

Análisis de Datos Ómicos - PEC1: Informe

Elisa Fuentes

2024-11-01

Tabla de contenidos

- Abstract.....pág. 1
- Objetivos del estudio.....pág. 2
- Materiales y métodos.....pág. 2
- Resultados.....pág. 7
- Discusión y limitaciones y conclusiones del estudio.....pág. 16
- Referenciaspág. 17
- Enlace al repositorio de Githubpág. 17

Abstract

El síndrome caquético está relacionado con un deterioro funcional del paciente que caracteriza a muchas patologías entre las que se incluyen el cáncer. Hemos explorado los datos disponibles de niveles de 63 metabolitos diferentes tomados a 77 pacientes con cáncer, de los que disponemos la información de si padecen o no síndrome caquético. El objetivo del análisis es entender si existen patrones metabólicos que diferencian los pacientes con síndrome caquético de los que no lo padecen. Hemos observado tendencias a patrones diferenciados en los niveles de los metabolitos explorados entre los pacientes que sufren caquexia de los que no. Para poder concluir y determinar de manera correcta cuáles son dichos patrones, hace falta un análisis de los datos más profundo.

Objetivos

El objetivo del análisis es explorar los datos de que disponemos para entender si existen patrones metabólicos que diferencian los pacientes con síndrome caquético de los que no lo padecen.

Materiales y métodos

Para realizar el estudio hemos utilizado la base de datos *human_cachexia.csv*, obtenidos de Github. Los datos con los que trabajamos son utilizados en varios tutoriales de MetaboAnalyst. Dicha base de datos contiene los contajes contra diferentes metabolitos y una variable de clasificación en dos grupos según los pacientes tienen caquexia o son pacientes control.

Además hemos creado un repositorio en Github para almacenar todos los datos del estudio y el código utilizado para explorar los datos, de manera que otras personas puedan reproducir el estudio. Además, para la exploración de los datos en R, hemos creado un objeto de tipo *SummarizedExperiment*. Finalmente, hemos explorado los datos. A continuación se detalla el proceso seguido para realizar dichas acciones.

1. Descarga de datos

Hemos descargado la base de datos *human_cachexia.csv*, encontrada dentro de la carpeta *2024-Cachexia* del repositorio de github <https://github.com/nutrimetabolomics/metaboData/>.

2. Creación del repositorio

Hemos creado un repositorio en Github llamado Fuentes_Raventós_Elisa_PEC1, donde hemos almacenado este informe, los datos y metadatos y el código de R generado para explorar los datos. Una vez creado, hemos copiado el link del repositorio de Github:

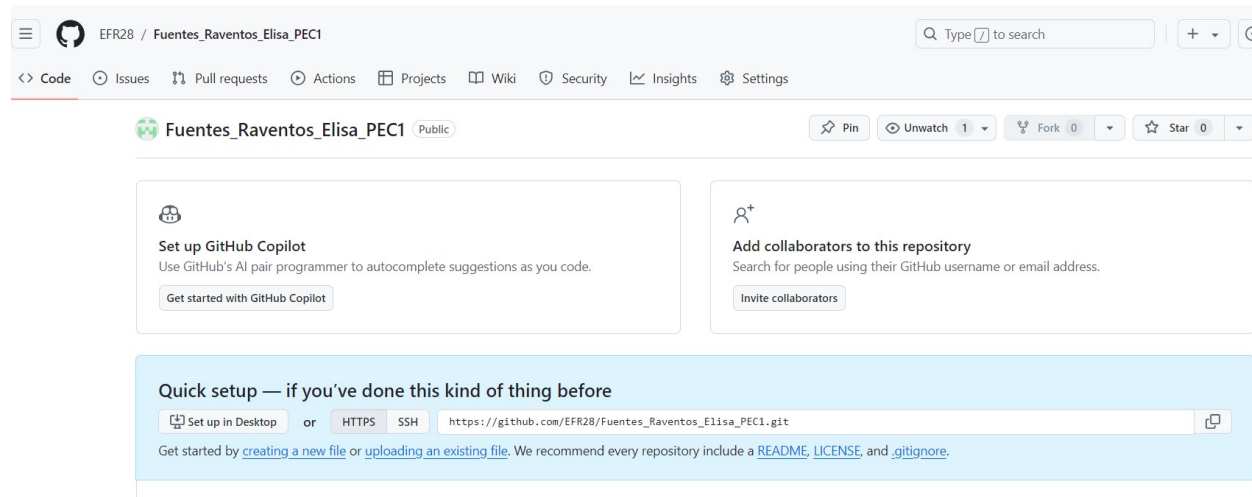
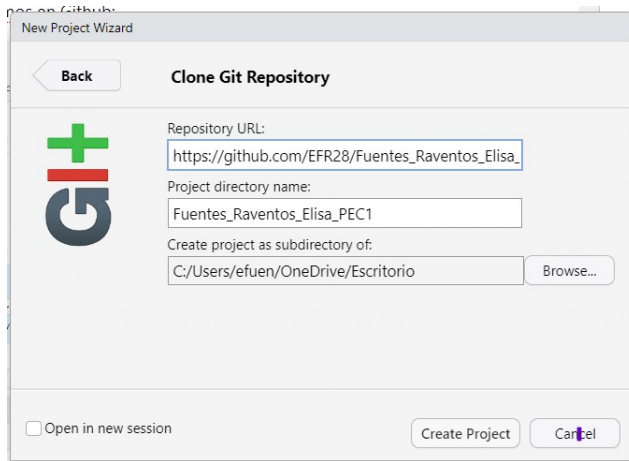


Figure 1: Pantalla del repositorio creado en Github

Hemos empezado un nuevo proyecto en R con control de versiones vinculado al repositorio que hemos creado (File > New Project > Version Control > Git):



3. Carga del dataset a R:

```
# Cargamos dataset
human_cachexia <- read.csv("C:/Users/efuen/OneDrive/Escritorio/UOC/Semestre 3/Datos ómicos/human_ca
```

4. Creación de un contenedor del tipo *SummarizedExperiment* y pre-procesamiento de los datos:

Para trabajar con un conjunto de datos ómicos, lo mejor es utilizar un objeto de tipo *SummarizedExperiment*, de Bioconductor. Este objeto nos permite almacenar los datos y los metadatos de manera conjunto y acceder a cada tipo de los datos cuando nos interese.

Antes de crear el objeto, hemos observado la estructura de la base de datos que tenemos, para saber si nos conviene manipularla.

```
# Cargamos la librería necesaria de Bioconductor
library(SummarizedExperiment)
```

```
# Observamos la estructura de nuestros datos
str(human_cachexia)
```

```
## 'data.frame': 77 obs. of 64 variables:
## $ Muscle.loss : Factor w/ 2 levels "cachexic","control": 1 1 1 1 1 1 1 1 1 1 ...
## $ 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 ...
## $ Quinolinolate        : 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 ...
```

Tenemos una base de datos de 77 pacientes de los cuales se han tomado los valores de 63 metabolitos. Además, tenemos una variable que nos indica si el paciente sufre del síndrome de caquexia o no (control). Dado que nuestro conjunto de datos no contiene valores ausentes (NA), no hemos necesitado lidiar con ellos ni hacer ningún tipo de imputación. Sin embargo, sí los hemos normalizado. Para ello, igual que para el resto de análisis, hemos utilizado el workflow *POMA* y las funciones de dicho paquete. El método de normalización elegido ha sido “log pareto”, que reduce el impacto de valores extremos, pero mantiene cierta sensibilidad a la variabilidad natural.

```
# Trasponemos la matriz para tener los datos dispuestos de tal manera que cada columna sea una muestra
datos_met <- t(as.matrix(human_cachexia)[,2:64])
```

```

variables <- rownames(datos_met)
datos_met <- apply(datos_met, 2, as.numeric)
rownames(datos_met) <- variables

samples <- as.matrix(t(as.matrix(human_cachexia))[1,])

colnames(samples)[1] <- "Group"

features <- as.matrix(rownames(datos_met))
colnames(features) <- "features"

se <- SummarizedExperiment(
  assays = list(counts = datos_met),
  rowData = features,
  colData = samples
)

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(1): features
## colnames(77): PIF_178 PIF_087 ... NETL_003_V1 NETL_003_V2
## colData names(1): Group

# Normalización del conjunto de datos
# Cargamos los paquetes
library(POMA)

## Welcome to POMA!
## Version 1.12.0
## POMAShiny app: https://github.com/pcastellanoescuder/POMAShiny
## For more detailed package information please visit https://pcastellanoescuder.github.io/POMA/

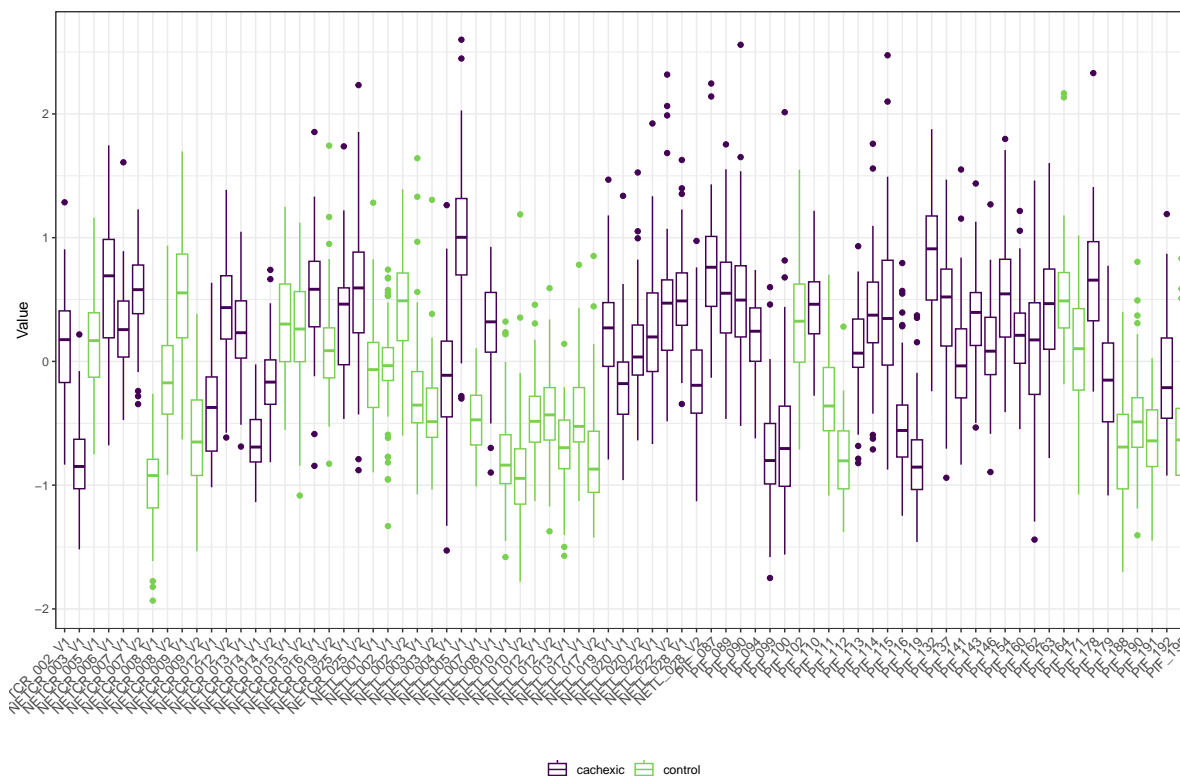
senorm <- se %>%
  PomaNorm(method = "log_pareto")

senorm

## class: SummarizedExperiment
## dim: 63 77
## metadata(0):
## assays(1): ''
## 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(1): Group

# Representamos diagramas de cajas
PomaBoxplots(senorm)

```



En la figura podemos ver la distribución de los valores para cada una de las muestras. Los extremos inferior y superior de la caja representa el rango intercuartil (Q1 y Q3, respectivamente) y la línea horizontal la mediana (Q2) de los datos. Las líneas verticales el rango donde residen todos los valores no-outliers y los puntos son los valores considerados outliers.

5. Análisis explorativo

Para el análisis explorativo de los datos, hemos hecho una descripción de los datos y realizado un correlograma, por el método de Pearson, entre los niveles de metabolitos. También hemos observado los niveles de los diferentes metabolitos en cada uno de los pacientes en un mapa de color, culsterizando por muestras y metabolitos para ver la estructura de agrupación y si existen patrones diferenciados entre los dos grupos de pacientes. Finalmente hemos realizado un análisis de componentes principales para reducir la dimensionalidad de los datos.

Hemos utilizado el workflow *POMA* de Bioconductor, y las funciones *cor()*, *prcomp()*, además de *pheatmap()* y *ggplot()* para las representaciones gráficas.

Resultados

Análisis descriptivo Antes de empezar a explorar y analizar los datos metabólicos, hemos hecho una descripción de los datos. Veamos la distribución de sujetos en cada uno de los grupos de clasificación de síndrome caquético.

```
# Distribución de los sujetos
table(colData(senorm)$Group)
```

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

De los 77 sujetos, 47 padecen el síndrome caquético y 30 pertenecen al grupo control.

No tenemos más información sobre los sujetos. Por lo tanto, hemos centrado el resto del análisis descriptivo en los datos metabólicos. Hemos obtenido un resumen estadístico de cada variable y una matriz de correlaciones entre los niveles de metabolitos.

```
# Calcular estadísticas descriptivas para la matriz de conteos
counts <- t(assay(senorm))
```

```
# Cálculo de estadísticas por columnas (por variable)
```

```
apply(counts, 2, function(x) {
  c(mean = mean(x),
    sd = sd(x),
    median = median(x),
    min = min(x),
    max = max(x))
})
```

```
##      X1.6.Anhydro.beta.D.glucose X1.Methylnicotinamide X2.Aminobutyrate
## mean      -1.298701e-05      1.298701e-05      -5.194805e-05
## sd         6.739040e-01      6.682508e-01      6.140366e-01
## median     -1.750000e-01     -1.900000e-02     -4.400000e-02
## min        -1.528000e+00     -1.074000e+00     -1.188000e+00
## max         1.559000e+00      2.134000e+00      1.875000e+00
##      X2.Hydroxyisobutyrate X2.Oxoglutarate X3.Aminoisobutyrate
## mean      5.832163e-18      2.597403e-05      2.330683e-18
## sd         5.683140e-01      7.444116e-01      7.306879e-01
## median     7.900000e-02      2.700000e-02     -1.290000e-01
## min        -1.254000e+00     -1.229000e+00     -1.247000e+00
## max         8.740000e-01      2.233000e+00      2.330000e+00
##      X3.Hydroxybutyrate X3.Hydroxyisovalerate X3.Indoxylsulfate
## mean      -5.194805e-05      0.0001038961     -6.493506e-05
## sd         6.417478e-01      0.6564560470      6.163675e-01
## median     -8.200000e-02     -0.02300000000     -3.700000e-02
## min        -1.130000e+00     -1.31600000000     -1.179000e+00
## max         1.700000e+00      1.63000000000      1.354000e+00
##      X4.Hydroxyphenylacetate Acetate Acetone Adipate
## mean      -1.298701e-05      1.298701e-05      1.298701e-05     -3.896104e-05
## sd         6.022347e-01      6.811242e-01      5.276679e-01      6.505833e-01
## median     -6.600000e-02      2.100000e-02     -6.900000e-02     -1.140000e-01
## min        -1.120000e+00     -1.383000e+00     -8.110000e-01     -1.101000e+00
## max         1.677000e+00      1.499000e+00      2.600000e+00      2.141000e+00
##      Alanine Asparagine Betaine Carnitine Citrate
```

## mean	2.597403e-05	1.298701e-05	3.896104e-05	-1.298701e-05	7.792208e-05
## sd	6.684138e-01	6.071105e-01	6.829191e-01	6.783170e-01	7.008191e-01
## median	7.200000e-02	-3.100000e-02	8.200000e-02	-1.100000e-01	1.640000e-01
## min	-1.485000e+00	-1.264000e+00	-1.822000e+00	-1.426000e+00	-1.933000e+00
## max	1.311000e+00	1.292000e+00	1.219000e+00	1.798000e+00	1.421000e+00
##	Creatine	Creatinine	Dimethylamine	Ethanolamine	Formate
## mean	2.597403e-05	-2.597403e-05	-3.896104e-05	3.896104e-05	-1.298701e-05
## sd	7.675744e-01	6.116548e-01	6.136830e-01	6.574268e-01	6.296059e-01
## median	-4.000000e-02	1.270000e-01	1.280000e-01	8.100000e-02	-6.000000e-03
## min	-1.449000e+00	-1.314000e+00	-1.273000e+00	-1.561000e+00	-1.776000e+00
## max	2.064000e+00	1.184000e+00	1.280000e+00	1.366000e+00	1.877000e+00
##	Fucose	Fumarate	Glucose	Glutamine	Glycine
## mean	1.298701e-05	-5.194805e-05	-2.597403e-05	1.298701e-05	3.896104e-05
## sd	6.285958e-01	5.962553e-01	7.116683e-01	6.769889e-01	6.952598e-01
## median	1.200000e-02	-1.210000e-01	-2.100000e-02	1.040000e-01	1.300000e-02
## min	-1.531000e+00	-8.840000e-01	-1.259000e+00	-1.328000e+00	-1.614000e+00
## max	1.309000e+00	2.028000e+00	2.246000e+00	1.391000e+00	1.424000e+00
##	Glycolate	Guanidoacetate	Hippurate	Histidine	Hypoxanthine
## mean	5.194805e-05	5.194805e-05	6.493506e-05	-2.597403e-05	-2.597403e-05
## sd	6.989588e-01	6.241967e-01	7.090309e-01	7.020959e-01	6.622653e-01
## median	9.500000e-02	5.800000e-02	-6.000000e-03	1.800000e-02	1.900000e-02
## min	-1.780000e+00	-1.405000e+00	-1.581000e+00	-1.496000e+00	-1.391000e+00
## max	1.154000e+00	1.551000e+00	1.684000e+00	1.481000e+00	1.245000e+00
##	Isoleucine	Lactate	Leucine	Lysine	Methylamine
## mean	-3.896104e-05	6.493506e-05	-5.194805e-05	5.194805e-05	-3.896104e-05
## sd	5.180965e-01	6.779047e-01	5.819229e-01	6.453025e-01	5.984954e-01
## median	2.000000e-02	4.700000e-02	4.700000e-02	1.600000e-02	1.100000e-01
## min	-8.800000e-01	-1.422000e+00	-1.256000e+00	-1.204000e+00	-1.221000e+00
## max	1.373000e+00	2.474000e+00	1.278000e+00	1.642000e+00	9.980000e-01
##	Methylguanidine	N.N.Dimethylglycine	O.Acetylcarnitine	Pantothenate	
## mean	-2.780908e-18	-2.597403e-05	2.597403e-05	-1.298701e-05	
## sd	6.252473e-01	6.584578e-01	6.633122e-01	6.755638e-01	
## median	-1.190000e-01	1.640000e-01	4.300000e-02	-5.000000e-03	
## min	-9.430000e-01	-1.519000e+00	-1.084000e+00	-1.217000e+00	
## max	1.810000e+00	1.262000e+00	2.021000e+00	2.167000e+00	
##	Pyroglutamate	Pyruvate	Quinolate	Serine	Succinate
## mean	2.274008e-19	2.597403e-05	-5.194805e-05	-6.493506e-05	-3.145594e-18
## sd	6.338714e-01	6.608226e-01	5.948206e-01	6.167560e-01	7.352142e-01
## median	6.400000e-02	3.600000e-02	3.000000e-02	1.900000e-02	3.200000e-02
## min	-1.279000e+00	-1.298000e+00	-1.527000e+00	-1.478000e+00	-1.422000e+00
## max	1.369000e+00	1.715000e+00	1.202000e+00	1.543000e+00	1.757000e+00
##	Sucrose	Tartrate	Taurine	Threonine	Trigonelline
## mean	5.550904e-18	3.896104e-05	1.895242e-18	1.953676e-18	-3.006197e-19
## sd	7.003420e-01	6.900970e-01	7.314271e-01	6.488856e-01	7.323955e-01
## median	-1.050000e-01	-1.300000e-01	-3.300000e-02	1.200000e-02	-6.100000e-02
## min	-1.172000e+00	-1.057000e+00	-1.571000e+00	-1.294000e+00	-1.451000e+00
## max	2.318000e+00	2.448000e+00	1.651000e+00	1.308000e+00	1.701000e+00
##	Trimethylamine.N.oxide	Tryptophan	Tyrosine	Uracil	
## mean	-3.896104e-05	-3.896104e-05	3.896104e-05	-3.864406e-18	
## sd	6.542661e-01	6.130499e-01	6.506711e-01	6.177013e-01	
## median	-5.000000e-03	6.000000e-03	8.500000e-02	8.600000e-02	
## min	-1.276000e+00	-1.129000e+00	-1.559000e+00	-1.274000e+00	
## max	1.759000e+00	1.205000e+00	1.537000e+00	1.387000e+00	
##	Valine	Xylose	cis.Aconitate	myo.Inositol	trans.Aconitate

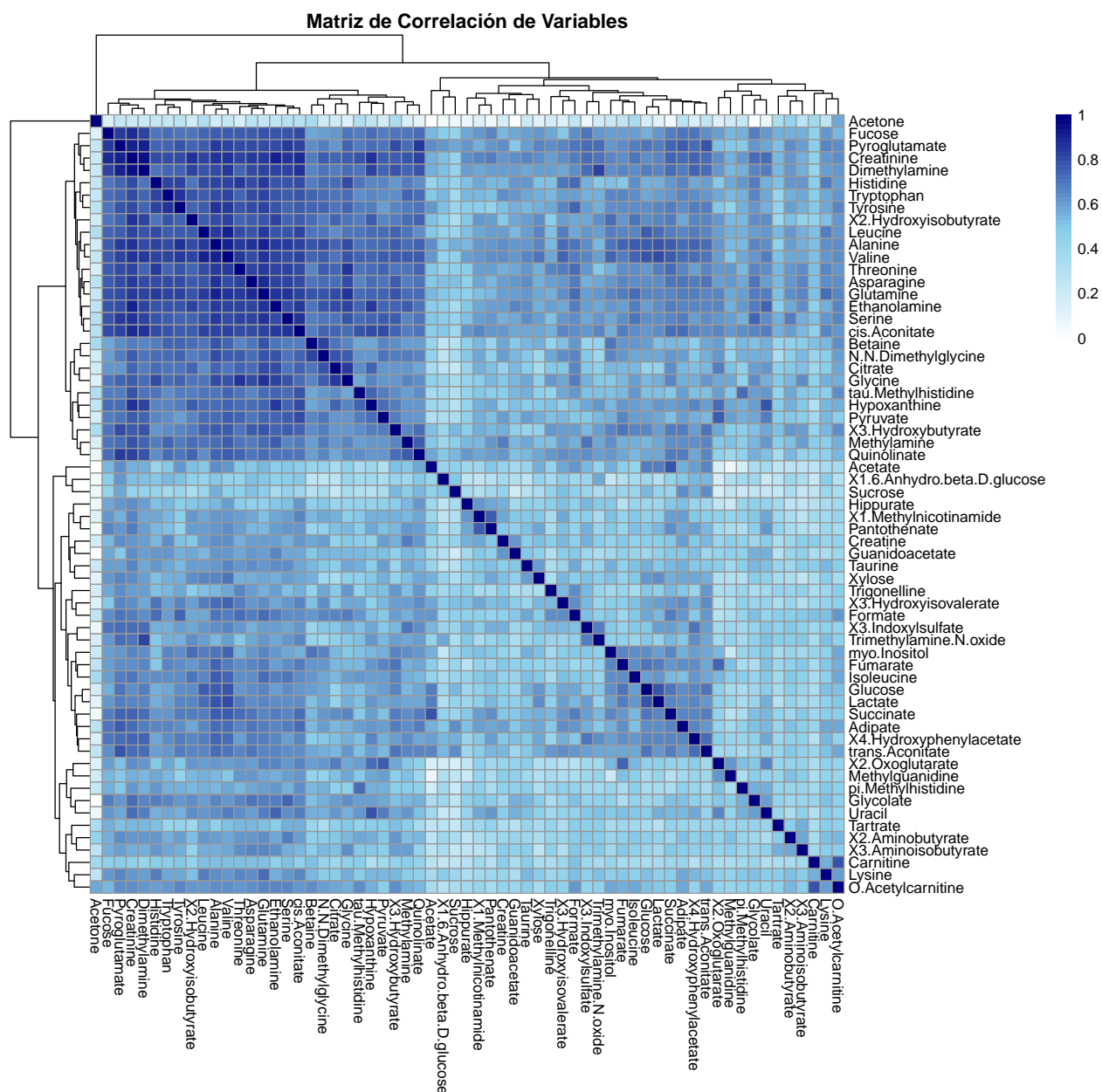

```
## mean      1.298701e-05  1.298701e-05  2.597403e-05  6.493506e-05  -2.597403e-05
## sd        6.003479e-01  6.189597e-01  7.079542e-01  6.898135e-01  6.169388e-01
## median    1.760000e-01 -6.500000e-02  1.000000e-01  2.900000e-02  -1.700000e-02
## min       -1.200000e+00 -1.142000e+00 -1.270000e+00 -1.129000e+00 -1.109000e+00
## max       1.302000e+00  2.559000e+00  1.733000e+00  1.527000e+00  1.431000e+00
##          pi.Methylhistidine tau.Methylhistidine
## mean      -8.718323e-18      -1.298701e-05
## sd         7.442465e-01       6.411401e-01
## median     -2.800000e-02       8.700000e-02
## min        -1.534000e+00      -1.299000e+00
## max         1.609000e+00       1.116000e+00
```

```
# Calcular la matriz de correlación
cor_matrix <- cor(counts, method = "pearson")

# Visualizar la matriz de correlación usando pheatmap
library(pheatmap)
```

```
## Warning: package 'pheatmap' was built under R version 4.3.3
```

```
pheatmap(cor_matrix,
          color = colorRampPalette(c("white", "skyblue", "navy"))(50),
          main = "Matriz de Correlación de Variables")
```



Claramente, algunos metabolitos correlacionan fuertemente entre sí. Por ejemplo, los metabolitos Fucosa, Piroglutamina, Creatina y Dimetilamina tienen una alta correlación positiva. También observamos que no existen correlaciones negativas entre los metabolitos.

Sin embargo, con este gráfico no podemos saber si los niveles de metabolitos están relacionados o no con tener o no caquexia. Por eso, hemos decidido explorar más los datos con un mapa de color de los contejes de cada uno de los metabolitos medidos en cada uno de los pacientes, indicando a qué grupo pertenece cada paciente.

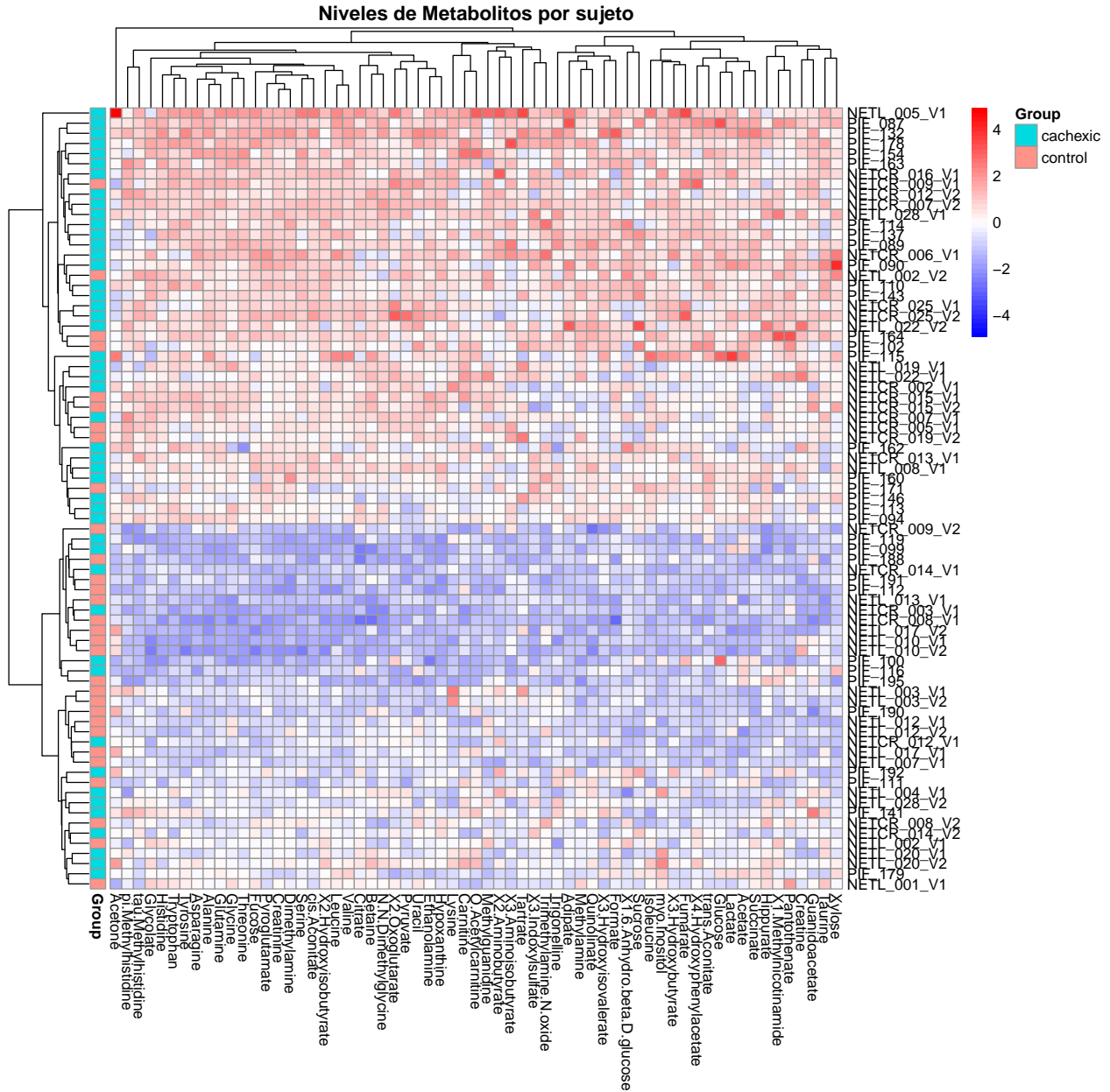
```
# Crear la anotación de los grupos desde rowData
annotation <- data.frame(Group = colData(senorm)$Group)
rownames(annotation) <- rownames(counts)

# Crear el heatmap con anotación
pheatmap(counts,
          annotation_row = annotation, # Anotación de grupo
```

```

scale = "column",
clustering_distance_cols = "euclidean",
clustering_method = "complete",
color = colorRampPalette(c("blue", "white", "red"))(50),
main = "Niveles de Metabolitos por sujeto"

```



En el heatmap podemos observar que, cuando clusterizamos por metabolitos, existen muestras con diferentes tendencias: los sujetos en las filas superiores tienden a valores más elevados de la mayor parte de los metabolitos, los de las filas intermedias tienden a valores más bajos en todos los metabolitos y los sujetos en las últimas filas tienen patrones mixtos.

Para continuar con el análisis, hemos realizado un Análisis de Componentes Principales mediante la descomposición en valores singulares de la matriz de datos con el objetivo de disminuir la dimensionalidad de los datos y revelar los patrones dominantes y las principales tendencias en los datos.

```

# Realizar el PCA
pca <- prcomp((counts), center = TRUE, scale. = FALSE)

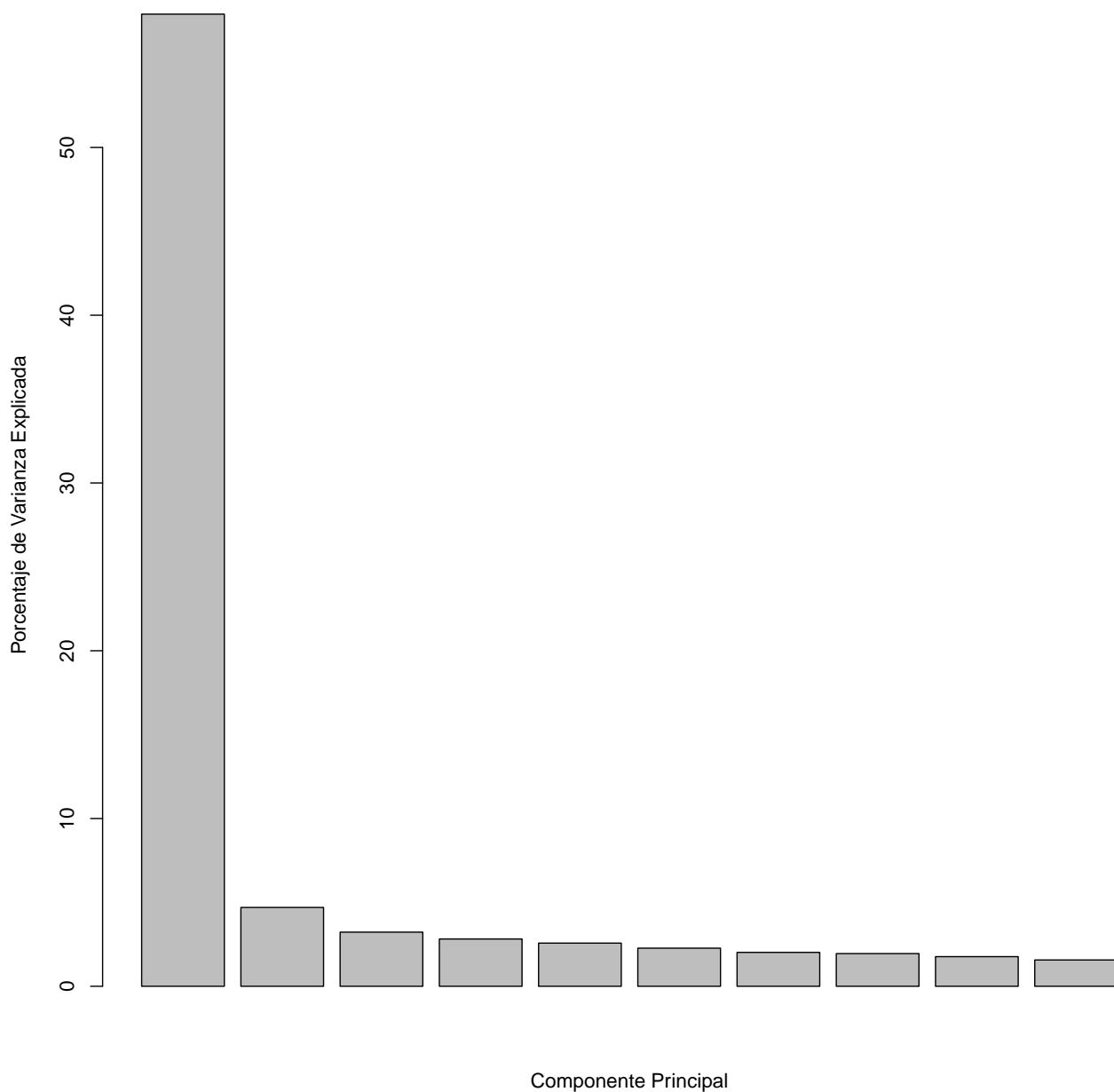
# Calcular la varianza explicada
var_exp <- pca$sdev^2 / sum(pca$sdev^2) * 100
var_exp

## [1] 57.944554790 4.698637386 3.226672232 2.819159580 2.573114507
## [6] 2.271989393 2.021206827 1.948507255 1.767519560 1.567457206
## [11] 1.383714168 1.284663476 1.235727803 1.178299575 1.028947544
## [16] 0.958189184 0.898967612 0.822138549 0.779024327 0.735575400
## [21] 0.707669671 0.653783193 0.578963360 0.557391080 0.510026551
## [26] 0.459472243 0.436810145 0.404425172 0.378972856 0.352405941
## [31] 0.347252265 0.311578890 0.299064492 0.276061353 0.243949376
## [36] 0.234544470 0.207001116 0.183435904 0.174207575 0.165058494
## [41] 0.148559028 0.140514645 0.120736746 0.110575581 0.105224609
## [46] 0.096627455 0.085627924 0.082024928 0.072540608 0.062840762
## [51] 0.059282976 0.052959644 0.052008206 0.037818306 0.032719669
## [56] 0.028050344 0.023027287 0.018243710 0.015390042 0.011226037
## [61] 0.007677239 0.005812826 0.004370909

# Graficar la varianza explicada por cada componente principal
barplot(var_exp[1:10], main = "Varianza Explicada por Componentes Principales",
        xlab = "Componente Principal", ylab = "Porcentaje de Varianza Explicada")

```

Varianza Explicada por Componentes Principales



```
loadings <- pca$rotation
(loadings[, 1:2])
```

##		PC1	PC2
##	X1.6.Anhydro.beta.D.glucose	0.09047086	0.166404329
##	X1.Methylnicotinamide	0.11493516	0.024120310
##	X2.Aminobutyrate	0.10170236	-0.084787999
##	X2.Hydroxyisobutyrate	0.12280797	-0.022674506
##	X2.Oxoglutarate	0.12592095	-0.301964048
##	X3.Aminoisobutyrate	0.11914017	-0.148393947
##	X3.Hydroxybutyrate	0.13436824	0.023336692
##	X3.Hydroxyisovalerate	0.11987088	0.166227393
##	X3.Indoxylsulfate	0.11102386	0.121145671

## X4.Hydroxyphenylacetate	0.11639944	0.138857555
## Acetate	0.10585043	0.364686751
## Acetone	0.03542815	-0.108464858
## Adipate	0.12976753	0.125154522
## Alanine	0.15797699	0.039647945
## Asparagine	0.13805757	-0.036621597
## Betaine	0.13465865	-0.102726096
## Carnitine	0.09576987	-0.070364207
## Citrate	0.13839209	-0.184777034
## Creatine	0.13392577	0.063714392
## Creatinine	0.14620076	-0.016513188
## Dimethylamine	0.14095201	0.011324089
## Ethanolamine	0.15074499	-0.093691763
## Formate	0.11991405	0.066432349
## Fucose	0.13643953	0.004934596
## Fumarate	0.11039883	-0.001727248
## Glucose	0.13476479	0.202651696
## Glutamine	0.15844563	-0.074341924
## Glycine	0.14594782	-0.090194070
## Glycolate	0.12491201	-0.135166319
## Guanidoacetate	0.10277306	-0.066258608
## Hippurate	0.10970348	0.122875039
## Histidine	0.15252837	-0.084553450
## Hypoxanthine	0.13672755	-0.114176473
## Isoleucine	0.09690615	-0.009240621
## Lactate	0.12729342	0.168039194
## Leucine	0.13014564	-0.011724569
## Lysine	0.11246755	-0.043834830
## Methylamine	0.12322799	0.030326423
## Methylguanidine	0.09493409	-0.207302970
## N.N.Dimethylglycine	0.13487524	-0.083953650
## O.Acetylcarnitine	0.11908786	-0.084037822
## Pantothenate	0.11599429	0.115071071
## Pyroglutamate	0.14609334	0.101514492
## Pyruvate	0.13337244	-0.122873327
## Quinolate	0.12259839	0.104276686
## Serine	0.13937939	-0.060770745
## Succinate	0.14686429	0.198603333
## Sucrose	0.09269789	0.300658043
## Tartrate	0.10112656	-0.108251577
## Taurine	0.12834742	-0.036095134
## Threonine	0.14678224	-0.050645191
## Trigonelline	0.12012737	0.085407334
## Trimethylamine.N.oxide	0.11257029	0.051507623
## Tryptophan	0.13322538	-0.039533524
## Tyrosine	0.14182793	0.012421308
## Uracil	0.11271277	-0.105969812
## Valine	0.14211927	0.031937709
## Xylose	0.10455772	0.155389682
## cis.Aconitate	0.16247946	-0.070441519
## myo.Inositol	0.12639576	0.073584247
## trans.Aconitate	0.12036402	0.183630748
## pi.Methylhistidine	0.11950405	-0.217884899
## tau.Methylhistidine	0.12687783	-0.136740675

La primera Componente Principal explica casi un 60% de los datos, mientras que la segunda un 4.7%. Es decir, las dos primeras componentes principales explican una alta proporción de la varianza explicada de los datos (casi un 65%).

En cuanto a la carga de los metabolitos a cada una de las dos primeras Componentes Principales, todos tienen un impacto positivo en PC1, mientras que para PC2, algunos metabolitos influyen de manera negativa.

```
# Gráfico por grupos
```

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.3.3
```

```
pca_data <- as.data.frame(pca$x)
```

```
pca_data$Group <- annotation$Group
```

```
ggplot(pca_data, aes(x = PC1, y = PC2, color = Group)) +  
  geom_point(size = 3) +  
  labs(title = "PCA de Muestras", x = "PC1", y = "PC2") +  
  theme_minimal() +  
  theme(legend.position = "bottom")
```



En la proyección de las muestras en la bidimensión PC1vsPC2, no se ven clústers muy pronunciados, pero sí existe una tendencia a la agrupación entre muestras del mismo grupo. Además, no existen puntos exageradamente alejados del resto.

Discusión, limitaciones y conclusiones

El síndrome caquético está relacionado con un deterioro funcional del paciente que caracteriza a muchas patologías entre las que se incluyen el cáncer. El desarrollo de la caquexia engloba alteraciones metabólicas, hormonales y de otros mediadores celulares que desencadenan un síndrome de consunción que no puede resolverse mediante la ingesta forzada de nutrientes.

Hemos explorado los datos disponibles de niveles de 63 metabolitos diferentes tomados a 77 pacientes con cáncer, de los que disponemos la información de si padecen o no síndrome caquético. El objetivo del análisis es entender si existen patrones metabólicos que diferencian los pacientes con síndrome caquético de los que no lo padecen.

Observando el correlograma, vemos que los niveles de algunos de los metabolitos correlacionan de manera contundente entre sí. Es decir, que cuando los niveles de unos están elevados, los de otros también. Este es el caso de metabolitos como Fucosa, Piroglutamina, Creatina y Dimetilamina.

Al explorar los patrones de niveles de metabolitos entre los pacientes teniendo en cuenta a qué grupo pertenecen, hemos observado que se podrían agrupar a los sujetos en tres grupos diferentes: los de niveles elevados, los de niveles bajos y los de niveles mixtos. Además, hemos visto que las muestras de pacientes caquéticos se concentran más en la parte superior del gráfico, donde los valores están más elevados. Por lo tanto, de manera preliminar, se intuye que valores más elevados de los metabolitos medidos podrían indicar síndrome caquético. Los metabolitos con niveles elevados entre los pacientes con caquexia podrían ser potenciales biomarcadores de dicho síndrome.

Finalmente, al reducir la dimensionalidad con el PCA, hemos visto que dos componentes principales son suficientes para capturar la mayor parte de la variabilidad de los datos, es decir que el conjunto de datos tiene una estructura que puede ser bien representada con menos dimensiones. Las cargas de los dos primeras componentes principales nos muestran que los metabolitos más relacionados (positivamente) con PC1 son el cis-aconitato, la glutamina y la alanina y que ningún metabolito aporta de manera negativa a esta componente. En cuanto a PC2, hay metabolitos relacionados positiva y negativamente y los de más peso son, respectivamente, el succinato y la pi-metil-histidina.

En la representación bidimensional de PC1 vs. PC2, se observan dos tendencias diferenciadas en función del grupo de las muestras, lo que podría indicar que sí existen patrones de niveles metabólicos que pueden ayudar a diagnosticar, o predecir, el síndrome caquético.

Este estudio presenta varias limitaciones. Para poder caracterizar mejor los patrones metabólicos deberían tenerse en cuenta más datos sobre los pacientes incluidos en el estudio. Por ejemplo, la variedad de patologías: Si bien el estudio se centra en pacientes con cáncer, la heterogeneidad en los tipos de cáncer y en su tratamiento podría influir en los niveles de metabolitos, dificultando la identificación de patrones específicos para el síndrome caquético. Tampoco se han tenido en cuenta las condiciones ambientales y dietéticas de los paciente, que afectan a los niveles de metabolitos y la interpretación de los resultados. Finalmente, haría falta un análisis estadístico, multivariado, más profundo de los datos, pues el análisis exploratorio de las correlaciones entre metabolitos no siempre capturan interacciones complejas que podrían ser relevantes para el síndrome caquético.

En conclusión, este estudio sugiere que existen patrones metabólicos diferenciadores entre pacientes con y sin síndrome caquético en el contexto del cáncer, evidenciados por la correlación entre metabolitos como Fucosa, Piroglutamina, Creatina y Dimetilamina. La clasificación de los pacientes en grupos de niveles elevados, bajos y mixtos, junto con el análisis de componentes principales que identifica metabolitos clave como cis-aconitato y glutamina, indica que un aumento en los metabolitos podría ser un indicador del síndrome caquético. Sin embargo, se reconocen limitaciones significativas, como la falta de información respecto a la heterogeneidad en los tipos de cáncer y los factores ambientales y dietéticos, que requieren un análisis más profundo en futuras investigaciones para validar estos hallazgos y su potencial como biomarcadores.

Referencias

FERRIOLS LISART F, TORDERA BAVIERA M. El síndrome caquético en el paciente oncológico: fisiopatología, manifestaciones clínicas y tratamiento farmacológico. Farm Hosp. 2003;27(5):308-10.

<https://bioconductor.org/packages/release/bioc/vignettes/SummarizedExperiment/inst/doc/SummarizedExperiment.html>

<https://www.bioconductor.org/packages/release/bioc/vignettes/POMA/inst/doc/POMA-workflow.html#the-poma-workflow>

Enlace al repositorio de Github

https://github.com/EFR28/Fuentes_Raventos_Elisa_PEC1.git