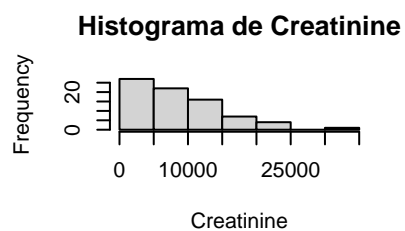
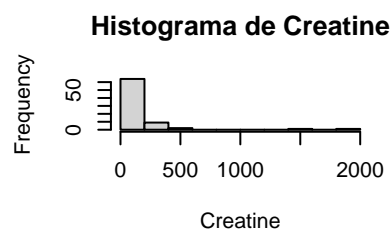


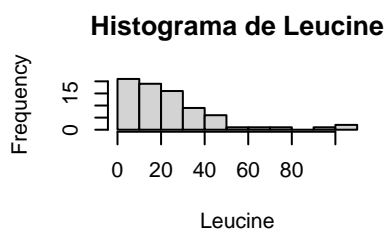
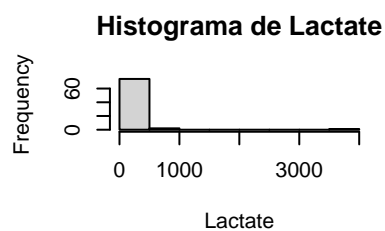
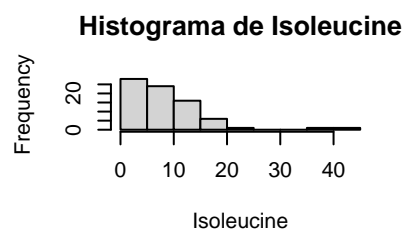
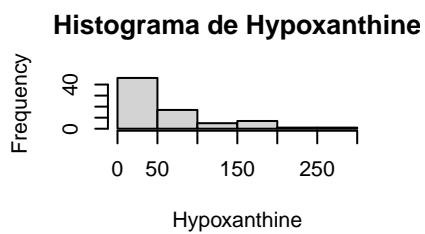
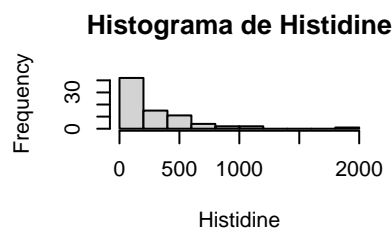
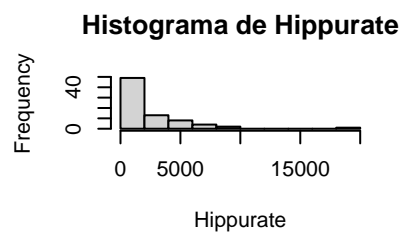
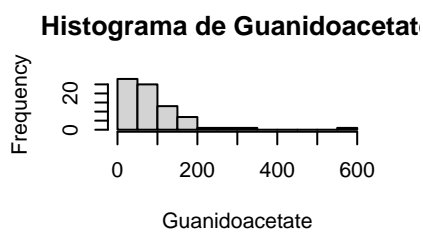
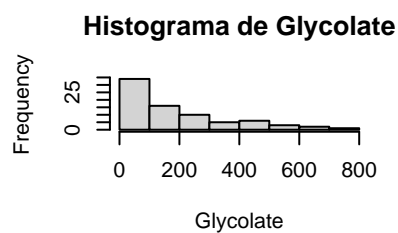
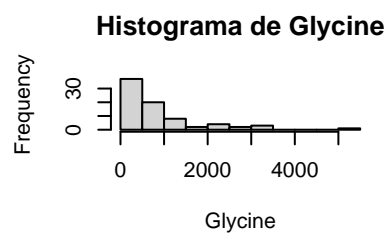
Histograma de Carnitine

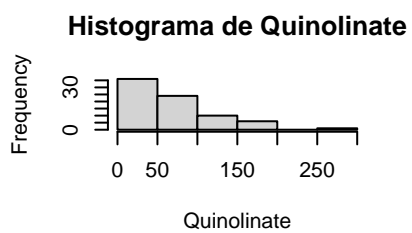
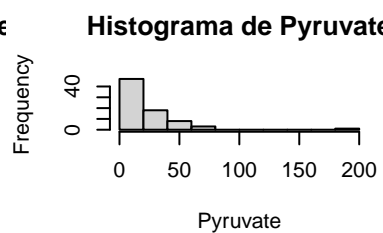
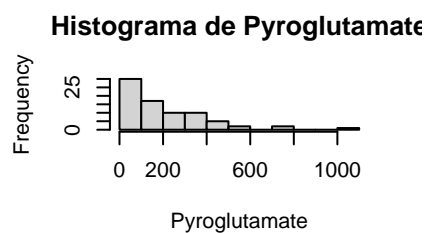
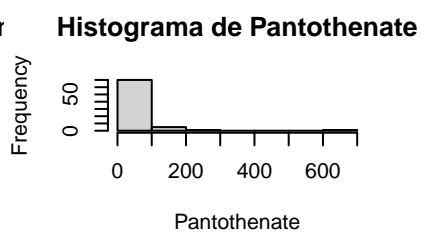
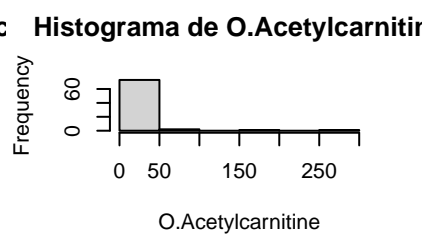
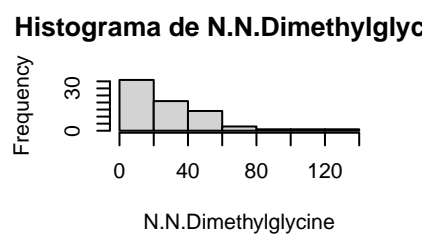
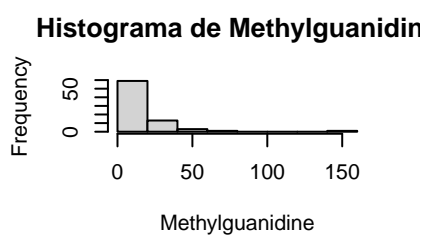
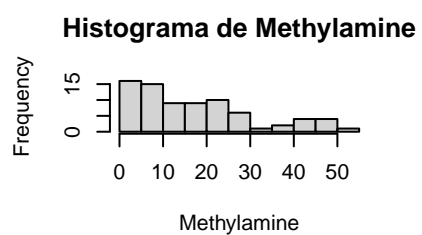
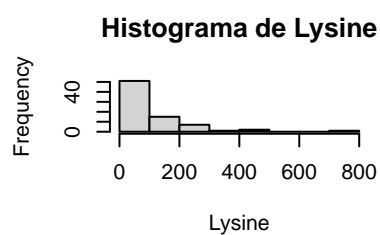


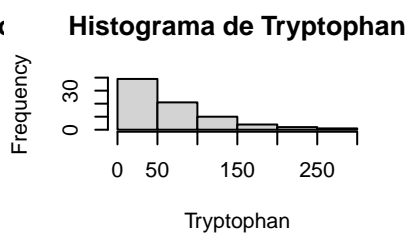
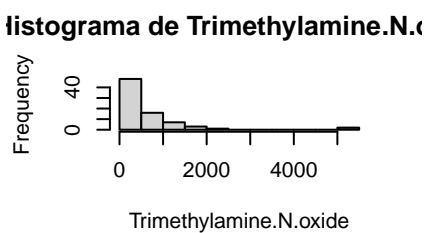
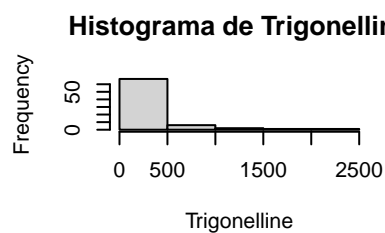
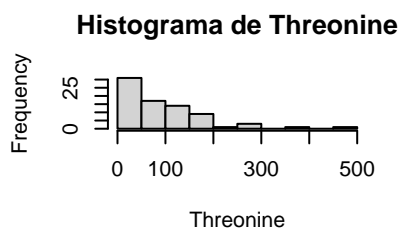
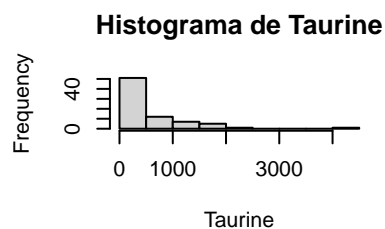
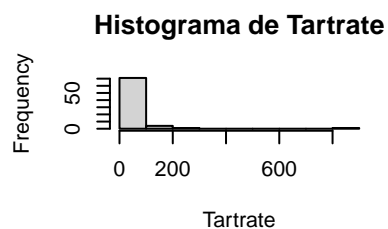
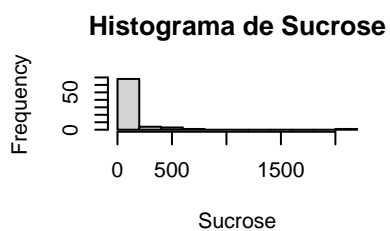
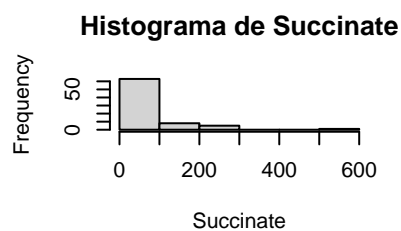
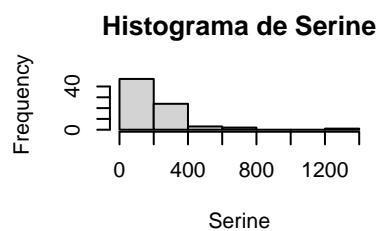
Histograma de Citrate

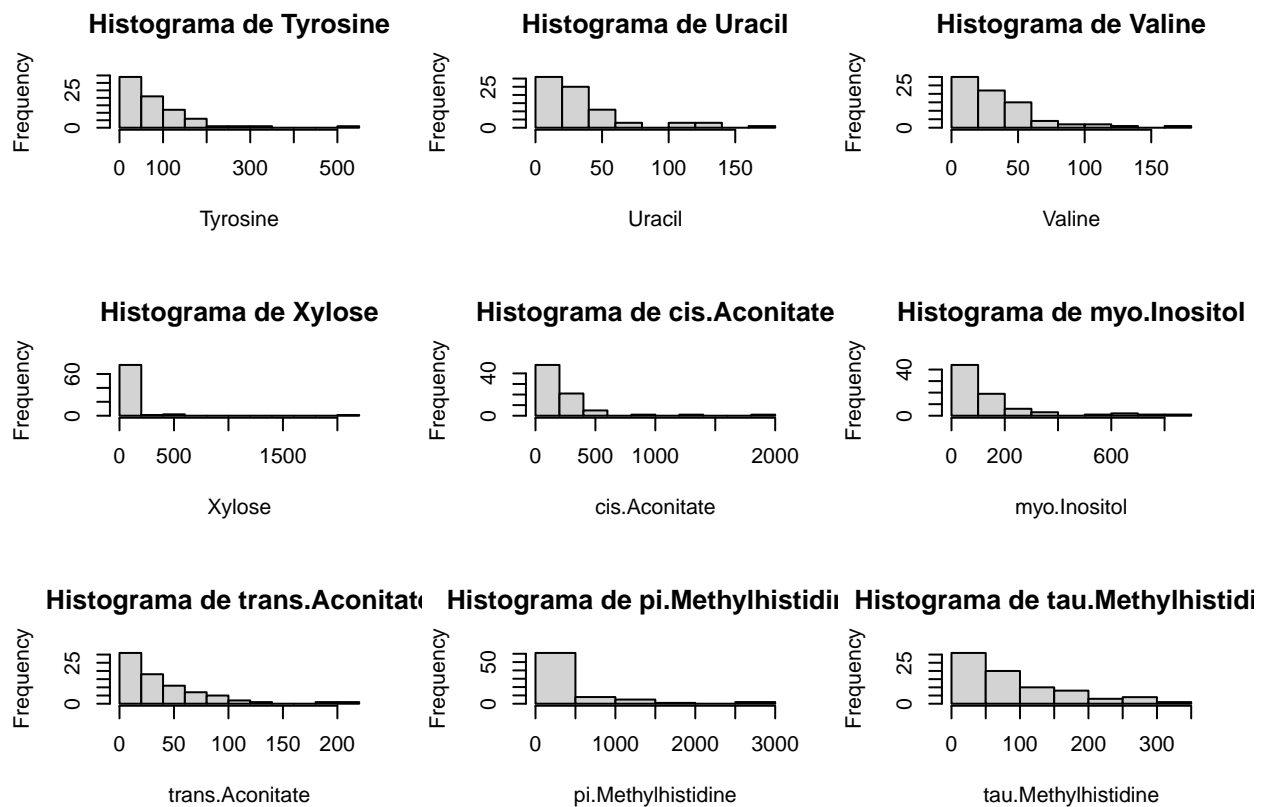












También podemos observar la distribución de la variable categórica, pérdida muscular que clasifica a los pacientes en dos grupos, en caquéticos y control.

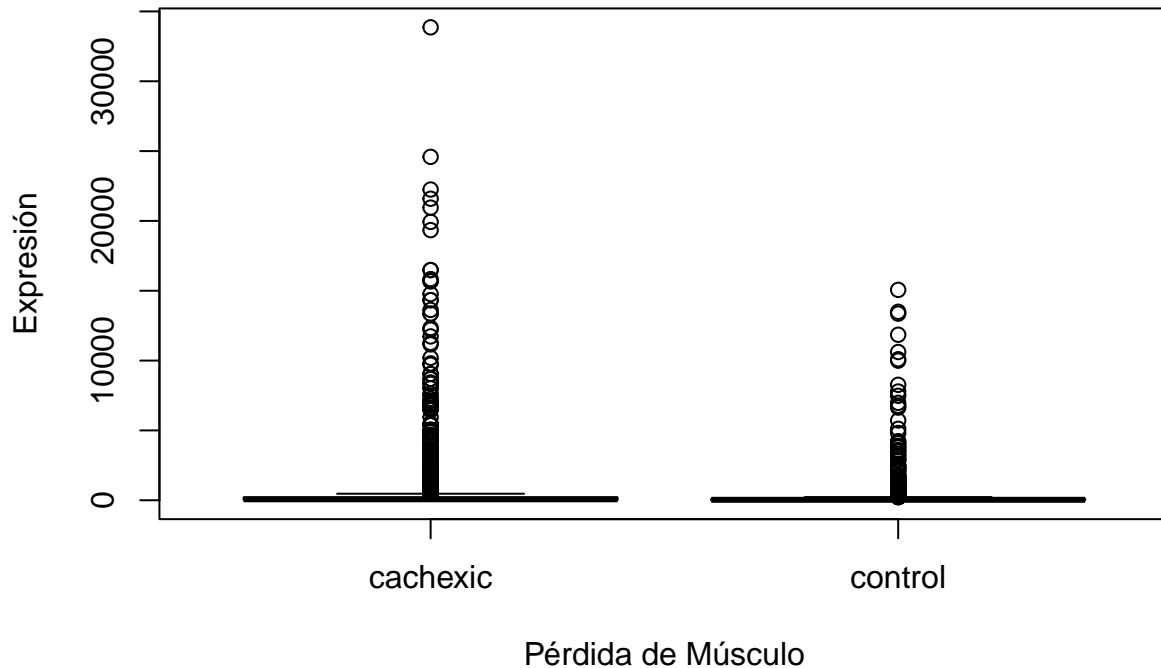
```
table(data$Muscle.loss)
```

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

Podemos realiza un boxplot, para visualizar mejor la distribución de metabolitos en los pacientes categorizados según pérdida de masa muscular. El diagrama de cajas nos permitier ver la mediana, los cuartiles y los valores atípicos de los datos, lo que facilita comparar visualmente estas estadísticas entre diferentes grupos de pacientes.

```
boxplot(assay(se) ~ rowData(se)$Muscle.loss, main="Boxplot de Expresión por Pérdida de Músculo", xlab="")
```

Boxplot de Expresión por Pérdida de Músculo



Podemos ver que en el grupo de caquexia presenta un grupo más grande de outliers, y se distribuye de forma amplia el rango de valores de expresión de los metabolitos comparado con el grupo control. Esto puede presentar un significado biológico que podría estar relacionado a la gravedad del estado de pérdida muscular en el cual no vemos en el grupo control (sanos).

A continuación realizaremos un Análisis de Componentes Principales (PCA) sobre nuestro conjunto de expresión. La PCA es útil para reducir la dimensionalidad de nuestros datos y nos permite visualizar patrones o agrupaciones que no serían tan sencillos de objetivar con los datos de alta dimensión.

Para ello preparamos primero nuestros datos. Convirtiendo nuestros datos de expresión en una matriz, para ello excluimos las dos primeras columnas que contienen información que no son valores de expresión como tal.

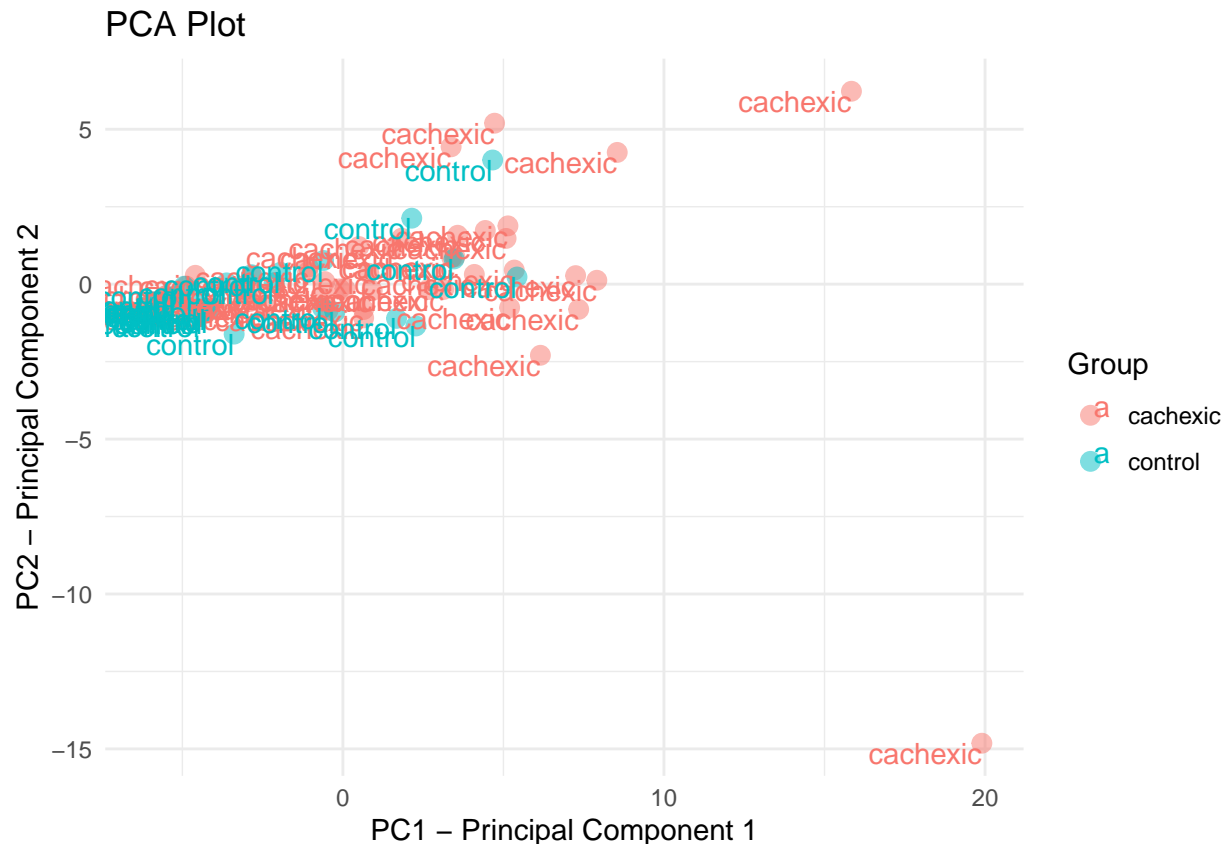
Posteriormente escalamos los datos para que tengan media cero y desviación estándar de uno. También podemos realizar la función `prcomp()` que nos aseguran que los datos están bien centrados y escalados.

```
expression_data <- as.matrix(data[, -c(1:2)])  
expression_scaled <- scale(expression_data)  
pca_result <- prcomp(expression_scaled, center = TRUE, scale. = TRUE)
```

Finalmente creamos el gráfico de puntos de los componentes principales, en este caso los dos primeros componentes principales que representan la mayor varianza en nuestro dataset.

```
# Convertir a dataframe  
pca_data <- data.frame(PC1 = pca_result$x[,1], PC2 = pca_result$x[,2], Group = data$Muscle.loss)
```

```
# Usar ggplot2 para visualizar
ggplot(pca_data, aes(x = PC1, y = PC2, color = Group)) +
  geom_point(alpha = 0.5, size = 3) +
  geom_text(aes(label = Group), vjust = 1, hjust = 1) +
  labs(x = "PC1 - Principal Component 1", y = "PC2 - Principal Component 2", title = "PCA Plot") +
  theme_minimal()
```



Vemos que la distribución de los puntos presenta una superposición clara entre los dos grupos categóricos del estado de masa muscular (caquexia y control), aunque podemos observar que hay algunos puntos de caquexia que están bastante alejados del grupo principal. La mayoría de los valores están agrupados en el centro, esto significa que hay similitudes en los dos primeros componentes principales.

Además podemos ver que la mayor varianza lo presenta el PC1 (componente principal 1), esto podría indicar la existencia de algún proceso biológico que diferencia nuestros dos grupos de estado de pérdida muscular.

Por otro lado, el PC2 (componente principal 2), muestra menos separación entre los grupos, por lo que parece contribuir menos en la diferencia entre estos.

5. Repositorio Github

La dirección (url) del repositorio: <https://github.com/Annecy123/Chen-Ye-Ana-PEC1->

6. Referencias

- Introducción a clases de datos tabulares en Bioconductor: https://github.com/ASPteaching/Omics_Data_Analysis-Case_Study_0-Introduction_to_BioC/blob/main/Introduction_2_Bioc_classes_4_tabular_data.pdf
- ExpressionSet Introduction: <https://www.bioconductor.org/packages/devel/bioc/vignettes/Biobase/inst/doc/ExpressionSetIntroduction.pdf>
- Casos prácticos en análisis ómico: <https://aspteaching.github.io/AMVCasos/>
- MetaboAnalyst para análisis y visualización de datos metabólicos: <https://www.metaboanalyst.ca/docs/RTutorial.xhtml>
- Descripción de un conjunto de datos de Cachexia en Metabolómica: <https://github.com/nutrimetabolomics/metaboData/blob/main/Datasets/2024-Cachexia/description.md?plain=1>