

Análisis de datos ómicos PEC 1

Pedro Belijar Sáez

Tabla de contenidos

Abstract	1
Objetivo del estudio	2
Materiales y métodos	2
Resultados y gráficos	2
Gráfico de distribución de los datos según 7 biomarcadores. Media t-student.	3
Análisis de los datos seleccionados.....	7
Identificación de otros valores de interés	12
Discusión y limitaciones y conclusiones del estudio.....	16
Repositorio github	16
Anexo I: Base de datos.....	17
Anexo II: Paquetes y librerías instaladas	17

```
options(repos = c(CRAN = "https://cloud.r-project.org"))
```

Abstract

Durante un estudio clínico se analizan un total 77 muestras de orina que pertenecen a pacientes con la enfermedad de la Caquexia (47 de ellos) relacionada con la pérdida de masa muscular y otros efectos degenerativos y otros 30 pacientes a modo de grupo de control. Se recogen hasta 61 factores que podrían estar relacionados con la aparición de la enfermedad, ser un detonante o un efecto de dicha enfermedad.

En nuestro estudio hemos seleccionado 7 biomarcadores relacionados, a priori, con la pérdida muscular y fuerza, pérdida de peso o actividades relacionadas con los procesos metabólicos. Durante el análisis de los datos hemos logrado valorar y graficar la relación existente entre los 7 biomarcadores anteriores y la enfermedad caquexia.

Finalmente, hemos identificado otros 3 biomarcadores (X1.6.Anhydro.beta.D.glucose, X2.Aminobutyrate X2.Hydroxyisobutyrate) como prioritarios para el estudio de la caquexia gracias a 2 procesos estadísticos: el análisis del p-valor a través de t-test y un análisis de los componentes principales (PCA).

Objetivo del estudio

Identificar si existen valores anómalos de algunos de los biomarcadores (7) que se relacionan con la enfermedad como son la creatina, creatinina, 3-Hidroxibutirato, acetona, Carnitina, citrato y lactato.

Realizar una comparativa entre el grupo control y el grupo de enfermos para identificar patrones.

Búsqueda de otros biomarcadores de vital importancia relacionado con la enfermedad de Caquexia.

Materiales y métodos

Para realizar el estudio partimos de la base de datos de human_cachexia: https://raw.githubusercontent.com/nutrimetabolomics/metaboData/refs/heads/main/Datasets/2024-Cachexia/human_cachexia.csv

Utilizaremos el programa R Studio y sus librerías estadísticas.

Resultados y gráficos

Obtención de datos sobre Caquexia

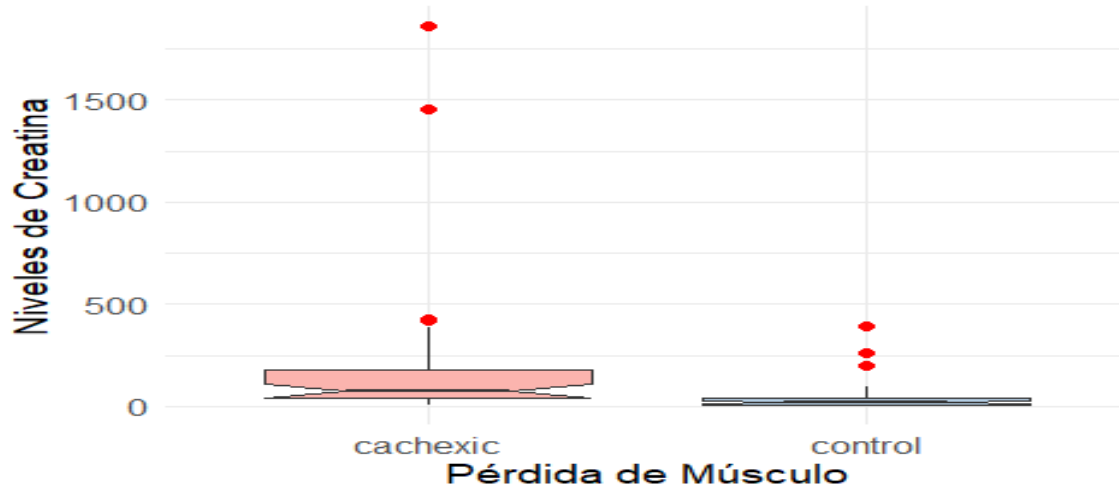
```
url <- "https://raw.githubusercontent.com/nutrimetabolomics/metaboData/refs/heads/main/Datasets/2024-Cachexia/human_cachexia.csv"
datos <- read.csv(url)
```

Creación de la dataset y de la clase SummarizedExperiment

```
se
##               class: SummarizedExperiment
##               dim:    77 64
##               metadata(0):
##               assays(1): counts
## rownames(77):  PIF_178 PIF_087 ... NETL_003_V1 NETL_003_V2
##               rowData names(2):  GeneID Description
## colnames(64):  Muscle.loss X1.6.Anhydro.beta.D.glucose ...
##               pi.Methylhistidine tau.Methylhistidine
## colData names(2): SampleID Condition
```

Gráfico de distribución de los datos según 7 biomarcadores. Media t-student.

ribución de Creatina según Estado de Pérdida d

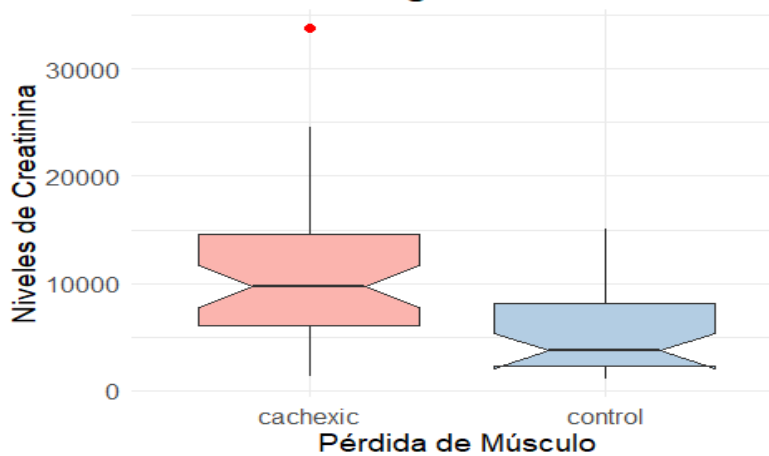


```
t_test_result <- t.test(datos$Creatine ~ datos$Muscle.loss)
print(t_test_result)
```

```
##
##              Welch          Two          Sample          t-test
##
##      data:          datos$Creatine          by          datos$Muscle.loss
##      t      =      2.3988,      df      =      55.284,      p-value      =      0.01985
## alternative hypothesis: true difference in means between group cachexic and group control is not e
qual
##      95          percent          confidence          interval:
##      20.3217          226.4964
##      estimates:
##      mean      in      group      sample      mean      in      group
##      174.91340      51.50433
```

Creatinina

ibución de Creatinina según Estado de Pérdida



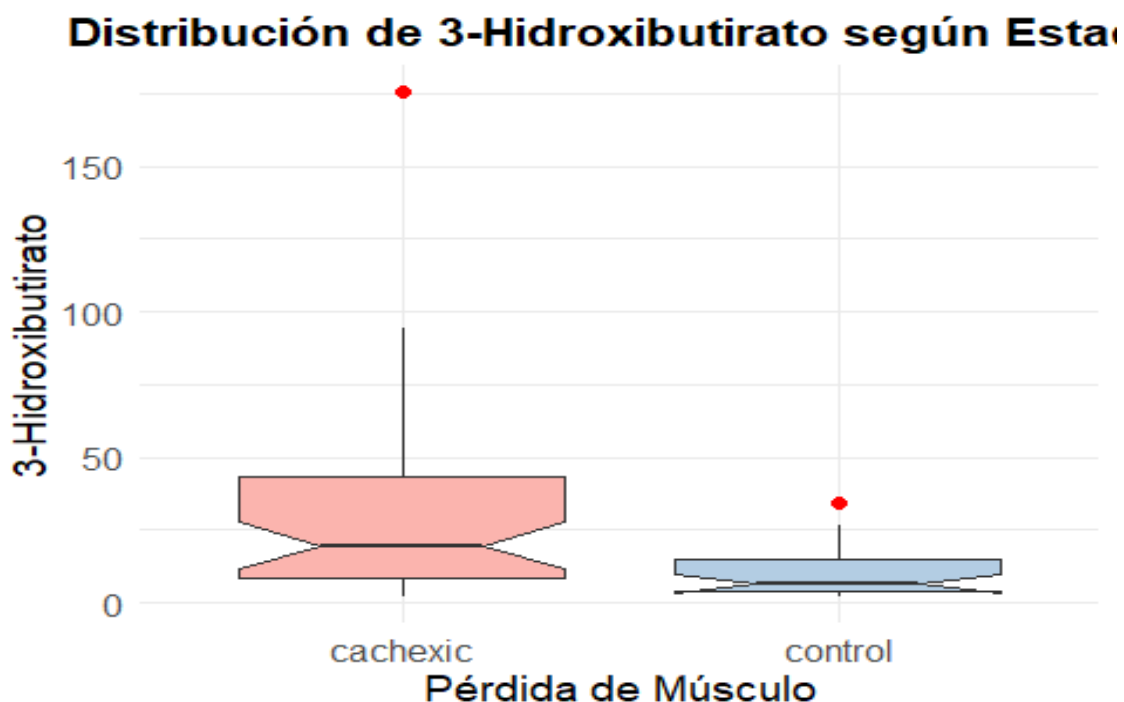
```

t_test_result <- t.test(datos$Creatinine ~ datos$Muscle.loss)
print(t_test_result)

##
##                               Welch          Two          Sample          t-test
##
## data:                  datos$Creatinine          by          datos$Muscle.loss
## t = 4.0209, df = 74.913, p-value = 0.000137
## alternative hypothesis: true difference in means between group cachexic and group control is not e
qual
## 95 percent to confidence interval:
## 2574.716
## sample estimates:
## mean in group cachexic mean in group control
## 10722.140 5619.175

```

X3.Hydroxybutyrate



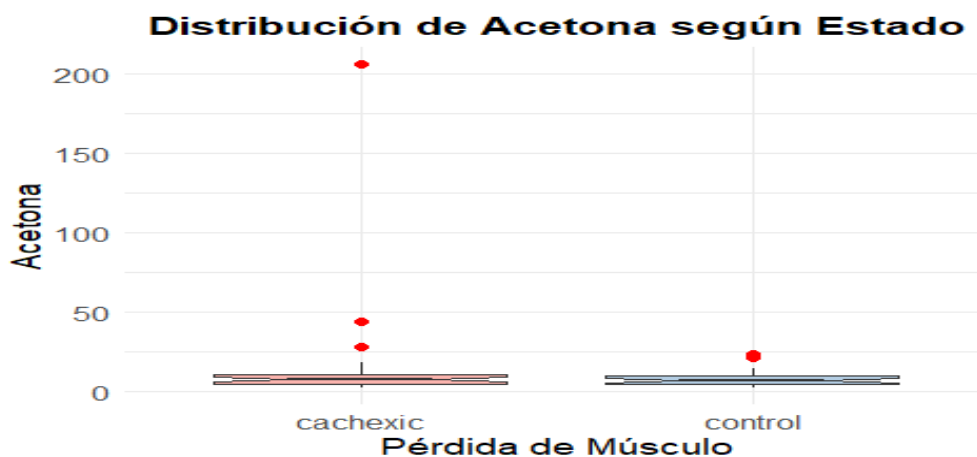
```

t_test_result <- t.test(datos$X3.Hydroxybutyrate ~ datos$Muscle.loss)
print(t_test_result)

##
##                               Welch          Two          Sample          t-test
##
## data:                  datos$X3.Hydroxybutyrate          by          datos$Muscle.loss
## t = 4.1072, df = 55.263, p-value = 0.0001335
## alternative hypothesis: true difference in means between group cachexic and group control is not e
qual
## 95 percent to confidence interval:
## 9.91565
## sample estimates:
## mean in group cachexic mean in group control
## 29.260638 9.898667

```

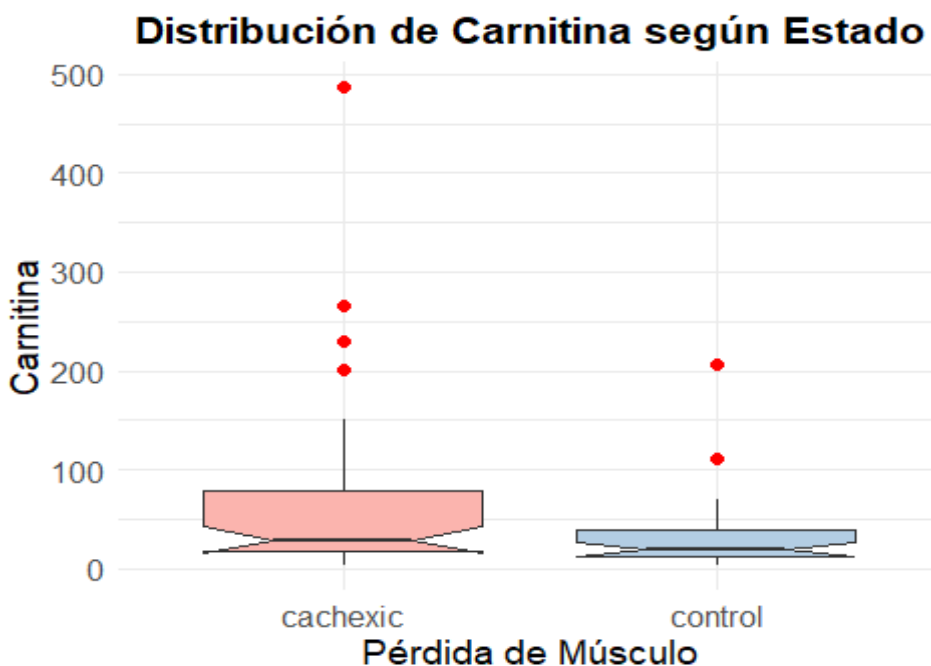
Acetona



```
t_test_result <- t.test(datos$Acetone ~ datos$Muscle.loss)
print(t_test_result)
```

```
##
##              Welch          Two          Sample          t-test
##
##      data:          datos$Acetone          by          datos$Muscle.loss
##      t      = 1.1077,      df      = 51.123,      p-value      = 0.2732
## alternative hypothesis: true difference in means between group cachexic and group control is not e
qual
##
##      95          percent          to          confidence          interval:
##
##      mean      in      group      sample      mean      in      group      estimates:
##      13.34638      cachexic      8.42000      control
```

Carnitina



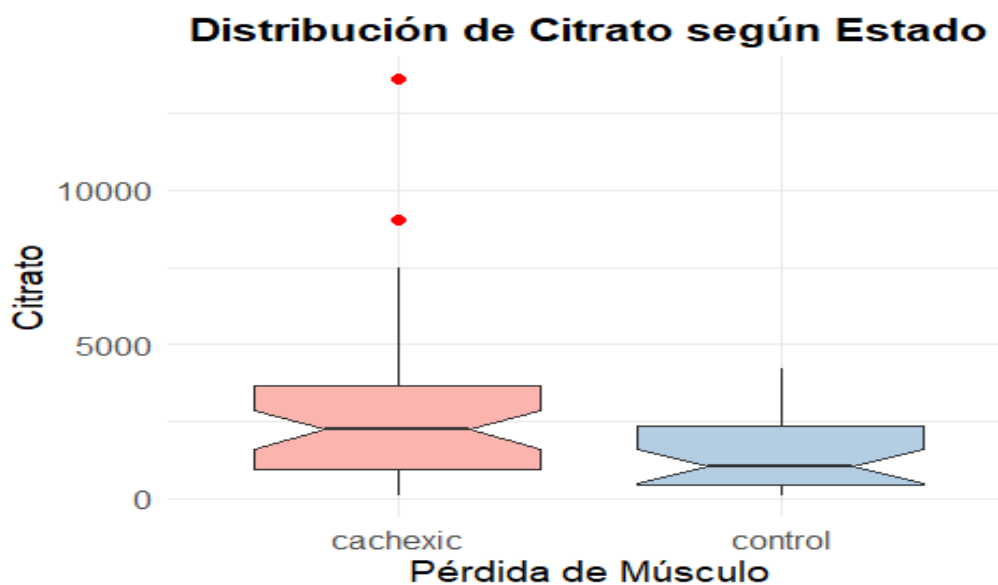
```

t_test_result <- t.test(datos$Carnitine ~ datos$Muscle.loss)
print(t_test_result)

##
##              Welch              Two              Sample              t-test
##
##      data:          datos$Carnitine      by          datos$Muscle.loss
##      t      =      2.1893,      df      =      69.708,      p-value      =      0.03193
## alternative hypothesis: true difference in means between group cachexic and group control is not e
qual
##              95              percent              to              confidence              interval:
##              2.862026
##
##      mean      in      group      sample      mean      in      group      estimates:
##      #      64.62213      cachexic      32.44367      control

```

Citrato



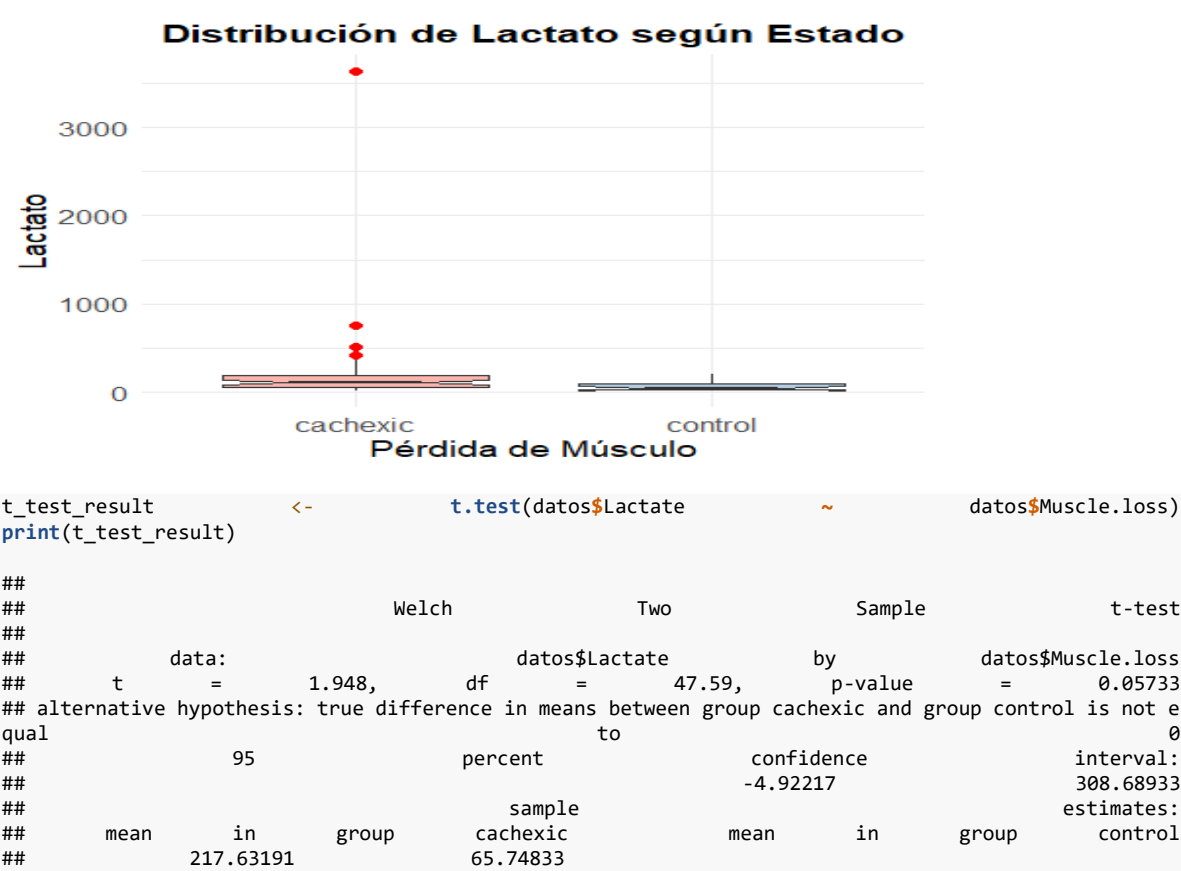
```

t_test_result <- t.test(datos$Citrate ~ datos$Muscle.loss)
print(t_test_result)

##
##              Welch              Two              Sample              t-test
##
##      data:          datos$Citrate      by          datos$Muscle.loss
##      t      =      2.8879,      df      =      73.123,      p-value      =      0.005099
## alternative hypothesis: true difference in means between group cachexic and group control is not e
qual
##              95              percent              to              confidence              interval:
##              386.1651
##
##      mean      in      group      sample      mean      in      group      estimates:
##      #      2720.853      cachexic      1474.719      control

```

Lactato



Análisis de los datos seleccionados

Acotación de los datos

```

datos_estudio <- datos %>%
  select(Muscle.loss, 9, 14, 19, 20, 21, 22, 37)
head(datos_estudio)

```

	Muscle.loss	X3.Hydroxybutyrate	Acetone	Carnitine	Citrate
## Creatine					
## PIF_178	cachexic	56.83	9.49	265.07	3714.50
196.37					
## PIF_087	cachexic	43.82	11.82	120.30	2617.57
212.72					
## PIF_090	cachexic	5.64	4.44	25.03	862.64
221.41					
## NETL_005_V1	cachexic	175.91	206.44	200.34	13629.61
85.63					
## PIF_115	cachexic	76.71	44.26	84.77	854.06
105.64					
## PIF_110	cachexic	31.82	14.44	40.04	1958.63

200.34			
##		Creatinine	Lactate
##	PIF_178	16481.60	106.70
##	PIF_087	15835.35	368.71
##	PIF_090	24587.66	749.95
##	NETL_005_V1	20952.22	368.71
##	PIF_115	6768.26	3640.95
##	PIF_110	15677.78	113.30

Visualización de los datos

```

f<-function(x){
  ifelse(
    hist(x,
    barplot(table(x))
  )
}
par(mfrow=c(3,3))
apply(datos_estudio,2,f)

```

##	Muscle.loss	X3.Hydroxybutyrate	Acetone	Carn
itine				
##	0.7	0.7	0.7	0
.7				
##	Citrate	Creatine	Creatinine	La
ctate				
##	0.7	0.7	0.7	0
.7				



Análisis de los PCA: Cálculo de matriz de varianza y correlaciones entre variables.

```

datosnum <- datos_estudio[apply(datos_estudio, is.numeric)]
datosnum <- scale(datosnum, center = TRUE, scale=FALSE)
apply(datosnum,2, mean)

## X3.Hydroxybutyrate      Acetone      Carnitine      Ci
trate
##      9.665339e-16      7.821857e-16      8.680601e-16      -9.13872
2e-14
##      Creatine      Creatinine      Lactate
##      1.703003e-15      -6.499274e-13      3.259478e-16

n<- dim(datos_estudio)[1]
S<-cov(datosnum)*(n-1)/n
show(S)

##      X3.Hydroxybutyrate      Acetone      Carnitine      Citrat
e
## X3.Hydroxybutyrate      677.4685      427.3044      620.0402      39865.7
6
## Acetone      427.3044      543.2439      447.4368      28786.7
0
## Carnitine      620.0402      447.4368      5396.7269      58061.5
3
## Citrate      39865.7633      28786.6983      58061.5324      4633054.5
0
## Creatine      620.5471      -105.2377      4355.6401      33836.2
4
## Creatinine      103627.6622      32323.2909      100661.3417      9156470.2
1
## Lactate      3790.5667      2084.6163      2836.1088      52416.5
5
##      Creatine      Creatinine      Lactate
## X3.Hydroxybutyrate      620.5471      103627.66      3790.567
## Acetone      -105.2377      32323.29      2084.616
## Carnitine      4355.6401      100661.34      2836.109
## Citrate      33836.2430      9156470.21      52416.545
## Creatine      73677.7407      234521.11      3158.188
## Creatinine      234521.1099      41414596.48      470068.289
## Lactate      3158.1876      470068.29      174442.000

R<-cor(datosnum)
show(R)

##      X3.Hydroxybutyrate      Acetone      Carnitine      Citrat
e
## X3.Hydroxybutyrate      1.00000000      0.70436117      0.32427249      0.7115768
6

```

```
## Acetone          0.70436117  1.00000000  0.26131782  0.5738002
1
## Carnitine        0.32427249  0.26131782  1.00000000  0.3671889
2
## Citrate          0.71157686  0.57380021  0.36718892  1.0000000
0
## Creatine         0.08783388 -0.01663435  0.21843318  0.0579135
6
## Creatinine       0.61866340  0.21549709  0.21292212  0.6610246
7
## Lactate          0.34868563  0.21414264  0.09243408  0.0583054
5
##                  Creatine  Creatinine  Lactate
## X3.Hydroxybutyrate 0.08783388  0.6186634  0.34868563
## Acetone           -0.01663435  0.2154971  0.21414264
## Carnitine          0.21843318  0.2129221  0.09243408
## Citrate            0.05791356  0.6610247  0.05830545
## Creatine            1.00000000  0.1342570  0.02785763
## Creatinine         0.13425704  1.0000000  0.17488748
## Lactate            0.02785763  0.1748875  1.00000000
```

Creatina y citrato muestran valores muy altos lo que indica una gran dispersión. Además, podemos añadir que el aumento de creatina disminuye el valor de acetona.

X3.Hydroxybutyrate y Acetona tiene un alto nivel de correlación entre ambas. Creatine con Acetone y Lactate con Creatine no presentan una correlación importante.

PCAs

```
EIG                                <- eigen(S)
show(EIG)

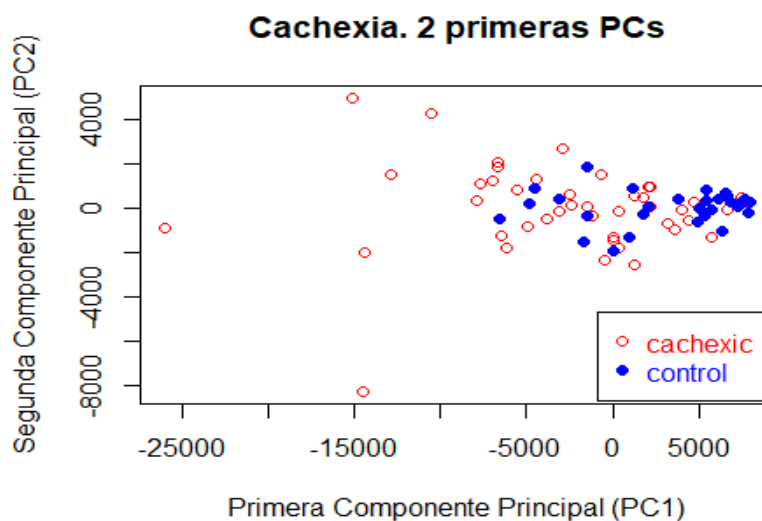
##                  eigen()                  decomposition
##                  $values
## [1] 4.357492e+07 2.481968e+06 1.680939e+05 7.245850e+04 4.387274e+03
## [6] 4.516335e+02 1.124756e+02
##
##                  $vectors
## [1,] -0.0025253308 -0.006053291 0.017712914 -0.002111101 -0.020537790
## [2,] -0.0008738129 -0.008294191 0.012995334 0.001880098 -0.012634176
## [3,] -0.0025551562 -0.013482316 0.015044468 -0.059375060 -0.997725191
## [4,] -0.2288962230 -0.973038210 0.020890854 -0.006242198 0.014666457
## [5,] -0.0054266785 0.008592476 0.001650884 -0.998182603 0.059309160
## [6,] -0.9733685494 0.228557571 -0.016126963 0.007161435 -0.001264002
## [7,] -0.0108197762 0.024434405 0.999295655 0.002801982 0.015124175
## [6] [7]
## [1,] -0.6455370872 0.7632161192
## [2,] -0.7630140251 -0.6460712069
```

```
##          [3,]                0.0233342996        -0.0077406749
##          [4,]                0.0106071879         0.0003894198
##          [5,]               -0.0014708599        -0.0023972414
##          [6,]               -0.0004187289        -0.0014020181
## [7,]  0.0207876367 -0.0050367514

eigenVecs1                                <- eig$vectors
PCAS1                                     <- datosnum %*% eigenVecs1
head(PCAS1)

##          [,1]      [,2]      [,3]      [,4]      [,5]      [,6]
## PIF_178    -7880.319    327.7705  -141.8491  -36.02949  -197.9547082   -4.95158
## PIF_087    -7002.719   1255.9738   105.1314  -40.76888  -63.6150984   -7.65129
## PIF_090   -15124.071   4974.9613   306.1016   30.98225    1.7953447    6.03587
## NETL_005_V1 -14503.952  -8294.2474   258.5259   49.32962   -1.1424496  -124.69926
## PIF_115      2191.717    978.6304  3484.7429   23.46122   -0.5157773   -1.19384
## PIF_110    -6695.453   1855.9196  -162.7298  -21.34653    2.6134181  -15.99143
##
##                                     [,7]
## PIF_178                          16.209281
## PIF_087                           5.015124
## PIF_090                         -33.514352
## NETL_005_V1                     -23.110573
## PIF_115                           5.234495
## PIF_110                        -3.934483
```

Los valores de creatinina son muy altos, lo que indica que tiene una gran influencia sobre el PCA 1.



Identificación de otros valores de interés

Vamos a proceder a identificar otros biomarcadores de interés

```
t_tests <- datos %>%
  summarise(across(where(is.numeric),
    ~t.test(. ~ Muscle.loss)$p.value))
print(t_tests)
```

	X1.6.Anhydro.beta.D.glucose	X1.Methylnicotinamide	X2.Aminobutyrate			
## 1	0.03531943	0.9435367	0.007859048			
	X2.Hydroxyisobutyrate	X2.Oxoglutarate	X3.Aminoisobutyrate	X3.Hydroxybutyrate		
## 1	0.004893295	0.1585788	0.1021946	0.00133459		
	X3.Hydroxyisovalerate	X3.Indoxylsulfate	X4.Hydroxyphenylacetate	Acetate		
## 1	0.003458034	0.004849645	0.5272919	0.001740727		
	Acetone	Adipate	Alanine	Asparagine	Betaine	Carnitine
## 1	0.2731685	0.007515595	0.0002960803	0.003566603	0.00217492	0.03192506
	Citrate	Creatine	Creatinine	Dimethylamine	Ethanolamine	Formate
## 1	0.005098507	0.01984702	0.0001369605	5.34899e-05	0.02011237	0.004909304
	Fucose	Fumarate	Glucose	Glutamine	Glycine	Glycolate
## 1	0.001863466	0.02620524	0.009239445	0.00033916	0.01304117	0.04940718
	Guanidoacetate	Hippurate	Histidine	Hypoxanthine	Isoleucine	Lactate
## 1	0.1165822	0.0101546	0.004541154	0.2421171	0.1087749	0.05733204
	Leucine	Lysine	Methylamine	Methylguanidine	N.N.Dimethylglycine	
## 1	2.662822e-05	0.3347563	0.001496975	0.202775	2.114015e-05	
	O.Acetylcarnitine	Pantothenate	Pyroglutamate	Pyruvate	Quinolinat	e
## 1	0.02418993	0.6085017	7.289814e-05	0.007357362	2.708732e-05	
	Serine	Succinate	Sucrose	Tartrate	Taurine	Threonine
## 1	0.0008173888	0.003676506	0.06558505	0.3758724	0.01418748	0.001183249
	Trigonelline	Trimethylamine.N.oxide	Tryptophan	Tyrosine	Uracil	
## 1	0.003219917	0.01586788	0.0008885986	0.004456642	0.5386741	
	Valine	Xylose	cis.Aconitate	myo.Inositol	trans.Aconitate	
## 1	1.574212e-05	0.1335317	0.0006190184	0.0003611776	0.01953314	
		pi.Methylhistidine	tau.Methylhistidine			
## 1	0.09958934	0.01710741				

Las variables con p-valores bajos indican diferencias significativas entre los grupos, lo cual podría señalar importancia en la cachexia.

Los valores identificados son: X1.6.Anhydro.beta.D.glucose - 0.0353
X2.Aminobutyrate - 0.00786 X2.Hydroxyisobutyrate - 0.00489

```
## Call:
## glm(formula = Muscle.loss ~ ., family = "binomial", data = datos_estandarizados)
##
## Coefficients:
## Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.454e+01 8.876e+04 0 1
## X1.6.Anhydro.beta.D.glucose -5.481e+00 5.508e+05 0 1
## X1.Methylnicotinamide 8.692e+01 1.923e+06 0 1
## X2.Aminobutyrate -2.993e+01 1.659e+06 0 1
## X2.Hydroxyisobutyrate 1.667e-01 2.097e+06 0 1
## X2.Oxoglutarate -1.733e+02 2.114e+06 0 1
## X3.Aminoisobutyrate -5.023e+01 1.232e+06 0 1
## X3.Hydroxybutyrate 2.313e+02 5.011e+06 0 1
## X3.Hydroxyisovalerate -1.371e+02 5.210e+05 0 1
## X3.Indoxylsulfate -5.519e+01 1.269e+06 0 1
## X4.Hydroxyphenylacetate 8.105e+01 7.312e+05 0 1
## Acetate -7.397e+01 8.845e+05 0 1
## Acetone -1.321e+02 1.363e+06 0 1
## Adipate 4.296e+01 8.139e+05 0 1
## Alanine 2.262e+01 1.256e+06 0 1
## Asparagine -9.753e+01 2.197e+06 0 1
## Betaine 3.812e+01 2.020e+06 0 1
## Carnitine -1.778e+01 5.882e+05 0 1
## Citrate -1.899e+01 6.701e+05 0 1
## Creatine 1.411e+01 2.801e+05 0 1
## Creatinine 9.665e+01 2.305e+06 0 1
## Dimethylamine -7.600e+01 4.132e+05 0 1
## Ethanolamine 3.214e+01 2.964e+06 0 1
## Formate -5.587e+01 2.108e+06 0 1
## Fucose -1.298e+02 9.607e+05 0 1
## Fumarate 1.149e+02 2.839e+06 0 1
## Glucose -2.904e+01 8.398e+05 0 1
## Glutamine -1.229e+02 1.817e+06 0 1
## Glycine 5.614e+01 8.142e+05 0 1
## Glycolate 1.889e+01 1.140e+06 0 1
## Guanidoacetate 1.284e+01 3.754e+05 0 1
## Hippurate -1.646e+00 5.821e+05 0 1
## Histidine -5.636e+01 2.189e+06 0 1
## Hypoxanthine -1.038e+02 1.222e+06 0 1
## Isoleucine -2.278e+01 1.108e+06 0 1
## Lactate 1.952e+01 1.359e+06 0 1
## Leucine -8.907e+01 5.680e+05 0 1
```

```

## Lysine          3.254e+01  5.541e+05      0      1
## Methylamine     2.375e+00  1.288e+06      0      1
## Methylguanidine  4.683e+00  1.167e+06      0      1
## N.N.Dimethylglycine -5.348e+01  1.173e+06      0      1
## O.Acetylcarnitine -7.945e+00  8.943e+05      0      1
## Pantothenate    -6.074e+01  2.042e+06      0      1
## Pyroglutamate    5.287e+00  3.334e+06      0      1
## Pyruvate         -8.311e-01  8.989e+05      0      1
## Quinolininate   -2.970e+01  2.070e+06      0      1
## Serine           -2.356e+01  7.922e+05      0      1
## Succinate        -4.938e+01  2.081e+06      0      1
## Sucrose          -6.871e+01  1.367e+06      0      1
## Tartrate         2.771e+01  5.923e+05      0      1
## Taurine          8.860e+01  1.164e+06      0      1
## Threonine        9.439e+01  1.690e+06      0      1
## Trigonelline     5.138e+01  1.597e+06      0      1
## Trimethylamine.N.oxide 1.980e+01  1.814e+06      0      1
## Tryptophan       -4.122e+01  7.509e+05      0      1
## Tyrosine         1.701e+02  3.123e+06      0      1
## Uracil           9.375e+01  5.281e+05      0      1
## Valine           7.175e+01  2.789e+06      0      1
## Xylose           -6.636e+00  1.286e+06      0      1
## cis.Aconitate     1.074e+02  1.391e+06      0      1
## myo.Inositol      -1.831e+01  1.497e+06      0      1
## trans.Aconitate    2.307e+01  8.176e+05      0      1
## pi.Methylhistidine -6.283e+01  3.208e+05      0      1
## tau.Methylhistidine 5.259e+01  7.032e+05      0      1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 1.0296e+02 on 76 degrees of freedom
## Residual deviance: 1.2833e-09 on 13 degrees of freedom
## AIC: 128
##
## Number of Fisher Scoring iterations: 25

```

Las variables con valores de importancia más altos en la salida son las que probablemente tengan mayor influencia en la caquexia.

Identificación PCAs

```

datos_num <- datos %>% select(where(is.numeric))
datos_escalado <- scale(datos_num)
pca <- prcomp(datos_escalado, center = TRUE, scale. = TRUE)
summary(pca)

## Importance of components:
## PC1 PC2 PC3 PC4 PC5 PC6
## Standard deviation 5.0467 2.2701 1.83311 1.74728 1.65906 1.6130 1.

```

```

47304
## Proportion of Variance 0.4043 0.0818 0.05334 0.04846 0.04369 0.0413 0.
03444
## Cumulative Proportion 0.4043 0.4861 0.53941 0.58787 0.63156 0.6729 0.
70730
##
PC8      PC9      PC10     PC11     PC12     PC13
PC14
## Standard deviation      1.36403 1.24275 1.20650 1.1584 1.05503 1.03620
0.9914
## Proportion of Variance 0.02953 0.02451 0.02311 0.0213 0.01767 0.01704
0.0156
## Cumulative Proportion 0.73683 0.76135 0.78445 0.8057 0.82342 0.84046
0.8561
##
PC15     PC16     PC17     PC18     PC19     PC20
PC21
## Standard deviation      0.96773 0.89551 0.86788 0.83041 0.8133 0.73918
0.72112
## Proportion of Variance 0.01487 0.01273 0.01196 0.01095 0.0105 0.00867
0.00825
## Cumulative Proportion 0.87093 0.88366 0.89562 0.90656 0.9171 0.92573
0.93399
##
PC22     PC23     PC24     PC25     PC26     PC27
PC28
## Standard deviation      0.71053 0.64606 0.63389 0.5830 0.5442 0.50539 0
.48743
## Proportion of Variance 0.00801 0.00663 0.00638 0.0054 0.0047 0.00405 0
.00377
## Cumulative Proportion 0.94200 0.94863 0.95500 0.9604 0.9651 0.96916 0
.97293
##
PC29     PC30     PC31     PC32     PC33     PC34
PC35
## Standard deviation      0.42674 0.42427 0.41483 0.38653 0.35092 0.32424
0.31646
## Proportion of Variance 0.00289 0.00286 0.00273 0.00237 0.00195 0.00167
0.00159
## Cumulative Proportion 0.97582 0.97867 0.98141 0.98378 0.98573 0.98740
0.98899
##
PC36     PC37     PC38     PC39     PC40     PC41
00027
## Cumulative Proportion 0.99659 0.9971 0.99753 0.9979 0.99830 0.9986 0.
99888
##
PC50     PC51     PC52     PC53     PC54     PC55
PC56
## Standard deviation      0.10759 0.10374 0.09853 0.08760 0.08258 0.08049
0.06927
## Proportion of Variance 0.00018 0.00017 0.00015 0.00012 0.00011 0.00010
0.00008
## Cumulative Proportion 0.99906 0.99923 0.99939 0.99951 0.99962 0.99972
0.99979
##
PC57     PC58     PC59     PC60     PC61     PC62

```

```

PC63
## Standard deviation      0.05937 0.05673 0.05088 0.04001 0.02972 0.02789
0.01876
## Proportion of Variance 0.00006 0.00005 0.00004 0.00003 0.00001 0.00001
0.00001
## Cumulative Proportion  0.99985 0.99990 0.99994 0.99997 0.99998 0.99999
1.00000

```

Discusión y limitaciones y conclusiones del estudio

Uno de los primeros problemas que se han planteado a la hora de realizar el trabajo es la gran cantidad de variables que podíamos estudiar, teniendo que seleccionar previamente aquellos biomarcadores que pensábamos que estaban relacionados directamente con la enfermedad.

En una primera fase, hemos dejando a un lado, seguramente, otros factores igual o más importantes para el análisis de la enfermedad. Por ello, hemos realizado un análisis PCA y estudio del p-valor con t de student para identificar otro factor determinante.

Con los resultados obtenidos y los gráficos generados hemos podido asegurar que:

- El valor medio de todos los valores valorados es mayor en pacientes con la enfermedad caquexia que en aquellos del grupo control.
- El grupo control presenta una menor dispersión de las variables analizadas.
- X3.Hydroxybutyrata y Acetona presenta una gran correlación en ambos grupos.
- Se identifican 3 biomarcadores con fuerte relación con los pacientes enfermos: X1.6.Anhydro.beta.D.glucose, X2.Aminobutyrate y X2.Hydroxyisobutyrate.
- El análisis de los dos primeros PCAs corrobora que el grupo de enfermos presenta mayor dispersión.

Repositorio github

Enlace al repositorio:

<https://github.com/belijar1985/BELIJAR-Saez-Pedro-PEC1.git>

Anexo I: Base de datos

Anexo II: Paquetes y librerías instaladas