Informe

Abel Pérez Barroso

2024-11-06

Contents

Abstract

La caquexia es un síndrome caracterizado por la pérdidad de peso, tanto masa muscular como de grasa, acompañada de debilidad extrema y fatiga.

En este informe se analizó el dataset *2024-Cachexia* para identificar diferencias significativas en las concentraciones de metabolitos entre el grupo control y el grupo con caquexia, con el objetivo de destacar potenciales biomarcadores asociados a la caquexia.

Primero, se empleó la estructura `SummarizedExperiment` para crear un contenedor que incluye tanto los datos como los metadatos, asegurando la integridad de la información para los siguientes análisis. Luego, se realizó una exploración inicial de los datos y un control de calidad preliminar, seguido de la normalización de las concentraciones de los metabolitos, tras lo cual se aplicó un segundo control de calidad sobre los datos normalizados.

Como paso final, se llevó a cabo un análisis de componentes principales (PCA) tanto para el dataset completo como para los datos segmentados en los grupos de caquexia y control. A través de gráficos de barras y otros métodos visuales, se lograron identificar diferencias en la distribución de las concentraciones de los metabolitos, destacando aquellos con mayor variabilidad entre ambos grupos.

Objetivo del estudio

El objetivo principal del estudio en el análisis del dataset de *2024-Cachexia* es identificar los metabolitos que más influyen en la caquexia. Para ello, se emplearán gráficos de barras, heatmaps y boxplots para observar las diferencias y la variabilidad de los metabolitos en el contexto del análisis de componentes principales (PCA).

Materiales y métodos

El análisis se basa en el dataset *2024-Cachexia*, obtenido de [https://github.com/nutrimetabolomics/metaboData/tree/main/Datasets/2024-Cachexia]. En este conjunto de datos se observan a 77 pacientes separados en dos grupos, uno con cachexia y un control, y las concentraciones de varios metabolitos en su cuerpo.

Herramientas utilizadas

Para el análisis de los datos se ha utilizado Rsdtudio y el lenguaje de programación R, junto con diferentes paquetes como:

- `readr`: paquete que nos permite la importación de datos en formato CSV.

- `Biobase`: para el manejo de objetos de clase `ExpressionSet`.

- `SummarizedExperiment`: paquete que nos permite integrar datos en un solo objeto.

- `ggplot2`: para la visualización de datos en formato de gráficos.

- `dplyr` y `tidyr`: para la manipulación y transformación de datos.

- `grDevices`: para la personalización de colores en heatmap.

Procedimiento del análisis

Como primer paso, se importaron los datos a R utilizando las funciones `setwd()` y `read_csv()` del paquete `readr`. Posteriormente, se creó un objeto `SummarizedExperiment` para organizar y estructurar los datos metabólicos junto con sus metadatos.

A continuación, se verificó la presencia de datos faltantes, que se eliminaron en caso de ser necesario, y se procedió con la normalización de los datos para asegurar su consistencia en los análisis posteriores.

Finalmente, se realizaron los análisis exploratorios y de componentes principales (PCA), y se visualizó la información mediante la generación de gráficos, incluyendo gráficos de barras, boxplots y heatmaps.

Métodos utilizados

1. **Importación y organización de los datos:**

- Los datos fueron importados y organizados utilizando las funciones `setwd()` para establecer el directorio de trabajo y `read_csv()` del paquete `readr` para cargar los datos en R. Con esta importación inicial se nos permite el inicio en el manejo de los datos.

2. **Creación del objeto `SummarizedExperiment`:**

- Los datos importados se encapsularon en un objeto `SummarizedExperiment`, que facilita el manejo de conjuntos de datos complejos agrupando las mediciones metabólicas en un único contenedor.

3. **Limpieza y prepocesamiento:**

- Se realizó una primera evaluación para observar si habia datos faltantes y su posterior eliminación si fuera necesario. La normalización de los datos se llevó a cabo para asegurar que las mediciones de los metabólitos fueran comparables entre sí y listas para su posterior análisis. Se hizo un control de calidad antes y después de su normalización.

4. **Análisis y visualización de datos:**

- Se emplearon técnicas de visualización y análisis a partir de PCA para observar los resultados del análisis.

5. **Resultados y conclusión:**

- Finalmente se observaron los resultados obtenidos y se obtuvo una conclusión respecto a estos.

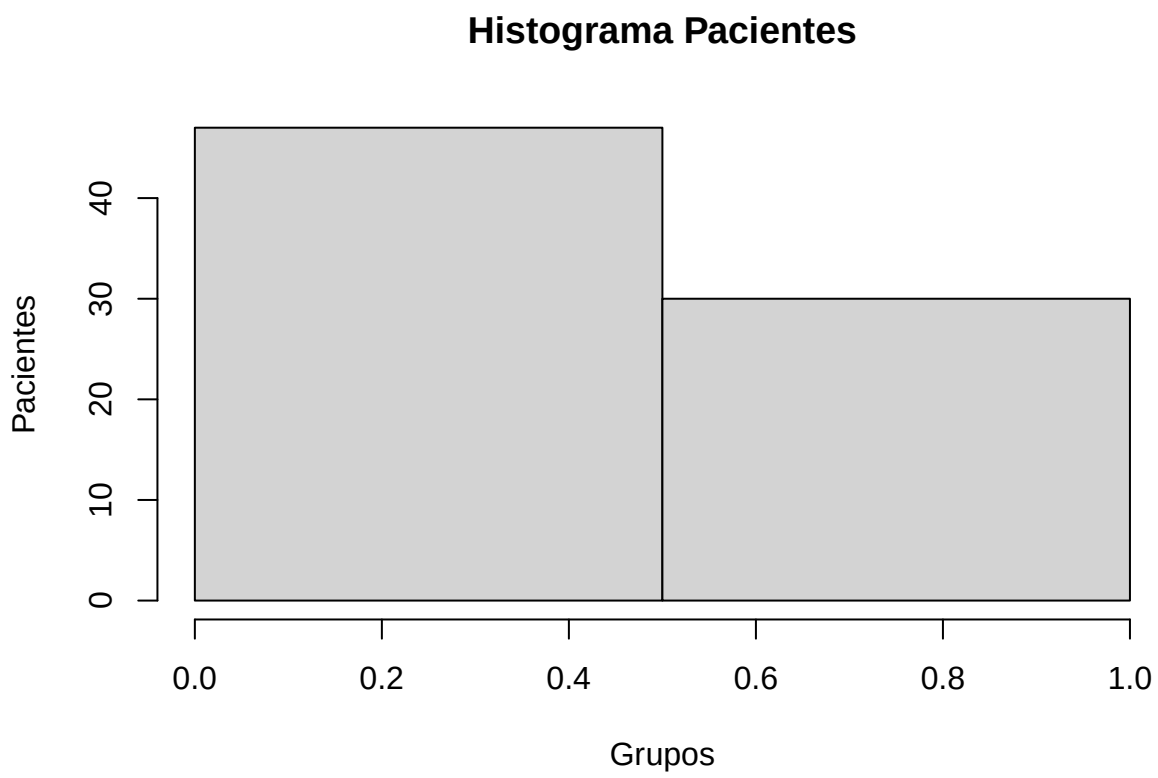
Resultados

Introducción a los resultados

A continuación, se procedera a observar los resultados obtenidos del análisis de datos y los procedimientos que se han utilizado como son el control de calidad (antes y después de la normalización), análisis de PCA y gráficos.

Observación preliminar

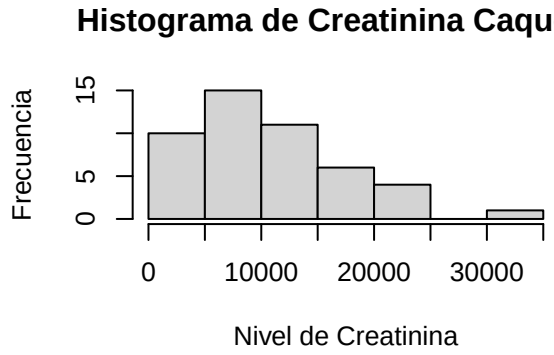
```
hist(Muscle_loss,
     main = "Histograma Pacientes",
     xlab = "Grupos",
     ylab = "Pacientes",
     breaks = 2)
```

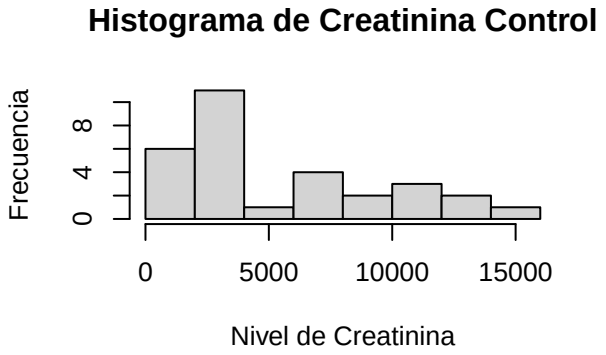


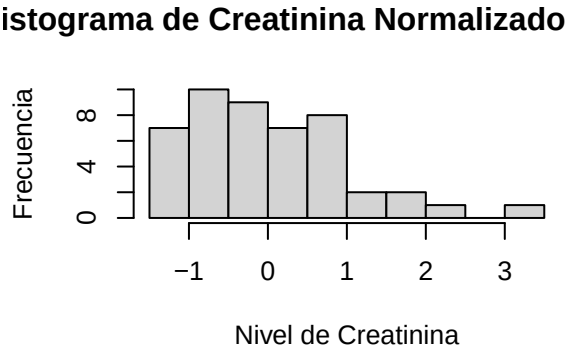
En el gráfico previamente mostrado se observa en un histograma con la cantidad de pacientes con cachexia y los del grupo control.

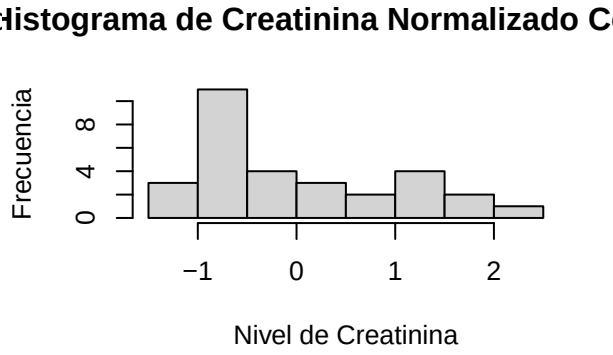
```
par(mfrow=c(2,2))
hist(datos_cachexia$Creatinine,
     main = "Histograma de Creatinina Caquexia",
     xlab = "Nivel de Creatinina",
```

```
     ylab = "Frecuencia",
     breaks = 10)
hist(datos_control$Creatinine,
     main = "Histograma de Creatinina Control",
     xlab = "Nivel de Creatinina",
     ylab = "Frecuencia",
     breaks = 10)
hist(datos_cachexia_norm$Creatinine,
     main = "Histograma de Creatinina Normalizado Caquexia",
     xlab = "Nivel de Creatinina",
     ylab = "Frecuencia",
     breaks = 10)
hist(datos_control_norm$Creatinine,
     main = "Histograma de Creatinina Normalizado Control",
     xlab = "Nivel de Creatinina",
     ylab = "Frecuencia",
     breaks = 10)
```









En los gráficos previamente mostrados se observan cuatro histogramas del metabolito creatinina, donde podemos observar los datos previos y posteriores a la normalizacion de los datos tanto del grupo control como del grupo con cachexia.

Al observar ambos gráficos de los grupos podemos ver que la normalización de los datos es correcta, puesto que estos siguen teniendo un gran parecido entre ellos.

Gracias a esta confirmación podemos comenzar con el análisis de los datos.

Observaciones y resultados

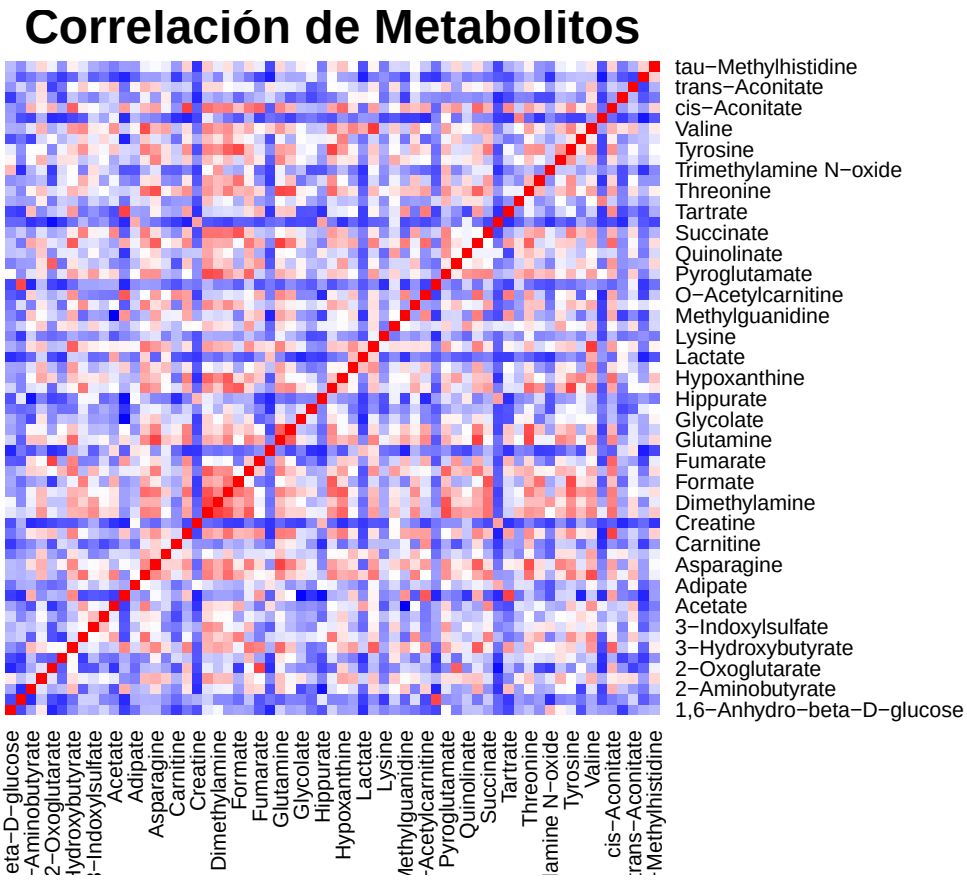
Correlación

En este análisis, hemos generado tres mapas de calor (heatmaps) que representan la matriz de correlación de diferentes metabolitos en tres conjuntos de datos:

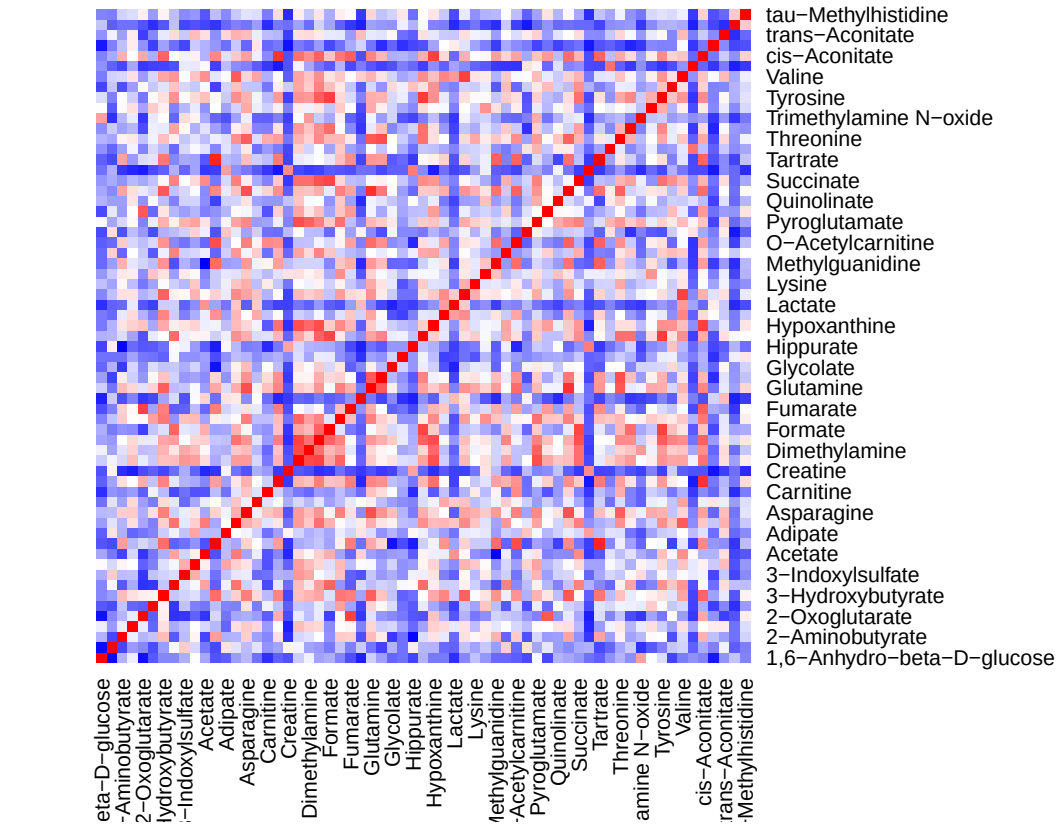
1. Todos los pacientes.
2. Pacientes con cachexia.
3. Pacientes control.

Cada heatmap permite visualizar cómo se relacionan entre sí los metabolitos dentro de cada grupo, lo que nos ayuda a entender patrones metabólicos.

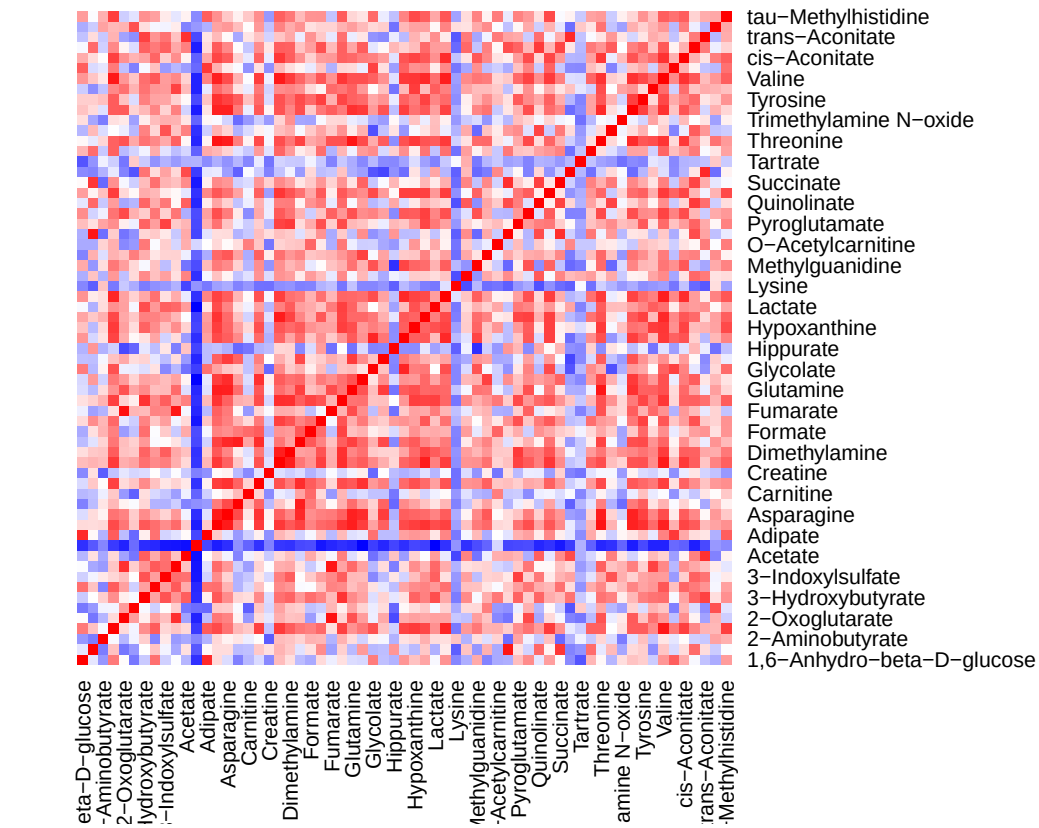
```
heatmap(cor_todo, main = "Correlación de Metabolitos",
        Colv = NA, Rowv = NA, scale = "none", margins = c(5, 5),col = paleta_cor)
```




```
heatmap(cor_cachexia, main = "Correlación de Metabolitos Cachexia",
        Colv = NA, Rowv = NA, scale = "none", margins = c(5, 5), col = paleta_cor)
```




```
heatmap(cor_control, main = "Correlación de Metabolitos Control",
        Colv = NA, Rowv = NA, scale = "none", margins = c(5, 5), col = paleta_cor)
```



En los tres heatmaps observamos diferentes patrones de correlación entre metabolitos. En el primer y segundo gráfico, que corresponden a todos los pacientes y a los pacientes con caquexia, respectivamente, predominan los tonos azulados, lo que sugiere correlaciones bajas o negativas entre los metabolitos. En cambio, el tercer heatmap, que representa al grupo de control, muestra tonos más rojizos, lo que indica correlaciones positivas más fuertes entre varios metabolitos.

Este patrón sugiere que, en el grupo de control, los metabolitos tienden a aumentar o disminuir en sincronía, reflejando un sistema metabólico más estable. En contraste, en el grupo de caquexia, las correlaciones son más bajas y variadas, lo que podría indicar una disrupción en la regulación metabólica. Esto implica que, mientras en los controles ciertos metabolitos aumentan junto con otros, en los pacientes con caquexia estos patrones son menos consistentes e incluso opuestos en algunos casos, reflejando posibles desbalances asociados con la enfermedad.

Análisis de componentes principales (PCA)

```
summary(pca_todo)
```

```
## Importance of components:
##                           PC1    PC2     PC3     PC4     PC5    PC6     PC7
## 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    PC42
## Standard deviation     0.2867 0.28435 0.26060 0.25353 0.24800 0.21896 0.19537
## Proportion of Variance 0.0013 0.00128 0.00108 0.00102 0.00098 0.00076 0.00061
## Cumulative Proportion  0.9903 0.99158 0.99266 0.99368 0.99465 0.99541 0.99602
##                           PC43   PC44    PC45   PC46    PC47   PC48    PC49
## Standard deviation     0.18914 0.1767 0.16864 0.1580 0.15287 0.1380 0.13101
## Proportion of Variance 0.00057 0.0005 0.00045 0.0004 0.00037 0.0003 0.00027
## 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
```

```
pca_todo$rotation[order(pca_todo$rotation[,1], decreasing = TRUE), 1]
```

```
##                 Creatinine                  Glutamine
##                  0.17549735                 0.17089565
##               Ethanolamine                  Asparagine
##                  0.17041813                 0.16916015
##                   Threonine                     Valine
##                  0.16845973                 0.16791310
##                      Alanine               cis-Aconitate
##                  0.16734332                 0.16611699
##                      Serine                     Fucose
##                  0.16496675                 0.16297413
##                    Tyrosine                    Leucine
##                  0.16204954                 0.16000426
##              Pyroglutamate          3-Hydroxybutyrate
##                  0.15962554                 0.15920496
##              Dimethylamine                   Histidine
##                  0.15905220                 0.15784801
##                     Citrate                Hypoxanthine
##                  0.15729614                 0.15450422
##                     Glycine                   Succinate
##                  0.14621127                 0.14547514
##                  Tryptophan       2-Hydroxyisobutyrate
##                  0.14292599                 0.14196456
```

```
##        N,N-Dimethylglycine                Methylamine
##                 0.13986242                 0.13408244
##                    Formate                 Isoleucine
##                 0.13354188                 0.13253073
##       3-Hydroxyisovalerate                   Pyruvate
##                 0.13137204                 0.12796424
##                   Fumarate                Trigonelline
##                 0.12617442                 0.12528021
##          O-Acetylcarnitine                    Betaine
##                 0.12430489                 0.12340673
##             trans-Aconitate                Quinolinate
##                 0.12237835                 0.12137345
##        tau-Methylhistidine           3-Indoxylsulfate
##                 0.11957240                 0.11955834
##                    Uracil             Methylguanidine
##                 0.11849005                 0.11636219
##     4-Hydroxyphenylacetate            2-Aminobutyrate
##                 0.11156540                 0.11064656
##                  Glycolate                    Acetate
##                 0.10987898                 0.10974322
##                    Taurine                  Tartrate
##                 0.10753145                 0.10752831
##                    Adipate                    Acetone
##                 0.10013298                 0.09241969
##         3-Aminoisobutyrate             2-Oxoglutarate
##                 0.08984882                 0.08826605
##              Guanidoacetate     Trimethylamine N-oxide
##                 0.08813351                 0.08774914
##                  Carnitine                     Lysine
##                 0.08598625                 0.07703294
## 1,6-Anhydro-beta-D-glucose               myo-Inositol
##                 0.07678198                 0.07609235
##                   Hippurate         pi-Methylhistidine
##                 0.07137427                 0.06520172
##       1-Methylnicotinamide                    Lactate
##                 0.06448034                 0.06170896
##                    Glucose                Pantothenate
##                 0.06024052                 0.05746210
##                    Xylose                    Sucrose
##                 0.04900948                 0.04240689
##                   Creatine
##                 0.04147524
```

```
pca_todo$rotation[order(pca_todo$rotation[,2], decreasing = TRUE), 1]
```

```
##                    Acetate       3-Hydroxyisovalerate
##                 0.10974322                 0.13137204
##                  Hippurate                  Succinate
##                 0.07137427                 0.14547514
##       1-Methylnicotinamide                Pantothenate
##                 0.06448034                 0.05746210
##                    Sucrose           3-Indoxylsulfate
##                 0.04240689                 0.11955834
##             trans-Aconitate                    Uracil
```

```
##                 0.12237835                 0.11849005
##               Quinolinate                     Xylose
##                 0.12137345                 0.04900948
##       N,N-Dimethylglycine                    Alanine
##                 0.13986242                 0.16734332
##    4-Hydroxyphenylacetate                    Glucose
##                 0.11156540                 0.06024052
##                 Creatinine                    Formate
##                 0.17549735                 0.13354188
##              Pyroglutamate                   Creatine
##                 0.15962554                 0.04147524
##                    Adipate                  Glycolate
##                 0.10013298                 0.10987898
##                     Valine               Trigonelline
##                 0.16791310                 0.12528021
##                    Taurine              Dimethylamine
##                 0.10753145                 0.15905220
## 1,6-Anhydro-beta-D-glucose                    Lactate
##                 0.07678198                 0.06170896
##             Guanidoacetate         pi-Methylhistidine
##                 0.08813351                 0.06520172
##                  Histidine     Trimethylamine N-oxide
##                 0.15784801                 0.08774914
##                Methylamine              myo-Inositol
##                 0.13408244                 0.07609235
##               Ethanolamine                    Tyrosine
##                 0.17041813                 0.16204954
##       2-Hydroxyisobutyrate                    Pyruvate
##                 0.14196456                 0.12796424
##               Hypoxanthine                    Betaine
##                 0.15450422                 0.12340673
##                    Leucine        tau-Methylhistidine
##                 0.16000426                 0.11957240
##                  Threonine                     Fucose
##                 0.16845973                 0.16297413
##                     Lysine                 Tryptophan
##                 0.07703294                 0.14292599
##                 Asparagine         3-Aminoisobutyrate
##                 0.16916015                 0.08984882
##                 Isoleucine                   Carnitine
##                 0.13253073                 0.08598625
##                    Glycine              cis-Aconitate
##                 0.14621127                 0.16611699
##                  Glutamine             2-Oxoglutarate
##                 0.17089565                 0.08826605
##                    Citrate          3-Hydroxybutyrate
##                 0.15729614                 0.15920496
##                     Serine            2-Aminobutyrate
##                 0.16496675                 0.11064656
##                   Fumarate          O-Acetylcarnitine
##                 0.12617442                 0.12430489
##            Methylguanidine                    Tartrate
##                 0.11636219                 0.10752831
##                    Acetone
```

```
##                 0.09241969
```

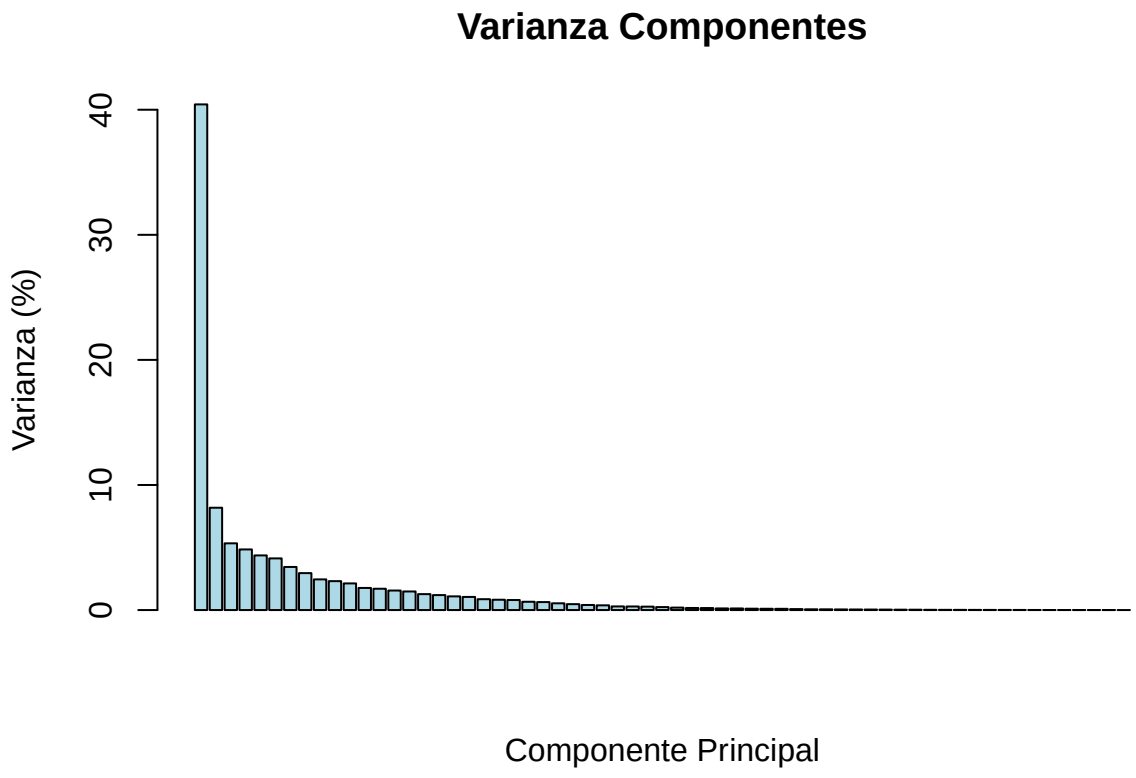
Al analizar las componentes principales (PCA) para reducir la dimensionalidad de los datos metabólicos, podemos observar los metabolitos que más contribuyen a la variabilidad de los datos en pacientes con caquexia.

En la primera componente principal (PC1), se destacan metabolitos con mayores cargas, incluyendo la alanina, creatinina, glutamina, valina y serina. Esto sugiere que estos metabolitos están fuertemente asociados con variaciones en el metabolismo de personas con caquexia. Por ejemplo, el aumento en creatinina y alanina es común en situaciones de atrofia o degradación muscular, características comunes en esta condición.

En la segunda componente principal (PC2) se observan otros metabolitos como el succinato, 3-hidroxisovalerato, 3-indoxilsulfato y cis-aconitato. Estos metabolitos están relacionados con procesos de energía celular. Las elevadas cargas de estos metabolitos indican una alteración en el metabolismo energético, un aspecto distintivo en el síndrome de la caquexia.

Observando ambas componentes principales, se evidencia que estos metabolitos están altamente relacionados con el metabolismo muscular y energético. Dado que la caquexia está caracterizada por estos desajustes metabólicos, los metabolitos antes mencionados podrían utilizarse como biomarcadores para identificar y monitorear a pacientes con esta condición.

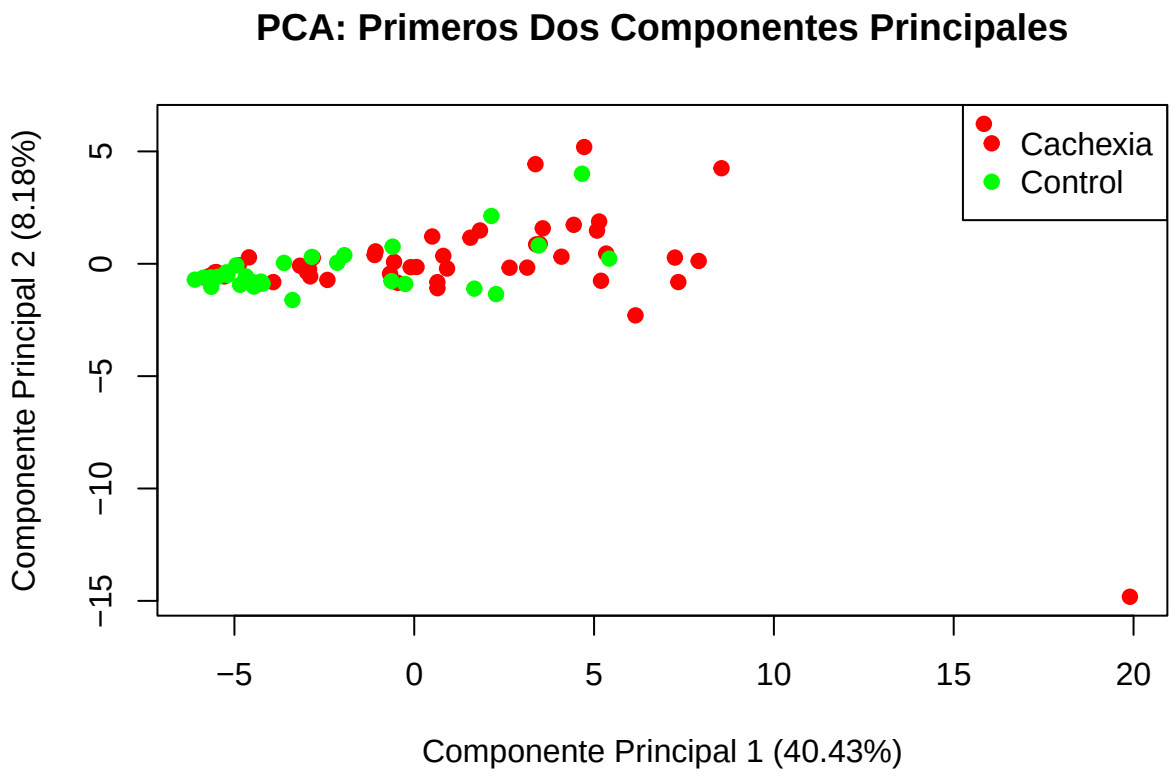
```
barplot(pca_todo$sdev^2 / sum(pca_todo$sdev^2) * 100,
        main = "Varianza Componentes",
        xlab = "Componente Principal",
        ylab = "Varianza (%)",
        col = "lightblue")
```



```
##  [1]  40.42679  48.60692  53.94070  58.78669  63.15568  67.28573  70.72992
##  [8]  73.68323  76.13469  78.44525  80.57526  82.34206  84.04637  85.60649
## [15]  87.09301  88.36593  89.56150  90.65608  91.70593  92.57322  93.39864
## [22]  94.19999  94.86253  95.50034  96.03993  96.51009  96.91552  97.29264
## [29]  97.58171  97.86743  98.14058  98.37774  98.57320  98.74008  98.89904
## [36]  99.02953  99.15787  99.26566  99.36769  99.46532  99.54142  99.60200
## [43]  99.65879  99.70833  99.75347  99.79310  99.83019  99.86043  99.88768
## [50]  99.90605  99.92314  99.93855  99.95073  99.96155  99.97184  99.97945
## [57]  99.98505  99.99016  99.99426  99.99680  99.99821  99.99944 100.00000
```

En el gráfico de barras observamos la varianza de las diferentes componentes principales junto a su varianza acumulada, indicando el porcentaje total de varianza al ir suamndo componentes.

```
plot(pca_scores[, 1:2],
     col = ifelse(human_cachexia_df$`Muscle loss` == "cachexic", "red", "green"),
     pch = 19,
     xlab = paste0("Componente Principal 1 (", round(var_todo[1], 2), "%)"),
     ylab = paste0("Componente Principal 2 (", round(var_todo[2], 2), "%)"),
     main = "PCA: Primeros Dos Componentes Principales")
legend("topright", legend = c("Cachexia", "Control"), col = c("red", "green"), pch = 19)
```



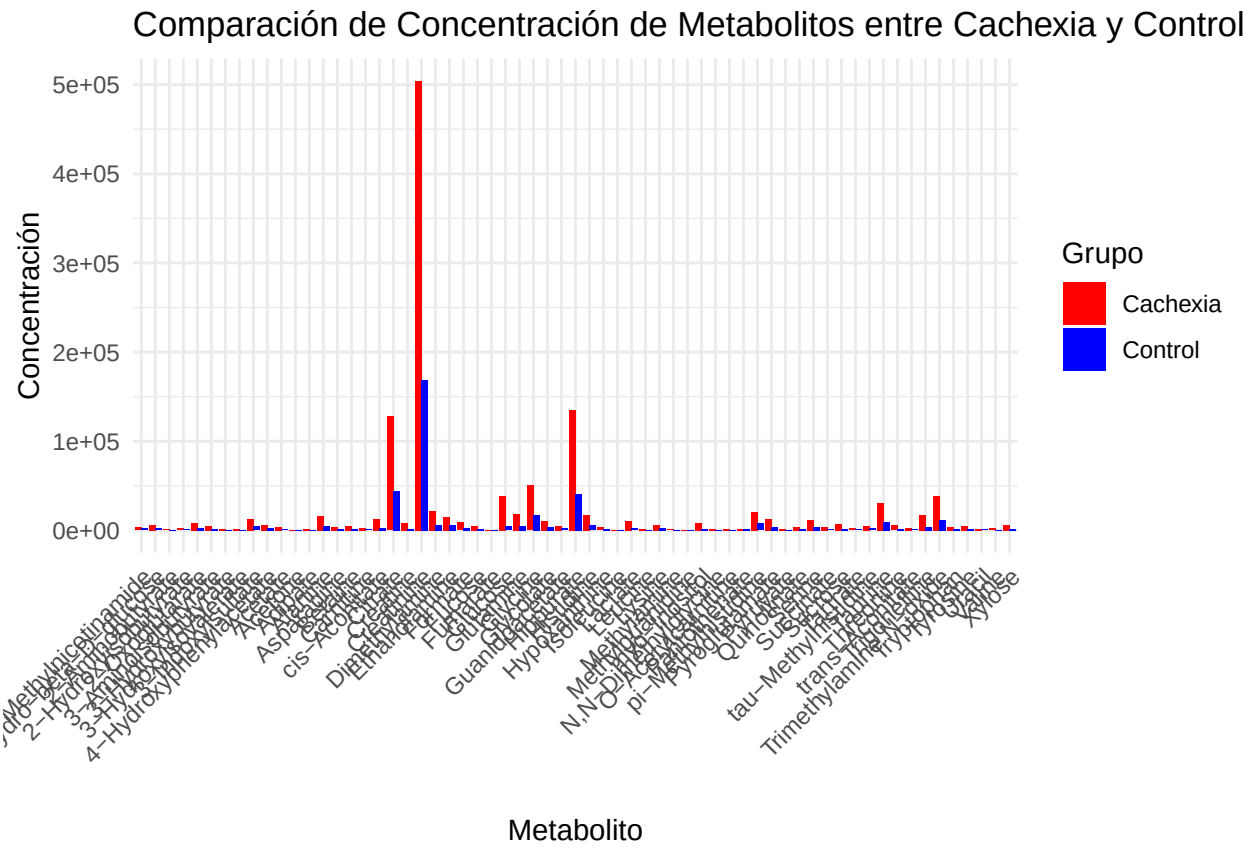
En el gráfico de componentes principales (PCA), se presenta una comparación visual entre los grupos de pacientes con caquexia y los controles.

Los puntos correspondientes a los pacientes control (marcados en verde) se encuentran fuertemente agrupados hacia la parte izquierda del gráfico. Esta agrupación indica que los individuos sin caquexia comparten perfiles metabólicos similares, lo que sugiere un estado metabólico relativamente homogéneo y saludable.

Por otro lado, los pacientes con caquexia (marcados en rojo) están distribuidos de manera más dispersa hacia la derecha. Esta dispersión podría reflejar una variabilidad en sus perfiles metabólicos, lo que es característico de la caquexia. La posición de los puntos rojos en esta área del gráfico sugiere que estos pacientes presentan niveles más altos de ciertos metabolitos que están relacionados con el metabolismo muscular y energético, los cuales se encuentran alterados en el contexto del síndrome de caquexia.

Comparativa de metabolitos caquexia y control

```
ggplot(comparacion_larga, aes(x = Metabolito, y = Cantidad, fill = Grupo)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(title = "Comparación de Concentración de Metabolitos entre Cachexia y Control",
       x = "Metabolito",
       y = "Concentración") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  scale_fill_manual(values = c("Cachexia" = "red", "Control" = "blue"))
```



En este gráfico, se presenta una comparativa de las cantidades normalizadas de los diferentes metabolitos entre el grupo de pacientes con caquexia y el grupo de control. A simple vista, se puede observar que las concentraciones de metabolitos en los pacientes con caquexia son significativamente más elevadas que en el grupo de control. Esta diferencia resalta el desajuste metabólico que el síndrome de caquexia provoca en el organismo humano, lo que puede tener implicaciones importantes para la salud y el manejo clínico de estos pacientes.

Discusión y limitaciones

En este estudio, hemos llevado a cabo un análisis de componentes principales (PCA) para explorar la variabilidad metabólica en pacientes con caquexia en comparación con un grupo control. A través de este enfoque, hemos identificado metabolitos clave que contribuyen a las diferencias observadas en los perfiles metabólicos entre los dos grupos. Sin embargo, es fundamental reconocer varias limitaciones que podrían afectar la interpretación de los resultados.

Primero, el tamaño de la muestra, aunque suficiente para algunas pruebas estadísticas, puede no ser representativo de la población general de pacientes con caquexia. Esto limita la generalización de los hallazgos. Además, el análisis se basa en datos obtenidos de un único momento, mientras que los organismos son dinámicos y varian en el tiempo.

Otra limitación se encuentra en la técnica de análisis utilizada. Aunque la PCA es efectiva para reducir la dimensionalidad y destacar patrones en los datos, no proporciona información sobre la causa o la interrelación entre los metabolitos.

Conclusión

A pesar de las limitaciones, el estudio ha permitido una exploración inicial de las diferencias metabólicas entre pacientes con caquexia y controles sanos. La identificación de metabolitos como la alanina, creatinina, y otros que se presentan en niveles elevados en el grupo de caquexia sugiere la existencia de un perfil metabólico distintivo asociado con esta condición.

Estos resultados respaldan la hipótesis de que el síndrome de caquexia altera el metabolismo musuclar y energético y abre puertas a futuras investigaciones sobre posibles biomarcadores para la identificación y monitoreo de la caquexia.

Repositorio de github

Dirección url: [https://github.com/xAbel95x/Perez-Barroso-Abel-PEC1.git]