

# Análisis de datos ómicos PEC1

María Ramírez García

## Descripción del proceso.

Para realizar esta PEC he buscado un dataset en metabolomics Workbench, filtrando por especie. Navegando con diferentes filtros de enfermedad y tipo de muestra, encontré un estudio que llamo mi atención, *Brain Metabolomics in Fragile X-Associated Tremor/Ataxia Syndrome (FXTAS)*. Me pareció interesante encontrar datos de metabolómica cerebral relacionados con una enfermedad rara como es FXTAS. El análisis de metabolitos y el estudio la mutación en el cromosoma X, son cruciales para poder entender mejor las causas y síntomas de este síndrome.

Mediante el ID **ST002825** del estudio en *metabolomics Workbench* he descargado los metadatos. Los diferentes datos del análisis se encuentran en un único archivo, en mi caso descargue el archivo con extensión txt, este fichero de texto estaba dividido en varias secciones, entre ellas secciones que contienen: datos relacionados con información del estudio, datos sobre las muestras, datos sobre los metabolitos estudiados o datos de los metabolitos en cada muestra.

Para trabajar con los datos de forma mas cómoda y su vez poder crear un objeto SummarizedExperiment, he creado varios archivos txt modificando el archivo original y dando lugar a tres archivos diferentes, que contienen la información de las muestras, la información de los metabolitos y el dataset. La modificación la he llevado a cabo partiendo del fichero original y eliminando las líneas no deseadas de cada fichero txt. Posteriormente he creado a partir de cada fichero txt un fichero .xlsx, con la finalidad de importar los datos y trabajar cómodamente con R. Una vez que he creado el objeto SummarizedExperiment y he comprobado su contenido y estructura he realizado un análisis exploratorio de los datos.

Estos archivos junto con SummarizedExperiment\_PEC1.Rda, el código en R (Ramírez\_García\_María\_PEC1.Rmd), un archivo descriptivo (README.md) y este informe se encuentran en el siguiente repositorio de GitHub:  
- <https://github.com/maria-r-g/Ramirez-Garcia-Maria-PEC1.git>

## Creación del SummarizedExperiment

El SummarizedExperiment está creado a partir con los archivos descritos al inicio de la PEC, a excepción de la sección que contiene los metadatos del estudio, los cuales han sido añadidos manualmente creando una lista y con la función metadata(). Como se menciona la parte del SummarizedExperiment que detalla la información de las muestras esta creada a partir del archivo SampleXf.xlsx con colData(). Las partes correspondientes al dataset y a la información de los metabolitos son creados de la misma forma pero con los archivos DataSetX.xlsx y metabolXf.xlsx y las funciones assays() y rowData() respectivamente.

```
# Leemos el archivo con los datos analizados, contendrá el DataSet.
datosXf <- read_xlsx("DataSetX.xlsx")

# Comprobamos la estructura de los datos y en que clase se han guardado.
head(datosXf)
```

```
## # A tibble: 6 x 83
```

```
## Sample      '1001-16-PF-CBL_057' '1002-18-WH-CBL_063' '1003-16-JSE-CBL_059'
## <chr>      <chr>                  <chr>                  <chr>
## 1 xylulose   1141                  3491                  1369
## 2 xylose     917                   6990                  2818
## 3 xylitol    3384                  18530                 5438
## 4 xanthosine 146                   761                   237
## 5 xanthine   14511                 29397                 9743
## 6 valine     159317                163558                74193
## # i 79 more variables: '1004-04-FS-CBL_040' <chr>, '1004-16-SJ-CBL_061' <chr>,
## #   '1004-19-EA-CBL_064' <chr>, '1006-05-KW-CBL_043' <chr>,
## #   '1007-05-HP-CBL_042' <chr>, '1007-05-LL-CBL_038' <chr>,
## #   '1008-20-RF-CBL_066' <chr>, '1009-20-DJ-CBL_068' <chr>,
## #   '1015-09-DK-CBL_051' <chr>, '1018-08-EB-CBL_047' <chr>,
## #   '1030-09-MM-CBL_053' <chr>, '1041-08-IM-CBL_049' <chr>,
## #   '1046-10-MR-CBL_055' <chr>, '105-03-AG-CBL_034' <chr>, ...
```

```
class(datosXf)
```

```
## [1] "tbl_df"      "tbl"        "data.frame"
```

```
dim(datosXf)
```

```
## [1] 501  83
```

```
# Guardamos los datos en un data.frame, para manipularlos mejor.
datosXf <- as.data.frame(datosXf)
# Asignamos la primera columna como nombres de filas
rownames(datosXf) <- datosXf[[1]]
datosXf <- datosXf[, -1] # Eliminamos la primera columna que ahora es redundante
# comprobamos el data.frame
datosXf[1:6,1:3]
```

```
##           1001-16-PF-CBL_057 1002-18-WH-CBL_063 1003-16-JSE-CBL_059
## xylulose           1141           3491           1369
## xylose             917            6990           2818
## xylitol            3384           18530          5438
## xanthosine         146             761            237
## xanthine           14511          29397           9743
## valine             159317         163558          74193
```

```
class(datosXf)
```

```
## [1] "data.frame"
```

```
dim(datosXf)
```

```
## [1] 501  82
```

```
# Convertimos los datos en una matriz para poder introducirlos al SummarizedExperiment
datosXf <- as.matrix(sapply(datosXf, as.numeric))
# Los datos han de ser una matriz numérica para poder añadirlos con assays()
class(datosXf)
```

```
## [1] "matrix" "array"
```

```
# Añadimos y modificamos los archivos correspondientes a las muestras.
SampleXf <- read_xlsx("SampleXf.xlsx")
SampleXf <- as.data.frame(SampleXf)
rownames(SampleXf) <- SampleXf[[1]]
SampleXf <- SampleXf[, -1]
# Comprobamos las muestras (columnas del dataset)
class(SampleXf)
```

```
## [1] "data.frame"
```

```
dim(SampleXf)
```

```
## [1] 82 2
```

```
SampleXf[1:3,]
```

```
##               treatment Raw file names and additional sample data
## 1001-16-PF-CBL_057 CBL - cases      RAW_FILE_NAME=220603bKCsa23_1
## 1002-18-WH-CBL_063 CBL - cases      RAW_FILE_NAME=220604bKCsa19_1
## 1003-16-JSE-CBL_059 CBL - cases      RAW_FILE_NAME=220604bKCsa10_1
```

```
# Añadimos y modificamos los archivos correspondientes a los metabolitos.
metabolXf <- read_xlsx("metabolXf.xlsx")
```

```
## New names:
## * ' ' -> '...1'
```

```
metabolXf <- as.data.frame(metabolXf)
rownames(metabolXf) <- metabolXf[[1]]
metabolXf <- metabolXf[, -1]
# Comprobamos los metabolitos (filas del dataset)
class(metabolXf)
```

```
## [1] "data.frame"
```

```
dim(metabolXf)
```

```
## [1] 501 7
```

```
metabolXf[1:3,]
```

```
##           ret.index quant mz  BB id
## xylulose    553450      173  31632
## xylose     546308      307 383564
## xylitol    567437      217   5857
##
## xylulose 85:1861.0 86:702.0 87:1148.0 88:324.0 89:10095.0 90:1041.0 91:452.0 94:60.0 95:96.0 96:155.0
## xylose
## xylitol
##
##           PubChem  KEGG                      InChI Key
## xylulose  439205  C00312  LQXVFWRQNMEDEE-PYHARJCCSA-N
## xylose    135191  C00181  SRBFZHDQGSBBOR-IOVATXLUSA-N
## xylitol    6912   C00379  HEBKCHPVOIAQTA-NGQZWHPSA-N
```

```
# Creamos el SummarizedExperiment
se <- SummarizedExperiment( assays = list(counts= datosXf), rowData = metabolXf, colData = SampleXf)
# Añadimos información del estudio manualmente.
metadata(se) <- list(
  Study_ID   = "ST002825",
  Study_Title = "Brain Metabolomics in Fragile X-Associated Tremor/Ataxia Syndrome (FXTAS)",
  Study_Summary = "The course of pathophysiological mechanisms involved in fragile X-associated tremor/",
  Institute = "University of California, Davis",
  Name_author = "Martínez-Cerdeño, Verónica",
  Email = "vmartinezcerdeno@ucdavis.edu",
  Submit_Date = "2023-08-22",
  Analysis_Type = "GC-MS",
  Release_Date = "2023-09-19",
  Subject_Type = "Human",
  Link_metabolomicsworkbench = "https://www.metabolomicsworkbench.org/data/DRCCMetadata.php?Mode=Study&S",
  Link_ncbi = "https://pmc.ncbi.nlm.nih.gov/articles/PMC10487256/"
)
# Guardamos el objeto se SummarizedExperiment en formato binario .Rda
save(se, file = "SummarizedExperiment_PEC1.Rda")
```

## Análisis exploratorio

A continuación, se muestran dos análisis exploratorios, el primero enfocado a la estructura e información del objeto `se` **SummarizedExperiment** a través de diferentes accesorios. Y el segundo enfocado en la exploración y relación de los datos obtenidos en el estudio.

### Análisis exploratorio, visión general del SummarizedExperiment

```
# Información de SummarizedExperiment
se
```

```
## class: SummarizedExperiment
## dim: 501 82
## metadata(12): Study_ID Study_Title ... Link_metabolomicsworkbench
```

```
## Link_ncbi
## assays(1): counts
## rownames(501): xylulose xylose ... 68 61
## rowData names(7): ret.index quant mz ... KEGG InChI Key
## colnames(82): 1001-16-PF-CBL_057 1002-18-WH-CBL_063 ... Pool_07 Pool_08
## colData names(2): treatment Raw file names and additional sample data
```

Extraemos información de SummarizedExperiment a través de diferentes accessor.

```
# Accedemos a los datos de expresión
class(assays(se)) # lista compuesta por los dataset, en esta ocasión solo uno.
```

```
## [1] "SimpleList"
## attr(,"package")
## [1] "S4Vectors"
```

```
dim(assay(se))
```

```
## [1] 501 82
```

```
class(assays(se)$counts)
```

```
## [1] "matrix" "array"
```

```
# Vemos el contenido de los datos de expresión (primeras filas y columnas)
assay(se)[1:6, 1:3]
```

```
##           1001-16-PF-CBL_057 1002-18-WH-CBL_063 1003-16-JSE-CBL_059
## xylulose           1141           3491           1369
## xylose             917           6990           2818
## xylitol            3384          18530           5438
## xanthosine         146            761            237
## xanthine           14511          29397           9743
## valine             159317         163558          74193
```

```
# Metadatos de las muestras
class(colData(se))
```

```
## [1] "DFrame"
## attr(,"package")
## [1] "S4Vectors"
```

```
head(colData(se))
```

```
## DataFrame with 6 rows and 2 columns
##           treatment Raw file names and additional sample data
##           <character>           <character>
## 1001-16-PF-CBL_057 CBL - cases      RAW_FILE_NAME=220603..
## 1002-18-WH-CBL_063 CBL - cases      RAW_FILE_NAME=220604..
## 1003-16-JSE-CBL_059 CBL - cases      RAW_FILE_NAME=220604..
## 1004-04-FS-CBL_040 CBL - cases      RAW_FILE_NAME=220603..
## 1004-16-SJ-CBL_061 CBL - cases      RAW_FILE_NAME=220604..
## 1004-19-EA-CBL_064 CBL - cases      RAW_FILE_NAME=220604..
```

```
# Metadatos de los metabolitos
class(rowData(se))
```

```
## [1] "DFrame"
## attr(,"package")
## [1] "S4Vectors"
```

```
head(rowData(se))
```

```
## DataFrame with 6 rows and 7 columns
##          ret.index  quant mz      BB id      mass spec  PubChem
##          <numeric> <numeric> <numeric>      <character> <numeric>
## xylulose      553450      173      31632 85:1861.0 86:702.0 8..      439205
## xylose        546308      307      383564 85:57.0 86:2871.0 87..      135191
## xylitol        567437      217        5857 85:22.0 87:53.0 88:9..      6912
## xanthosine     926133      325      104466 85:67.0 86:155.0 87:..      64959
## xanthine       701688      353        1669 85:2517.0 86:3710.0 ..      1188
## valine         313502      144          3 85:2124.0 86:7630.0 ..      6287
##          KEGG          InChI Key
##          <character>      <character>
## xylulose      C00312 LQXVFWRQNMEDEE-PYHAR..
## xylose        C00181 SRBFZHDQGSBBOR-IOVAT..
## xylitol        C00379 HEBKCHPVOIAQTA-NGQZW..
## xanthosine     C01762 UBORTCNDUKBEOP-UUOKF..
## xanthine       C00385 LRFVITYWQQMYALW-UHFFF..
## valine         C00183 KZSNJWFQEVHDMF-BYPYZ..
```

```
# Visualizamos los metadatos añadidos
metadata(se)
```

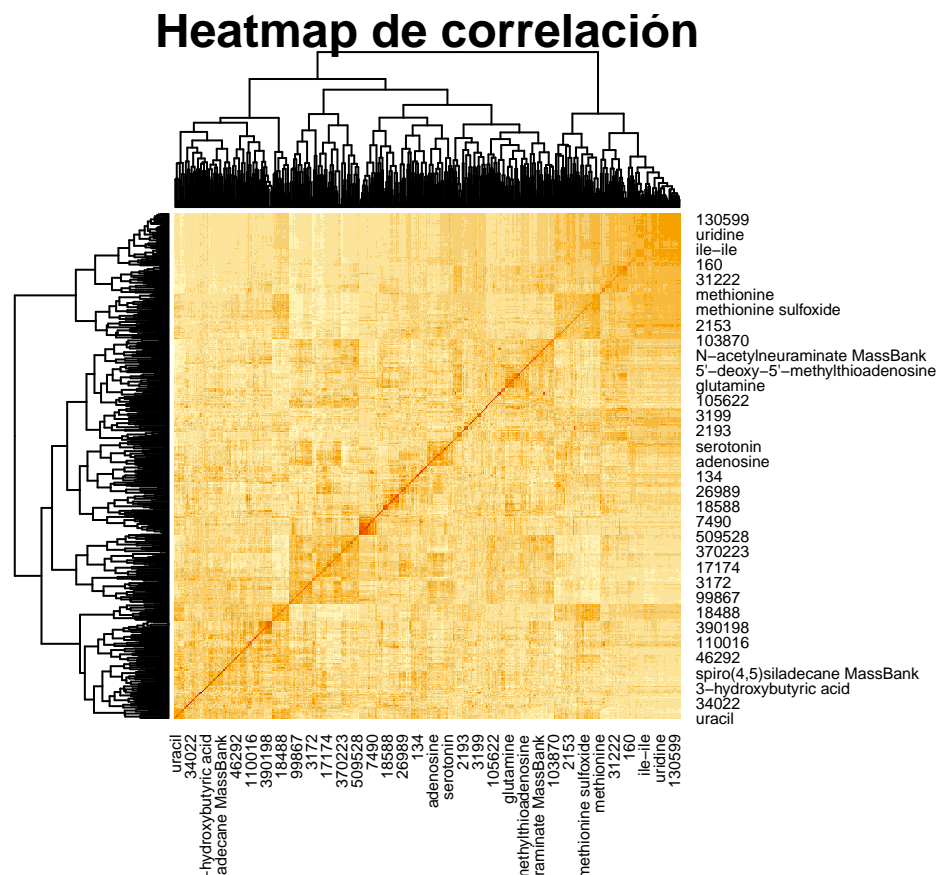
```
## $Study_ID
## [1] "ST002825"
##
## $Study_Title
## [1] "Brain Metabolomics in Fragile X-Associated Tremor/Ataxia Syndrome (FXTAS)"
##
## $Study_Summary
## [1] "The course of pathophysiological mechanisms involved in fragile X-associated tremor/ataxia syndr
##
## $Institute
## [1] "University of California, Davis"
##
## $Name_author
## [1] "Martínez-Cerdeño, Verónica"
##
## $Email
## [1] "vmartinezcerdeno@ucdavis.edu"
##
## $Submit_Date
## [1] "2023-08-22"
##
## $Analysis_Type
```

```
## [1] "GC-MS"
##
## $Release_Date
## [1] "2023-09-19"
##
## $Subject_Type
## [1] "Human"
##
## $Link_metabolomicsworkbench
## [1] "https://www.metabolomicsworkbench.org/data/DRCCMetadata.php?Mode=Study&StudyID=ST002825&StudyType=GC-MS"
##
## $Link_ncbi
## [1] "https://pmc.ncbi.nlm.nih.gov/articles/PMC10487256/"
```

## Análisis exploratorio, visión general de los datos.

Creamos un gráfico Heatmap, para que nos ayude a visualizar la expresión de los metabolitos en las muestras. Con él podemos observar la correlación entre las muestras de manera más visual. En los datos las muestras están ordenadas según la región analizada y dentro de esta, según si son casos o controles.

```
# Análisis de correlación
# Extraemos la matriz de datos y la transponemos para que las muestras sean filas y hacemos la correlación
cor_matrix <- cor(t(assay(se)))
# Visualización con un Heatmap
heatmap(cor_matrix, main = "Heatmap de correlación")
```



Como podemos observar la información que nos muestra el Heatmap de correlación no es de mucha utilidad. No podemos obtener conclusiones claras del grafico por varias razones, la principal seria la gran cantidad de muestras y metabolitos que hemos utilizado, aunque las muestras estén ordenadas por tipo y casos no las hemos agrupado lo que dificulta la visualización.

En el caso de los metabolitos pasa algo parecido, en los datos encontramos todos los metabolitos encontrados, lo ideal sería filtrar por relevancia para la enfermedad, esto lo podemos conseguir con ayuda de estudios anteriores y realizando otros análisis para averiguar cuales de los metabolitos tienen una expresión diferencial según si es control o caso dentro de cada tipo de muestra.

```
# Análisis del cerebelo (CBL)
```

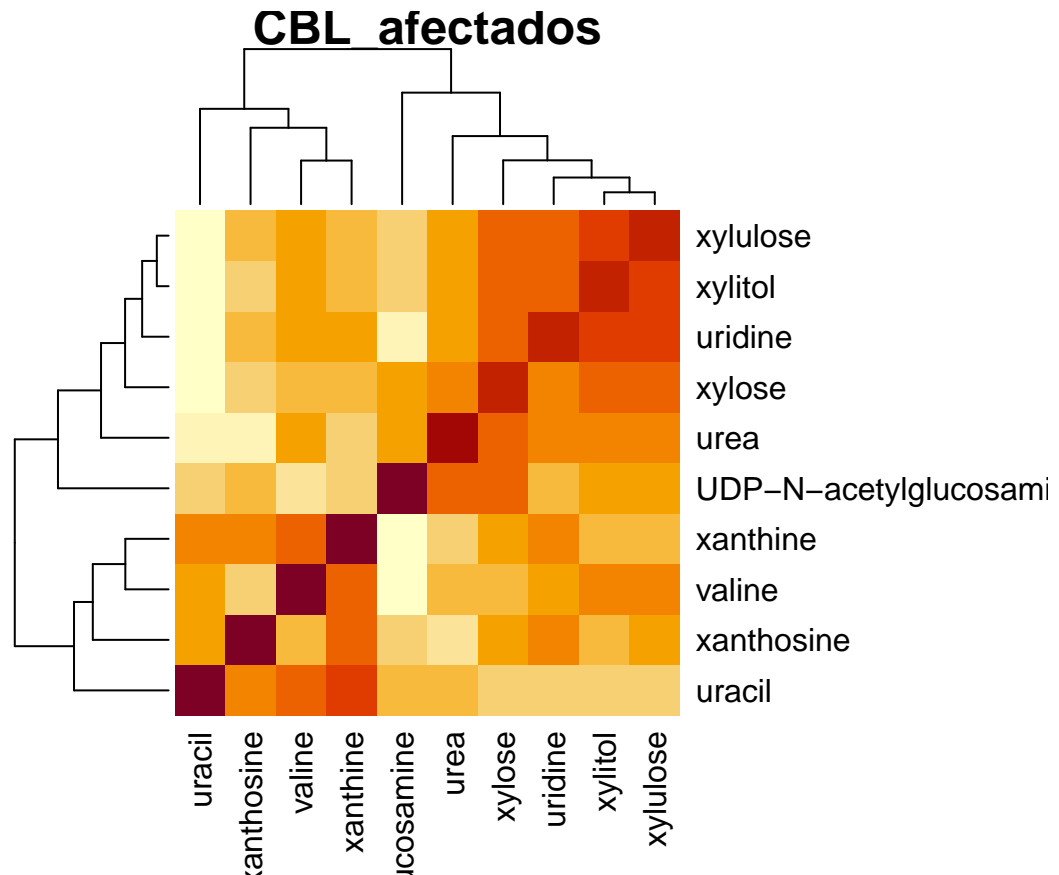
```
# Con la exploración de los datos sabemos que las muestras tomadas del CB son las columnas del 1:38  
# (1:20 casos y 21:38 controles).
```

```
# Análisis de correlación en casos afectados
```

```
CBL_casos<- cor(t(assay(se)[1:10,1:20]))
```

```
# Visualización
```

```
heatmap(CBL_casos, main = "CBL_afectados")
```

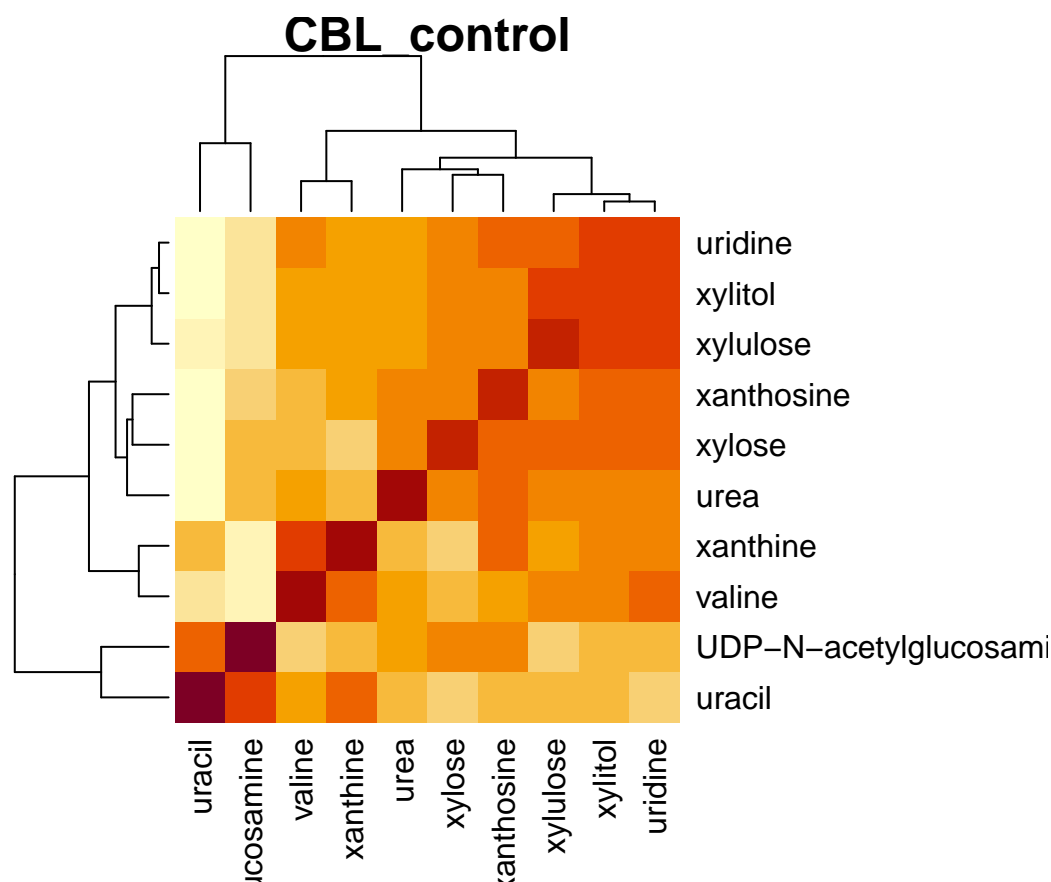


```
# Análisis de correlación en casos control
```

```
CBL_control<- cor(t(assay(se)[1:10,21:38]))
```

```
heatmap(CBL_control, main = "CBL_control")
```





Se filtran los metabolitos para una mejor visualización, pero no sigue un criterio claro, se debería realizar el filtro después de un análisis más profundo de los datos, para seleccionar los metabolitos que sean más significativos. Tampoco hemos realizado ninguna normalización de datos, ni análisis para determinar si las diferencias son significativas.

## Interpretación biológica

En el estudio “*Brain Metabolomics in Fragile X-Associated Tremor/Ataxia Syndrome (FXTAS)*” se realiza un análisis de perfiles metabólicos en muestras post mortem de individuos con FXTAS.

El síndrome FXTAS es un trastorno neurodegenerativo causado por una repetición expandida de trinucleótidos CGG en la región promotora del gen *FMR1*, que codifica para la proteína FMRP, la cual está implicada en el desarrollo cognitivo. El desarrollo de FXTAS esta influenciado por el número de expansiones del triplete CGG que conforman la permutación de *FMR1*, aumentando la tasa de conversión con la edad.

## Materiales y métodos

Los autores del estudio usan un enfoque de metabolómica no dirigida utilizando cromatografía de gases acoplada con espectrometría de masas (GC-MS) para un perfil metabólico global. Las muestras analizadas son de la corteza cerebral (*inferior temporal gyrus (ITG)*) y del cerebelo (*CB*) humana post mortem. La longitud de las repetición CGG se obtuvo de sangre y se determino mediante PCR y Southern blot.

## Participantes

Se analizaron 25 cerebros con FXTAS y 19 controles emparejados por edad y sexo. Los casos de FXTAS estaban avanzados en la progresión de la enfermedad y tenían una presentación clínica variable. El diagnóstico de FXTAS fue confirmado con la presencia de la premutación de *FMR1*.

## Resultados

En el artículo identifican 191 metabolitos primarios de los 513 que se detectaron. De los cuales 22 metabolitos mostraban cambios entre grupos de ITG, mientras que en el CB mostraron cambios 21 metabolitos. La citidina y el ácido glicérico mostraron una regulación negativa en el ITG de los pacientes con FXTAS, en comparación con los controles. En el CB, la fructosa-1-fosfato, el 1,5-anhidroglucitol, el ácido 2-hidroxibutanoico y la glucosa-6-fosfato mostraron una regulación positiva, mientras que la oleamida mostró una regulación negativa.

## Discusión general

Con los datos obtenidos del dataset se pueden realizar diversos análisis y plantear varias conclusiones sobre que metabolitos están más afectados en FXTAS, esto lo podemos saber gracias al artículo que publican los autores del estudio. Sin embargo, a nivel personal no he sabido analizar correctamente los datos para llegar a determinar resultados significativos.

## Bibliografía

- Salcedo-Arellano, M. J., Johnson, M. D., McLennan, Y. A., Hwang, Y. H., Juarez, P., McBride, E. L., Pantoja, A. P., Tassone, F., Hagerman, R. J., & Martínez-Cerdeño, V. (2023). Brain metabolomics in fragile X-associated tremor/ataxia syndrome (FXTAS). *Cells*, 12(17), 2132. <https://doi.org/10.3390/cells12172132>
- Metabolomics Workbench. (2023). Brain metabolomics in fragile X-associated tremor/ataxia syndrome (FXTAS). <https://www.metabolomicsworkbench.org/data/DRCCMetadata.php?Mode=Study&StudyID=ST002825&StudyType=MS&ResultType=1>