

Neuronal and glial DNA methylation and gene expression changes in early epileptogenesis

Erika Nathalia Ordoñez Guzmán

2025-02-07

Introducción:

Según los datos la Organización Mundial de la Salud proporcionados en 2024: “50 millones de personas padecen epilepsia” además “se estima que el 70% de las personas con epilepsia podrían vivir sin convulsiones si se diagnosticaran y trataran adecuadamente”. Este es un trastorno que afecta principalmente a las células que conforman el cerebro. El artículo: “Neuronal and glial DNA methylation and gene expression changes in early epileptogenesis” trabaja con células neuronales y de glía como lo es la microglía. Este padecimiento se caracteriza por descargas eléctricas en el cerebro y afectan negativamente a las personas que la padecen. Estudios recientes, resaltan que la metilación tiene un papel importante en este proceso, por lo que el equipo de trabajo dirigido por Toni C Berger, son pioneros en este campo de estudio.

Por la naturaleza del proceso de metilación, es un proceso heredable, modificable y estable, pero temporal, además de voluble y susceptible a cambios por influencia ambiental. Por eso, en el artículo se experimenta con células neuronales de ratón, en la inducción a la metilación con ácido kainico, que son el grupo experimental, mientras que el control se trató con solución salina. Se buscó encontrar si existe una diferencia significativa de expresión en ambas condiciones. Con este artículo, se busca encontrar el efecto que tiene la metilación en la patología y así, estar un paso más cerca de encontrar agentes terapéuticos para contrarrestar su efecto y apoyar a la salud mundial.

El set de datos obtenido fue obtenido desde recount3, cuenta con 28 muestras. Los datos demuestran las lecturas del mDNA que se dieron como consecuencia de la metilación y secuenciación de mRNA (mRNAseq).

Objetivos del proyecto:

- Encontrar blancos terapéuticos en células neuronales y glía que juegan un papel crucial en la epilepsia.
- Analizar el resultado del mRNAseq para comprobar la diferencia de expresión y correlacionarlo con la epilepsia.

Resultados:

```
# Explorar objeto de rse
rse_gene_SRP223512$sra.sample_attributes[1:5]

## [1] "neun;;negative|source_name;;Hippocampus|tissue;;Hippocampus|treatment;;Kainate"
## [2] "neun;;positive|source_name;;Hippocampus|tissue;;Hippocampus|treatment;;Kainate"
## [3] "neun;;negative|source_name;;Hippocampus|tissue;;Hippocampus|treatment;;Sham"
## [4] "neun;;positive|source_name;;Hippocampus|tissue;;Hippocampus|treatment;;Sham"
## [5] "neun;;negative|source_name;;Hippocampus|tissue;;Hippocampus|treatment;;Kainate"
```

```
hist(rse_gene_SRP223512$assigned_gene_prop)
```

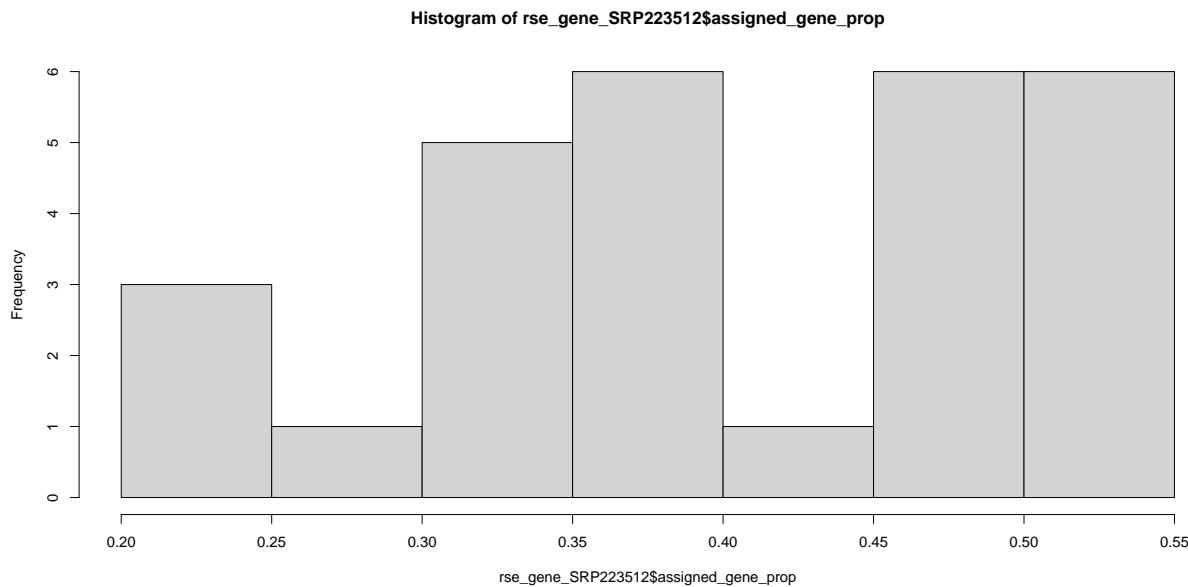


Figure 1: Frecuencias de la proporción de genes

En la figura 1 se muestra la frecuencia de las muestras según la proporción de los genes. Sirve para visualizar las muestras que se pueden conservar para el análisis.

```
# Se muestra de manera más numérica el resultado del histograma  
table(rse_gene_SRP223512$assigned_gene_prop < 0.3)
```

```
##  
## FALSE TRUE  
##    24     4
```

```
table(rse_gene_SRP223512$assigned_gene_prop < 0.4)
```

```
##  
## FALSE TRUE  
##    13     15
```

```
table(rse_gene_SRP223512$assigned_gene_prop < 0.5)
```

```
##  
## FALSE TRUE  
##     6    22
```

Como podemos ver con los resultados anteriores, el mejor umbral para obtener la cantidad de muestras para trabajar es la proporción de 0.3. Debido a que no buscamos un umbral muy estricto y quedarnos sin una buena cantidad de muestras para analizar.

```
rse_gene_SRP223512 <- rse_gene_SRP223512[, rse_gene_SRP223512$assigned_gene_prop > 0.3]
```

Resumen de los niveles medios de expresión de genes:

```
summary(gene_means)
```

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	0.00	0.50	2.88	96.98	33.42	132204.88

Se eligió un umbral de 0.1 para la elegir los genes con los que se trabajará en el análisis. El resto de genes se eliminan. De decir, que los genes que tienen una media menor a 0.1 se eliminan.

```
rse_gene_SRP223512 <- rse_gene_SRP223512[gene_means > 0.1, ]  
dim(rse_gene_SRP223512)
```

```
## [1] 49447    24
```

Porcentaje de genes conservados:

```
round(nrow(rse_gene_SRP223512) / nrow(rse_gene_SRP223512_unfiltered) * 100, 2)
```

```
## [1] 89.22
```

```
ggplot(as.data.frame(colData(rse_gene_SRP223512)), aes(y=assigned_gene_prop, x=treatment)) +  
  geom_boxplot() +  
  theme_bw(base_size = 20) +  
  ylab("Assigned Gene Prop") +  
  xlab("Treatment")
```

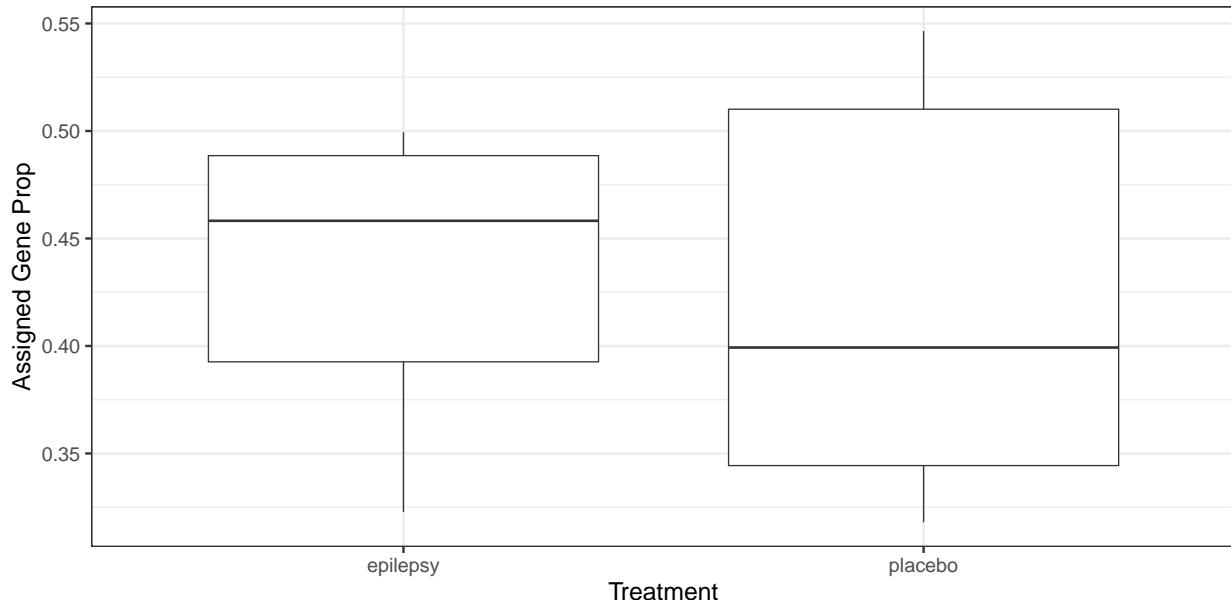


Figure 2: Boxplot con información sobre la proporción de genes con respecto a cada tratamiento

En la figura 2, se evidencia que la media de la proporción de genes es mayor en el tratamiento con ácido kainico.

```

# Modelo estadístico:
mod <- model.matrix(
  ~ treatment + sra_attribute.neun + assigned_gene_prop,
  data = colData(rse_gene_SRP223512)
)

# Se muestran los nombres de las columnas para evitar errores.
colnames(mod)

## [1] "(Intercept)"           "treatmentplacebo"
## [3] "sra_attribute.neunpositive" "assigned_gene_prop"

# Comprobación del modelo estadístico:
# Para visualizar el modelo:
# Data frame para la visualización.
sampleData <- data.frame(
  treatment = rse_gene_SRP223512$treatment,
  sra_attribute.neun = rse_gene_SRP223512$sra_attribute.neun
)
# Creación de las imágenes:
vd <- ExploreModelMatrix::VisualizeDesign(
  sampleData = sampleData,
  designFormula = ~ 0 + treatment + sra_attribute.neun,
  textSizeFitted = 3
)
cowplot::plot_grid(plotlist=vd$plot, nrow = 1)

```

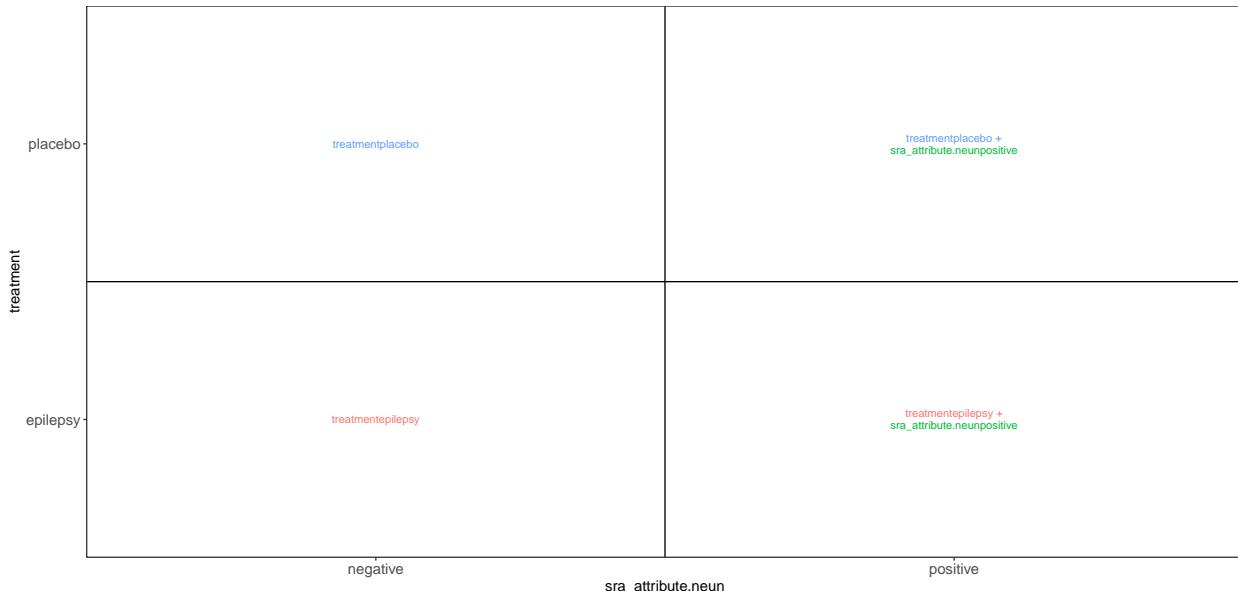


Figure 3: Cuadro con la comprobación del modelo de regresión lineal aplicado anteriormente

En el cuadro anterior, podemos darnos cuenta que el modelo fue correctamente planteado, esto se puede notar porque, si restamos la columna titulada negative a la titulada positive el resultado es

sra_attribute.neunpositive, lo que quiere decir que tuvo un efecto el receptor NeuN en las neuronas. Si se resta la fila de placebo a la de epilepsia en la columna positive, nos afirma que efectivamente, la fila de placebo es el control experimental mientras que la segunda fila de epilepsia afirma que ese es el grupo al que se aplicó el ácido kainico y por tanto el grupo experimental.

NOTA: Para recrear la imagen, no se aplicó el intercepto y tampoco se contempló la variable continua, que en este caso es la proporción de genes. Pues se sabe cual es esta variable y para visualizar de mejor manera se obvió. Por otro lado, esta variable si fue implementada en la construcción del modelo, junto con la aplicación del intercepto, para tener un modelo más confiable.

```
# Usar limma para el análisis de expresión diferencial:  
library("limma")  
# Crear el objeto DGEList  
# Se produce la gráfica para visualizar los datos.
```

```
vGene <- voom(dge, mod, plot=T)
```

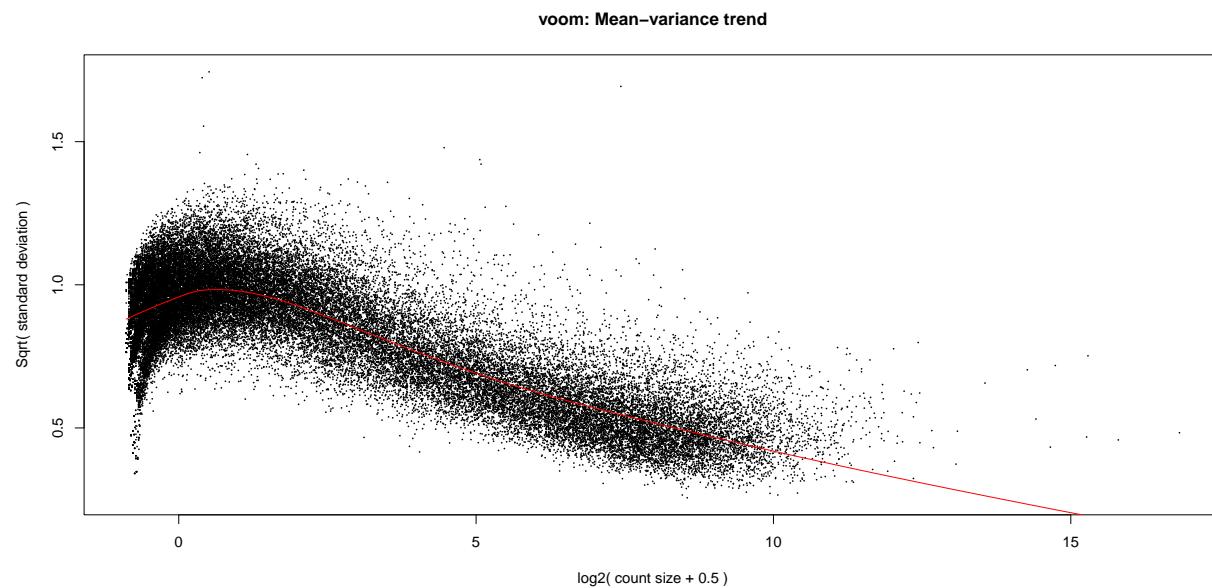


Figure 4: Gráfica para ver la desviación estándar

La gráfica anterior muestra los genes, es decir, cada punto es un gen. Muestra la buena calidad de los datos.

```
# Utiliza el modelo bayesiano para el análisis de expresión diferencial  
# Además, ajusta el modelo lineal  
eb_results <- eBayes(lmFit(vGene))  
  
# Selecciona los genes con expresión diferencial significativa.  
de_results <- topTable(  
  eb_results,  
  coef=2,  
  number=nrow(rse_gene_SRP223512),  
  sort.by="none"  
)
```

```

# Pequeña exploración y visualización de los datos.
dim(de_results)

## [1] 49447    17

head(de_results)

##           source type bp_length phase          gene_id
## ENSMUSG00000079800.2 ENSEMBL gene      1271      NA ENSMUSG00000079800.2
## ENSMUSG00000095092.1 ENSEMBL gene      366       NA ENSMUSG00000095092.1
## ENSMUSG00000079192.2 ENSEMBL gene      255       NA ENSMUSG00000079192.2
## ENSMUSG00000079794.2 ENSEMBL gene      255       NA ENSMUSG00000079794.2
## ENSMUSG00000094799.1 ENSEMBL gene      366       NA ENSMUSG00000094799.1
## ENSMUSG00000095250.1 ENSEMBL gene      366       NA ENSMUSG00000095250.1
##             gene_type gene_name level mgi_id havana_gene tag
## ENSMUSG00000079800.2 protein_coding AC125149.3     3 <NA> <NA> <NA>
## ENSMUSG00000095092.1 protein_coding AC125149.5     3 <NA> <NA> <NA>
## ENSMUSG00000079192.2 protein_coding AC125149.1     3 <NA> <NA> <NA>
## ENSMUSG00000079794.2 protein_coding AC125149.2     3 <NA> <NA> <NA>
## ENSMUSG00000094799.1 protein_coding AC125149.4     3 <NA> <NA> <NA>
## ENSMUSG00000095250.1 protein_coding AC133103.4     3 <NA> <NA> <NA>
##            logFC AveExpr          t   P.Value adj.P.Val
## ENSMUSG00000079800.2 -0.79211380 -1.339359 -1.83477639 0.07752263 0.3033380
## ENSMUSG00000095092.1  0.35531783 -2.827025  1.01138413 0.32076462 0.6599825
## ENSMUSG00000079192.2 -0.01990891 -1.833560 -0.05463589 0.95682874 0.9886429
## ENSMUSG00000079794.2 -0.60064052 -2.628905 -1.64792109 0.11089901 0.3774520
## ENSMUSG00000094799.1  0.41875215 -2.449011  1.52330036 0.13925145 0.4300254
## ENSMUSG00000095250.1  0.19371546 -1.691030  0.56942454 0.57375106 0.8360616
##           B
## ENSMUSG00000079800.2 -4.497655
## ENSMUSG00000095092.1 -5.630542
## ENSMUSG00000079192.2 -6.028328
## ENSMUSG00000079794.2 -4.808607
## ENSMUSG00000094799.1 -5.015600
## ENSMUSG00000095250.1 -5.907050

# Selección de genes con una expresión diferencial significativa
table(de_results$adj.P.Val < 0.05)

## 
## FALSE  TRUE
## 43793 5654

# Visualización de datos - plotMA
plotMA(eb_results, coef=2)

```

Cada punto en la gráfica anterior muestra un gen, se perciben más genes que se mencionen cerca del cero, es decir, que no tienen una expresión significativa. Hay más genes que se tienen un log fold-change con expresión negativa.

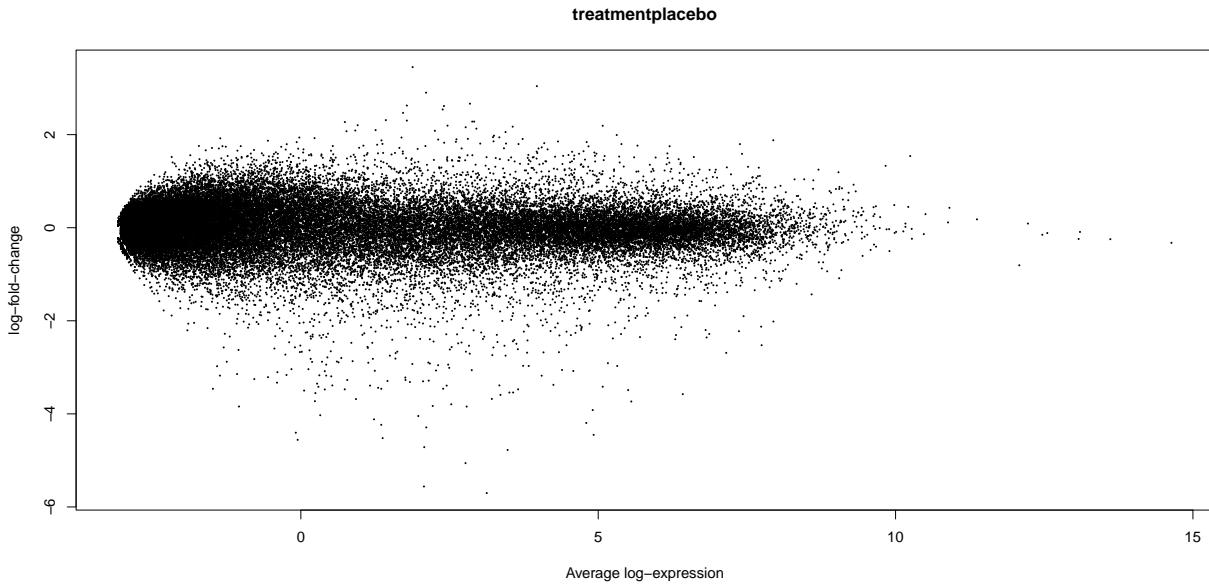


Figure 5: plotMA para evaluar la relación entre la expresión media y el cambio en la expresión

```
# Visualización de datos - Volcano:
volcanoplot(eb_results, coef=2, highlight=4, names=de_results$gene_name)
```

La imagen muestra 4 genes particulares que se encuentran diferencialmente expresados en color azul.

```
# Extraer los 50 genes con expresión más significativa
exprs_heatmap <- vGene$E[rank(de_results$P.Value) <= 50, ]

# Data frame
df <- as.data.frame(colData(rse_gene_SRP223512)[, c("treatment", "sra_attribute.neun")])

# Nombre de las columnas del data frame
colnames(df) <- c("treatment", "Neun/marcador")

# Permanecemos con los nombres dados por ENSEMBL.
original_names <- rownames(exprs_heatmap)

# Cambiamos los nombres de las filas para integrarlas al heatmap.
rownames(exprs_heatmap) <- rowRanges(rse_gene_SRP223512)$gene_name[
  match(rownames(exprs_heatmap), rowRanges(rse_gene_SRP223512)$gene_id)
]

# Heatmap
# Si no se encuentra instalado el paquete, correr la siguiente línea:
# install.packages("pheatmap")
library("pheatmap")
pheatmap(
  exprs_heatmap,
  cluster_rows = T,
```

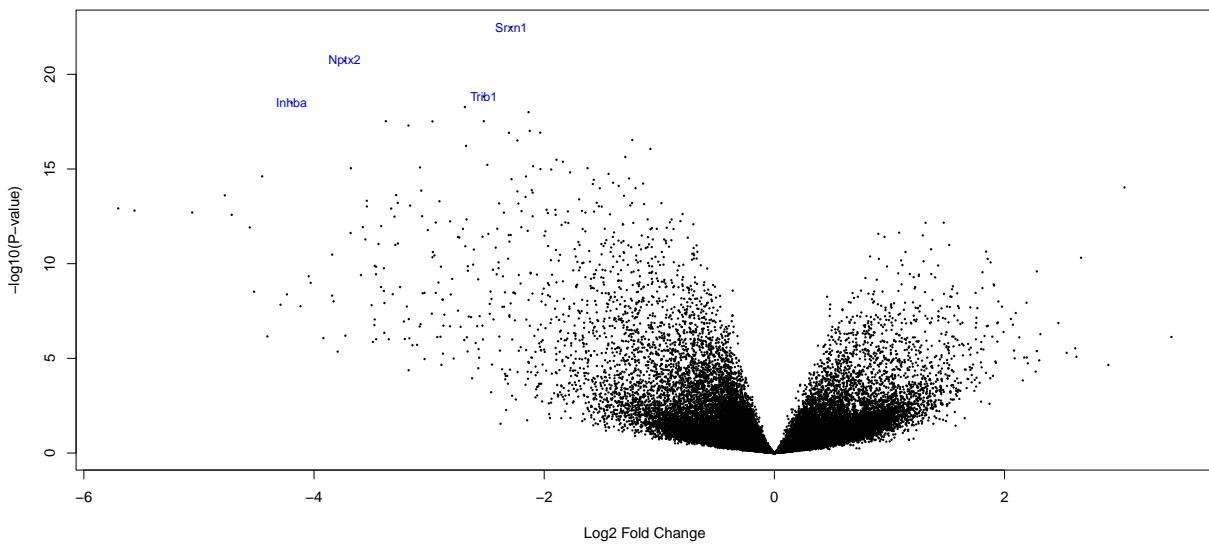


Figure 6: Los genes con expresión diferencial significativa resaltados en color azul

```

cluster_cols = T,
show_rownames = T,
show_colnames = F,
annotation_col = df
)

```

En el heatmap anterior se puede notar que las áreas que presentan un color roja más intenso son las que tienen una expresión diferencial más elevadas que el resto y corresponden al grupo experimental, es decir, son neuronas que se trataron con ácido kainico, aunque cabe resaltar que también se presentan células de glía que muestran con epilepsia y que no muestran signos del marcador.

```

library("RColorBrewer")

# PCA - con el tratamiento: Kainate/Sham
col.group <- as.factor(df$treatment)
levels(col.group) <- brewer.pal(nlevels(col.group), "Set1")

## Warning in brewer.pal(nlevels(col.group), "Set1"): minimal value for n is 3, returning requested pal

col.group <- as.character(col.group)

plotMDS(vGene$E, labels = df$treatment, col=col.group)

```

En el PCA anterior se muestran los grupos experimentales en rojo y en azul los controles. La imagen nos permite apreciar que no hay una marcada diferencia en la expresión dentro de estos dos grupos.

```

# PCA - con el Neun: positivo/negativo
col.neun <- df$`Neun/marcador`
levels(col.neun) <- brewer.pal(nlevels(col.neun), "Dark2")

```

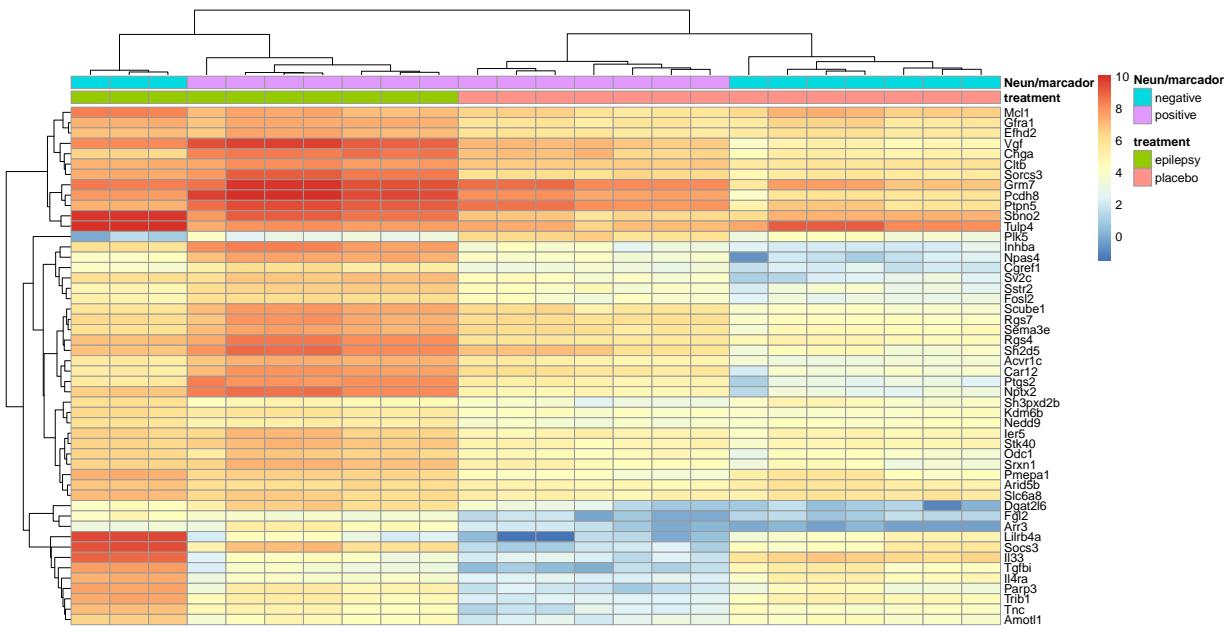


Figure 7: Heatmap con los 50 genes con expresión más significativa

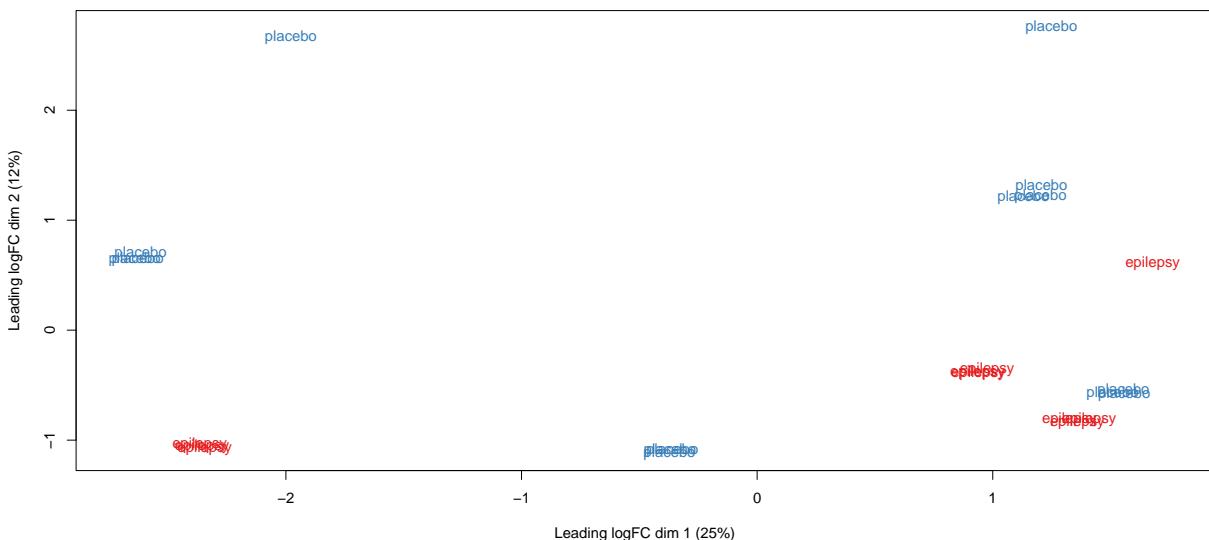


Figure 8: PCA de expresión en grupos experimentales

```

## Warning in brewer.pal(nlevels(col.neun), "Dark2"): minimal value for n is 3, returning requested pal

col.neun <- as.character(col.neun)
plotMDS(vGene$E, labels = df$`Neun/marcador`, col=col.neun)

```

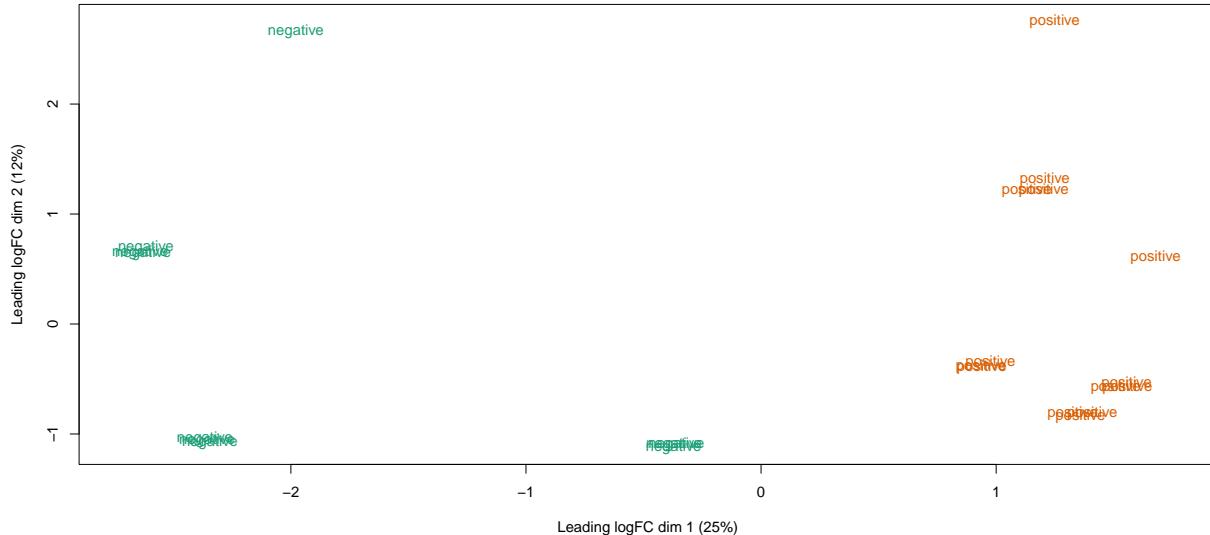


Figure 9: PCA de expresión en grupos con el marcador Neun

En el PCA anterior se muestran diferentes muestras, algunas presentan el marcador fluorescente NeuN para identificar a las neuronas de la glía, por lo que las que se presentan como positive corresponden a las neuronas y las que se presentan como negative corresponden a las células de la glía. En conclusión nos permite ver que efectivamente existe una expresión diferencial entre ambos grupos.

Análisis de resultados:

El artículo “Neuronal and glial DNA methylation and gene expression changes in early epileptogenesis” menciona que se encontró la expresión diferencial de los siguientes genes que están relacionados con la epilepsia, como son: HDAC11, SPP1, GAL, DRD1 y SV2C. Sin embargo, en un análisis de expresión realizado, para mostrar los genes que presentan una expresión diferencial significativa se llevó a cabo el siguiente volcánico plot.

```

# Código para descargar el paquete de EnhancedVolcano en caso de no tenerlo descargado
#if (!requireNamespace('BiocManager', quietly=TRUE))
# install.packages('BiocManager')

# BiocManager::install('EnhancedVolcano')

# Cargar el paquete:
library(EnhancedVolcano)

## Cargando paquete requerido: ggrepel

```

```
EnhancedVolcano(de_results,
  lab=de_results$gene_name,
  x='logFC',
  y='P.Value')
```

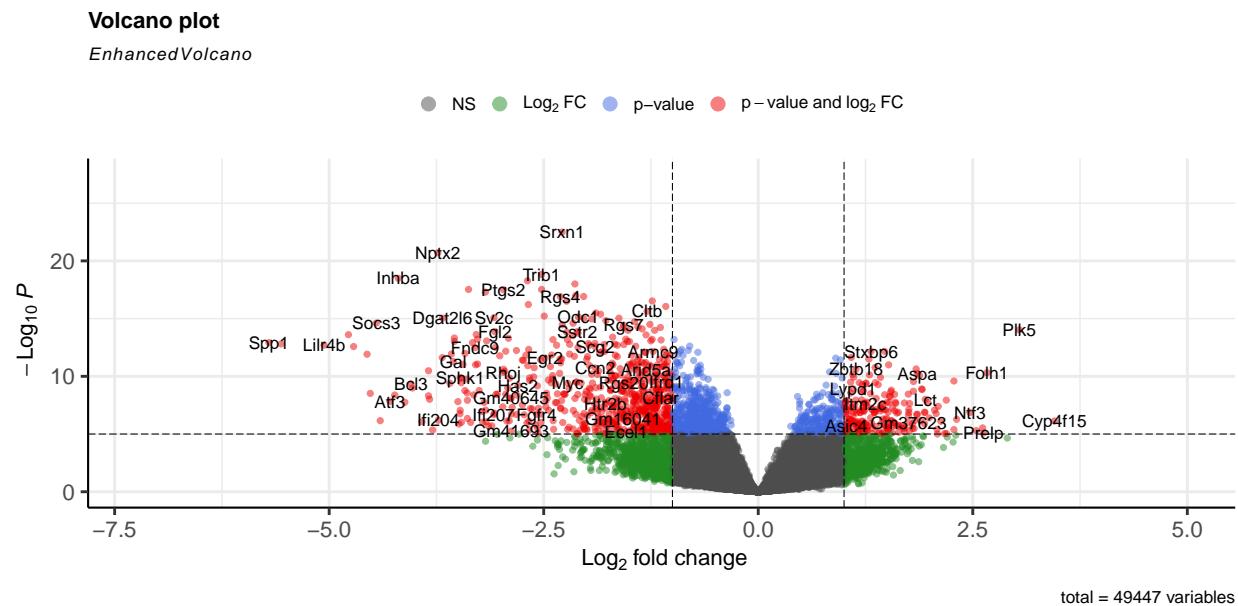


Figure 10: Los genes con expresión diferencial significativa resaltados en color azul

En la imagen nos podemos dar cuenta que de la lista de genes relacionados con la epilepsia resalta uno, por ser, además, el gen que presentó la expresión diferencial más elevada, este es el SPP1.

En las siguientes líneas de código presento la metodología para encontrar los genes que presentan mayor expresión, aunado con los valores que corresponden a los genes mencionados en el artículo.

NOTA: Se eligieron únicamente 4 genes extremos tanto positivos como negativos, únicamente para ilustrar que los genes mencionados en el artículo, salvo uno, no se encuentran entre los primeros cuatro con mayor expresión.

```
# Nuevo data frame con los genes especificados en el artículo
new_data_frame <- de_results[de_results$gene_name %in% c('Spp1', 'Hdac11', 'Gal', 'Drd1', 'Sv2c'),]

# Se muestra el data frame para hacer evidente que los genes si se encuentran en el análisis
new_data_frame
```

```
##          source type bp_length phase      gene_id
## ENSMUSG00000021478.6 HAVANA gene     3628    NA ENSMUSG00000021478.6
## ENSMUSG00000051111.16 HAVANA gene    9700    NA ENSMUSG00000051111.16
## ENSMUSG00000024907.7 HAVANA gene     788    NA ENSMUSG00000024907.7
## ENSMUSG00000029304.14 HAVANA gene    1648    NA ENSMUSG00000029304.14
## ENSMUSG00000034245.10 HAVANA gene    3849    NA ENSMUSG00000034245.10
##          gene_type gene_name level      mgi_id
## ENSMUSG00000021478.6 protein_coding Drd1      2 MGI:99578
## ENSMUSG00000051111.16 protein_coding Sv2c      2 MGI:1922459
## ENSMUSG00000024907.7 protein_coding Gal       2 MGI:95637
```

```

## ENSMUSG00000029304.14 protein_coding      Spp1      2  MGI:98389
## ENSMUSG00000034245.10 protein_coding      Hdac11     2  MGI:2385252
##                                     havana_gene      tag      logFC
## ENSMUSG00000021478.6  OTTMUSG00000067515.1 <NA> -1.8488683
## ENSMUSG00000051111.16 OTTMUSG00000035052.4 overlapping_locus -3.0789058
## ENSMUSG00000024907.7  OTTMUSG00000073812.1 <NA> -3.5538469
## ENSMUSG00000029304.14 OTTMUSG00000024389.2 <NA> -5.7014831
## ENSMUSG00000034245.10 OTTMUSG00000017285.4 <NA>  0.6899831
##             AveExpr      t      P.Value      adj.P.Val      B
## ENSMUSG00000021478.6  3.9746343 -12.504383 8.904495e-13 4.403005e-10 19.220141
## ENSMUSG00000051111.16 4.5810680 -16.705312 8.285062e-16 1.781180e-12 25.987226
## ENSMUSG00000024907.7  0.2442884 -11.578616 5.223026e-12 1.897759e-09 16.420331
## ENSMUSG00000029304.14 3.1255768 -13.612553 1.207208e-13 9.166402e-11 20.954115
## ENSMUSG00000034245.10 5.2091069   4.828541 4.769245e-05 1.145337e-03 1.562235

```

Se buscan los genes con mayor expresión diferencial negativa
`head(de_results[order(de_results$logFC),], 4)`

```

##           source type bp_length phase      gene_id
## ENSMUSG00000029304.14 HAVANA gene    1648      NA ENSMUSG00000029304.14
## ENSMUSG00000005892.4  HAVANA gene    1384      NA ENSMUSG00000005892.4
## ENSMUSG00000112023.1 HAVANA gene    5439      NA ENSMUSG00000112023.1
## ENSMUSG00000112148.1 HAVANA gene    5227      NA ENSMUSG00000112148.1
##           gene_type gene_name level      mgi_id
## ENSMUSG00000029304.14 protein_coding Spp1      2  MGI:98389
## ENSMUSG00000005892.4  protein_coding Trh      2  MGI:98823
## ENSMUSG00000112023.1 HAVANA gene    Lilr4b     2 MGI:102702
## ENSMUSG00000112148.1 HAVANA gene    Lilrb4a    2 MGI:102701
##           havana_gene      tag      logFC AveExpr
## ENSMUSG00000029304.14 OTTMUSG00000024389.2 <NA> -5.701483 3.125577
## ENSMUSG00000005892.4  OTTMUSG00000024415.1 <NA> -5.560045 2.069902
## ENSMUSG00000112023.1 HAVANA gene    OTTMUSG00000016502.6 overlapping_locus -5.058832 2.768749
## ENSMUSG00000112148.1 HAVANA gene    OTTMUSG00000063893.5 <NA> -4.775678 3.476774
##           t      P.Value      adj.P.Val      B
## ENSMUSG00000029304.14 -13.61255 1.207208e-13 9.166402e-11 20.95411
## ENSMUSG00000005892.4  -13.46298 1.569585e-13 1.124801e-10 20.09540
## ENSMUSG00000112023.1 -13.33173 1.979776e-13 1.313323e-10 20.45552
## ENSMUSG00000112148.1 -14.54465 2.466394e-14 2.594803e-11 22.64820

```

A su vez también se buscan los que cuentan con mayor expresión diferencial positiva:
`tail(de_results[order(de_results$logFC),], 4)`

```

##           source type bp_length phase      gene_id
## ENSMUSG00000001773.14 HAVANA gene    3210      NA ENSMUSG00000001773.14
## ENSMUSG00000022132.15 HAVANA gene    2591      NA ENSMUSG00000022132.15
## ENSMUSG00000035486.14 HAVANA gene    2678      NA ENSMUSG00000035486.14
## ENSMUSG00000073424.9  HAVANA gene    2282      NA ENSMUSG00000073424.9
##           gene_type gene_name level      mgi_id
## ENSMUSG0000001773.14 protein_coding Folh1      2 MGI:1858193
## ENSMUSG00000022132.15 protein_coding Cldn10     2 MGI:1913101
## ENSMUSG00000035486.14 protein_coding Plk5       2 MGI:3026984
## ENSMUSG00000073424.9  protein_coding Cyp4f15     2 MGI:2146921
##           havana_gene tag      logFC AveExpr      t

```

```

## ENSMUSG0000001773.14 OTTMUSG00000037777.2 <NA> 2.664941 2.843227 10.475187
## ENSMUSG00000022132.15 OTTMUSG00000023307.3 <NA> 2.902231 2.105378 5.108548
## ENSMUSG00000035486.14 OTTMUSG00000019121.1 <NA> 3.042446 3.968115 15.132444
## ENSMUSG00000073424.9 OTTMUSG00000035749.3 <NA> 3.450722 1.879848 6.390376
##          P.Value      adj.P.Val
## ENSMUSG0000001773.14 4.894297e-11 1.216122e-08 15.178312
## ENSMUSG00000022132.15 2.243085e-05 6.110955e-04 2.731081
## ENSMUSG00000035486.14 9.431937e-15 1.195849e-11 23.365595
## ENSMUSG00000073424.9 7.433850e-07 3.687592e-05 5.942051

#Nuevo data frame
new_data_frame <- rbind(new_data_frame, de_results[de_results$gene_name %in% c('Spp1', 'Trh', 'Lilr4b', 'Lilrb4a')])

# Ordenar los valores de acuerdo con su expresión diferencial:
new_data_frame <- new_data_frame[order(new_data_frame$logFC), ]
new_data_frame

```

	source	type	bp_length	phase	gene_id
## ENSMUSG00000029304.14	HAVANA	gene	1648	NA	ENSMUSG00000029304.14
## ENSMUSG00000029304.141	HAVANA	gene	1648	NA	ENSMUSG00000029304.14
## ENSMUSG00000005892.4	HAVANA	gene	1384	NA	ENSMUSG00000005892.4
## ENSMUSG000000112023.1	HAVANA	gene	5439	NA	ENSMUSG000000112023.1
## ENSMUSG000000112148.1	HAVANA	gene	5227	NA	ENSMUSG000000112148.1
## ENSMUSG00000024907.7	HAVANA	gene	788	NA	ENSMUSG00000024907.7
## ENSMUSG00000051111.16	HAVANA	gene	9700	NA	ENSMUSG00000051111.16
## ENSMUSG00000021478.6	HAVANA	gene	3628	NA	ENSMUSG00000021478.6
## ENSMUSG00000034245.10	HAVANA	gene	3849	NA	ENSMUSG00000034245.10
## ENSMUSG00000001773.14	HAVANA	gene	3210	NA	ENSMUSG00000001773.14
## ENSMUSG00000022132.15	HAVANA	gene	2591	NA	ENSMUSG00000022132.15
## ENSMUSG00000035486.14	HAVANA	gene	2678	NA	ENSMUSG00000035486.14
## ENSMUSG00000073424.9	HAVANA	gene	2282	NA	ENSMUSG00000073424.9
	gene_type	gene_name	level	mgi_id	
## ENSMUSG00000029304.14	protein_coding	Spp1	2	MGI:98389	
## ENSMUSG00000029304.141	protein_coding	Spp1	2	MGI:98389	
## ENSMUSG00000005892.4	protein_coding	Trh	2	MGI:98823	
## ENSMUSG000000112023.1	protein_coding	Lilr4b	2	MGI:102702	
## ENSMUSG000000112148.1	protein_coding	Lilrb4a	2	MGI:102701	
## ENSMUSG00000024907.7	protein_coding	Gal	2	MGI:95637	
## ENSMUSG00000051111.16	protein_coding	Sv2c	2	MGI:1922459	
## ENSMUSG00000021478.6	protein_coding	Drd1	2	MGI:99578	
## ENSMUSG00000034245.10	protein_coding	Hdac11	2	MGI:2385252	
## ENSMUSG00000001773.14	protein_coding	Folh1	2	MGI:1858193	
## ENSMUSG00000022132.15	protein_coding	Cldn10	2	MGI:1913101	
## ENSMUSG00000035486.14	protein_coding	Plk5	2	MGI:3026984	
## ENSMUSG00000073424.9	protein_coding	Cyp4f15	2	MGI:2146921	
	havana_gene	tag	logFC		
## ENSMUSG00000029304.14	OTTMUSG00000024389.2	<NA>	-5.7014831		
## ENSMUSG00000029304.141	OTTMUSG00000024389.2	<NA>	-5.7014831		
## ENSMUSG00000005892.4	OTTMUSG00000024415.1	<NA>	-5.5600446		
## ENSMUSG000000112023.1	OTTMUSG00000016502.6	overlapping_locus	-5.0588322		
## ENSMUSG000000112148.1	OTTMUSG00000063893.5	<NA>	-4.7756783		
## ENSMUSG00000024907.7	OTTMUSG00000073812.1	<NA>	-3.5538469		
## ENSMUSG00000051111.16	OTTMUSG00000035052.4	overlapping_locus	-3.0789058		
## ENSMUSG00000021478.6	OTTMUSG00000067515.1	<NA>	-1.8488683		

```

## ENSMUSG00000034245.10 OTTMUSG00000017285.4 <NA> 0.6899831
## ENSMUSG0000001773.14 OTTMUSG00000037777.2 <NA> 2.6649405
## ENSMUSG00000022132.15 OTTMUSG00000023307.3 <NA> 2.9022308
## ENSMUSG00000035486.14 OTTMUSG00000019121.1 <NA> 3.0424465
## ENSMUSG00000073424.9 OTTMUSG00000035749.3 <NA> 3.4507222
##          AveExpr      t    P.Value   adj.P.Val       B
## ENSMUSG00000029304.14 3.1255768 -13.612553 1.207208e-13 9.166402e-11 20.954115
## ENSMUSG00000029304.141 3.1255768 -13.612553 1.207208e-13 9.166402e-11 20.954115
## ENSMUSG00000005892.4  2.0699021 -13.462984 1.569585e-13 1.124801e-10 20.095402
## ENSMUSG000000112023.1 2.7687486 -13.331725 1.979776e-13 1.313323e-10 20.455518
## ENSMUSG000000112148.1 3.4767741 -14.544646 2.466394e-14 2.594803e-11 22.648196
## ENSMUSG00000024907.7  0.2442884 -11.578616 5.223026e-12 1.897759e-09 16.420331
## ENSMUSG00000051111.16 4.5810680 -16.705312 8.285062e-16 1.781180e-12 25.987226
## ENSMUSG00000021478.6  3.9746343 -12.504383 8.904495e-13 4.403005e-10 19.220141
## ENSMUSG00000034245.10 5.2091069  4.828541 4.769245e-05 1.145337e-03 1.562235
## ENSMUSG00000001773.14 2.8432269  10.475187 4.894297e-11 1.216122e-08 15.178312
## ENSMUSG00000022132.15 2.1053776  5.108548 2.243085e-05 6.110955e-04 2.731081
## ENSMUSG000000035486.14 3.9681152  15.132444 9.431937e-15 1.195849e-11 23.365595
## ENSMUSG00000073424.9   1.8798477  6.390376 7.433850e-07 3.687592e-05 5.942051

```

Ahora para comprobar la posición en la que se encuentran los genes que están relacionados con la epilepsia, se ordenará la tabla de acuerdo a su expresión diferencial.

```

de_results <- de_results[order(de_results$logFC), ]
# Data frame ordenado
head(de_results)

##           source type bp_length phase      gene_id
## ENSMUSG00000029304.14 HAVANA gene     1648      NA ENSMUSG00000029304.14
## ENSMUSG00000005892.4 HAVANA gene     1384      NA ENSMUSG00000005892.4
## ENSMUSG000000112023.1 HAVANA gene     5439      NA ENSMUSG000000112023.1
## ENSMUSG000000112148.1 HAVANA gene     5227      NA ENSMUSG000000112148.1
## ENSMUSG00000062593.17 HAVANA gene     1990      NA ENSMUSG00000062593.17
## ENSMUSG00000001131.11 HAVANA gene     888       NA ENSMUSG00000001131.11
##          gene_type gene_name level      mgi_id
## ENSMUSG00000029304.14 protein_coding Spp1     2 MGI:98389
## ENSMUSG00000005892.4 protein_coding Trh      2 MGI:98823
## ENSMUSG000000112023.1 protein_coding Lilr4b    2 MGI:102702
## ENSMUSG000000112148.1 protein_coding Lilrb4a   2 MGI:102701
## ENSMUSG00000062593.17 protein_coding Gm49339  2 MGI:6121530
## ENSMUSG00000001131.11 protein_coding Timp1    2 MGI:98752
##          havana_gene      tag      logFC
## ENSMUSG00000029304.14 OTTMUSG00000024389.2 <NA> -5.701483
## ENSMUSG00000005892.4  OTTMUSG00000024415.1 <NA> -5.560045
## ENSMUSG000000112023.1 OTTMUSG00000016502.6 overlapping_locus -5.058832
## ENSMUSG000000112148.1 OTTMUSG00000063893.5 <NA> -4.775678
## ENSMUSG00000062593.17 OTTMUSG00000016501.5 <NA> -4.715456
## ENSMUSG00000001131.11 OTTMUSG00000017184.1 <NA> -4.558336
##          AveExpr      t    P.Value   adj.P.Val       B
## ENSMUSG00000029304.14 3.1255768 -13.61255 1.207208e-13 9.166402e-11 20.95411
## ENSMUSG00000005892.4  2.0699021 -13.46298 1.569585e-13 1.124801e-10 20.09540
## ENSMUSG000000112023.1 2.7687486 -13.33173 1.979776e-13 1.313323e-10 20.45552
## ENSMUSG000000112148.1 3.4767741 -14.54465 2.466394e-14 2.594803e-11 22.64820

```

```

## ENSMUSG00000062593.17 2.07463215 -13.17397 2.622919e-13 1.601179e-10 20.11330
## ENSMUSG00000001131.11 -0.05339836 -12.33644 1.218858e-12 5.795084e-10 17.86926

# Comprobación de la posición que tiene cada gen en la tabla principal.
# Gene Hdac11
which(de_results$gene_name=='Hdac11')

## [1] 46259

# Gene Spp1
which(de_results$gene_name=='Spp1')

## [1] 1

# Gene Gal
which(de_results$gene_name=='Gal')

## [1] 27

# Gene Drd1
which(de_results$gene_name=='Drd1')

## [1] 298

# Gene Sv2c
which(de_results$gene_name=='Sv2c')

## [1] 66

```

Estos últimos resultados demuestran que aunque no fueron los genes que mostraron mayor expresión, si tuvieron una diferencia. Y que además, el PCA que de la figura 9 que ambos tipos de marcadores mostraron expresión diferencial, es decir, tanto neuronas como glía, además, como se menciona en el artículo, su propósito se cumplió. Se demostró con los datos que se encuentra una metilación diferencial.

Por último, para ir un poco más allá en el análisis, se realizó un análisis de enriquecimiento con relación a la ontología génica (GO) de NCBI. De esta manera analizar de una manera más detallada los genes y su función biológica según su pvalue ajustado.

```

# Correr las siguientes líneas en caso de que no se encuentre instalado el paquete de GEOquery
#if (!require("BiocManager", quietly = TRUE))
#  install.packages("BiocManager")

#BiocManager::install("GEOquery")

#if (!requireNamespace("BiocManager", quietly = TRUE)) {
#  install.packages("BiocManager")
#}

```

```

#BiocManager::install("clusterProfiler")

# Cargar el paquete:
library(GEOquery)

## Setting options('download.file.method.GEOquery'='auto')

## Setting options('GEOquery.inmemory.gpl'=FALSE)

library(clusterProfiler)

## 

## clusterProfiler v4.14.4 Learn more at https://yulab-smu.top/contribution-knowledge-mining/
## 
## Please cite:
## 
## Guangchuang Yu, Li-Gen Wang, Yanyan Han and Qing-Yu He.
## clusterProfiler: an R package for comparing biological themes among
## gene clusters. OMICS: A Journal of Integrative Biology. 2012,
## 16(5):284-287

## 
## Adjuntando el paquete: 'clusterProfiler'

## The following object is masked from 'package:IRanges':
## 
##     slice

## The following object is masked from 'package:S4Vectors':
## 
##     rename

## The following object is masked from 'package:stats':
## 
##     filter

library(org.Mm.eg.db)

## Cargando paquete requerido: AnnotationDbi

## 
## Adjuntando el paquete: 'AnnotationDbi'

## The following object is masked from 'package:clusterProfiler':
## 
##     select

##

```

```

# Para hacer un análisis de enriquecimiento
ego <- enrichGO(gene = de_results$gene_name,
                 OrgDb = org.Mm.eg.db,
                 keyType = "SYMBOL", # Corregido: "keyType" en minúsculas
# Queremos analizar los procesos biológicos.
                 ont = "BP",
                 pAdjustMethod = "BH",
                 pvalueCutoff = 0.05)

# Para lo que es necesario cambiar el formato en que se encuentran los identificadores que regresa recor
# Se eliminan los puntos decimales para evitar problemas en el análisis.
ensembl_ids <- rownames(assay(rse_gene_SRP223512))
ensembl_ids_clean <- sub("\\.*", "", ensembl_ids)

library(AnnotationDbi)

# Mapear identificadores
gene_symbols <- mapIds(
  org.Mm.eg.db,
  keys = ensembl_ids_clean,
  column = 'SYMBOL',
  keytype = 'ENSEMBL',
  multiVals = 'first'
)

## 'select()' returned 1:many mapping between keys and columns

# Pequeño filtrado para evitar problemas posteriores.
important_genes <- na.omit(gene_symbols)

# Una vez más se hace el análisis de enriquecimiento pero con datos limpios y filtrados, se conserva el
ego <- enrichGO(
  gene=important_genes,
  OrgDb = org.Mm.eg.db,
  keyType = 'SYMBOL',
  ont = 'BP',
  pAdjustMethod = 'BH',
  pvalueCutoff = 0.05
)

# Se muestra el dotplot para observar los resultados.
dotplot(ego, showCategory=20)

# Se cambian los ids que habíamos estado trabajando a ids de entrez.
genes_entrez <- bitr(important_genes, fromType = 'SYMBOL', toType = "ENTREZID", OrgDb = org.Mm.eg.db)

## 'select()' returned 1:1 mapping between keys and columns

# Enriquecimiento de rutas metabólicas.
kegg <- enrichKEGG(gene = genes_entrez$ENTREZID, organism = 'mmu')

## Reading KEGG annotation online: "https://rest.kegg.jp/link/mmu/pathway"...

```

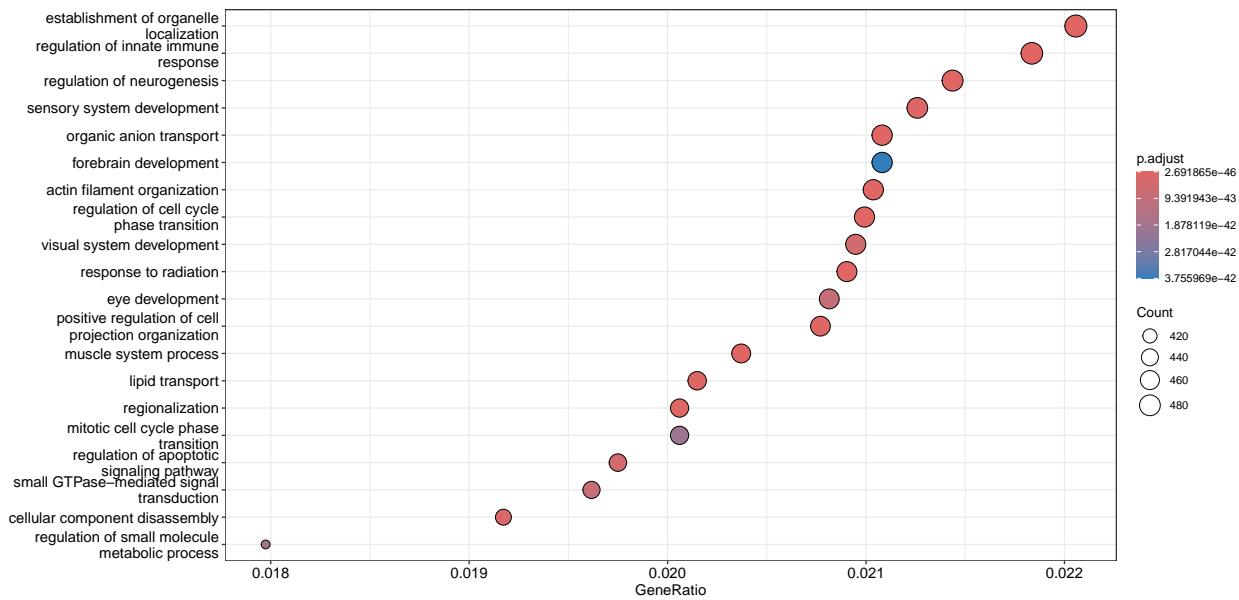


Figure 11: Enriquecimiento para analizar procesos biológicos

```
## Reading KEGG annotation online: "https://rest.kegg.jp/list/pathway/mmu"...
```

```
dotplot(kegg, showCategory=15)
```

Conclusión:

Como se mencionó con anterioridad visualmente con los PCA, se presentó expresión diferencial tanto en neuronas, positivas para el marcador NeuN y glía, negativas para el marcador. Al mismo tiempo, el PCA donde se encuentran las muestras de placebo y epilepsia, demuestra que las muestras sin importar si se encuentran en el grupo control (placebo) o experimental (epilepsia) no mostraron diferencias entre los grupos con respecto a su expresión. Salvo en las primeras dimensiones. El resto y la mezcla entre las muestras puede deberse a que no hay cambios drásticos en la expresión. Es decir, que existe mucha variabilidad entre las muestras. Sin mencionar que en el grupo control, al que no se indujo a la epilepsia existían tanto células neuronales como células de la glía, mismo caso ocurrió con las muestras experimentales, por lo que no se ve clara la separación entre ambos grupos.

Con todo esto podemos concluir que efectivamente existe un cambio en la expresión tanto en glía como en neuronas. Por otro lado, existen genes que no están directamente relacionados con la epilepsia, pero que mostraron niveles altos de expresión y aunque no se compara mucho la expresión con los directamente relacionados, si presentaron expresión diferencial.

Referencias:

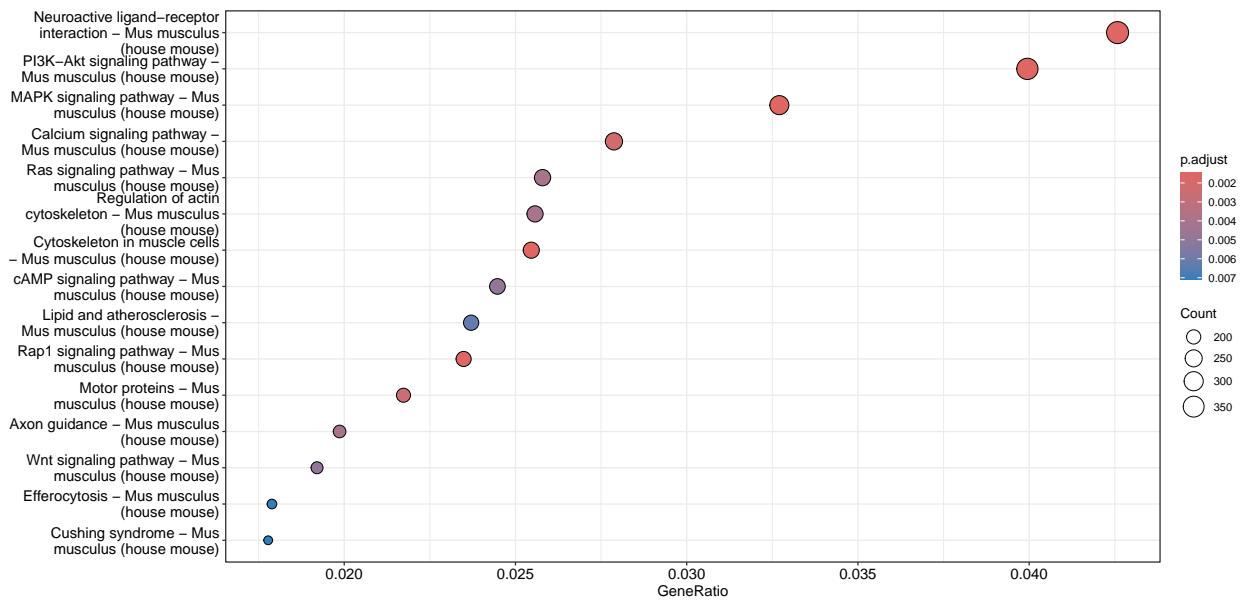


Figure 12: Enriquecimiento de acuerdo a las rutas metabólicas