

- **Breve introducción a SummarizedExperiment.**

Lo primero que debemos hacer es crear el contenedor, para ello vamos a emplear la herramienta SummarizedExperiment. Este tipo de clase se utiliza para almacenar matrices rectangulares. Este SummarizedExperiment se basa en que cada columna contendrá las muestras, mientras que las filas contendrán lo que nos interesa analizar. A su vez, estas son almacenadas empleando un dataframe.

Lo primero que se ha hecho ha sido instalar el paquete desde bioconductor, y leer cómo funciona, como se tienen que organizar los datos, así como que partes de la tabla son los que podemos utilizar como metadatos (es decir, la ID del paciente, los diferentes metabolitos que se han medido, si son sanos o enfermos...).

- **Creación del contenedor.**

En primer lugar, se ha descargado la carpeta que había disponible en el repositorio de Github y se ha escogido los datos procedentes del experimento 2024-Cachexia.

Se ha preparado el entorno de trabajo, con la apertura del fichero, un vistazo rápido de como están organizados los datos, cuantas columnas hay, cuantas filas, que contiene cada una de esas filas o columnas. Las dos primeras columnas al contener la ID del paciente, y el estado (si es enfermo o sano), se ha aplicado como metadatos, así como las distintas variables que se han analizado (los distintos metabolitos). Esto está compilado en la variable Coldata.

Se ha aislado posteriormente los datos de expresión, que serian el resultado de las medidas de cada uno de los distintos metabolitos. Lo que se ha denominado expressio_data.

Debido a que los datos no estaban organizados de la forma correcta, se ha tenido que transponer los datos de expresión.

Finalmente se ha aplicado la sintaxis del SummarizedExperiment. Aquí he de remarcar, que tengo un fallo en el código, porque todo sale correcto, pero no me aparece ningún metadato.

- **Exploración de los datos.**

Para la exploración de los datos nos debemos plantear las siguientes preguntas:

1. Cual es la pregunta biológica que subyace al estudio:

Para ello, tenemos que definir primero que es la enfermedad que se está estudiando. La caquexia es un síndrome metabólico complejo asociado a una enfermedad subyacente y que se caracteriza por la pérdida de músculo con o sin pérdida de masa grasa, estando este síndrome asociado por ejemplo al cáncer.

En base a esto, el principal objetivo que perseguía este estudio investigar si se podría detectar la pérdida de músculo en los pacientes con cáncer.

2. Diseño experimental

El diseño experimental que se siguió fue la obtención de muestras de orina, tanto de pacientes, como de personas sanas (47 enfermos de cancer, y 30 controles o sanos). Se uso como criterio de exclusión aquellos que estaban en un proceso activo de radiación o bien un proceso tumoral en el riñón fueron excluidos.

Por otro lado, se utilizo NMR para poder analizar los distintos metabolitos presentes en la orina. Solo se incluyeron en este estudio, aquellos metabolitos que estaban en común con los analistas empleados.

3. Obtención de los datos crudos.

Para la obtención de las muestras se recogió la orina de los grupos experimentales, sin privarles ni de comida ni de bebida. Las muestras fueron procesadas con un reactivo particular. Posteriormente fueron analizadas mediante la resonancia magnética nuclear.

4. Procesamiento y exploración de datos

Para la exploración de los datos, en primer lugar, se han visto el número de variables a estudiar. se ha analizado las diferentes medias de los distintos metabolitos, se ha comprobado el tipo de datos que se ha obtenido.

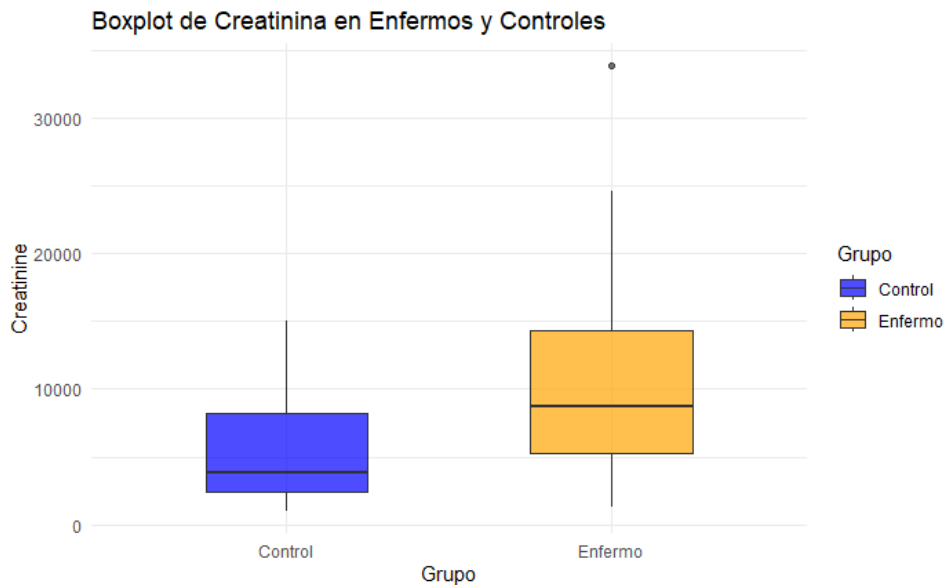
[1] "Patient.ID"	"Muscle.loss"	"x1.6.Anhydro.beta.D.glucose"
[4] "x1.Methylnicotinamide"	"x2.Aminobutyrate"	"x2.Hydroxyisobutyrate"
[7] "x2.Oxoglutarate"	"x3.Aminoisobutyrate"	"x3.Hydroxybutyrate"
[10] "x3.Hydroxyisovalerate"	"x3.Indoxylsulfate"	"x4.Hydroxyphenylacetate"
[13] "Acetate"	"Acetone"	"Adipate"
[16] "Alanine"	"Asparagine"	"Betaine"
[19] "Carnitine"	"Citrate"	"Creatine"
[22] "Creatinine"	"Dimethylamine"	"Ethanolamine"
[25] "Formate"	"Fucose"	"Fumarate"
[28] "Glucose"	"Glutamine"	"Glycine"
[31] "Glycolate"	"Guanidoacetate"	"Hippurate"
[34] "Histidine"	"Hypoxanthine"	"Isoleucine"
[37] "Lactate"	"Leucine"	"Lysine"
[40] "Methylamine"	"Methylguanidine"	"N.N.Dimethylglycine"
[43] "O.Acetylcarnitine"	"Pantothenate"	"Pyroglutamate"
[46] "Pyruvate"	"Quinolate"	"Serine"
[49] "Succinate"	"Sucrose"	"Tartrate"
[52] "Taurine"	"Threonine"	"Trigonelline"
[55] "Trimethylamine.N.oxide"	"Tryptophan"	"Tyrosine"
[58] "Uracil"	"Valine"	"Xylose"
[61] "cis.Aconitate"	"myo.Inositol"	"trans.Aconitate"
[64] "pi.Methylhistidine"	"tau.Methylhistidine"	

Estas son todas las variables que están incluidas en el estudio. La segunda de ellas "Muscle.loss" es la que nos clasifica entre persona con cáncer o persona control. Por otro lado, todas las demás, salvo Patient ID, son los distintitos metabolitos que han sido analizados.

5. Análisis de esos datos

En base a los datos que se han obtenido, en primer lugar, se ha visto cuales son aquellos datos que entre el control y el enfermo más cambian entre sí, obteniéndose que es la creatinina. Para confirmar esto, se ha realizado un test de normalidad (Saphiro test), en el que se observan que ambos grupos experimentales presentan una distribución normal. (creatinine_enfermo, p-value = 0.006585); (creatinine_control, p-value = 0.00483).

Posteriormente, para poder ver si realmente existen diferencias significativas entre estos grupos, se ha procedido con un t-test, obteniéndose un p-value = 0.0005389. Lo que sugiere que realmente, si que existen diferencias en los niveles de creatinina en orina entre el grupo control y el grupo categorizado como Cachexia. Para clarificar esto, se ha graficado empleando un boxplot. En ellos se ve como en el enfermo hay una mayor cantidad de creatinina en la orina que en el control.



Sin embargo, la creatinina es un marcador de daño muscular, ampliamente empleado para evaluar la destrucción tisular. Sin embargo, puede no ser el más efectivo. Por eso, en otro análisis se vio cuales eran las otras dos variables que más cambiaban, siendo "Hippurate" y "Citrate".

En el caso del Hippurate, vemos como también hay diferencias significativas entre ambos grupos, con un p-value = 0.01858. Finalmente, el citrato, se ve que también hay un aumento significativo en el grupo Cachexia en comparación al grupo control (p-value = 0.01071).

6. Interpretación de los resultados.

Tal y como podemos observar, hay una alteración en la creatinina. La creatinina es un biomarcador ampliamente usado para estudiar la destrucción del tejido muscular. En este caso observamos como en el caso de los pacientes afectados de la patología, hay un incremento significativo de la creatinina en la orina en comparación al control. Este resultado tiene sentido, puesto que esta enfermedad cursa con una destrucción del músculo. Por otro lado, el citrato, también puede ser un indicador de una destrucción muscular, por lo que un aumento en estos biomarcadores sugieren un incremento en la destrucción del músculo, lo que contesta a la pregunta inicial.

Ignacio Babiloni Chust, PEC1

Para la creación del repositorio, se ha creado uno nuevo en Github, y añadido los archivos que se solicitaban.

Este es el enlace para acceder al mismo:

https://github.com/Nachobabiloni/Babiloni-Chust-Ignacio-PEC1_Public.git

<https://github.com/Nachobabiloni/Babiloni-Chust-Ignacio-PEC1.git>