

PEC 1 Análisis de Datos Ómicos

Joaquín Astete Frugoni

Abril 2020

Contents

1 Abstract	1
1.1 Repostorio GitHub	1
2 Objetivos	1
3 Materiales y Métodos	1
3.1 Datos	1
3.2 Diseño del experimento	1
3.3 Preparación de datos	2
3.4 Análisis de datos	5
4 Resultados	6
5 Discusión	11
6 Apéndice	11
Referencias	16

1 Abstract

La Bronquiolitis Obliterante (BO) es una patología que se caracteriza por la fibrosis progresiva e irreversible de la pequeña vía aérea que produce la obstrucción de la misma. En el presente trabajo se analiza la expresión diferencial de genes en un modelo de BO en ratas expuestas a 2,3-Pentadiona. Se detectaron diferencias en la expresión de genes involucrados en mecanismos inflamatorios, vías de la matriz extracelular y aparato ciliar. Un análisis específico de estos mecanismos moleculares podrían ser útil para estrategias terapéuticas de esta patología.

1.1 Repostorio GitHub

<https://github.com/joacoastete/PEC-1-analisis-datos-omicos.git>

2 Objetivos

- Utilizar una base de datos real de **GEO** para aplicar los conocimientos aprendidos en la asignatura **Análisis de Datos Ómicos**
- Análisis de genes diferencialmente expresados en muestras de un modelo de BO en ratas
- Establecer que vías moleculares pueden estar involucradas en la fisiopatogenia de la BO

3 Materiales y Métodos

3.1 Datos

Los datos utilizados para la realización del presente trabajo fueron obtenidos de la base de datos **Gene Expression Omnibus (GEO)**, se ha escogido la serie **GSE52761** que se corresponde con el estudio “*Gene expression in 2,3-Pentanedione Induced Bronchiolitis Obliterans*”(Morgan et al. 2015). Dicho estudio analiza la expresión génica de un modelo en ratas de Bronquiolitis Obliterante al exponerlas a vapores de 2,3-Pentadiona comparados a un grupo control. Luego del análisis histopatológico de los pulmones de las ratas expuestas, se dividieron las muestras según si presentaban fibrosis o no presentaban fibrosis. Se han utilizado arrays de la marca Affymetrix® modelo *Rat Genome 230 2.0 Array*. Se han generado las covariantes a partir de la información ofrecida en la web de **GEO**.

3.2 Diseño del experimento

Es un diseño de un factor con tres niveles. El factor es el tejido pulmonar y los niveles: Fibrotico, No fibrotico y control. Se obtuvieron 6 muestras para el grupo control, 3 muestras para el grupo de Fibrosis y 3 muestras para el grupo de No fibrosis. En el trabajo original se realizaron las siguientes comparaciones:

- Fibrosis-Control
- No Fibrosis-Control

En el presente trabajo añadiré la comparación:

- Fibrosis-No fibrosis

Ya que sería interesante analizar si se puede obtener información de genes que podrían intervenir en la protección contra la fibrosis en las muestras expuestas a 2,3-Pentadiona.

3.3 Preparación de datos

Los datos se cargarán y analizarán principalmente con las herramientas del proyecto **Bioconductor**. Una vez cargados los archivos *.CEL* de los arrays y generado el objeto **ExpressionSet**. Se analizó la calidad de los datos a través del paquete **arrayQualityMetrics**.

	array	sampleNames	*1	*2	*3	group	shortname
<input type="checkbox"/>	1	control.1			x	control	control.1
<input type="checkbox"/>	2	control.2				control	control.2
<input type="checkbox"/>	3	non.fibrotic.1				non.fibrotic	non.fibrotic.1
<input type="checkbox"/>	4	non.fibrotic.2			x	non.fibrotic	non.fibrotic.2
<input type="checkbox"/>	5	non.fibrotic.3				non.fibrotic	non.fibrotic.3
<input type="checkbox"/>	6	fibrotic.1				fibrotic	fibrotic.1
<input type="checkbox"/>	7	fibrotic.2				fibrotic	fibrotic.2
<input type="checkbox"/>	8	fibrotic.3				fibrotic	fibrotic.3
<input type="checkbox"/>	9	control.3			x	control	control.3
<input type="checkbox"/>	10	control.4				control	control.4
<input type="checkbox"/>	11	control.5				control	control.5
<input type="checkbox"/>	12	control.6				control	control.6

Figure 1: tabla calidad de arrays

En la tabla anterior se puede observar que la calidad de los arrays son aceptables ya que no se observa que ningún array tenga más de una columna de método de detección de *outlier* marcada.

Distribucion de valores de intensidad

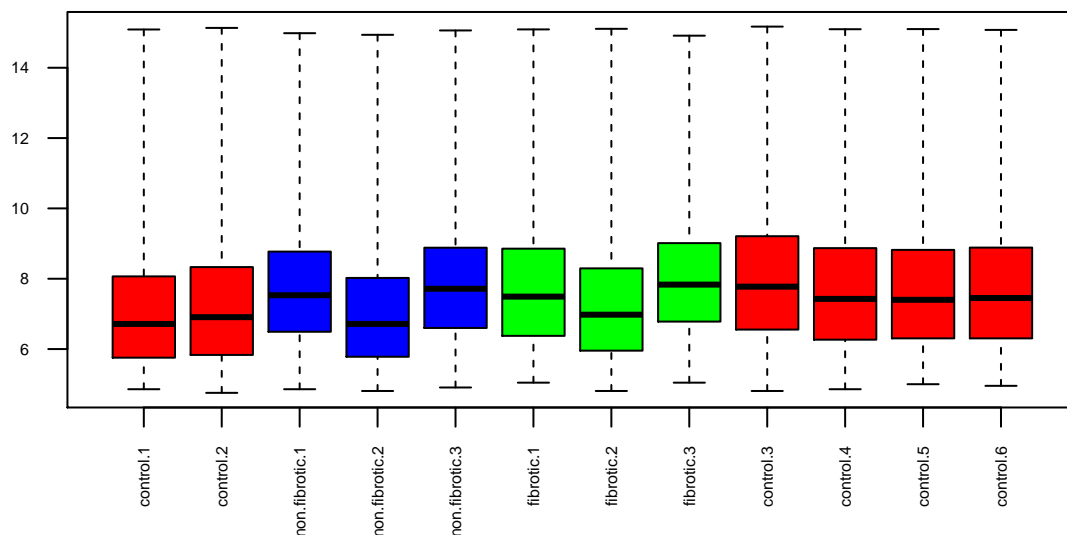


Figure 2: boxplot de intensidades

Si bien en el bloxplot anterior cierta variabilidad entre las intensidades no se observa ningún array que pudiera resultar defectuoso. En el siguiente gráfico se analizan los dos Componentes Principales de los datos

Analisis de Componente Principal

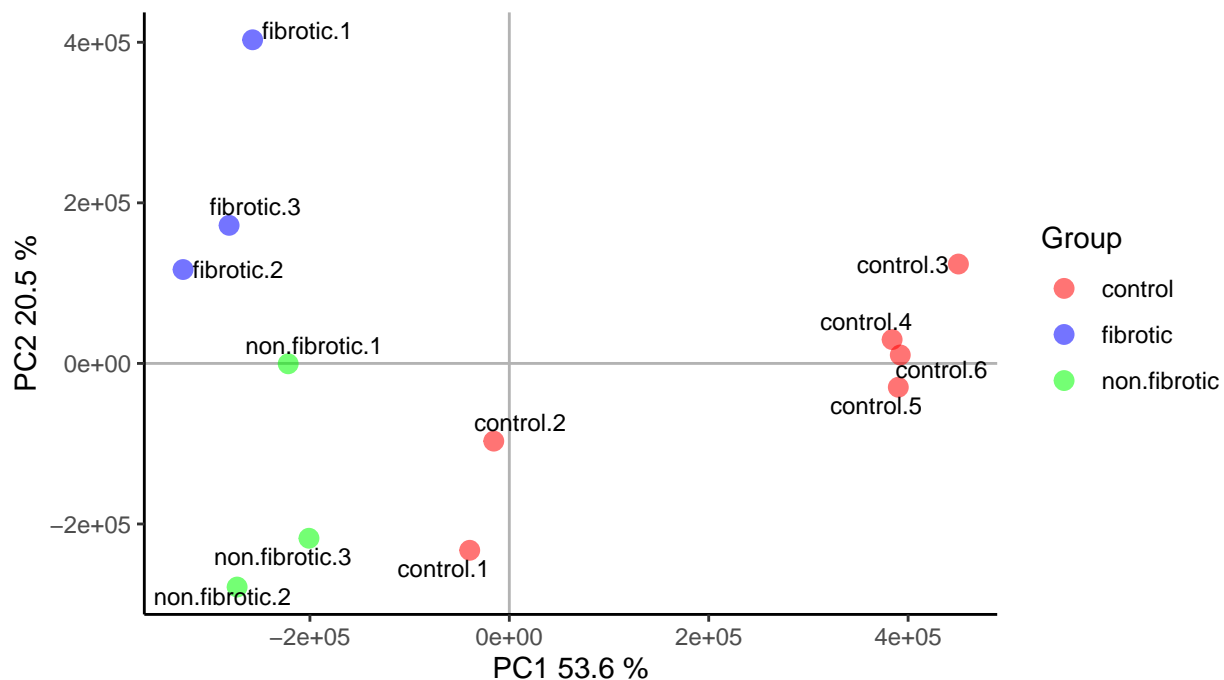


Figure 3: Grafico de Componente principal

El primer Componente Principal explica un 53% de la variabilidad que se corresponde con la exposición a 2,3-PD, mientras que el segundo Componente Principal explica un 20% de la variabilidad, parecería corresponder

con la presencia de fibrosis en el tejido pulmonar. Se observan dos arrays del grupo control que se localizarían en la zona de expuestos al 2,3-PD, por lo que habría que corroborar si están bien etiquetados o si hubo algún problema con dichas muestras.

3.3.1 Normalización de datos

Para la normalización de datos se utilizó el método *Robust Multichip Analysis* (Irizarry et al. 2003)

Distribucion de valores de intensidad: datos normalizados

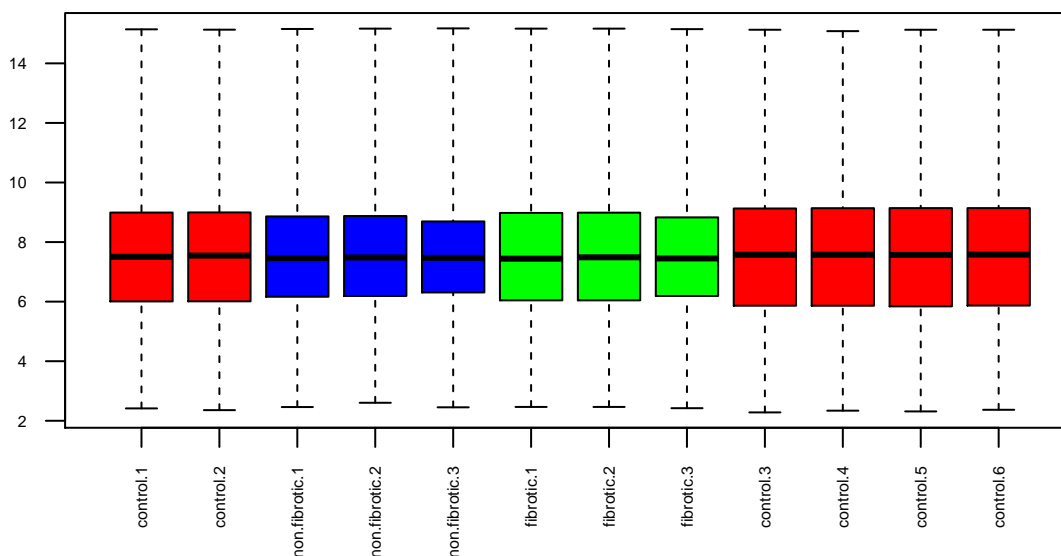


Figure 4: boxplot de intensidades-datos normalizados

En el boxplot anterior se observa que el proceso de normalización a funcionado correctamente ya que cada caja del boxplot son casi identicas. Lo que se corroborará con el análisis de `arrayQualityMetrics`:

	array	sampleNames	*1	*2	*3	group	shortname	treatment
<input type="checkbox"/>	1	control.1				control	control.1	no
<input type="checkbox"/>	2	control.2				control	control.2	no
<input type="checkbox"/>	3	non.fibrotic.1				non.fibrotic	non.fibrotic.1	yes
<input type="checkbox"/>	4	non.fibrotic.2				non.fibrotic	non.fibrotic.2	yes
<input type="checkbox"/>	5	non.fibrotic.3				non.fibrotic	non.fibrotic.3	yes
<input type="checkbox"/>	6	fibrotic.1				fibrotic	fibrotic.1	yes
<input type="checkbox"/>	7	fibrotic.2				fibrotic	fibrotic.2	yes
<input type="checkbox"/>	8	fibrotic.3				fibrotic	fibrotic.3	yes
<input type="checkbox"/>	9	control.3				control	control.3	no
<input type="checkbox"/>	10	control.4				control	control.4	no
<input type="checkbox"/>	11	control.5				control	control.5	no
<input type="checkbox"/>	12	control.6				control	control.6	no

Figure 5: tabla calidad de arrays-datos normalizados

Ningún array ha sido marcado como posible *outlier*.

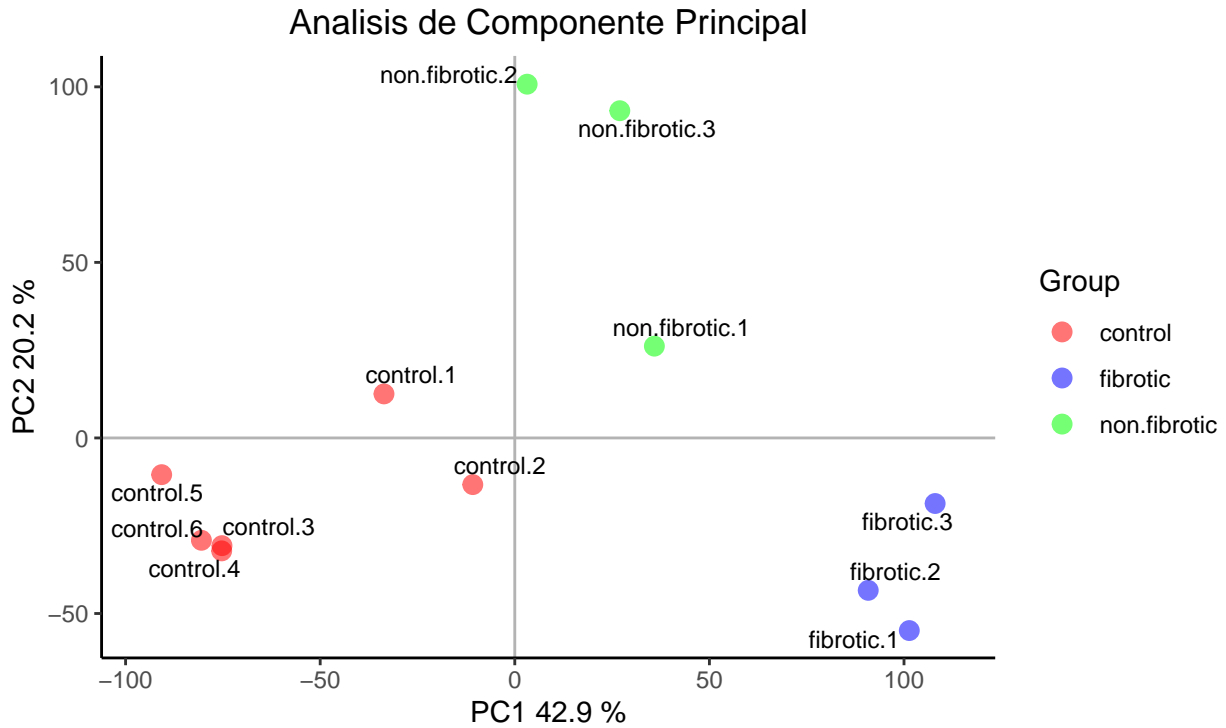


Figure 6: Grafico de Componente principal-datos normalizados

En el gráfico de los Componentes Principales de los datos normalizados se observa que la variabilidad por la exposición a 2,3-PD ha disminuido escasamente. Habría que corroborar si hay algún problema con la muestra correspondiente con el array *control.1* ya que se aleja bastante de las demás muestras del grupo control.

3.3.2 Filtrado de genes

Se utilizó el paquete **genefilter** para filtrar aquellos genes cuya variabilidad podría ser aleatoria, también se han filtrados sondas que no tenían anotación asociada. De esta manera se puede aumentar la potencia del estudio al disminuir el número de comparaciones.

Se han filtrado 27498 genes, quedando en el conjunto de datos un total de 3601 genes para analizar.

3.4 Análisis de datos

Para la selección de genes diferencialmente expresados he utilizado el método de *Modelos lineales para Microarrays* a través del paquete **limma**. Para el diseño de la matriz de contrastes, he establecido los siguientes contrastes:

- Fibrosis-Control
- No Fibrosis-Control
- Fibrosis-No fibrosis

Se ha utilizado la función **topTable** para obtener la listas de genes de cada comparación, ordenados según su valor de significación estadística. Para el ajuste de los *p-valores* para comparaciones múltiples se ha utilizado el método de *False discovery Rate* propuesto por Benjamini and Hochberg (Benjamini and Hochberg 1995). Para las anotaciones de los genes de las sondas empleadas, se ha utilizado la base de datos **rat2302.db** correspondiente al modelo de microarray utilizado. Para la selección y visualización de los genes diferencialmente expresados se ha utilizado un valor de corte de *FDR* de 0.05 y de *LogFC* de 1. Para la selección de genes que cambian en cada comparación se ha utilizado la función **decideTests**. Para el análisis

de la significación biológica se ha utilizado el método de *Gene Set Enrichment Analysis* a través del paquete **reactomePA**. Se ha establecido como valor de corte de FDR <0.15 y como universo se ha establecido todo los genes que tengan por lo menos una anotación de *Gene Ontology*.

4 Resultados

Al realizar los contrastes entre los diferentes grupos, se seleccionaron los genes que hayan obtenido un valor de FDR menor a 0.05. Para la diferentes comparaciones se obtuvieron los siguientes totales de genes diferencialmente expresados:

- Fibrosis-Control: 2350
- No Fibrosis-Control: 1393
- Fibrosis-No fibrosis: 1313

En la siguiente tabla se puede observar el total de genes *up-regulated* y *down-regulated* para cada comparación.

Table 1: Genes especialmente diferenciados entre comparaciones FDR <0.05 y logFC <1

	FIBvsCONTROL	NONFIBvsCONTROL	FIBvsNONFIB
Down	1591	1024	715
NotSig	1251	2208	2288
Up	759	369	598

A continuación se muestra un diagrama de Venn con los genes seleccionados en común entre las tres comparaciones

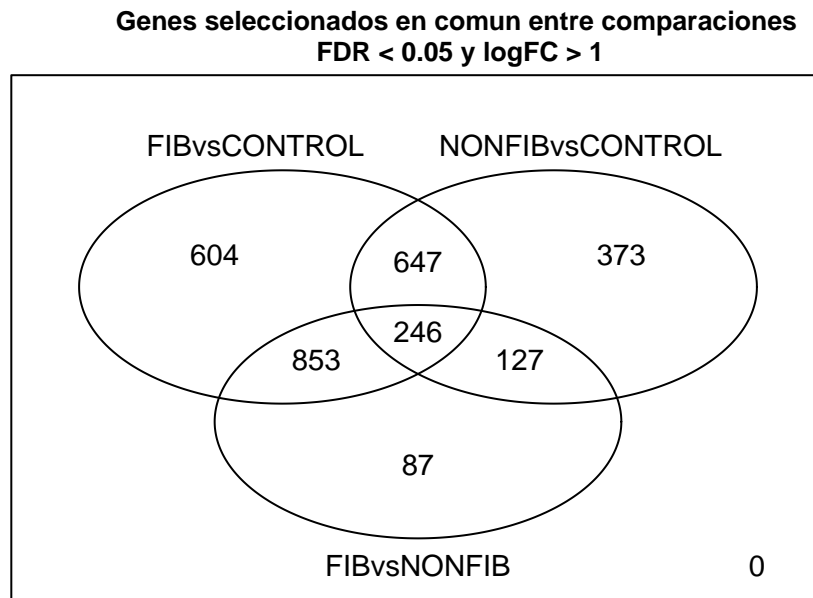


Figure 7: Diagrama de Venn - genes en comun

En las siguientes tablas se muestran los diez primeros genes ordenados según su *p-valor ajustado* con su correspondiente anotación de cada contraste.

Table 2: Genes diferencialmente expresados: Fibrosis-Control

PROBEID	SYMBOL	ENTREZID	GENENAME	adj.P.Val
1371124_a_at	Kng2l1	25087	kininogen 2-like 1	1e-07
1381039_at	Dnah1	171339	dynein, axonemal, heavy chain 1	1e-07
1384500_at	Ccdc30	362580	coiled-coil domain containing 30	1e-07
1385134_at	T2	681288	brachyury 2	1e-07
1376124_at	Iqub	296936	IQ motif and ubiquitin domain containing	1e-07
1377721_at	Pacrg	499021	parkin coregulated	1e-07
1380342_at	LOC680885	680885	hypothetical protein LOC680885	1e-07
1386931_at	Tnni3	29248	troponin I3, cardiac type	1e-07
1382956_at	LOC361016	361016	similar to RIKEN cDNA 4933406L09	1e-07
1375775_at	Odf3b	500917	outer dense fiber of sperm tails 3B	1e-07

Table 3: Genes diferencialmente expresados: No Fibrosis-Control

PROBEID	SYMBOL	ENTREZID	GENENAME	adj.P.Val
1386023_at	Lgi1	252892	leucine-rich, glioma inactivated 1	0.0000088
1369136_at	Cyp2a3	24299	cytochrome P450, family 2, subfamily a, polypeptide 3	0.0001328
1394786_at	Sorl1	300652	sortilin related receptor 1	0.0001328
1368344_at	Gad1	24379	glutamate decarboxylase 1	0.0001880
1382956_at	LOC361016	361016	similar to RIKEN cDNA 4933406L09	0.0001920
1378430_at	Moxd1	294119	monooxygenase, DBH-like 1	0.0002068
1368016_at	Pecr	113956	peroxisomal trans-2-enoyl-CoA reductase	0.0002853
1369381_a_at	Slc15a1	117261	solute carrier family 15 member 1	0.0002853
1378340_at	Cfap299	685158	cilia and flagella associated protein 299	0.0003088
1368236_at	Mep1a	25684	meprin A subunit alpha	0.0003169

Table 4: Genes diferencialmente expresados: Fibrosis-No Fibrosis

PROBEID	SYMBOL	ENTREZID	GENENAME	adj.P.Val
1371124_a_at	Kng2l1	25087	kininogen 2-like 1	1.40e-05
1376124_at	Iqub	296936	IQ motif and ubiquitin domain containing	1.40e-05
1385134_at	T2	681288	brachyury 2	1.40e-05
1386931_at	Tnni3	29248	troponin I3, cardiac type	1.40e-05
1391228_at	Rsph1	361818	radial spoke head component 1	1.40e-05
1384855_at	Cfap126	498278	cilia and flagella associated protein 126	1.45e-05
1375032_at	Kndc1	361672	kinase non-catalytic C-lobe domain containing 1	1.56e-05
1384789_at	Tekt4	302991	tektin 4	1.82e-05
1388224_at	Sec14l3	64543	SEC14-like lipid binding 3	2.20e-05
1372190_at	Aqp4	25293	aquaporin 4	2.36e-05

En el *heatmap* siguiente, se muestra los genes seleccionados con un *FDR* menor a 0.05 y un *logFC* mayor a 1, con los genes y muestras agrupados según similitud.

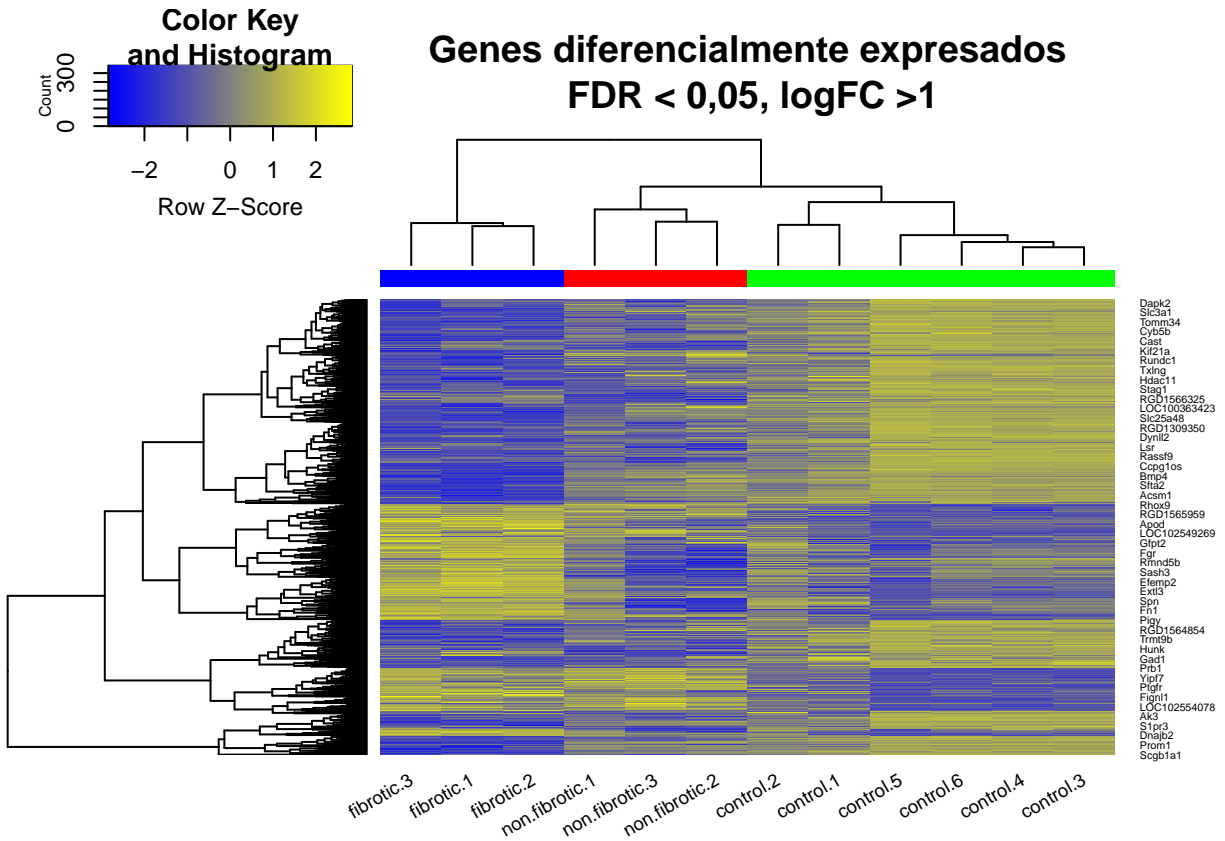


Figure 8: Heatmap de expresion con los genes y muestras agrupado por similitud

Para el análisis de significación biológica se ha utilizado el método **Gene Set Enrichment Analysis** a través de la herramienta *Reactome Pathway Analysis*. A continuación se muestra, una tabla y un gráfico de tipo *Barplot* con las primeras vías celulares ordenadas según su *p-valor ajustado* para cada comparación realizada

Table 5: Reactome Pathway Analysis para la comparacion FIBvsCONTROL

ID	Description	GeneRatio	BgRatio	pvalue	p.adjust
R-RNO-5620924	Intraflagellar transport	28/1196	47/7447	0.00e+00	0.0000000
R-RNO-1852241	Organelle biogenesis and maintenance	65/1196	182/7447	0.00e+00	0.0000000
R-RNO-5617833	Cilium Assembly	59/1196	160/7447	0.00e+00	0.0000000
R-RNO-1474244	Extracellular matrix organization	58/1196	210/7447	1.17e-05	0.0027965
R-RNO-70895	Branched-chain amino acid catabolism	11/1196	18/7447	1.87e-05	0.0035785

Reactome Pathway Analysis para la comparacion FIBvsCONTROL

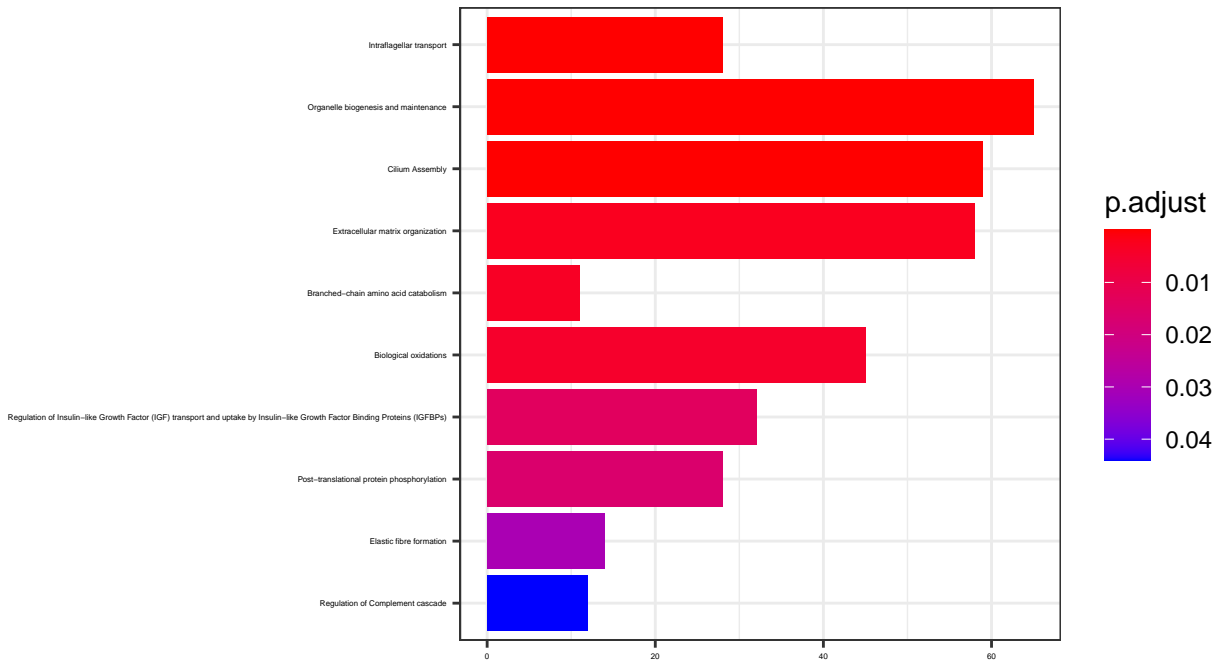


Figure 9: Barplot Reactome - FIBvsCONTROL

Table 6: Reactome Pathway Analysis para la comparacion NONFIBvsCONTROL

ID	Description	GeneRatio	BgRatio	pvalue	p.adjust
R-RNO-1852241	Organelle biogenesis and maintenance	52/900	182/7447	0.00e+00	0.0000010
R-RNO-5617833	Cilium Assembly	46/900	160/7447	0.00e+00	0.0000038
R-RNO-5620924	Intraflagellar transport	21/900	47/7447	0.00e+00	0.0000076
R-RNO-1428517	The citric acid (TCA) cycle and respiratory electron transport	30/900	120/7447	6.47e-05	0.0152798
R-RNO-70895	Branched-chain amino acid catabolism	9/900	18/7447	9.29e-05	0.0175441

Reactome Pathway Analysis para la comparacion NONFIBvsCONTROL

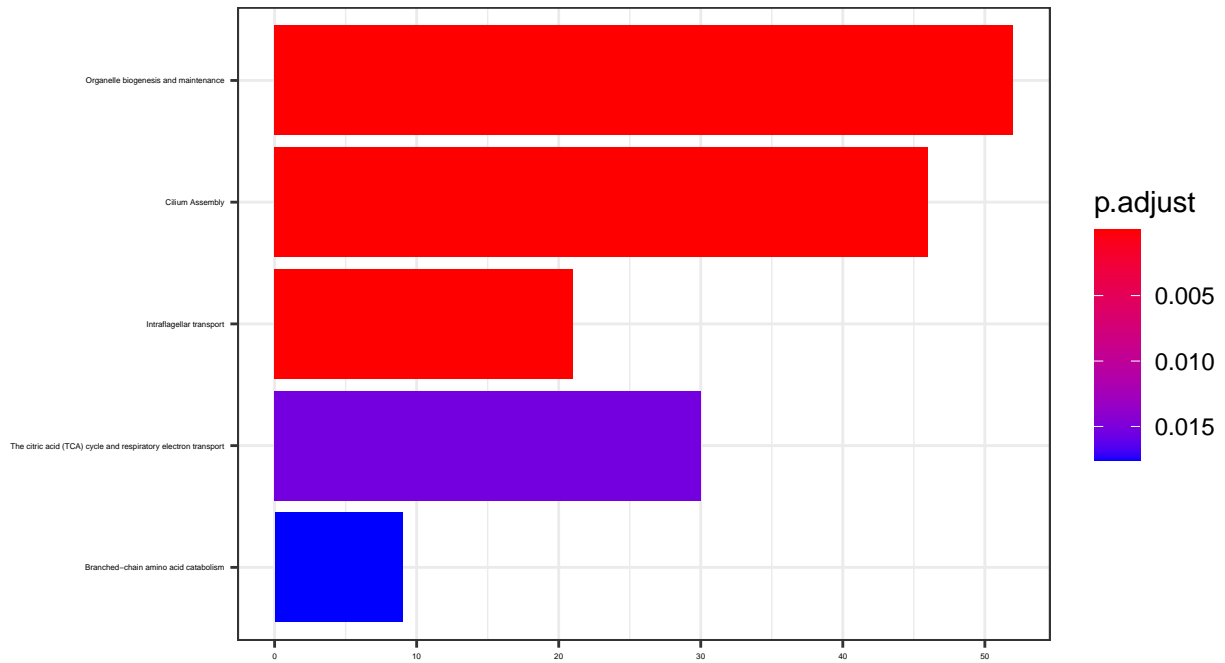


Figure 10: Barplot Reactome - NONFIBvsCONTROL

Table 7: Reactome Pathway Analysis para la comparacion FIBvsNONFIB

ID	Description	GeneRatio	BgRatio	pvalue	p.adjust
R-RNO-5620924	Intraflagellar transport	24/694	47/7447	0e+00	0.0000000
R-RNO-5617833	Cilium Assembly	42/694	160/7447	0e+00	0.0000001
R-RNO-1852241	Organelle biogenesis and maintenance	42/694	182/7447	0e+00	0.0000045
R-RNO-1474244	Extracellular matrix organization	46/694	210/7447	0e+00	0.0000045
R-RNO-202427	Phosphorylation of CD3 and TCR zeta chains	8/694	11/7447	7e-07	0.0001162

Reactome Pathway Analysis para la comparacion FIBvsNONFIB

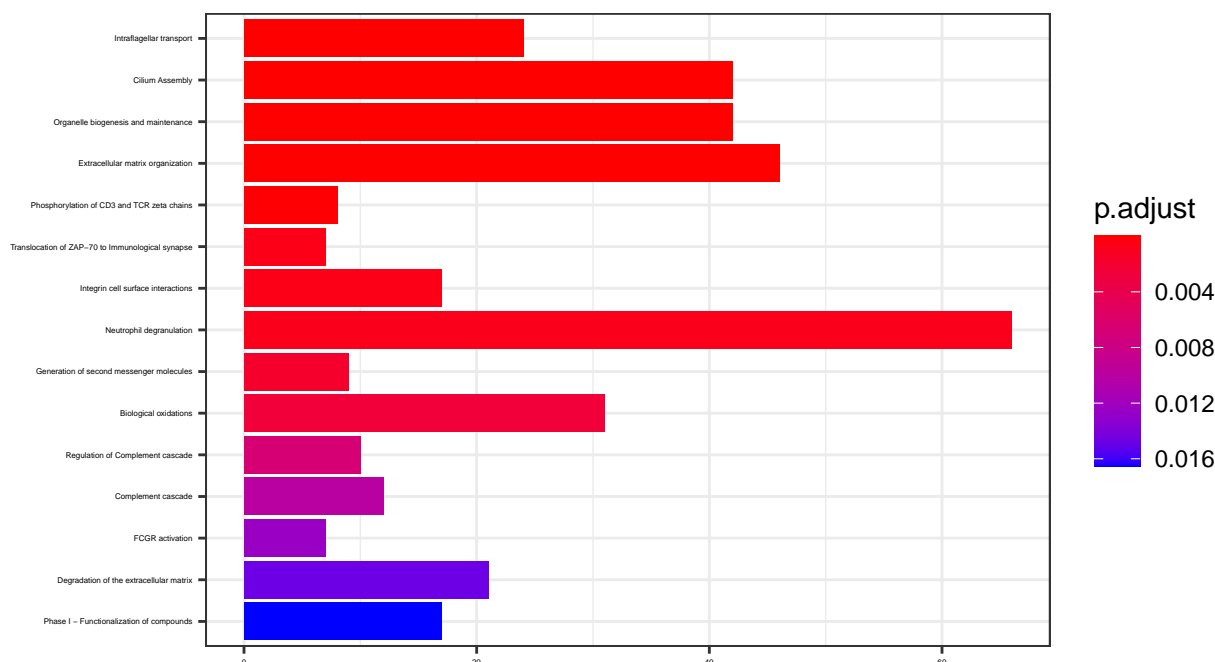


Figure 11: Barplot Reactome - FIBvsNONFIB

Se observa que las vías implicadas con el sistema ciliar y con la matriz extracelular están presente en los tres contrastes. Cabe destacar que en los contrastes de las muestras con fibrosis, también se ven implicadas vías relacionadas con mecanismos inflamatorios (degranulación de neutrofilos, activación del complemento, señalización de citoquinas).

5 Discusión

En el presente trabajo se analizó las alteraciones de expresión génica que se pueden encontrar en un modelo en ratas de Bronquiolitis Obliterante. La principal debilidad de este trabajo es el pequeño muestral ya que solo se obtuvieron 3 muestras de tejido fibrótico. También habría que reconsiderar dos muestras de tejidos control que durante el procesamiento de datos se observó que podría ser posibles muestras atípicas.

En el trabajo se detectaron alrededor de 2000 genes expresados diferencialmente en el tejido fibroso. Se mostró que vías relacionadas con mecanismos inflamatorios podrían estar implicadas en la fisiopatología de la Bronquiolitis Obliterante según el modelo animal empleado, por lo que habría que ampliar el estudio de estas vías para corroborar dicha hipótesis.

6 Apéndice

Código R utilizado para realizar el análisis

```
library(knitr)
library(colorspace)
library(gplots)
library(ggplot2)
library(ggrepel)
library(htmlTable)
```

```

library(devtools)
library(BiocManager)
library(Biobase)
library(oligo)
library(arrayQualityMetrics)
library(pvca)
library(limma)
library(genefilter)
library(rat2302.db)
library(annotate)
library(org.Rn.eg.db)
library(ReactomePA)
library(reactome.db)
library(magrittr)
library(kableExtra)

# Genero el 'ExpressionSet'
celFiles <- list.celfiles("./Data", full.names = TRUE)
my.targets <- read.AnnotatedDataFrame(file.path("./Data",
  "targets.csv"), header = TRUE, row.names = 1, sep = ";")
rawData <- read.celfiles(celFiles, phenoData = my.targets)
rownames(pData(rawData)) <- my.targets@data$shortname
colnames(rawData) <- rownames(pData(rawData))
myInfo <- new("MIAME", name = "Joaquin Astete Frugoni",
  lab = "UOC-ADO", contact = "jastete0@uoc.edu",
  title = "PEC 1")
experimentData(rawData) <- myInfo

# Utilizo la funcion 'arrayQualityMetrics' para la busqueda de outliers
arrayQualityMetrics(rawData, outdir = file.path("./arrayquality.raw"),
  force = T)
knitr::include_graphics("figures/quality-raw.png")

# Genero graficos para valorar la calidad de los datos

boxplot(rawData, cex.axis = 0.5, las = 2, which = "all",
  col = c(rep("red", 2), rep("blue", 3), rep("green",
  3), rep("red", 4)),
  main = "Distribucion de valores de intensidad")

plotPCA3 <- function(datos, labels, factor, scale,
  colores, size = 1.5, glineas = 0.25) {
  data <- prcomp(t(datos), scale = scale)
  dataDf <- data.frame(data$x)
  Group <- factor
  loads <- round(data$sdev^2/sum(data$sdev^2) * 100,
    1)
  p1 <- ggplot(dataDf, aes(x = PC1, y = PC2)) + theme_classic() +
    geom_hline(yintercept = 0, color = "gray70") +
    geom_vline(xintercept = 0, color = "gray70") +
    geom_point(aes(color = Group), alpha = 0.55,
      size = 3) + coord_cartesian(xlim = c(min(data$x[,

```

```

1]) - 5, max(data$x[, 1]) + 5)) + scale_fill_discrete(name = "Grupo")
p1 + geom_text_repel(aes(y = PC2 + 0.25, label = labels),
  segment.size = 0.25, size = size) + labs(x = c(paste("PC1",
  loads[1], "%")), y = c(paste("PC2", loads[2],
  "%")))) + ggtitle(paste("Análisis de Componente Principal")) +
  theme(plot.title = element_text(hjust = 0.5))
  scale_color_manual(values = colores)
}
plotPCA3(exprs(rawData), labels = pData(rawData)$shortname,
  factor = pData(rawData)$group, scale = FALSE, size = 3,
  colores = c("red", "blue", "green"))

# Normalizo los datos con el metodo 'RMA'
eset_rma <- rma(rawData)

# Valoro la calidad de los datos normalizados

boxplot(eset_rma, cex.axis = 0.5, las = 2, col = c(rep("red",
2), rep("blue", 3), rep("green", 3), rep("red",
4)),
  main = "Distribucion de valores de intensidad: datos normalizados")
arrayQualityMetrics(eset_rma, outdir = file.path("./arrayquality.rma"),
  force = T)
knitr::include_graphics("figures/quality-rma.png")

plotPCA3(exprs(eset_rma), labels = pData(rawData)$shortname,
  factor = pData(rawData)$group, scale = FALSE, size = 3,
  colores = c("red", "blue", "green"))

# Realizo el filtraje de genes
annotation(eset_rma) <- "rat2302.db"
filtered <- nsFilter(eset_rma, require.entrez = TRUE,
  remove.dupEntrez = TRUE, var.filter = TRUE, var.func = IQR,
  var.cutoff = 0.75, filterByQuantile = TRUE, feature.exclude = "^AFFX")
eset_filtered <- filtered$eset

# Establezco la matriz de diseño y de contraste

designMat <- model.matrix(~0 + group, pData(eset_filtered))
colnames(designMat) <- c("control", "fibrotic", "non.fibrotic")
cont.matrix <- makeContrasts(FIBvsCONTROL = fibrotic -
  control, NONFIBvsCONTROL = non.fibrotic - control,
  FIBvsNONFIB = fibrotic - non.fibrotic, levels = designMat)

# Genero el modelo lineal
fit <- lmFit(eset_filtered, designMat)
fit.main <- contrasts.fit(fit, cont.matrix)
fit.main <- eBayes(fit.main)
res <- decideTests(fit.main, method = "separate", adjust.method = "fdr",
  p.value = 0.05, lfc = 1)
sum.res.rows <- apply(abs(res), 1, sum)
res.selected <- res[sum.res.rows != 0, ]
res.sum <- (summary(res))

```

```

# Tabla que muestra los genes diferenciados de cada contraste
kable(res.sum, "latex", booktabs = T,
      caption = " Genes especialmente diferenciados entre comparaciones\n
      FDR<0.05 y logFC <1") %>%
kable_styling(latex_options = "HOLD_position")

# Diagrama de Venn para los genes en comun entre contrasts
vennDiagram(res.selected[, 1:3], cex = 0.9)
title("Genes seleccionados en comun entre comparaciones\n FDR < 0.05 y logFC > 1",
      cex.main = 0.8)

# Genero las topTable
topTab_FIBvsCONTROL <- topTable(fit.main, number = nrow(fit.main),
  coef = "FIBvsCONTROL", adjust = "fdr")
topTab_NONFIBvsCONTROL <- topTable(fit.main, number = nrow(fit.main),
  coef = "NONFIBvsCONTROL", adjust = "fdr")
topTab_FIBvsNONFIB <- topTable(fit.main, number = nrow(fit.main),
  coef = "FIBvsNONFIB", adjust = "fdr")

# Agrega las anotaciones a las topTables
annotatedTopTable <- function(topTab, anotPackage) {
  topTab <- cbind(PROBEID = rownames(topTab), topTab)
  myProbes <- rownames(topTab)
  thePackage <- eval(parse(text = anotPackage))
  geneAnots <- select(thePackage, myProbes, c("SYMBOL",
    "ENTREZID", "GENENAME"))
  annotatedTopTab <- merge(x = geneAnots, y = topTab,
    by.x = "PROBEID", by.y = "PROBEID")
  annotatedTopTab <- annotatedTopTab[with(annotatedTopTab,
    order(adj.P.Val)), ]

  return(annotatedTopTab)
}

topAnnotated_FIBvsCONTROL <- annotatedTopTable(topTab_FIBvsCONTROL,
  anotPackage = "rat2302.db")
topAnnotated_NONFIBvsCONTROL <- annotatedTopTable(topTab_NONFIBvsCONTROL,
  anotPackage = "rat2302.db")
topAnnotated_FIBvsNONFIB <- annotatedTopTable(topTab_FIBvsNONFIB,
  anotPackage = "rat2302.db")

# Genero tablas con un seleccion de genes por contraste
col.sel <- c("PROBEID", "SYMBOL", "ENTREZID", "GENENAME",
  "adj.P.Val")
kable(topAnnotated_FIBvsCONTROL[1:10, col.sel], "latex",
  booktabs = T, row.names = FALSE,
  caption = "Genes diferencialmente expresados: Fibrosis-Control") %>%
kable_styling(latex_options = "hold_position")
kable(topAnnotated_NONFIBvsCONTROL[1:10, col.sel],
  "latex", booktabs = T, row.names = FALSE,

```

```

caption = "Genes diferencialmente expresados: No Fibrosis-Control") %>%
kable_styling(latex_options = "HOLD_position")
kable(topAnnotated_FIBvsNONFIB[1:10, col.sel], "latex",
booktabs = T, row.names = FALSE,
caption = "Genes diferencialmente expresados: Fibrosis-No Fibrosis") %>%
kable_styling(latex_options = "HOLD_position")

# Genero un Heatmap
probesInHeatmap <- rownames(res.selected)
HMdata <- exprs(eset_filtered)[rownames(exprs(eset_filtered)) %in%
  probesInHeatmap, ]
geneSymbols <- select(rat2302.db, rownames(HMdata),
  c("SYMBOL"))
SYMBOLS <- geneSymbols$SYMBOL
rownames(HMdata) <- SYMBOLS
my_palette <- colorRampPalette(c("blue", "yellow"))(n = 299)
heatmap.2(HMdata, Rowv = TRUE, Colv = TRUE, dendrogram = "both",
  main = "Genes diferencialmente expresados\n FDR < 0,05, logFC >1",
  scale = "row", col = my_palette, sepcolor = "white",
  sepwidth = c(0.05, 0.05), cexRow = 0.5, cexCol = 0.9,
  key = TRUE, keysize = 1.8, density.info = "histogram",
  ColSideColors = c(rep("green", 2), rep("red", 3),
    rep("blue", 3), rep("green", 4)), tracecol = NULL,
  srtCol = 30)

# Aplica el Reactome Pathway Analysis a cada contraste y genera
# Una tabla y barplot para cada resultado

listOfTables <- list(FIBvsCONTROL = topTab_FIBvsCONTROL,
  NONFIBvsCONTROL = topTab_NONFIBvsCONTROL,
  FIBvsNONFIB = topTab_FIBvsNONFIB)
listOfSelected <- list()
for (i in 1:length(listOfTables)) {
  topTab <- listOfTables[[i]]
  whichGenes <- topTab["adj.P.Val"] < 0.15
  selectedIDs <- rownames(topTab)[whichGenes]
  EntrezIDs <- select(rat2302.db, selectedIDs, c("ENTREZID"))
  EntrezIDs <- EntrezIDs$ENTREZID
  listOfSelected[[i]] <- EntrezIDs
  names(listOfSelected)[i] <- names(listOfTables)[i]
}
mapped_genes2GO <- mappedkeys(org.Rn.egGO)
mapped_genes2KEGG <- mappedkeys(org.Rn.egPATH)
mapped_genes <- union(mapped_genes2GO, mapped_genes2KEGG)

comparisonsNames <- names(listOfSelected)
universe <- mapped_genes
for (i in 1:length(listOfSelected)) {
  genesIn <- listOfSelected[[i]]
  comparison <- comparisonsNames[i]
  enrich.result <- enrichPathway(gene = genesIn,
    pvalueCutoff = 0.05, readable = T, pAdjustMethod = "BH",

```

```

    organism = "rat", universe = universe)
tab <- as.data.frame(enrich.result)[1:5, 1:6]
tabla <- kable(tab, "latex", booktabs = T, row.names = FALSE,
  caption = paste0("Reactome Pathway Analysis para la comparacion ",
    comparison)) %>% kable_styling(latex_options = c("scale_down",
  "HOLD_position"), full_width = F)
print(tabla)
cat("\n")
print(barplot(enrich.result,
  title = paste0("Reactome Pathway Analysis para la comparacion\n ",
    comparison),
  showCategory = 15, font.size = 3))
}

```

Referencias

- Benjamini, Yoav, and Yosef Hochberg. 1995. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Source Journal of the Royal Statistical Society. Series B (Methodological)* 57 (1): 289–300. <http://www.jstor.org/stable/2346101> <http://www.jstor.org/http://www.jstor.org/action/showPublisher?publisherCode=black>.
- Irizarry, Rafael A., Bridget Hobbs, Francois Collin, Yasmin D. Beazer-Barclay, Kristen J. Antonellis, Uwe Scherf, and Terence P. Speed. 2003. "Exploration, normalization, and summaries of high density oligonucleotide array probe level data." *Biostatistics* 4 (2): 249–64. doi:10.1093/biostatistics/4.2.249.
- Morgan, Daniel L, B Alex Merrick, Kevin E Gerrish, Patricia S Stockton, Yu Wang, Julie F Foley, William M Gwinn, et al. 2015. "Gene Expression in Obliterative Bronchiolitis-Like Lesions in 2, 3-Pentanedione-Exposed Rats." *PloS One* 10 (2). Public Library of Science.
- Villa, M Carmen Ruiz de, and Àlex Sànchez. 2013. *Genómica Funcional Y análisis de Microarrays*. UOC.