

# PEC2-Análisis de Datos de Ultrasecuenciación

## Análisis de Datos Ómicos

Alba Moya Garcés

14 de junio, 2020

## Contents

<b>1</b>	<b>Resumen</b>	<b>2</b>
<b>2</b>	<b>Objetivos</b>	<b>2</b>
<b>3</b>	<b>Material</b>	<b>2</b>
3.1	Software . . . . .	2
3.2	Datos . . . . .	2
<b>4</b>	<b>Métodos</b>	<b>3</b>
4.1	Preparación del área de trabajo: . . . . .	3
4.2	Instalación de paquetes en R . . . . .	3
4.3	Lectura y selección de los datos . . . . .	4
4.4	Objeto DESeqDataSet . . . . .	5
4.5	Transformación de los datos . . . . .	6
4.6	Control de calidad de los datos . . . . .	6
4.6.1	Gráfico de densidad . . . . .	6
4.6.1.1	Boxplot . . . . .	9
4.6.2	Análisis de Componentes Principales . . . . .	9
4.6.3	Clúster Jerárquico . . . . .	9
4.7	Filtraje no específico de datos . . . . .	9
4.8	Análisis de Expresión Diferencial . . . . .	14
4.8.1	Genes diferenciados NIT vs. ELI . . . . .	14
4.8.1.1	Anotación de los genes expresados genéticamente . . . . .	16
4.8.1.2	Análisis de Significación Biológica . . . . .	16
4.8.2	Genes diferenciados NIT vs. SFI . . . . .	17
4.8.2.1	Anotación de los genes expresados genéticamente . . . . .	18

4.8.2.2	Análisis de Significación Biológica . . . . .	19
4.8.3	Genes diferenciados ELI vs. SFI . . . . .	20
4.8.3.1	Anotación de los genes expresados genéticamente . . . . .	22
4.8.3.2	Análisis de Significación Biológica . . . . .	22
<b>5</b>	<b>Resumen de resultados</b>	<b>25</b>
	<b>Bibliografía</b>	<b>26</b>

## 1 Resumen

Se analizaron los datos de expresión RNAseq provenientes de un análisis del tiroides donde se comparan tres tipos de infiltración en un total de 292 muestras.

## 2 Objetivos

El objetivo de este estudio es examinar las diferencias en la expresión génica de las muestras y analizar su significancia biológica.

## 3 Material

El código completo para desarrollar este análisis, o cualquier otro a partir de su adaptación, así como todos los archivos de resultados generados, puede descargarse del siguiente repositorio de *GitHub*:

<https://github.com/albamgarces/analisis-de-datos-de-RNA-seq.git>.

### 3.1 Software

Se realizó este análisis utilizando el lenguaje R version 3.6.3 (2020-02-29) R en la interfaz RStudio versión 1.1.456 y las librerías desarrolladas para este tipo de análisis por el proyecto Bioconductor. El programa estadístico R se puede descargar desde la página web del [proyecto CRAN](#) (The Comprehensive R Archive Network) siguiendo las indicaciones. R-Studio puede descargarse desde su página web <https://www.rstudio.com/>.

Finalmente, las librerías adicionales necesarias para llevar a cabo este análisis se obtuvieron del proyecto Bioconductor versión 3.10, el cuál se instala junto con algunos paquetes básicos mediante el siguiente código:

```
if (!requireNamespace("BiocManager", quietly = TRUE))
  install.packages("BiocManager")
BiocManager::install()
```

### 3.2 Datos

Los archivos `counts.csv` y `targets.csv` contienen la información de un estudio obtenido del repositorio del proyecto [GTEx](#) (Genotype-Tissue Expression). Encontramos los datos de expresión (RNA-seq) pertenecientes a un análisis del tiroide donde se comparan tres tipos de infiltración en 292 muestras:

- tejidos no infiltrados (NIT): 236 muestras
- infiltración focalizada (SFI): 42 muestras
- infiltración linfoide extensiva (ELI): 14 muestras

## 4 Métodos

### 4.1 Preparación del área de trabajo:

Para llevar a cabo el análisis, se debe gestionar una gran cantidad de archivos entre aquellos que ocupan los datos originales y los generados durante su análisis. Es por ello que siempre se debería comenzar creando las carpetas necesarias para simplificar la ruta de trabajo. Se recomienda generar una **carpeta principal** con el nombre de nuestro proyecto en cuyo interior alojaremos una carpeta con los archivos de **datos** y otra con los **resultados** generados del análisis

Estas carpetas las genereamos rápidamente desde el explorador de archivos o la consola de cualquiera de los sistemas operativos usuales. Desde R también podemos generar estas subcarpetas mediante el siguiente código:

```
setwd(".")
dir.create("data")
dir.create("results")
```

### 4.2 Instalación de paquetes en R

A continuación se muestran los paquetes necesarios para este estudio que requieren instalación:

```
#### UNCOMMENT IF INSTALL REQUIRES ####
##install.packages("readr")
##install.packages("sampling")
## install.packages("knitr")
##install.packages("tables")
#install.packages("cluster")
## install.packages("gplots")
## install.packages("ggplot2")
##install.packages("enrichR")
# install.packages("ggrepel")
## install.packages("BiocManager")
##BiocManager::install("Biobase")
##BiocManager::install("DESeq")
##BiocManager::install("mixOmics")
##BiocManager::install("AnnotationDbi")
##BiocManager::install("org.Hs.eg.db")
# BiocManager::install("oligo")
# BiocManager::install("arrayQualityMetrics")
# BiocManager::install("pvca")
# BiocManager::install("pacman")
# BiocManager::install("geneplotter")
# BiocManager::install("org.Dm.eg.db")
# BiocManager::install("limma")
# BiocManager::install("genefilter")
# BiocManager::install("drosophila2.db")
# BiocManager::install("ReactomePA")
```

### 4.3 Lectura y selección de los datos

Importamos los archivos proporcionados a R. El archivo `targets` contiene las 292 muestras identificadas por un número según sean provenientes de tejidos NIT (1), SFI (2) o ELI (3).

Table 1: Fragmento de la tabla de datos `targets`

Sample_Name	Grupo_analisis	molecular_data_type	sex	Group
GTEX-111CU-0226-SM-5GZXC	1	Allele-Specific Expression	male	NIT
GTEX-111FC-1026-SM-5GZX1	1	RNA Seq (NGS)	male	NIT
GTEX-111VG-0526-SM-5N9BW	3	RNA Seq (NGS)	male	ELI
GTEX-111YS-0726-SM-5GZY8	1	Allele-Specific Expression	male	NIT
GTEX-1122O-0226-SM-5N9DA	1	RNA Seq (NGS)	female	NIT
GTEX-1128S-0126-SM-5H12S	1	Allele-Specific Expression	female	NIT

El archivo `counts` contempla las 292 muestras como variables y nos informa del número de veces que se ha detectado cada uno de los 56202 genes identificados en la primera columna.

Table 2: Fragmento de la tabla de datos `count`

	GTEX.111CU.0226.SM.5GZXC	GTEX.111FC.1026.SM.5GZX1
ENSG00000223972.4	7	0
ENSG00000227232.4	401	1064
ENSG00000243485.2	4	0
ENSG00000237613.2	2	0
ENSG00000268020.2	0	0
ENSG00000240361.1	0	1

Con el fin de simplificar el análisis, se decidió seleccionar aleatoriamente 10 muestras de cada tipo de tejido. El código para realizar esta extracción es el que se indica a continuación.

```
#selección de muestras en el archivo targets
set.seed(1234)
selected <- strata(targets, c("Group"), size=c(10,10,10), method = "srswor")
muestreo <- getdata(targets, selected)
muestreo <- muestreo[,1:10]
```

Table 3: Selección de diez muestras de cada tipo

	Sample_Name	molecular_data_type	sex	Group
36	GTEX-11TTK-0826-SM-5N9EG	RNA Seq (NGS)	female	NIT
107	GTEX-13O61-0226-SM-5KM52	RNA Seq (NGS)	male	NIT
129	GTEX-144GL-1226-SM-5O9A4	RNA Seq (NGS)	male	NIT
139	GTEX-14753-0926-SM-5Q5BI	RNA Seq (NGS)	male	NIT
164	GTEX-P4QS-2626-SM-2I3EV	Allele-Specific Expression	male	NIT
165	GTEX-P4QT-2626-SM-2I3FM	Allele-Specific Expression	female	NIT
172	GTEX-Q2AI-0326-SM-2I3EK	Allele-Specific Expression	male	NIT
180	GTEX-QV44-0826-SM-2S1RG	Allele-Specific Expression	male	NIT
190	GTEX-RNOR-0926-SM-2TF56	RNA Seq (NGS)	female	NIT
209	GTEX-T8EM-0226-SM-3DB7C	RNA Seq (NGS)	male	NIT

	Sample_Name	molecular_data_type	sex	Group
29	GTEX-11NV4-0626-SM-5N9BR	RNA Seq (NGS)	male	ELI
100	GTEX-13NZ9-1126-SM-5MR37	RNA Seq (NGS)	male	ELI
146	GTEX-14ABY-0926-SM-5Q5DY	Allele-Specific Expression	male	ELI
147	GTEX-14AS3-0226-SM-5Q5B6	RNA Seq (NGS)	female	ELI
149	GTEX-14BMU-0226-SM-5S2QA	Allele-Specific Expression	female	ELI
167	GTEX-PLZ4-1226-SM-2I5FE	RNA Seq (NGS)	female	ELI
186	GTEX-R55G-0726-SM-2TC6J	RNA Seq (NGS)	female	ELI
211	GTEX-TMMY-0826-SM-33HB9	Allele-Specific Expression	female	ELI
251	GTEX-YFC4-2626-SM-5P9FQ	Allele-Specific Expression	female	ELI
253	GTEX-YJ89-0726-SM-5P9F7	RNA Seq (NGS)	male	ELI
14	GTEX-11DXY-0426-SM-5H12R	RNA Seq (NGS)	male	SFI
21	GTEX-11EQ8-0826-SM-5N9FG	Allele-Specific Expression	male	SFI
22	GTEX-11EQ9-0626-SM-5A5K1	RNA Seq (NGS)	male	SFI
23	GTEX-11GS4-0826-SM-5986J	RNA Seq (NGS)	male	SFI
90	GTEX-13FXS-0726-SM-5LZXJ	RNA Seq (NGS)	male	SFI
98	GTEX-13NYC-2426-SM-5MR3K	RNA Seq (NGS)	male	SFI
185	GTEX-R55E-0826-SM-2TC5M	Allele-Specific Expression	male	SFI
199	GTEX-S341-0226-SM-5S2VG	RNA Seq (NGS)	female	SFI
224	GTEX-WYVS-0326-SM-3NM9V	RNA Seq (NGS)	female	SFI
261	GTEX-ZE7O-1126-SM-57WC8	Allele-Specific Expression	female	SFI

```
#selección de muestras en el archivo counts
count_subset <- counts[, c(muestreo$ID_unit)]
```

Table 4: Fragmento tabla counts de muestras seleccionadas

	NIT1	NIT2	NIT3
ENSG00000223972.4	3	2	3
ENSG00000227232.4	1241	624	913
ENSG00000243485.2	1	1	1
ENSG00000237613.2	2	1	0
ENSG00000268020.2	0	2	1

Con los datos de conteo de estas 30 muestras seleccionadas continuaremos el resto del análisis.

#### 4.4 Objeto DESeqDataSet

Preparamos el objeto DESeqDataSet necesario para parte del análisis de los datos. Creamos la tabla de metadatos del experimento que nos permitirá trabajar con las funciones de Bioconductor más fácilmente.

```
#eliminamos el número de versión de los genes
rownames(count_subset) <- gsub("\\.[0-9]*$", "", rownames(count_subset))
```

```
samples <- colnames(count_subset)
ID_unit <- selected$ID_unit
Sex <- muestrero$sex
Group <- selected$Group
experiment.metadata <- data.frame(ID_unit, Group, Sex)
```

```
experiment.metadata$Group<-as.factor(experiment.metadata$Group)
rownames(experiment.metadata)<-samples
```

```
dds<-DESeqDataSetFromMatrix(countData = count_subset,
                             colData = muestreo,
                             design = ~ Group)
dds
```

```
FALSE class: DESeqDataSet
FALSE dim: 56202 30
FALSE metadata(1): version
FALSE assays(1): counts
FALSE rownames(56202): ENSG00000223972 ENSG00000227232 ... ENSG00000210195
FALSE ENSG00000210196
FALSE rowData names(0):
FALSE colnames(30): NIT1 NIT2 ... SFI9 SFI10
FALSE colData names(10): Experiment SRA_Sample ... Group ID_unit
```

## 4.5 Transformación de los datos

Para poder explorar y visualizar los datos en su conjunto, es esencial tener todos los valores en una escala comparable. Transformaremos los datos a escala  $\log_2$  para normalizar las distribuciones. Se utiliza este método de transformación porque facilita su conversión a la escala original.

Tal y como recomienda González (Toulouse, 2014), sumaremos 1 a cada registro de conteo para evitar los valores 0 de algunas condiciones:

$$y = \log_2(K + 1)$$

La transformación logarítmica nos sirve para que los valores de conteo más pequeños se aproximen hacia la media de expresión de todos los genes, obteniendo unos datos transformados aproximadamente homocedásticos, con varianzas no distintas significativamente. Podemos ver el efecto de la transformación de los datos tomando de ejemplo una de las muestras. En la figura 1 observamos como los datos transformados permiten apreciar las posibles diferencias entre las muestras.

## 4.6 Control de calidad de los datos

La evaluación de la calidad de los datos es esencial para cualquier tipo de análisis de datos. Esto nos permitirá detectar errores técnicos que puedan manifestarse en la toma de datos para poder detectar los genes expresados diferencialmente.

Una primera aproximación para verificar la calidad de nuestros datos es comparar las distintas muestras entre sí, pudiendo comprobar si aquellas pertenecientes al mismo grupo presentan más similitudes entre sí que con otros grupos.

### 4.6.1 Gráfico de densidad

El gráfico de densidad de Kernel (figura 2) nos permitirá hacernos una idea de la similitud entre las diferentes distribuciones, las cuales no indican la presencia de ninguna muestra defectuosa, con la salvedad de una de las muestras ELI que parece seguir una distribución diferente a la del resto.

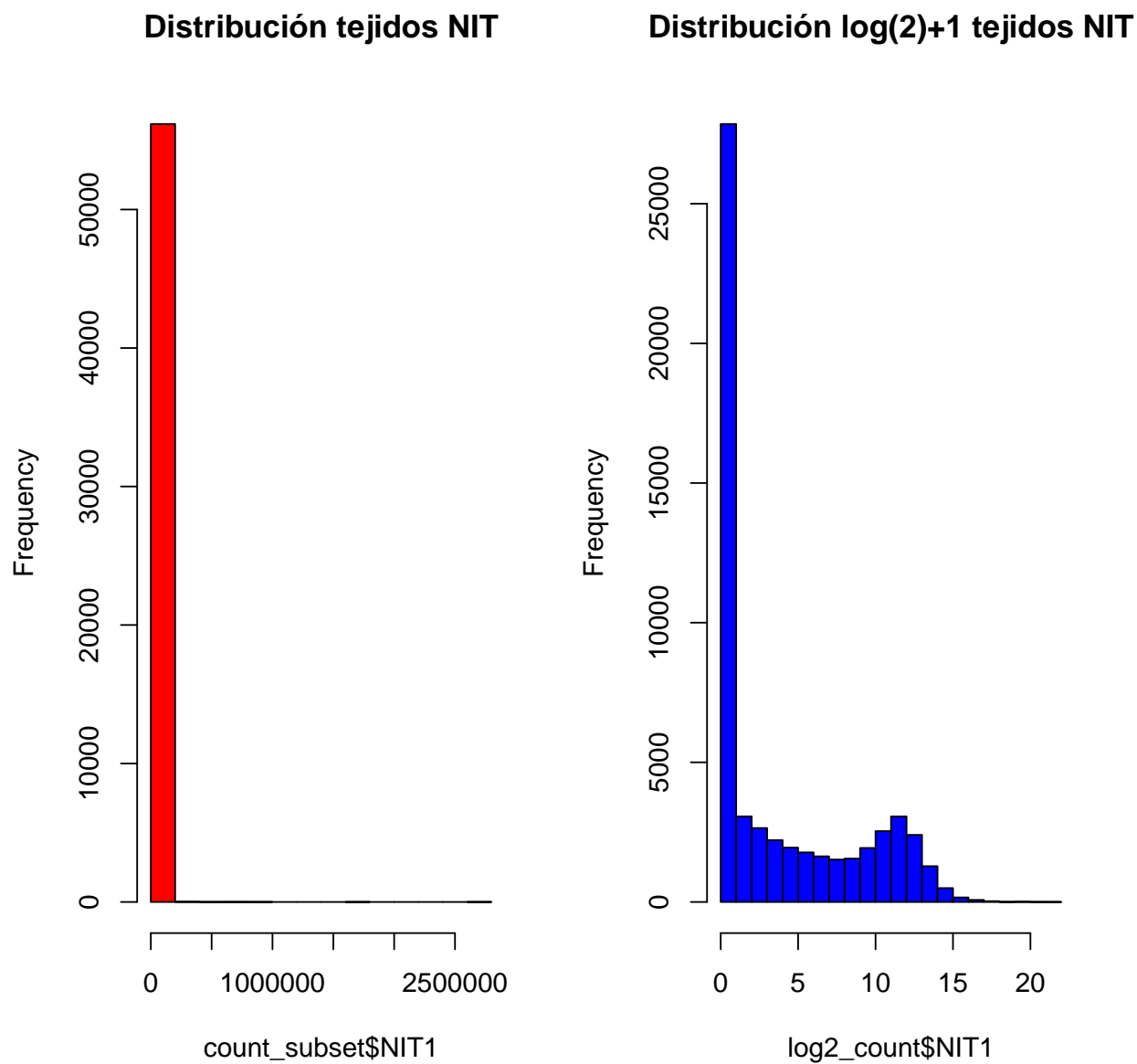


Figure 1: Distribución de las expresiones de una de las muestras de tejido NIT. Podemos ver el efecto de la transformación de los datos mediante  $\log(2)+1$

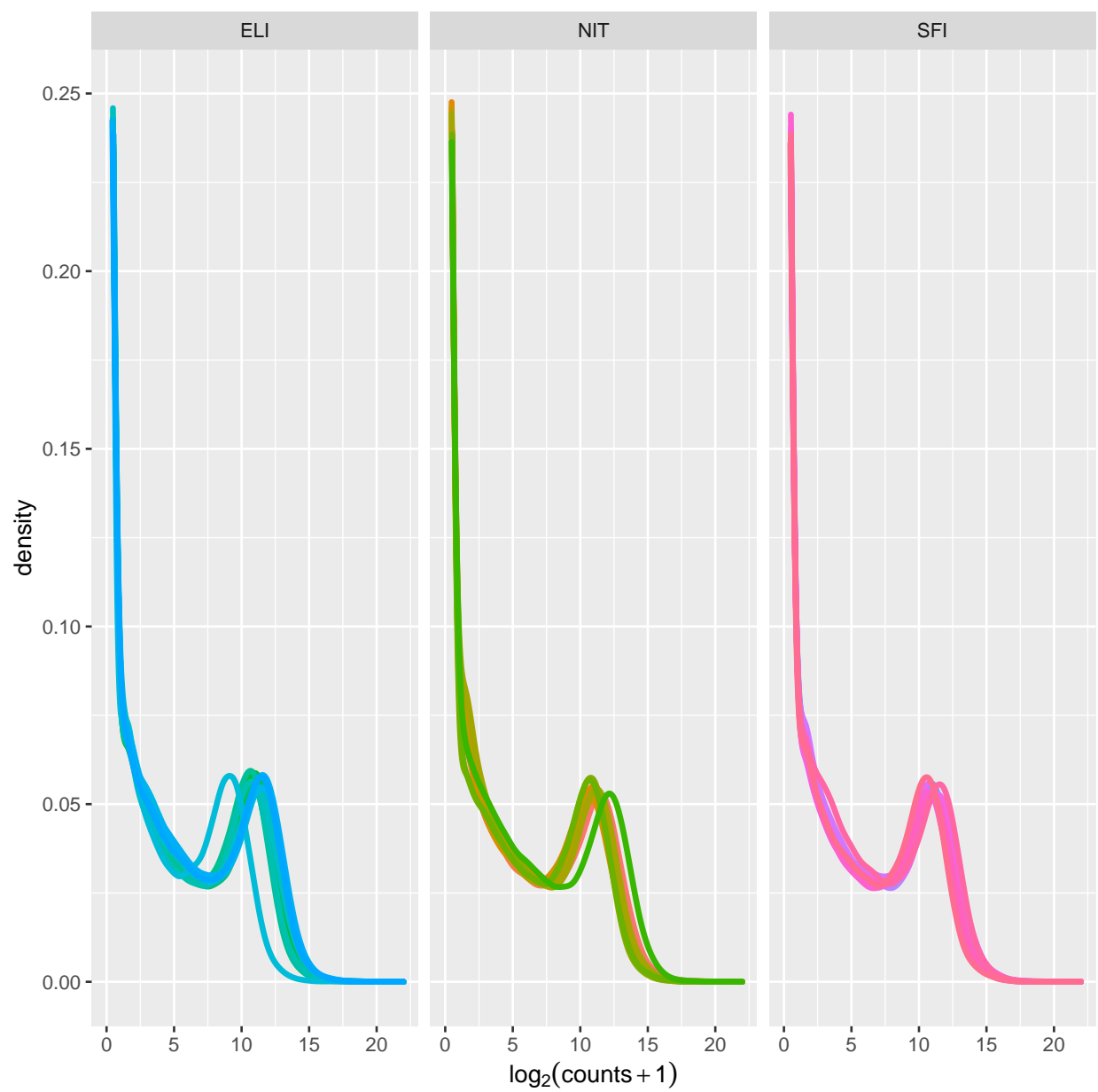


Figure 2: Curvas de densidad de la expresión de los genes de las diferentes muestras separadas por grupos.



#### 4.6.1.1 Boxplot

El diagrama de cajas también no mostrará la distribución de las intensidades, en la figura 3 se pueden apreciar pequeñas variaciones esperables en los datos normalizados. Podemos destacar la muestra ELI7 que presenta una distribución de la intensidad de expresión de los genes que parece diferente a la del resto del grupo.

#### 4.6.2 Análisis de Componentes Principales

Mediante el análisis de componentes principales (ACP) podemos detectar si las muestras se agrupan entre otras muestras del mismo grupo o si no hay una clara correspondencia entre ellas. Que las muestras no se agrupen por “familias” podría ser debido al efecto *batch* por defectos técnicos. El principal objetivo del ACP es reducir las dimensiones del conjunto de datos y poder determinar las características clave de los diferentes grupos. Teniendo en cuenta que nuestros datos están agrupados en tres grupos, correspondientes a los tres tipos de tejido, cabe esperar una clara diferenciación en las componentes.

Observando en el gráfico ACP de la figura 4, la distribución de las dos primeras componentes de la expresión de cada gen sobre cada muestra no permite observar una clara separación en el espacio bidimensional definido por las dos primeras componentes de las muestras en los tres grupos.

Hay que tener en cuenta que estas dos componentes principales no llegan a explicar el 70% de la variabilidad, pero aún así podemos apreciar que las muestras ELI se comportan de forma diferente a las SFI o NIT, salvo alguna muestra que se sale del patrón (ELI6, SFI5 y SFI9). Esto no tiene por qué interpretarse como un problema en la calidad de los datos.

#### 4.6.3 Clúster Jerárquico

Otra forma de asegurarnos que las muestra se agrupan según los grupos experimentales, es mediante un clúster jerárquico que nos agrupa las muestras por grado de similaridad. Mediante un dendograma, se va mostrando a qué valor se produce la unión de los grupos y qué grupos se unen. Si las muestras se unen por condiciones experimentales, es un buen indicador de la calidad de las mismas, si no, no es necesariamente reflejo de algún problema.

En el caso de los datos del experimento, en la figura 5 encontramos una tendenciá de las muestras a agruparse y situarse cercanas a muestras de us mismo grupo, pero hay algunos caso en los que no es así.

### 4.7 Filtraje no específico de datos

El filtraje no específico nos permite realizar una criba de genes con poca variabilidad entre las condiciones o con otra característica que lleve a que nos interese eliminarlos.

Una forma de realizar este filtrado es eliminando aquellos con un recuento total de lecturas menor a un umbral dado.

Reducimos a 46707 genes disponibles para analizar.

```
FALSE class: DESeqDataSet
FALSE dim: 46707 30
FALSE metadata(1): version
FALSE assays(1): counts
FALSE rownames(46707): ENSG00000223972 ENSG00000227232 ... ENSG00000210195
FALSE ENSG00000210196
FALSE rowData names(0):
FALSE colnames(30): NIT1 NIT2 ... SFI9 SFI10
FALSE colData names(10): Experiment SRA_Sample ... Group ID_unit
```

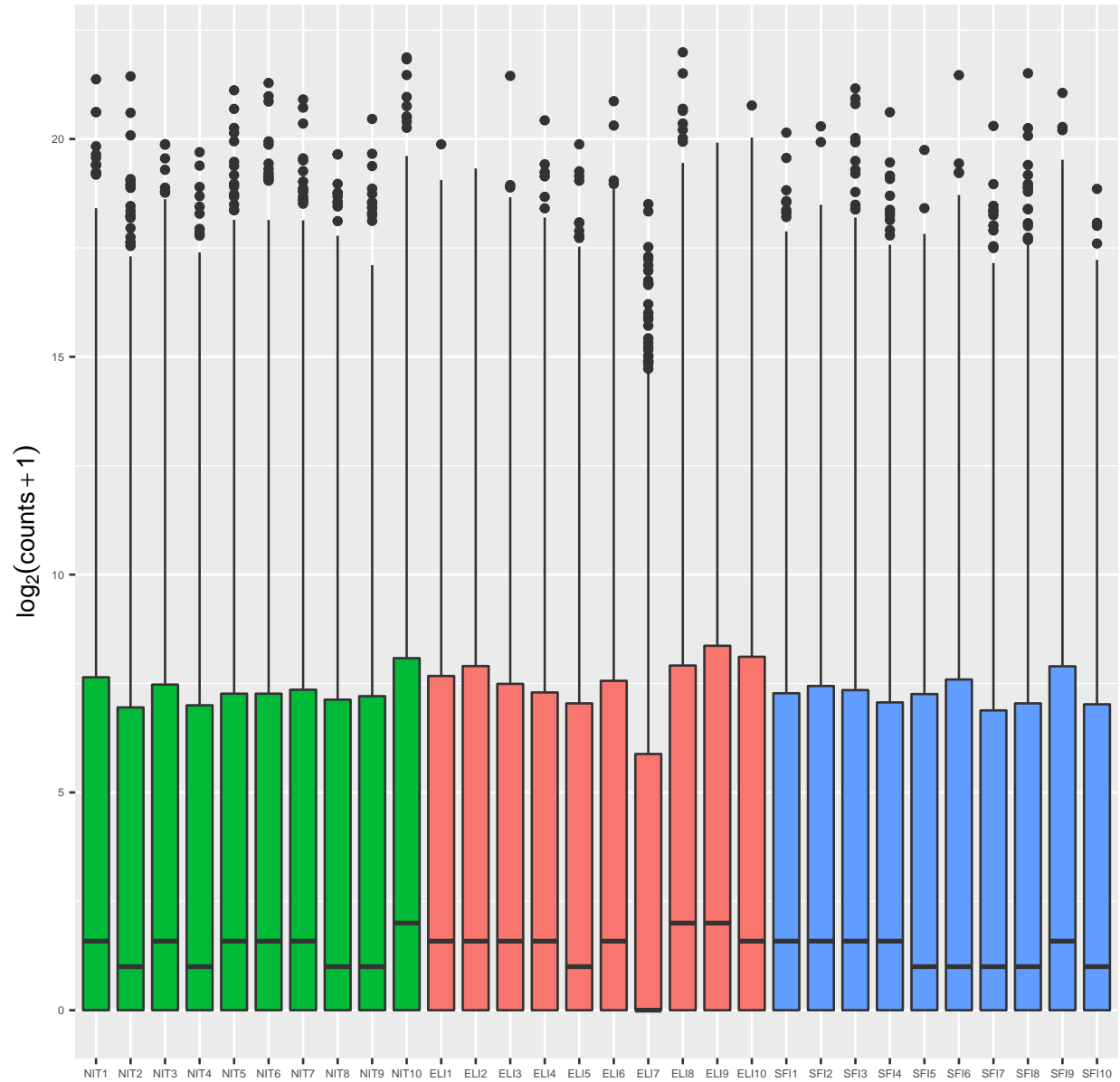


Figure 3: Diagrama de cajas que muestran la distribución de los diferentes tipos de tejido. Se puede observar cierta heterogeneidad, sobretodo en la muestra 7 del tipo tisular NIT, pero no parece que sea sistemática ni que sea producido por muestras defectuosas.

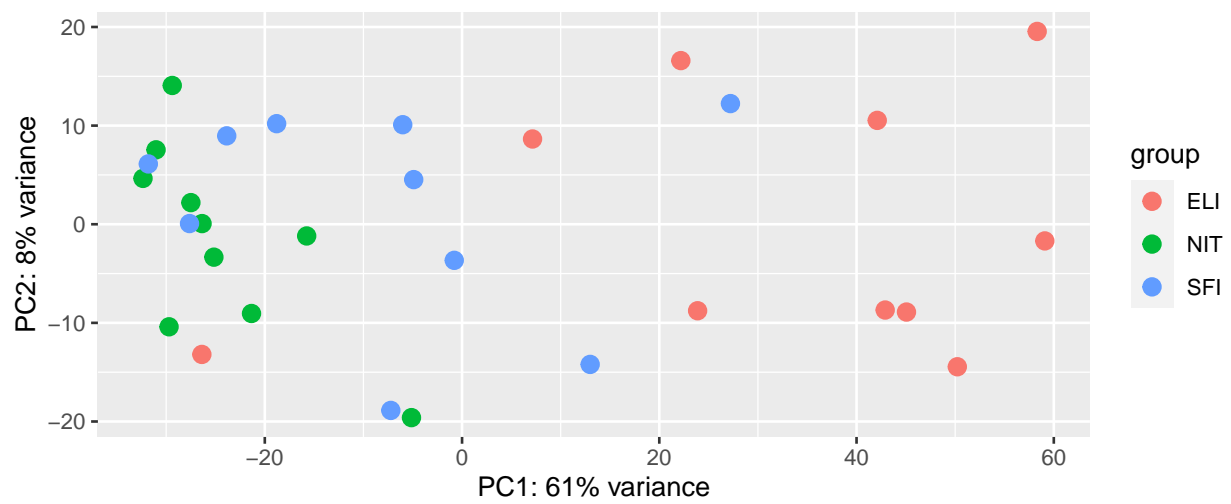


Figure 4: Dos primeras componentes principales de los datos transformados utilizando como variables la expresión de los genes en las muestras

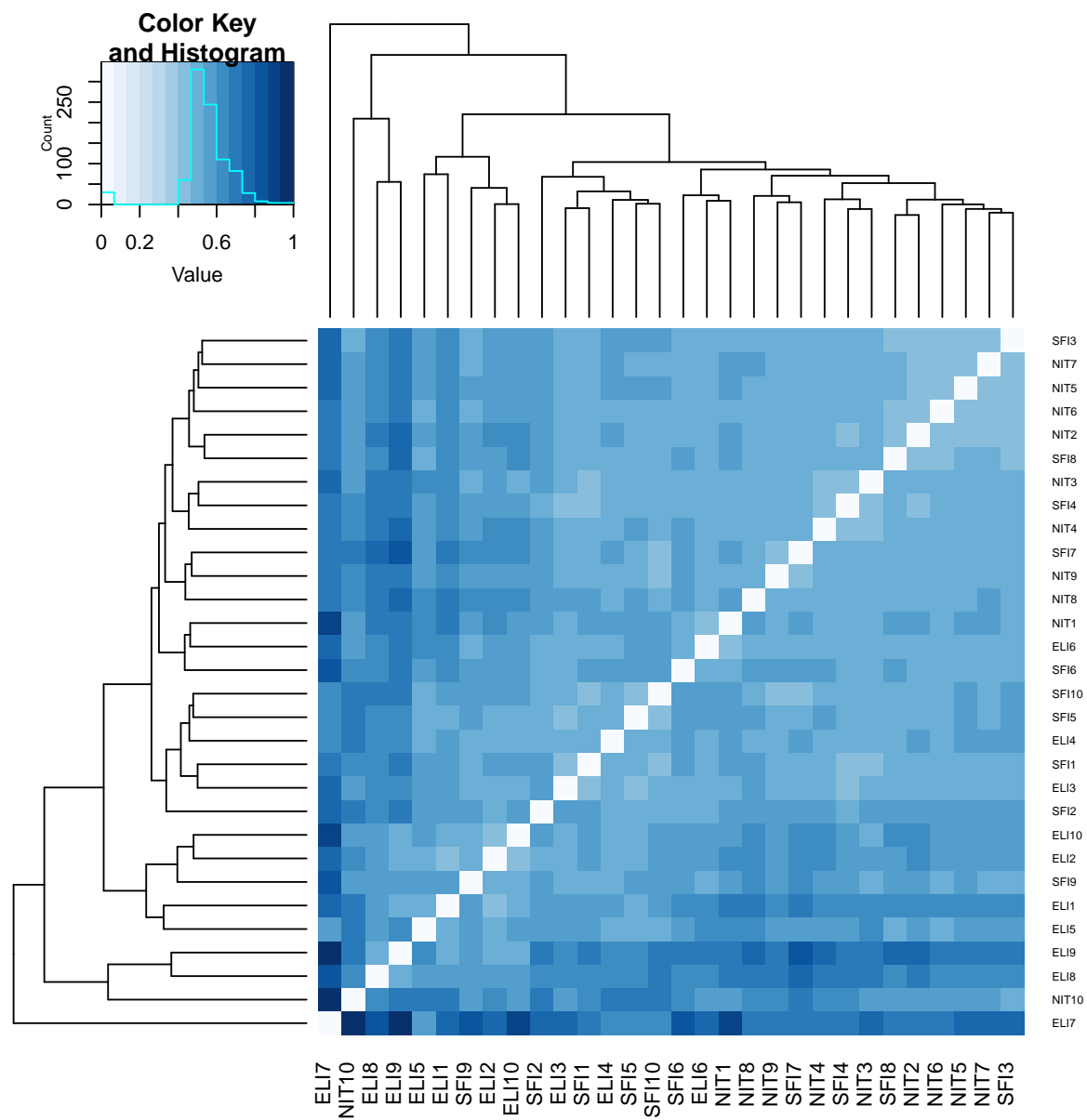


Figure 5: Mapa de colores (heatmap) y dendrograma de las distancias euclídeas entre las distintas muestras.

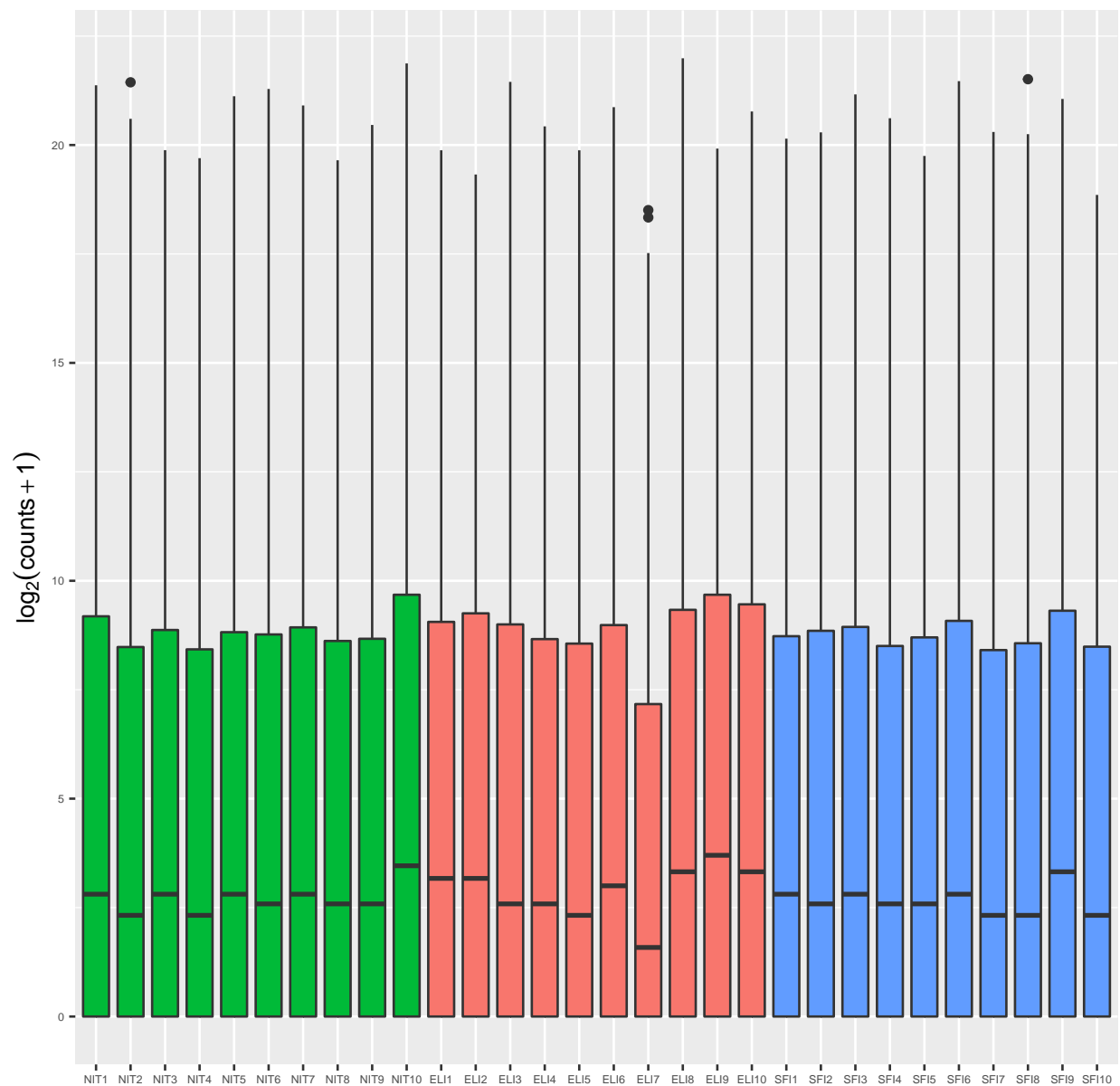


Figure 6: Diagrama de cajas de la distribución de los datos tras el filtrado de genes con baja expresión. Podemos ver que apenas varía el patrón, aunque sí que mejoran las medias de aquellas muestras conflictivas antes del filtrado.

## 4.8 Análisis de Expresión Diferencial

La función `DESeq2`, además de filtrar los genes cuya expresión no se ha detectado, nos permite también eliminar del análisis aquellos genes con una expresión especialmente baja, lo que nos facilita la tarea a la hora de realizar las comparaciones entre los grupos.

```
dds_filt <- DESeq(dds_filt, parallel = TRUE)
#dds_filt
```

Aplicando la función `results()` al objeto `DESeqDataSet` se obtienen los resultados en base a la etiqueta proporcionada, entre ellos el logaritmo en base dos y el p-valor de la comparación entre los grupos a partir del test de Wald. También calcula el error estándar del log2 (`lfcSE`) y `baseMean`, que es la media de los valores de contejo de cada gen normalizados (`baseMean`).

En el caso de que no haya contejo para algún gen, no se computa ninguna de las variables mencionadas anteriormente y aparece como valor no disponible (NA).

Debido al gran número de genes en el análisis, `DESeq2` ejecuta un ajuste del p-valor por el método Benjamin-Hochberg en el que el umbral de este p-valor ajustado hace referencia al porcentaje de falsos negativos positivos permitidos en el total de genes seleccionados como diferencialmente expresados. Por tanto, para un p-valor ajustado de 0.1 se espera encontrar hasta un 10% de falsos positivos en los genes seleccionados.

### 4.8.1 Genes diferenciados NIT vs. ELI

```
NITvsELI <- results(dds_filt, contrast = c("Group", "NIT", "ELI"))
summary(NITvsELI)
```

```
FALSE
FALSE out of 46704 with nonzero total read count
FALSE adjusted p-value < 0.1
FALSE LFC > 0 (up)      : 894, 1.9%
FALSE LFC < 0 (down)    : 3107, 6.7%
FALSE outliers [1]      : 0, 0%
FALSE low counts [2]    : 14491, 31%
FALSE (mean count < 1)
FALSE [1] see 'cooksCutoff' argument of ?results
FALSE [2] see 'independentFiltering' argument of ?results
```

Table 5: Cuadro resumen comparación genes diferencialmente expresados entre grupo NIT y ELI. La primera fila indica el número total de genes con expresión diferencial (p-valor ajustado <0.1); la segunda fila indica el número de genes con un p-valor inferior a 0.01; las dos siguientes filas corresponden a los genes significantes con una fuerte sub y sobre expresión, respectivamente; y la última fila, el número máximo de falsos positivos que podemos encontrar entre la selección de genes.

NITvsELI	
Genes expr diferencial	4001
p-valor<0.01	3759
Sub-expresados	3107
Sobre-expresados	894
max falsos positivos	401

Aparte de los p-valor ajustados, se tuvo en cuenta el log2FoldChange para evaluar la sub o sobre expresión de los genes diferencialmente expresados. Según se definieron los niveles del factor grupo, contrastando NIT frente a ELI, los genes que presentan un log2FoldChange negativo están subexpresados en el grupo NIT con respecto al de ELI y sobreexpresados en caso de ser positivos.

A continuación se muestra los seis primeros genes diferencialmente sub y sobre-expresados con p-valor más pequeño.

#### FALSE Genes sub-expresados

FALSE log2 fold change (MLE): Group NIT vs ELI

FALSE Wald test p-value: Group NIT vs ELI

FALSE DataFrame with 6 rows and 6 columns

FALSE		baseMean	log2FoldChange	lfcSE	
FALSE		<numeric>	<numeric>	<numeric>	
FALSE	ENSG00000110777	1991.25426316944	-5.58477787402009	0.540187109355479	
FALSE	ENSG00000230006	839.494909820157	-4.95157745790839	0.513984147918807	
FALSE	ENSG00000137265	1004.9044570547	-5.06932413820878	0.536827165851127	
FALSE	ENSG00000156738	5226.53051460219	-7.42853733796805	0.792444488575697	
FALSE	ENSG00000177455	1253.05359529907	-7.58056514972693	0.828192669917017	
FALSE	ENSG00000211640	1071.51473491082	-8.10581679435911	0.888630991507211	
FALSE		stat	pvalue	padj	
FALSE		<numeric>	<numeric>	<numeric>	
FALSE	ENSG00000110777	-10.3385989359937	4.71378536144042e-25	1.51859309204165e-20	
FALSE	ENSG00000230006	-9.63371628864044	5.76075125985982e-22	6.18627875292147e-18	
FALSE	ENSG00000137265	-9.44312147499371	3.6183621183966e-21	2.91422885015662e-17	
FALSE	ENSG00000156738	-9.37420531666483	6.96988233289328e-21	4.4908345847298e-17	
FALSE	ENSG00000177455	-9.15314204662845	5.5301194987254e-20	2.96930549618229e-16	
FALSE	ENSG00000211640	-9.12169041123672	7.39615502434293e-20	3.4039218609176e-16	

#### FALSE Genes sobre-expresados

FALSE log2 fold change (MLE): Group NIT vs ELI

FALSE Wald test p-value: Group NIT vs ELI

FALSE DataFrame with 6 rows and 6 columns

FALSE		baseMean	log2FoldChange	lfcSE	
FALSE		<numeric>	<numeric>	<numeric>	
FALSE	ENSG00000134443	15.8565450697657	20.2083121835446	2.04136651422826	
FALSE	ENSG00000100003	1122.5042755828	0.632683355208256	0.126699269468941	
FALSE	ENSG00000116885	660.814701209956	0.69797982502106	0.13983357502677	
FALSE	ENSG00000259942	72.4629194385833	2.88682039778111	0.616190488405295	
FALSE	ENSG00000132840	2000.67844157124	1.22558079660444	0.266954784142074	
FALSE	ENSG00000140254	6881.17911360331	0.773511876554707	0.169702849568479	
FALSE		stat	pvalue	padj	
FALSE		<numeric>	<numeric>	<numeric>	
FALSE	ENSG00000134443	9.89940417004654	4.18762427701157e-23	6.74542518541024e-19	
FALSE	ENSG00000100003	4.99358329262783	5.926920694304e-07	2.19221213648332e-05	
FALSE	ENSG00000116885	4.99150382794288	5.99109953074955e-07	2.21340897342463e-05	
FALSE	ENSG00000259942	4.6849480024468	2.80031014987974e-06	8.76722952269445e-05	
FALSE	ENSG00000132840	4.59096771965768	4.41195576566025e-06	0.000130879895899181	
FALSE	ENSG00000140254	4.55803705430755	5.16339293624087e-06	0.000150401326251298	

#### 4.8.1.1 Anotación de los genes expresados genéticamente

Con el fin de obtener el nombre de los genes seleccionados como diferencialmente expresados a partir del identificador de Ensembl, se empleó el paquete `AnnotationDbi`. Este paquete permite cargar los datos de anotación del genoma de *Homo sapiens*, buscar los identificadores de Ensembl y sus correspondientes nombres.

Con la función `mapIds` añadimos nuevas columnas a nuestra tabla de resultados.

```
NITvsELIsig$symbol <- mapIds(org.Hs.eg.db,
                             keys=row.names(NITvsELIsig),
                             keytype = "ENSEMBL",
                             column = "SYMBOL",
                             multiVals = "first")
NITvsELIsig$entrez <- mapIds(org.Hs.eg.db,
                             keys=row.names(NITvsELIsig),
                             keytype = "ENSEMBL",
                             column = "ENTREZID",
                             multiVals = "first")
```

Obtenemos la lista ordenada por p-valor de los genes expresados diferencialmente.

```
FALSE
FALSE
FALSE DataFrame with 6 rows and 2 columns
FALSE          symbol      entrez
FALSE      <character> <character>
FALSE ENSG00000110777    POU2AF1    5450
FALSE ENSG00000134443      GRP      2922
FALSE ENSG00000230006 ANKRD36BP2   645784
FALSE ENSG00000137265     IRF4     3662
FALSE ENSG00000156738    MS4A1     931
FALSE ENSG00000177455     CD19     930
```

#### 4.8.1.2 Análisis de Significación Biológica

Una vez anotados los genes seleccionados, se buscó su significación biológica empleando el paquete `enrichR`. Se eligió la base de datos Gene Ontology de procesos biológicos ya que nos ofrece una visión general de los procesos biológicos en los que puede estar implicado el gen.

Buscaremos la significación biológica del gen más significativamente sobre expresado y el más sub expresado.

```
FALSE Uploading data to Enrichr... Done.
FALSE Querying GO_Biological_Process_2015... Done.
FALSE Parsing results... Done.
```

Table 6: Gen sub-expresado en NIT sobre ELI

TCL1A
stem cell maintenance ( <a href="#">GO:0019827</a> )
multicellular organismal development ( <a href="#">GO:0007275</a> )



```
FALSE Uploading data to Enrichr... Done.
FALSE Querying GO_Biological_Process_2015... Done.
FALSE Parsing results... Done.
```

Table 7: Gen sobre-expresado en NIT sobre ELI

SERPINA9
negative regulation of endopeptidase activity ( <a href="#">GO:0010951</a> )
negative regulation of peptidase activity ( <a href="#">GO:0010466</a> )
negative regulation of proteolysis ( <a href="#">GO:0045861</a> )
negative regulation of protein maturation ( <a href="#">GO:1903318</a> )
negative regulation of protein processing ( <a href="#">GO:0010955</a> )
regulation of endopeptidase activity ( <a href="#">GO:0052548</a> )
negative regulation of hydrolase activity ( <a href="#">GO:0051346</a> )
regulation of peptidase activity ( <a href="#">GO:0052547</a> )

#### 4.8.2 Genes diferenciados NIT vs. SFI

```
NITvsSFI <- results(dds_filt, contrast = c("Group", "NIT", "SFI"))
summary(NITvsSFI)
```

```
FALSE
FALSE out of 46704 with nonzero total read count
FALSE adjusted p-value < 0.1
FALSE LFC > 0 (up)      : 29, 0.062%
FALSE LFC < 0 (down)    : 162, 0.35%
FALSE outliers [1]      : 0, 0%
FALSE low counts [2]    : 21735, 47%
FALSE (mean count < 5)
FALSE [1] see 'cooksCutoff' argument of ?results
FALSE [2] see 'independentFiltering' argument of ?results
```

Table 8: Cuadro resumen comparación genes diferencialmente expresados entre grupo NIT y SFI. La primera fila indica el número total de genes con expresión diferencial (p-valor ajustado <0.1); la segunda fila indica el número de genes con un p-valor inferior a 0.01; las dos siguientes filas corresponden a los genes significantes con una fuerte sub y sobre expresión, respectivamente; y la última fila, el número máximo de falsos positivos que podemos encontrar entre la selección de genes.

	NITvsSFI
Genes expr diferencial	191
p-valor<0.01	191
Sub-expresados	162
Sobre-expresados	29
max falsos positivos	20

FALSE Genes sub-expresados

FALSE log2 fold change (MLE): Group NIT vs SFI

FALSE Wald test p-value: Group NIT vs SFI

FALSE DataFrame with 6 rows and 6 columns

FALSE		baseMean	log2FoldChange	lfcSE	
FALSE		<numeric>	<numeric>	<numeric>	
FALSE	ENSG00000242534	68.8062058155926	-7.60489387051801	1.3611897843628	
FALSE	ENSG00000211619	50.3577620744547	-7.29896265739475	1.34164261403951	
FALSE	ENSG00000187862	99.0542735333606	-3.08451821190126	0.610069295792068	
FALSE	ENSG00000211895	58858.0169865427	-3.26027068485198	0.672618315247166	
FALSE	ENSG00000117322	1108.11606315774	-4.59121963661748	0.966533409027904	
FALSE	ENSG00000160223	1324.07293431035	-1.38304717732764	0.291237342697968	
FALSE		stat	pvalue	padj	
FALSE		<numeric>	<numeric>	<numeric>	
FALSE	ENSG00000242534	-5.5869460363883	2.31097495674225e-08	0.000577096666197675	
FALSE	ENSG00000211619	-5.44031814509717	5.31855061581831e-08	0.000664074229891075	
FALSE	ENSG00000187862	-5.0560128712863	4.28112452294304e-07	0.00356360805289779	
FALSE	ENSG00000211895	-4.8471333755667	1.25258220917587e-06	0.00781987073188499	
FALSE	ENSG00000117322	-4.75019238210827	2.03223219464435e-06	0.00851379555016678	
FALSE	ENSG00000160223	-4.74886621514724	2.04560200628707e-06	0.00851379555016678	

FALSE Genes sobre-expresados

FALSE log2 fold change (MLE): Group NIT vs SFI

FALSE Wald test p-value: Group NIT vs SFI

FALSE DataFrame with 6 rows and 6 columns

FALSE		baseMean	log2FoldChange	lfcSE	
FALSE		<numeric>	<numeric>	<numeric>	
FALSE	ENSG00000149968	101.836960902911	4.91047166653758	1.08358880532895	
FALSE	ENSG00000139531	2272.76763791309	0.484160730239652	0.117633516307453	
FALSE	ENSG00000173928	207.16318339967	0.572685853377995	0.143784151749068	
FALSE	ENSG00000197124	805.822730528455	0.780396743602256	0.196967899885907	
FALSE	ENSG00000155636	947.707954626443	0.437888713371884	0.115067728001132	
FALSE	ENSG00000119440	22.4242639266271	1.55668019545022	0.412138916118979	
FALSE		stat	pvalue	padj	
FALSE		<numeric>	<numeric>	<numeric>	
FALSE	ENSG00000149968	4.53167441596711	5.85179927656316e-06	0.012291057576337	
FALSE	ENSG00000139531	4.1158399870852	3.85771879848339e-05	0.024701270214289	
FALSE	ENSG00000173928	3.98295532860567	6.8063552975708e-05	0.0314756119427663	
FALSE	ENSG00000197124	3.96205038513534	7.43088533210081e-05	0.032769370754103	
FALSE	ENSG00000155636	3.80548674227389	0.000141525595396067	0.0430018925155565	
FALSE	ENSG00000119440	3.77707645302981	0.00015868003386021	0.0460762535529902	

#### 4.8.2.1 Anotación de los genes expresados genéticamente

```
require(AnnotationDbi)
require(org.Hs.eg.db)
NITvsSFI$g$symbol <- mapIds(org.Hs.eg.db,
                             keys=row.names(NITvsSFI$g),
                             keytype = "ENSEMBL",
                             column = "SYMBOL",
```

```

                                multiVals = "first")
NITvsSFIsig$entrez <- mapIds(org.Hs.eg.db,
                                keys=row.names(NITvsSFIsig),
                                keytype = "ENSEMBL",
                                column = "ENTREZID",
                                multiVals = "first")

```

Obtenemos la lista ordenada por p-valor de los genes expresados diferencialmente.

```

FALSE
FALSE
FALSE DataFrame with 6 rows and 2 columns
FALSE              symbol      entrez
FALSE      <character> <character>
FALSE ENSG00000242534      NA        NA
FALSE ENSG00000211619      NA        NA
FALSE ENSG00000187862    TTC24    164118
FALSE ENSG00000211895      NA        NA
FALSE ENSG00000117322     CR2      1380
FALSE ENSG00000160223   ICOSLG    23308

```

#### 4.8.2.2 Análisis de Significación Biológica

```

FALSE Uploading data to Enrichr... Done.
FALSE Querying GO_Biological_Process_2015... Done.
FALSE Parsing results... Done.

```

Table 9: Gen sub-expresado en NIT sobre SFI

CR2
B cell proliferation ( <a href="#">GO:0042100</a> )
B cell differentiation ( <a href="#">GO:0030183</a> )
lymphocyte proliferation ( <a href="#">GO:0046651</a> )
mononuclear cell proliferation ( <a href="#">GO:0032943</a> )
leukocyte proliferation ( <a href="#">GO:0070661</a> )
B cell activation ( <a href="#">GO:0042113</a> )
complement activation, classical pathway ( <a href="#">GO:0006958</a> )
lymphocyte differentiation ( <a href="#">GO:0030098</a> )
complement activation ( <a href="#">GO:0006956</a> )
protein activation cascade ( <a href="#">GO:0072376</a> )
leukocyte differentiation ( <a href="#">GO:0002521</a> )
humoral immune response ( <a href="#">GO:0006959</a> )
lymphocyte activation ( <a href="#">GO:0046649</a> )
immune response-activating cell surface receptor signaling pathway ( <a href="#">GO:0002429</a> )
leukocyte activation ( <a href="#">GO:0045321</a> )
immune response-activating signal transduction ( <a href="#">GO:0002757</a> )
immune response-regulating cell surface receptor signaling pathway ( <a href="#">GO:0002768</a> )
activation of immune response ( <a href="#">GO:0002253</a> )

```

FALSE Uploading data to Enrichr... Done.

```

```
FALSE Querying GO_Biological_Process_2015... Done.
FALSE Parsing results... Done.
```

Table 10: Gen sobre-expresado en NIT sobre SFI

PLA2G2D
phosphatidylinositol acyl-chain remodeling ( <a href="#">GO:0036149</a> )
phosphatidylglycerol acyl-chain remodeling ( <a href="#">GO:0036148</a> )
phosphatidylserine acyl-chain remodeling ( <a href="#">GO:0036150</a> )
phosphatidylethanolamine acyl-chain remodeling ( <a href="#">GO:0036152</a> )
phosphatidylcholine acyl-chain remodeling ( <a href="#">GO:0036151</a> )
phosphatidylserine metabolic process ( <a href="#">GO:0006658</a> )
phosphatidic acid metabolic process ( <a href="#">GO:0046473</a> )
phosphatidic acid biosynthetic process ( <a href="#">GO:0006654</a> )
phosphatidylglycerol metabolic process ( <a href="#">GO:0046471</a> )
alditol phosphate metabolic process ( <a href="#">GO:0052646</a> )
phosphatidylcholine metabolic process ( <a href="#">GO:0046470</a> )
ethanolamine-containing compound metabolic process ( <a href="#">GO:0042439</a> )
cellular biogenic amine metabolic process ( <a href="#">GO:0006576</a> )
cellular amine metabolic process ( <a href="#">GO:0044106</a> )
phosphatidylinositol metabolic process ( <a href="#">GO:0046488</a> )
amine metabolic process ( <a href="#">GO:0009308</a> )
glycerophospholipid biosynthetic process ( <a href="#">GO:0046474</a> )
cellular modified amino acid metabolic process ( <a href="#">GO:0006575</a> )
glycerolipid biosynthetic process ( <a href="#">GO:0045017</a> )
phospholipid biosynthetic process ( <a href="#">GO:0008654</a> )
glycerophospholipid metabolic process ( <a href="#">GO:0006650</a> )
lipid catabolic process ( <a href="#">GO:0016042</a> )
phospholipid metabolic process ( <a href="#">GO:0006644</a> )
glycerolipid metabolic process ( <a href="#">GO:0046486</a> )
alcohol metabolic process ( <a href="#">GO:0006066</a> )
inflammatory response ( <a href="#">GO:0006954</a> )
cellular amino acid metabolic process ( <a href="#">GO:0006520</a> )
organophosphate biosynthetic process ( <a href="#">GO:0090407</a> )
organic hydroxy compound metabolic process ( <a href="#">GO:1901615</a> )
lipid biosynthetic process ( <a href="#">GO:0008610</a> )

#### 4.8.3 Genes diferenciados ELI vs. SFI

```
ELIvsSFI <- results(dds_filt, contrast = c("Group", "ELI", "SFI"))
summary(ELIvsSFI)
```

```
FALSE
FALSE out of 46704 with nonzero total read count
FALSE adjusted p-value < 0.1
FALSE LFC > 0 (up)      : 2299, 4.9%
FALSE LFC < 0 (down)    : 790, 1.7%
FALSE outliers [1]      : 0, 0%
FALSE low counts [2]    : 17207, 37%
FALSE (mean count < 2)
```

FALSE [1] see 'cooksCutoff' argument of ?results  
 FALSE [2] see 'independentFiltering' argument of ?results

Table 11: Cuadro resumen comparación genes diferencialmente expresados entre grupo NIT y SFI. La primera fila indica el número total de genes con expresión diferencial (p-valor ajustado <0.1); la segunda fila indica el número de genes con un p-valor inferior a 0.01; las dos siguientes filas corresponden a los genes significantes con una fuerte sub y sobre expresión, respectivamente; y la última fila, el número máximo de falsos positivos que podemos encontrar entre la selección de genes.

ELIvsSFI	
Genes expr diferencial	3089
p-valor<0.01	3040
Sub-expresados	790
Sobre-expresados	2299
max falsos positivos	309

FALSE Genes sub-expresados

```
FALSE log2 fold change (MLE): Group ELI vs SFI
FALSE Wald test p-value: Group ELI vs SFI
FALSE DataFrame with 6 rows and 6 columns
FALSE      baseMean  log2FoldChange  lfcSE
FALSE      <numeric>      <numeric>      <numeric>
FALSE ENSG00000134443 15.8565450697657 -22.1702642318013 2.03767193470189
FALSE ENSG00000110680 1524.15718722265 -9.03670047408347 1.3864478618371
FALSE ENSG00000157613 621.105440586409 -1.29951752446805 0.231983330437944
FALSE ENSG00000163873 98.219371846124 -2.03124213294135 0.37159308760737
FALSE ENSG00000148180 86359.9907134648 -1.09277772175022 0.214580986924887
FALSE ENSG00000120332 254.082508768811 -2.82533748190365 0.559161327175679
FALSE      stat      pvalue      padj
FALSE      <numeric>      <numeric>      <numeric>
FALSE ENSG00000134443 -10.8801931529006 1.4325907970823e-27 4.22614285139278e-23
FALSE ENSG00000110680 -6.51787977234821 7.13081088391602e-11 1.23740541809131e-07
FALSE ENSG00000157613 -5.60177113594663 2.12172579066725e-08 6.98499807090941e-06
FALSE ENSG00000163873 -5.46630763779864 4.59506574609575e-08 1.24361871109931e-05
FALSE ENSG00000148180 -5.0926120594866 3.53163850009618e-07 5.87652989083844e-05
FALSE ENSG00000120332 -5.05281274757398 4.35350667986137e-07 6.68898161749533e-05
```

FALSE Genes sobre-expresados

```
FALSE log2 fold change (MLE): Group ELI vs SFI
FALSE Wald test p-value: Group ELI vs SFI
FALSE DataFrame with 6 rows and 6 columns
FALSE      baseMean  log2FoldChange  lfcSE
FALSE      <numeric>      <numeric>      <numeric>
FALSE ENSG00000165066 17.4059665395142 4.36277747849577 0.574499110423329
FALSE ENSG00000127074 149.780226398877 4.57333387368789 0.625010133150488
FALSE ENSG00000101057 361.887972325873 3.92366655996305 0.554211309128116
```

```
FALSE ENSG00000262488 34.0867785159697 4.57801526361421 0.650250273978417
FALSE ENSG00000112486 231.809768156334 2.45529864160126 0.350046350421062
FALSE ENSG00000110777 1991.25426316944 3.75650864675105 0.539038828753813
FALSE                                stat                                pvalue                                padj
FALSE                                <numeric>                        <numeric>                        <numeric>
FALSE ENSG00000165066 7.59405436725739 3.10048212139964e-14 4.57321112906447e-10
FALSE ENSG00000127074 7.31721556358629 2.53168999181303e-13 2.48949515861615e-09
FALSE ENSG00000101057 7.07973023166156 1.44435351797148e-12 1.06521071950396e-08
FALSE ENSG00000262488 7.04038959584684 1.91703054848798e-12 1.13104802360791e-08
FALSE ENSG00000112486 7.01421008574391 2.31251699782214e-12 1.13698752392922e-08
FALSE ENSG00000110777 6.96890176805186 3.19424627011127e-12 1.34614664240404e-08
```

#### 4.8.3.1 Anotación de los genes expresados genéticamente

```
require(AnnotationDbi)
require(org.Hs.eg.db)
ELIvsSFI$symbol <- mapIds(org.Hs.eg.db,
                           keys=row.names(ELIvsSFI),
                           keytype = "ENSEMBL",
                           column = "SYMBOL",
                           multiVals = "first")
ELIvsSFI$entrez <- mapIds(org.Hs.eg.db,
                           keys=row.names(ELIvsSFI),
                           keytype = "ENSEMBL",
                           column = "ENTREZID",
                           multiVals = "first")
```

Obtenemos la lista ordenada por p-valor de los genes expresados diferencialmente.

```
FALSE
FALSE
FALSE DataFrame with 6 rows and 2 columns
FALSE      symbol      entrez
FALSE      <character> <character>
FALSE ENSG00000134443      GRP      2922
FALSE ENSG00000165066      NKX6-3    157848
FALSE ENSG00000127074      RGS13     6003
FALSE ENSG00000101057      MYBL2     4605
FALSE ENSG00000262488      NA        NA
FALSE ENSG00000112486      CCR6     1235
```

#### 4.8.3.2 Análisis de Significación Biológica

```
FALSE Uploading data to Enrichr... Done.
FALSE Querying GO_Biological_Process_2015... Done.
FALSE Parsing results... Done.
```

Table 12: Gen sub-expresado en ELI sobre SFI

---

AVP

---

regulation of female receptivity ([GO:0045924](#))

regulation of transmission of nerve impulse ([GO:0051969](#))  
 hyperosmotic salinity response ([GO:0042538](#))  
 positive regulation of amino acid transport ([GO:0051957](#))  
 female mating behavior ([GO:0060180](#))  
 aggressive behavior ([GO:0002118](#))  
 maternal behavior ([GO:0042711](#))  
 parental behavior ([GO:0060746](#))  
 regulation of glutamate secretion ([GO:0014048](#))  
 positive regulation of fatty acid biosynthetic process ([GO:0045723](#))  
 protein kinase C signaling ([GO:0070528](#))  
 sodium-independent organic anion transport ([GO:0043252](#))  
 positive regulation of systemic arterial blood pressure ([GO:0003084](#))  
 negative regulation of neurological system process ([GO:0031645](#))  
 grooming behavior ([GO:0007625](#))  
 negative regulation of release of cytochrome c from mitochondria ([GO:0090201](#))  
 ERK1 and ERK2 cascade ([GO:0070371](#))  
 regulation of amino acid transport ([GO:0051955](#))  
 multicellular organismal reproductive behavior ([GO:0033057](#))  
 hyperosmotic response ([GO:0006972](#))  
 regulation of renal sodium excretion ([GO:0035813](#))  
 positive regulation of organic acid transport ([GO:0032892](#))  
 positive regulation of amine transport ([GO:0051954](#))  
 response to salt stress ([GO:0009651](#))  
 regulation of excretion ([GO:0044062](#))  
 response to testosterone ([GO:0033574](#))  
 positive regulation of vasoconstriction ([GO:0045907](#))  
 positive regulation of fatty acid metabolic process ([GO:0045923](#))  
 negative regulation of mitochondrion organization ([GO:0010823](#))  
 regulation of fatty acid biosynthetic process ([GO:0042304](#))  
 multi-organism reproductive behavior ([GO:0044705](#))  
 vasoconstriction ([GO:0042310](#))  
 positive regulation of blood pressure ([GO:0045777](#))  
 multicellular organismal water homeostasis ([GO:0050891](#))  
 reproductive behavior ([GO:0019098](#))  
 water homeostasis ([GO:0030104](#))  
 response to nicotine ([GO:0035094](#))  
 social behavior ([GO:0035176](#))  
 intraspecies interaction between organisms ([GO:0051703](#))  
 water transport ([GO:0006833](#))  
 regulation of release of cytochrome c from mitochondria ([GO:0090199](#))  
 regulation of organic acid transport ([GO:0032890](#))  
 fluid transport ([GO:0042044](#))  
 regulation of neurological system process ([GO:0031644](#))  
 positive regulation of lipid biosynthetic process ([GO:0046889](#))  
 regulation of vasoconstriction ([GO:0019229](#))  
 regulation of amine transport ([GO:0051952](#))  
 regulation of blood vessel size ([GO:0050880](#))  
 regulation of tube size ([GO:0035150](#))  
 response to osmotic stress ([GO:0006970](#))  
 positive regulation of cAMP biosynthetic process ([GO:0030819](#))  
 negative regulation of cysteine-type endopeptidase activity involved in apoptotic process ([GO:0043154](#))  
 negative regulation of cysteine-type endopeptidase activity ([GO:2000117](#))

regulation of fatty acid metabolic process ([GO:0019217](#))  
 positive regulation of blood circulation ([GO:1903524](#))  
 positive regulation of cAMP metabolic process ([GO:0030816](#))  
 multi-organism behavior ([GO:0051705](#))  
 positive regulation of peptidyl-serine phosphorylation ([GO:0033138](#))  
 positive regulation of cyclic nucleotide biosynthetic process ([GO:0030804](#))  
 positive regulation of purine nucleotide biosynthetic process ([GO:1900373](#))  
 positive regulation of nucleotide biosynthetic process ([GO:0030810](#))  
 regulation of anion transport ([GO:0044070](#))  
 vascular process in circulatory system ([GO:0003018](#))  
 positive regulation of cyclic nucleotide metabolic process ([GO:0030801](#))  
 regulation of mitochondrion organization ([GO:0010821](#))  
 regulation of peptidyl-serine phosphorylation ([GO:0033135](#))  
 regulation of cAMP biosynthetic process ([GO:0030817](#))  
 positive regulation of lipid metabolic process ([GO:0045834](#))  
 response to ketone ([GO:1901654](#))  
 positive regulation of cell growth ([GO:0030307](#))  
 response to ethanol ([GO:0045471](#))  
 response to alkaloid ([GO:0043279](#))  
 regulation of lipid biosynthetic process ([GO:0046890](#))  
 regulation of cAMP metabolic process ([GO:0030814](#))  
 regulation of cyclic nucleotide biosynthetic process ([GO:0030802](#))  
 regulation of nucleotide biosynthetic process ([GO:0030808](#))  
 regulation of purine nucleotide biosynthetic process ([GO:1900371](#))  
 multi-multicellular organism process ([GO:0044706](#))  
 positive regulation of homeostatic process ([GO:0032846](#))  
 circulatory system process ([GO:0003013](#))  
 regulation of blood pressure ([GO:0008217](#))  
 regulation of cyclic nucleotide metabolic process ([GO:0030799](#))  
 multi-organism reproductive process ([GO:0044703](#))  
 positive regulation of cytosolic calcium ion concentration ([GO:0007204](#))  
 regulation of ion homeostasis ([GO:2000021](#))  
 MAPK cascade ([GO:0000165](#))  
 cytosolic calcium ion homeostasis ([GO:0051480](#))  
 regulation of cellular ketone metabolic process ([GO:0010565](#))  
 locomotory behavior ([GO:0007626](#))  
 regulation of cysteine-type endopeptidase activity involved in apoptotic process ([GO:0043281](#))  
 positive regulation of ion transport ([GO:0043270](#))  
 negative regulation of apoptotic signaling pathway ([GO:2001234](#))  
 positive regulation of growth ([GO:0045927](#))  
 regulation of cysteine-type endopeptidase activity ([GO:2000116](#))  
 signal transduction by phosphorylation ([GO:0023014](#))  
 regulation of blood circulation ([GO:1903522](#))  
 negative regulation of endopeptidase activity ([GO:0010951](#))  
 negative regulation of peptidase activity ([GO:0010466](#))  
 regulation of anatomical structure size ([GO:0090066](#))  
 negative regulation of organelle organization ([GO:0010639](#))  
 regulation of lipid metabolic process ([GO:0019216](#))  
 positive regulation of secretion by cell ([GO:1903532](#))  
 cellular calcium ion homeostasis ([GO:0006874](#))  
 cellular divalent inorganic cation homeostasis ([GO:0072503](#))  
 positive regulation of secretion ([GO:0051047](#))



---

## AVP

---

calcium ion homeostasis ([GO:0055074](#))  
response to alcohol ([GO:0097305](#))  
divalent inorganic cation homeostasis ([GO:0072507](#))  
negative regulation of proteolysis ([GO:0045861](#))  
regulation of homeostatic process ([GO:0032844](#))  
negative regulation of protein processing ([GO:0010955](#))  
negative regulation of protein maturation ([GO:1903318](#))  
regulation of cell growth ([GO:0001558](#))  
organic anion transport ([GO:0015711](#))  
negative regulation of hydrolase activity ([GO:0051346](#))  
regulation of endopeptidase activity ([GO:0052548](#))  
regulation of apoptotic signaling pathway ([GO:2001233](#))  
cellular metal ion homeostasis ([GO:0006875](#))  
single-organism behavior ([GO:0044708](#))  
response to steroid hormone ([GO:0048545](#))  
regulation of system process ([GO:0044057](#))  
regulation of peptidase activity ([GO:0052547](#))  
generation of precursor metabolites and energy ([GO:0006091](#))  
cellular cation homeostasis ([GO:0030003](#))  
cellular ion homeostasis ([GO:0006873](#))  
metal ion homeostasis ([GO:0055065](#))  
anion transport ([GO:0006820](#))  
cellular chemical homeostasis ([GO:0055082](#))  
cation homeostasis ([GO:0055080](#))  
negative regulation of cellular component organization ([GO:0051129](#))  
multicellular organismal reproductive process ([GO:0048609](#))  
behavior ([GO:0007610](#))

---

FALSE Uploading data to Enrichr... Done.  
FALSE Querying GO\_Biological\_Process\_2015... Done.  
FALSE Parsing results... Done.

Table 13: Gen sobre-expresado en ELI sobre SFI

GRP
neuropeptide signaling pathway ( <a href="#">GO:0007218</a> )

## 5 Resumen de resultados

Los controles de calidad llevados a cabo antes de comenzar el análisis, han asegurado una buena calidad de los datos de trabajo, a pesar de presentar ciertas problemáticas achacadas a fallos de manipulación de los datos.

Tras el procesado de los datos, se han podido detectar una serie de genes diferencialmente expresados, como podemos ver en la tabla resumen

Table 14: Tabla resumen con el número de genes diferenciados en su expresión según las condiciones

	NITvsELI	NITvsSFI	ELIvsSFI
Genes expr diferencial	4001	191	3089
p-valor<0.01	3759	191	3040
Sub-expresados	3107	162	790
Sobre-expresados	894	29	2299
max falsos positivos	401	20	309

La siguiente tabla muestra la lista de todos los archivos de resultados generados durante el análisis y que pueden encontrarse en el repositorio [github](#).

Table 15: Listado de archivos de resultados generados en el presente análisis

Archivos de resultados
ELIvsSFIover.csv
ELIvsSFIsig_ann.csv
ELIvsSFIunder.csv
NITvsELIover.csv
NITvsELIsig_ann.csv
NITvsELIunder.csv
NITvsSFIover.csv
NITvsSFIsig_ann.csv
NITvsSFIunder.csv

## Bibliografía

- Benjamini, Yoav, and Yosef Hochberg. 1995. “Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing.” *Journal of the Royal Statistical Society: Series B (Methodological)* 57 (1). Wiley Online Library: 289–300.
- González, Ignacio. 2014. “Statistical Analysis of Rna-Seq Data.” url<http://www.nathalievieialaneix.eu/doc/pdf/tutorial-rnaseq.pdf>.
- Ritchie, Matthew E., Belinda Phipson, Di Wu, Yifang Hu, Charity W. Law, Wei Shi, and Gordon K. Smyth. 2015. “limma powers differential expression analyses for RNA-sequencing and microarray studies.” *Nucleic Acids Research* 43 (7): e47–e47. doi:[10.1093/nar/gkv007](https://doi.org/10.1093/nar/gkv007).
- RStudio Team. 2016. *RStudio: Integrated Development Environment for R*. Boston, MA: RStudio, Inc. <http://www.rstudio.com/>.