

# ballester\_gabriel\_ADO\_PEC1

Gabriel Ballester Lozano

4/21/2020

## Contents

<b>Abstract</b>	<b>1</b>
<b>Objetivos</b>	<b>1</b>
<b>Materiales, Métodos y Resultados</b>	<b>2</b>
Control de calidad . . . . .	2
Normalización de los datos . . . . .	5
Control de calidad de los datos normalizados . . . . .	5
Filtraje no específico . . . . .	8
Identificación de genes diferencialmente expresados . . . . .	8
Filtrado de los genes menos variables . . . . .	8
Matriz de diseño . . . . .	11
Identificación de genes diferencialmente expresados . . . . .	12
Anotación de los resultados obtenidos . . . . .	13
Comparación entre distintas comparaciones . . . . .	13
Análisis de significación biológica . . . . .	17
<b>Discusión</b>	<b>19</b>
<b>Conclusión</b>	<b>20</b>

## Abstract

En el presente trabajo como primera PEC de la asignatura “Análisis de Datos Ómicos”. La neumonía adquirida en la comunidad es una enfermedad generalizada con morbilidad y mortalidad significativas. Los macrófagos alveolares son células pulmonares residentes en los tejidos que juegan un papel crucial en la inmunidad innata contra las bacterias que causan neumonía. Se pretende analizar cómo los macrófagos alveolares de ratones muestran características adaptativas después de la resolución de la neumonía bacteriana. Los perfiles transcriptómicos revelaron que los macrófagos alveolares de ratones que se recuperaron de la neumonía tenían nuevas actividades de referencia y respuestas alteradas a la infección.

## Objetivos

El estudio que he escogido para realizar la presente PEC es el estudio GSE133975, publicado el 17 de abril de 2020 y titulado “Reprogrammed alveolar macrophages after pneumonia recovery”.

En el estudio usaron microarrays para detallar la expresión génica de macrófagos alveolares de ratones infectados con la bacteria que causa la neumonía denominada *pneumococcus* -aunque en las identificaciones de las muestra contendrá el valor de SP3-, que se dividen en dos grupos (cada uno con su control); ratones “naive” y ratones ya infectados y recuperados pasado un mes. A ambos ratones se les tomó una muestra de pulmón para el análisis de macrófagos, que son las células objetivo de este estudio.

- Los objetivos del presente trabajo son:
  - Realizar un análisis completo de los datos contenidos de los Arrays publicados en el estudio “Reprogrammed alveolar macrophages after pneumonia recovery”.
  - Analizar las comparaciones de perfiles de expresión génica en macrófagos alveolares de ratones infectados con SP3.
  - Comparar los resultados obtenidos en el presente estudio con algunos de los del estudio de referencia.

## Materiales, Métodos y Resultados

El organismo de estudio es *Mus musculus* y el diseño del estudio es comparativo, cuyo diseño se basa en 16 Arrays de Affymetrix clasificados en dos factores; experimental y tipo de célula. En el factor experimental tiene dos niveles que codifican a los Arrays control (ratones inyectados con suero solamente) y Arrays con SP3 a los que se les ha inyectado la bacteria.

El presente archivo y sus resultados se hallan disponibles en el repositorio de Github [https://github.com/GABRIELBALLESTER/Ballester\\_Gabriel\\_ADO\\_PEC1.git](https://github.com/GABRIELBALLESTER/Ballester_Gabriel_ADO_PEC1.git).

Para obtener los datos del estudio se ha preferido realizar un archivo de targets propio para simplificarlo dada la extensión de la lista de phenodata:

```
> #elist <- getGEO("GSE133975")
> 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)
> my.targets@data$ShortName->rownames(pData(rawData))
> colnames(rawData) <-rownames(pData(rawData))
```

La plataforma de Affymetrix utilizada para este estudio es [MoGene-2\_0-st] Affymetrix Mouse Gene 2.0 ST Array mogene20st\_Mm\_ENTREZG\_17.1.0 basada en oligonucleótidos *in situ*. Por ello, cabe esperar que el paquete de anotaciones de bases de datos de bioconductor “mogene21sttranscriptcluster.db” sea el adecuado para las anotaciones de los genes correspondientes.

## Control de calidad

Al crear el directorio “arrayQualityMetrics\_report\_for\_rawData” se observa en el archivo index.html (Figura 1) que tan sólo hay una marca en algunos arrays por lo que se puede decidir seguir adelante con todos;

```
> #arrayQualityMetrics(rawData)
```

Sin embargo, para un control de calidad de los datos brutos más exhaustivo, se procede con el análisis de componentes principales o PCA (Figura 2);

Tras ver estos resultados se comprueban que los etiquetados están correctamente, puesto que hay dos Arrays que se separan bastante del resto (E.C.3 y E.SP3.4), ambos pertenecientes al tipo “experienced” pero no del mismo grupo experimental. Estos resultados también se encuentran en la carpeta “arrayQualityMetrics\_report\_for\_rawData”. Quizá deberían eliminarse de los posteriores análisis puesto que estos dos Arrays se encuentran a ambos lados del eje que explica el 49.6% de la variabilidad, no obstante se continua teniéndolos en cuenta y efectuando el análisis de boxplot (Figura 3).

En el Boxplot se observa, una vez más, que E.C.3 y E.SP3.4 son con diferencia los que más variabilidad contienen. Lo siguiente sería, por tanto, ver si esta variabilidad es fruto de un error técnico o si se pueden realizar comparaciones con la normalización de los datos.

	array	sampleNames	*1	*2	*3	Group	Experimental	Cell.type	ShortName
<input type="checkbox"/>	1	N.C.1				N.C	Control	Naive	N.C.1
<input type="checkbox"/>	2	N.C.2				N.C	Control	Naive	N.C.2
<input checked="" type="checkbox"/>	3	N.C.3			x	N.C	Control	Naive	N.C.3
<input type="checkbox"/>	4	N.C.4				N.C	Control	Naive	N.C.4
<input type="checkbox"/>	5	N.SP3.1				N.SP3	SP3	Naive	N.SP3.1
<input type="checkbox"/>	6	N.SP3.2				N.SP3	SP3	Naive	N.SP3.2
<input type="checkbox"/>	7	N.SP3.3				N.SP3	SP3	Naive	N.SP3.3
<input checked="" type="checkbox"/>	8	N.SP3.4			x	N.SP3	SP3	Naive	N.SP3.4
<input type="checkbox"/>	9	E.C.1				E.C	Control	Experienced	E.C.1
<input type="checkbox"/>	10	E.C.2				E.C	Control	Experienced	E.C.2
<input checked="" type="checkbox"/>	11	E.C.3	x			E.C	Control	Experienced	E.C.3
<input checked="" type="checkbox"/>	12	E.C.4			x	E.C	Control	Experienced	E.C.4
<input type="checkbox"/>	13	E.SP3.1				E.SP3	SP3	Experienced	E.SP3.1
<input checked="" type="checkbox"/>	14	E.SP3.2		x		E.SP3	SP3	Experienced	E.SP3.2
<input checked="" type="checkbox"/>	15	E.SP3.3			x	E.SP3	SP3	Experienced	E.SP3.3
<input checked="" type="checkbox"/>	16	E.SP3.4			x	E.SP3	SP3	Experienced	E.SP3.4

Figure 1: Tabla resumen de los datos del estudio

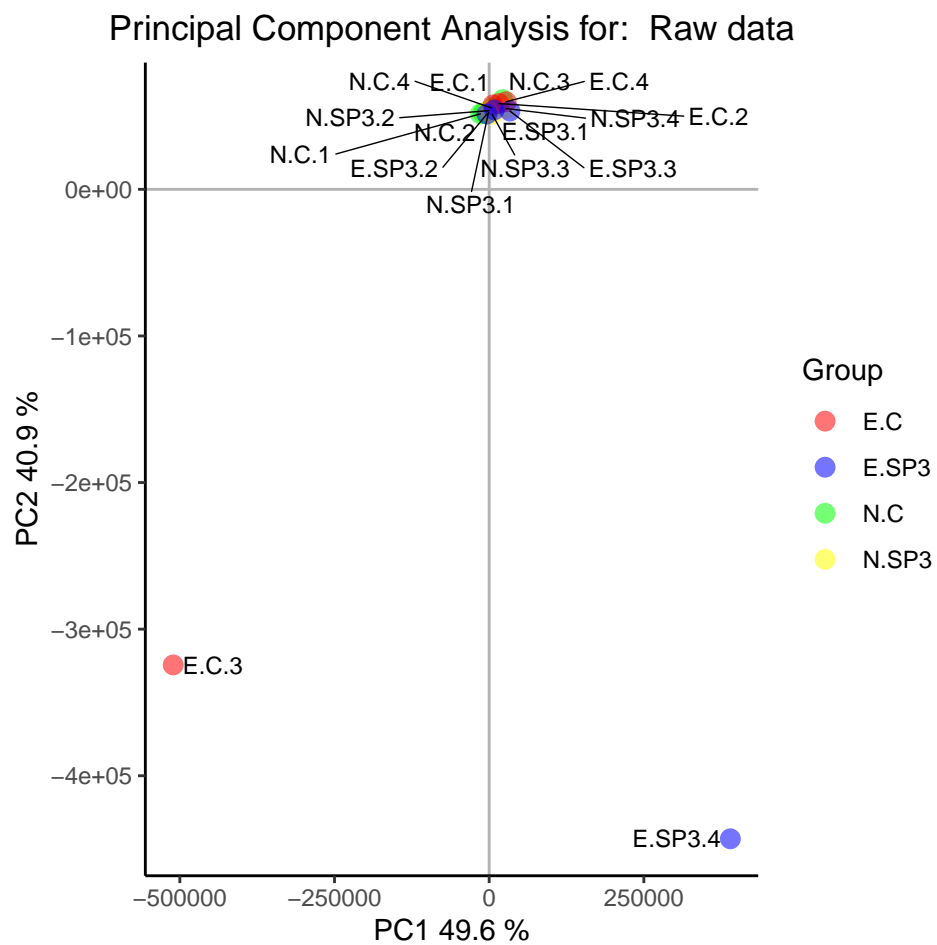


Figure 2: Análisis de los Componentes Principales de los Arrays con datos brutos

## Distribución de intensidad de los Arrays

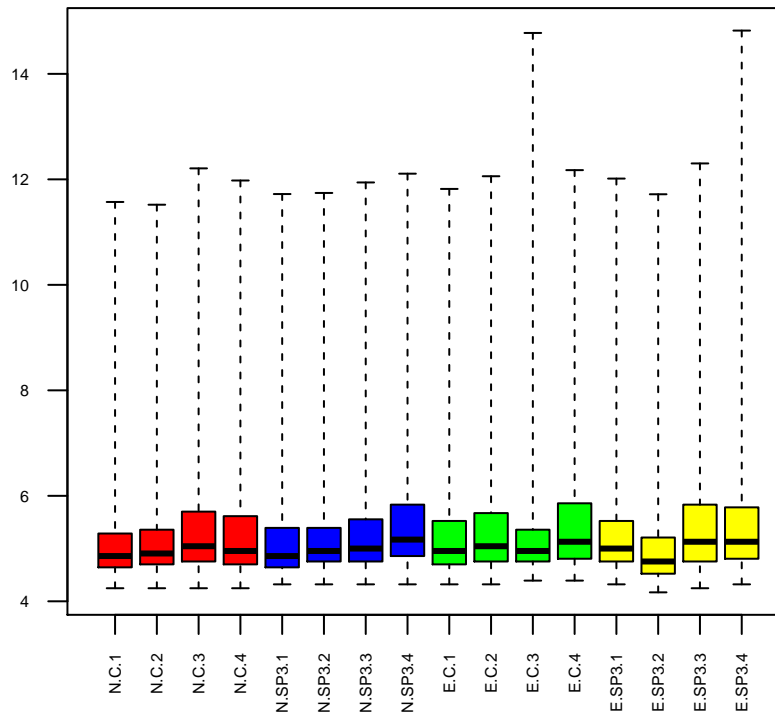


Figure 3: Boxplot de los Arrays con datos brutos

## Normalización de los datos

Para efectuar la normalización se empleó el método más indicado para microarrays de Affymetrix, el RMA de Bioconductor que es el método estándar:

```
> eset_rma <- rma(rawData)
```

```
Background correcting  
Normalizing  
Calculating Expression
```

## Control de calidad de los datos normalizados

Al efectuar el análisis de control de calidad de los datos una vez normalizados (Figura 4) se observa que se verifica la disposición de “outlier” el Array E.C.3. Sin embargo, el array E.SP3.4 ya no aparece en la lista. Estos resultados pueden corroborarse con otro Análisis de Componentes Principales para los datos estandarizados (Figura 5).

```
> #arrayQualityMetrics(eset_rma, outdir = file.path("./results", "QCDir.Norm"), force=TRUE)
```

Con los datos normalizados se observa que el eje que explica el 22.9% de la variabilidad separa a los controles de los casos inoculados con SP3. Además se puede apreciar que el Array E.C.3 sigue estando bastante aislado del resto de sus pares. Por otra parte el eje que explica un 18.2% de la variabilidad de los Arrays también separa, aunque con una menor claridad, los casos “naive” (no expuestos previamente a la bacteria) de los experimentados (ratones inoculados con SP3 y recuperados previamente al experimento).

En este análisis se pueden apreciar mucha más similitud en las intensidades de todos los Arrays (a excepción una vez mas del Array E.C.3).

	array	sampleNames	<u>*1</u>	<u>*2</u>	<u>*3</u>	Group	Experimental	Cell.type	ShortName
<input type="checkbox"/>	1	N.C.1				N.C	Control	Naive	N.C.1
<input type="checkbox"/>	2	N.C.2				N.C	Control	Naive	N.C.2
<input type="checkbox"/>	3	N.C.3				N.C	Control	Naive	N.C.3
<input type="checkbox"/>	4	N.C.4				N.C	Control	Naive	N.C.4
<input type="checkbox"/>	5	N.SP3.1				N.SP3	SP3	Naive	N.SP3.1
<input type="checkbox"/>	6	N.SP3.2				N.SP3	SP3	Naive	N.SP3.2
<input type="checkbox"/>	7	N.SP3.3				N.SP3	SP3	Naive	N.SP3.3
<input type="checkbox"/>	8	N.SP3.4				N.SP3	SP3	Naive	N.SP3.4
<input type="checkbox"/>	9	E.C.1				E.C	Control	Experienced	E.C.1
<input type="checkbox"/>	10	E.C.2				E.C	Control	Experienced	E.C.2
<input checked="" type="checkbox"/>	11	E.C.3	x	x		E.C	Control	Experienced	E.C.3
<input type="checkbox"/>	12	E.C.4				E.C	Control	Experienced	E.C.4
<input type="checkbox"/>	13	E.SP3.1				E.SP3	SP3	Experienced	E.SP3.1
<input checked="" type="checkbox"/>	14	E.SP3.2		x		E.SP3	SP3	Experienced	E.SP3.2
<input type="checkbox"/>	15	E.SP3.3				E.SP3	SP3	Experienced	E.SP3.3
<input type="checkbox"/>	16	E.SP3.4				E.SP3	SP3	Experienced	E.SP3.4

Figure 4: Tabla resumen de los datos del estudio normalizados

### Principal Component Analysis for: Datos normalizados

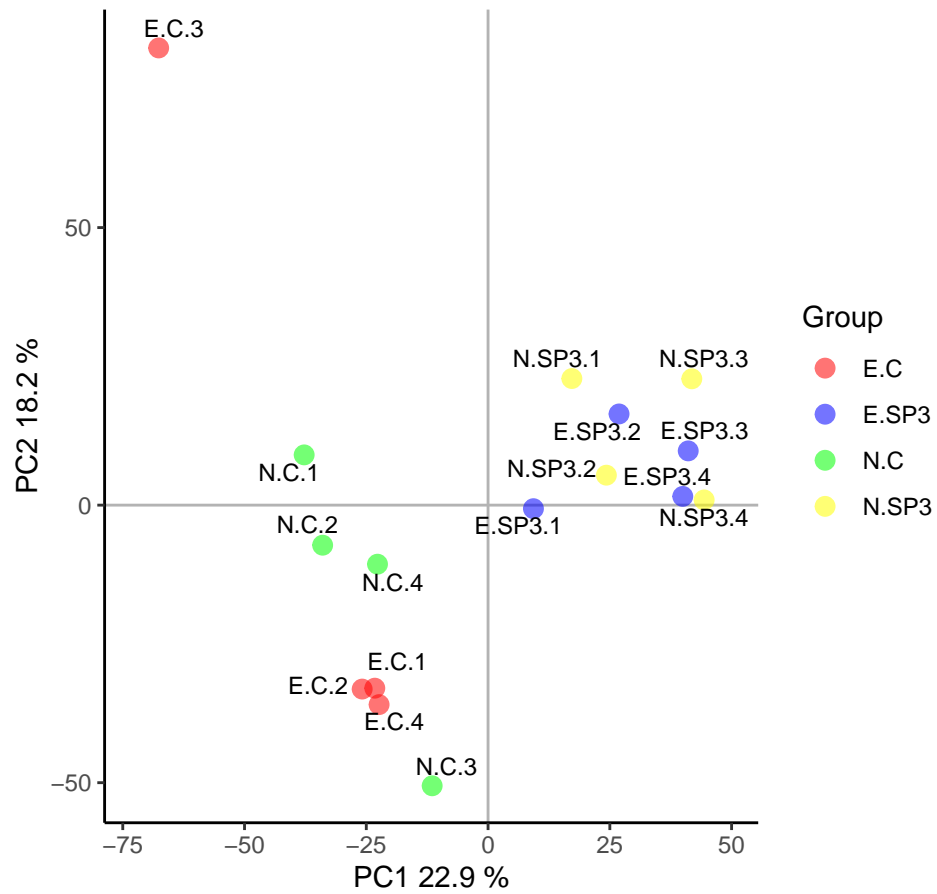


Figure 5: Análisis de Componentes Principales de los datos normalizados

## Datos normalizados

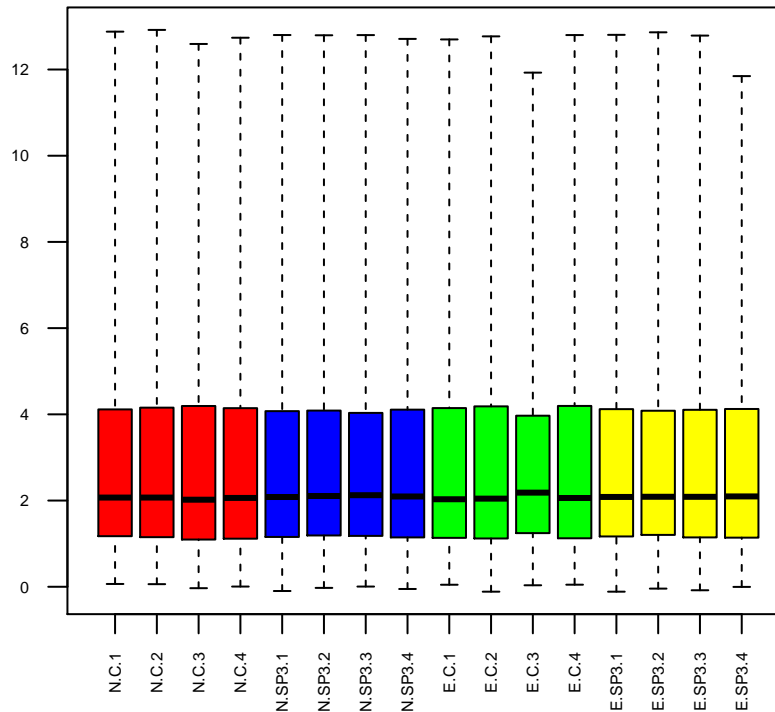


Figure 6: Boxplot de los datos normalizados

### Filtraje no específico

El análisis del posible ruido de fondo que puedan tener los diferentes Arrays del estudio se ha efectuado mediante el análisis de componentes principales de variación (Figura 7) con un threshold de 0.6.

Como se puede observar, existe una mayor variación no biológica en este caso que es la residual. Esto hace suponer que deberían eliminarse al menos los Arrays que más ruido puedan presentar (por ejemplo el Array E.C.3). En segundo lugar hemos corroborado que es la variabilidad explicada por el grupo experimental (diferencias entre control y experienced). Y por último, un 16% de la variabilidad sería explicada por el factor de interés en este estudio que no es otro que la diferencia entre las células “naive” y las experimentadas.

### Identificación de genes diferencialmente expresados

Tras los resultados del apartado anterior, hay serias sospechas de que la variabilidad total de las muestras pueda enmascarar las diferencias entre los distintos grupos experimentales. Por este motivo se procede a realizar un análisis de la variabilidad de todos los genes para ver qué porcentaje de genes pueden mostrar una variabilidad distinta a la variabilidad genérica de las muestras (Figura 8).

De esta manera se puede apreciar que existe un gran número de genes que presentan una desviación estándar mayor a 1.0, aunque no llega a ser el 5% de los genes.

### Filtrado de los genes menos variables

El siguiente paso a realizar es tratar de eliminar todos aquellos genes que puedan provocar una mayor distorsión a la hora de realizar los análisis comparativos, para ello se utilizó el siguiente script:

```
> annotation(eset_rma) <- "mogene21sttranscriptcluster.db"
> filtered <- nsFilter(eset_rma,
```



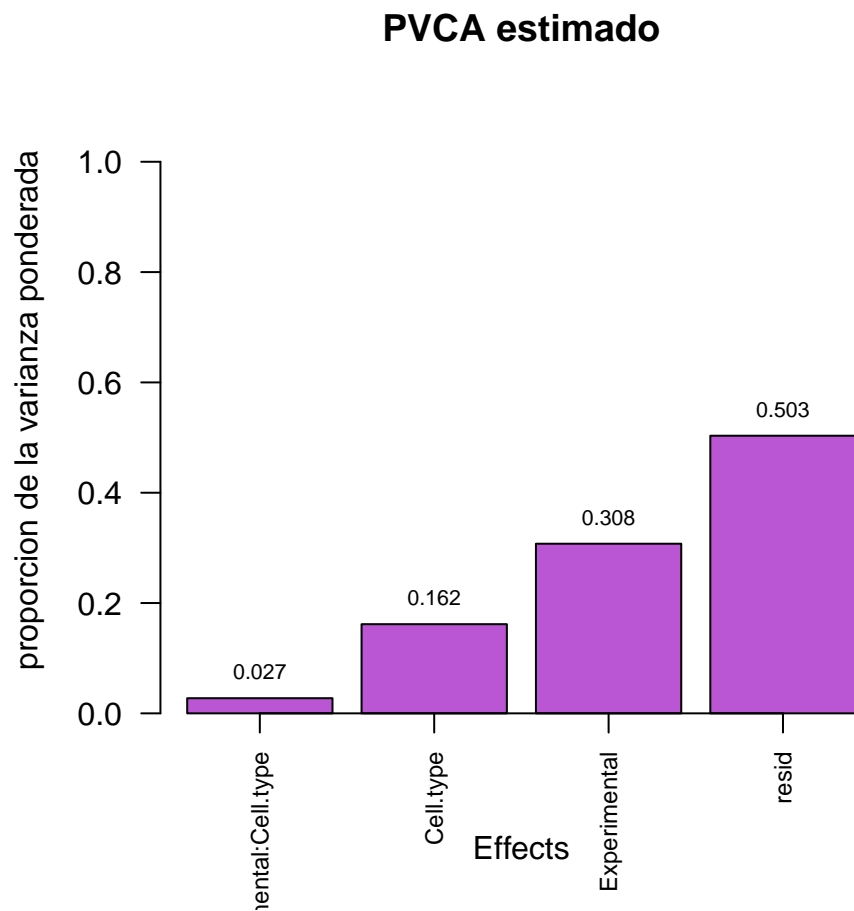


Figure 7: Importancia relativa de los factores Grupo experimental y Tipo celular

## Distribución de la variabilidad de los genes

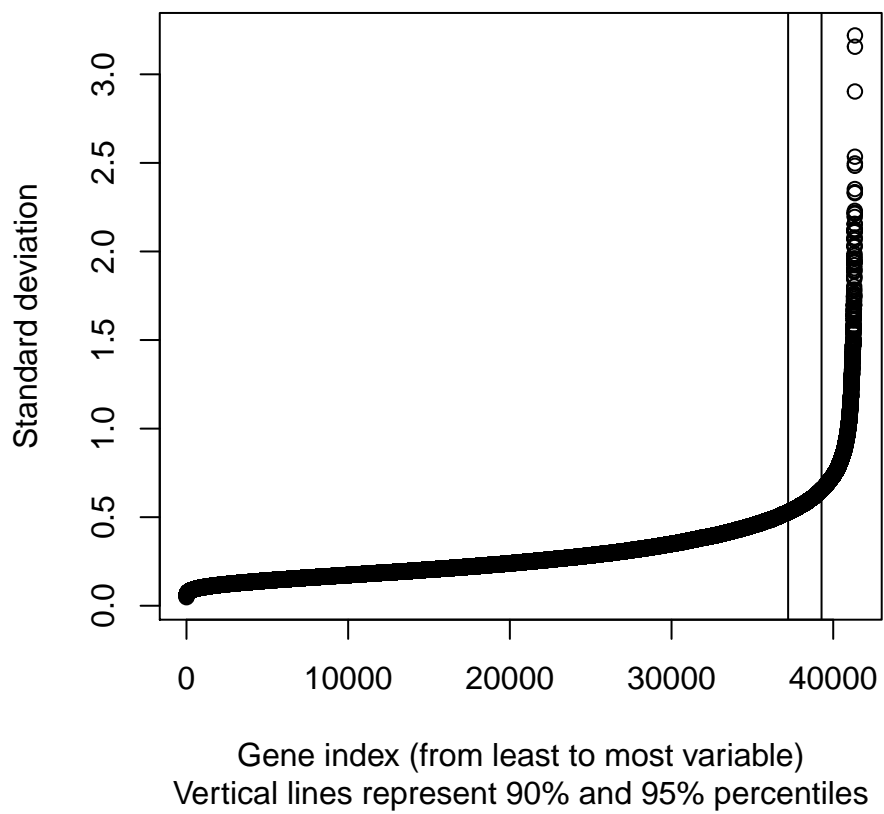


Figure 8: Valores de desviaciones estándar de todas las muestras para todos los genes

```
+ require.entrez = TRUE, remove.dupEntrez = TRUE,
+ var.filter=TRUE, var.func=IQR, var.cutoff=0.75,
+ filterByQuantile=TRUE, feature.exclude = "^AFFX")
```

```
$numDupsRemoved
[1] 671
```

```
$numLowVar
[1] 17973
```

```
$numRemoved.ENTREZID
[1] 16710
```

Con lo que, con esta función se han filtrado los genes, eliminando un total de 16710 lo cual es bastante considerable y de suponer teniendo en cuenta los análisis previos. Los archivos resultantes de este análisis se guardaron en archivos .csv del siguiente modo;

```
> write.csv(exprs(eset_rma), file="./results/normalized.Data.csv")
> write.csv(exprs(eset_filtered), file="./results/normalized.Filtered.Data.csv")
> save(eset_rma, eset_filtered, file="./results/normalized.Data.Rda")
```

## Matriz de diseño

En el presente estudio se pretende una comparación de los macrófagos *naive* de los macrófagos experimentados. Además, cada grupo experimental cuenta con su control, por lo que la matriz de diseño elaborada es la siguiente:

	E.C	E.SP3	N.C	N.SP3
GSM3931536.CEL	0	0	1	0
GSM3931537.CEL	0	0	1	0
GSM3931538.CEL	0	0	1	0
GSM3931539.CEL	0	0	1	0
GSM3931540.CEL	0	0	0	1
GSM3931541.CEL	0	0	0	1
GSM3931542.CEL	0	0	0	1
GSM3931543.CEL	0	0	0	1
GSM3931544.CEL	1	0	0	0
GSM3931545.CEL	1	0	0	0
GSM3931546.CEL	1	0	0	0
GSM3931547.CEL	1	0	0	0
GSM3931548.CEL	0	1	0	0
GSM3931549.CEL	0	1	0	0
GSM3931550.CEL	0	1	0	0
GSM3931551.CEL	0	1	0	0

```
attr("assign")
[1] 1 1 1 1
attr("contrasts")
attr("contrasts")$Group
[1] "contr.treatment"
```

En este estudio y dado el diseño del mismo, cabría esperar también anañizar el efecto memoria de los macrófagos (llamado aquí MEM) que establecen los macrófagos al experimentar previamente la infección del SP3. No obstante el principal objetivo es examinar qué efecto tiene la infección de SP3 en la expresión de genes de los macrófagos de ambos grupos:

```
> cont.matrix <- makeContrasts (Infec.E = E.C-E.SP3,
+                               Infec.N = N.C-N.SP3,
```

```

+ MEM = (E.C-E.SP3) - (N.C-N.SP3),
+ levels=designMat)
> print(cont.matrix)

```

```

      Contrasts
Levels Infec.E Infec.N MEM
E.C      1      0      1
E.SP3    -1      0     -1
N.C       0      1     -1
N.SP3     0     -1      1

```

## Identificación de genes diferencialmente expresados

Con el fin de obtener la identificación de genes diferencialmente expresados, se ha obtenido un listado de los mismos según la matriz de diseño y de contrastes. Para ello se han comprobado las pruebas de significación para cada gen y cada comparación:

```

> fit<-lmFit(eset_filtered, designMat)
> fit.main<-contrasts.fit(fit, cont.matrix)
> fit.main<-eBayes(fit.main)
> class(fit.main)

```

```

[1] "MArrayLM"
attr(,"package")
[1] "limma"

```

Y con la función “Toplevel” se podrán generar para cada contraste una lista de genes ordenados de mayor a menor diferencia de expresión:

Para los macrófagos experimentados;

	logFC	AveExpr	t	P.Value	adj.P.Val	B
17438987	-3.778553	6.270982	-26.41976	2.946531e-15	1.765267e-11	23.58938
17344309	-3.538061	9.426964	-17.26318	3.197329e-12	9.577600e-09	17.90463
17527016	-2.919567	5.690399	-14.45400	5.468506e-11	1.092061e-07	15.33612
17487457	-2.111985	5.969117	-13.82153	1.106635e-10	1.510013e-07	14.68127
17213192	-2.171798	6.812150	-13.67856	1.302712e-10	1.510013e-07	14.52887
17376153	-2.767861	4.017678	-13.54891	1.512281e-10	1.510013e-07	14.38923

Para los macrófagos “naive”;

	logFC	AveExpr	t	P.Value	adj.P.Val	B
17438987	-3.688876	6.270982	-25.79274	4.393042e-15	2.631871e-11	23.31176
17246091	-4.993739	4.321504	-14.87421	3.473426e-11	8.496398e-08	15.75903
17213192	-2.331613	6.812150	-14.68512	4.254581e-11	8.496398e-08	15.57222
17527016	-2.866409	5.690399	-14.19083	7.308629e-11	1.094650e-07	15.07134
17344309	-2.824384	9.426964	-13.78095	1.158905e-10	1.388599e-07	14.64171
17376153	-2.685823	4.017678	-13.14733	2.419115e-10	2.415486e-07	13.95072

Para el contraste a priori de menor interés, la diferencia entre los macrófagos experimentados y “naive”, o efecto memoria (MEM);

	logFC	AveExpr	t	P.Value	adj.P.Val	B
17447013	-1.966180	2.703197	-6.432623	6.154179e-06	0.03496504	1.19491666
17467415	1.756991	2.401136	6.102050	1.167252e-05	0.03496504	0.90511732
17221627	-2.509283	3.825326	-5.265138	6.291006e-05	0.12563140	0.08863356
17352517	-1.563086	1.970251	-5.079013	9.256584e-05	0.13864049	-0.10935220
17366992	-1.283150	3.172075	-4.628628	2.392466e-04	0.25879676	-0.61258183
17211369	-2.250983	3.568376	-4.591066	2.591855e-04	0.25879676	-0.65604486

## Anotación de los resultados obtenidos

Con el fin de correlacionar las etiquetas de los genes o ID. establecidas por Affimetrix con los *gene symbol* establecidos para la descripción de cada uno de los genes, se realizaron las siguientes instrucciones con el paquete de anotaciones *mogene21sttranscriptcluster.db*:

```
> topAnnotated_Infec.E <- annotatedTopTable(topTab_Infec.E,  
+ anotPackage="mogene21sttranscriptcluster.db")  
> topAnnotated_Infec.N <- annotatedTopTable(topTab_Infec.N,  
+ anotPackage="mogene21sttranscriptcluster.db")  
> topAnnotated_MEM <- annotatedTopTable(topTab_MEM,  
+ anotPackage="mogene21sttranscriptcluster.db")  
> write.csv(topAnnotated_Infec.E, file="./results/topAnnotated_Infec.E.csv")  
> write.csv(topAnnotated_Infec.N, file="./results/topAnnotated_Infec.N.csv")  
> write.csv(topAnnotated_MEM, file="./results/topAnnotated_MEM.csv")  
> show(head(topAnnotated_Infec.N[1:5,1:4]))
```

	PROBEID	SYMBOL	ENTREZID	GENENAME
1	17210912	Rb1cc1	12421	RB1-inducible coiled-coil 1
2	17211000	Rrs1	59014	ribosome biogenesis regulator 1
3	17211004	Adhfe1	76187	alcohol dehydrogenase, iron containing, 1
4	17211043	Sgk3	170755	serum/glucocorticoid regulated kinase 3
5	17211090	Cspp1	211660	centrosome and spindle pole associated protein 1

Y con el fin de visualizar los datos comparados de una manera más simple se diseñó un gráfico volcanoPlot (Figura 9) a modo de ejemplo de las diferencias de memoria entre los macrófagos infectados:

Estas diferencias encontradas en el volcanoPlot no son determinantes en este estudio, puesto que se centra en las expresiones de los macrófagos *naive* y experimentados principalmente.

## Comparación entre distintas comparaciones

En el estudio resulta ser interesante saber qué genes se han seleccionado en cada comparación. A veces, los genes biológicamente relevantes serán aquellos que se seleccionan en uno de ellos pero no en otros. Aunque el principal motivo de este estudio es la comparación entremacrófagos *naive* y experimentados principalmente, puede resultar de interés qué sucede en el resto de comparaciones con el efecto memoria (MEM). Por todo ello, se trató de ver si hay genes que están seleccionados entre las combinaciones de las comparaciones efectuadas del presente estudio.

Para ello se contó con el valor +1 para los sobreexpresados y el valor -1 para los infraexpresados el punto de corte para el análisis se define como “FDR < 0.1” y “logFC > 1” (+1 para valores de t-test > 0, FDR < punto de corte y -1 para valores de t-test < 0, FDR < punto de corte) tomando valores 0 para no diferencias significativas:

	Infec.E	Infec.N	MEM
Down	294	253	1
NotSig	5532	5471	5989
Up	165	267	1

Existe tan sólo un gen sobreexpresado y otro infraexpresado que comparten el efecto memoria con los macrófagos experimentados, esto es muy informativo y llama bastante la atención. Otra manera más simplificada y visual de representar esta información sería mediante un Diagrama de Venn (Figura 10):

En este diagrama se observa el número de genes expresados diferentemente de manera significativa, pero no se ven si están infra o sobreexpresados. Para ver la información con más detalle se procedió a realizar un Heatmap (Figura 11):

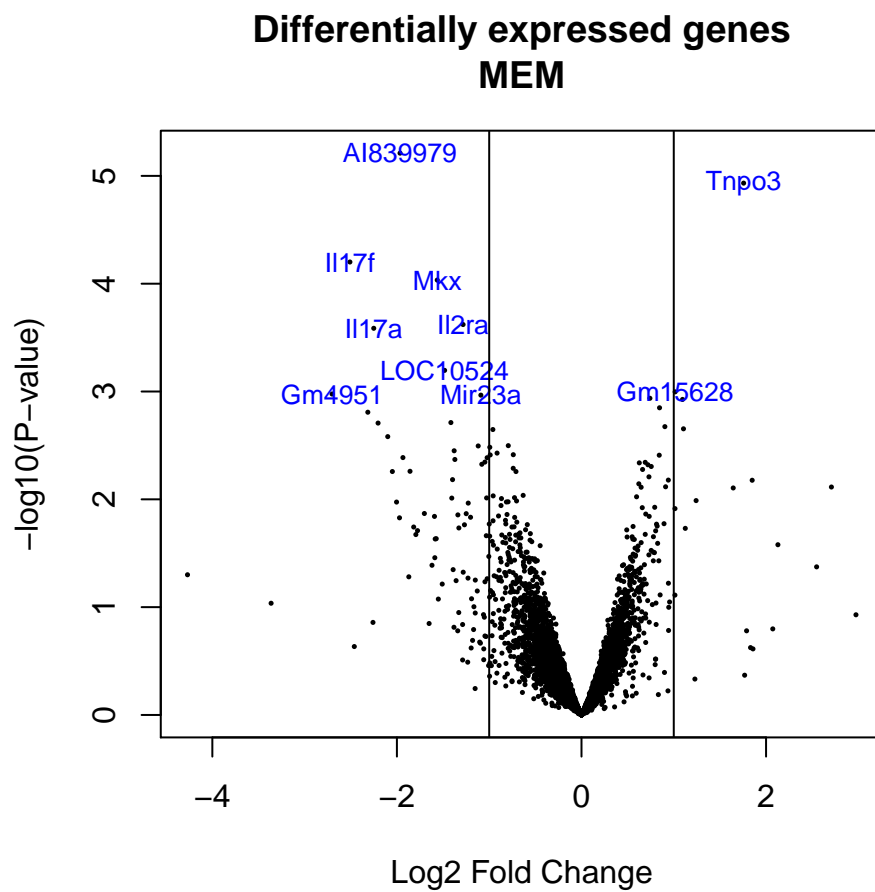


Figure 9: VolcanoPlot de las comparaciones entre macrófagos naive y experimentados. Los nombres de los 10 genes que se hallan en las primeras posiciones de la Toplevel se resaltan en azul

**Genes en común entre los tres grupos definidos**  
**Genes seleccionados mediante  $FDR < 0.1$  y  $\log FC >$**

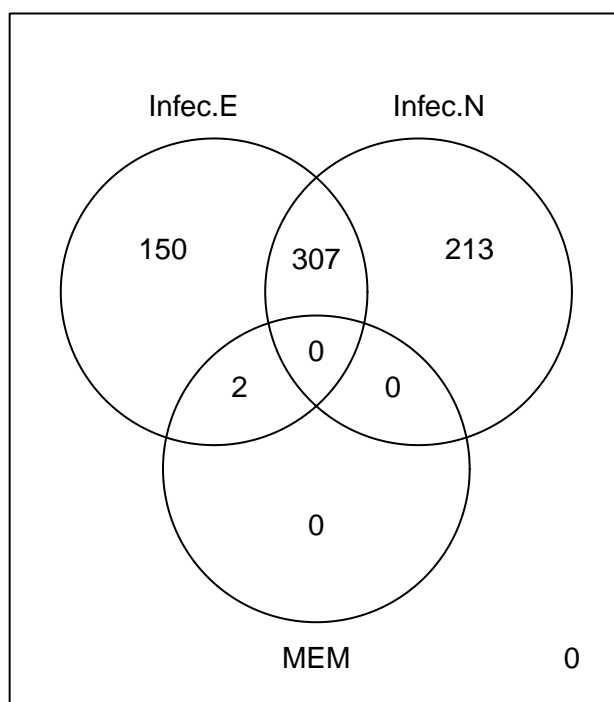


Figure 10: Genes en común visualizados en un Diagrama de Venn cuya expresión varía significativamente entre los tres contrastes establecidos en el estudio

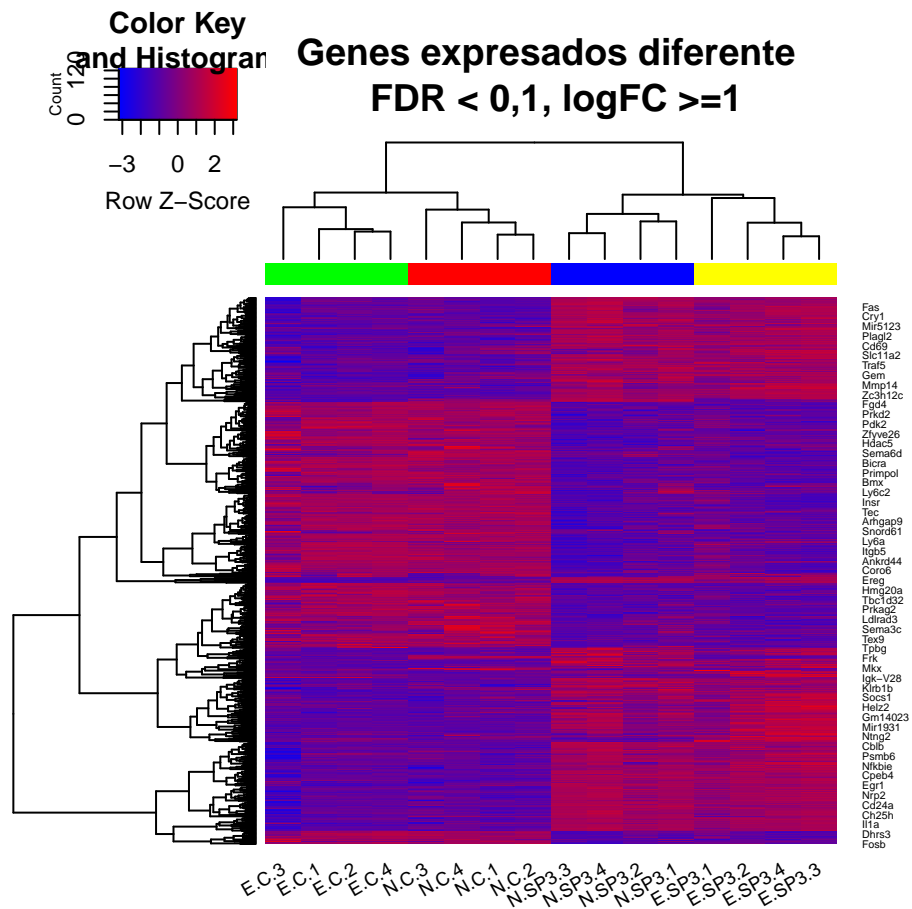


Figure 11: Heatmap de similaridad para expresión de genes (filas) y muestras (columnas)



## Análisis de significación biológica

Para tratar de vislumbrar qué vía metabólica tienen en común los gens que están diferenciados significativamente entre los grupos analizados, se trabajó con el paquete **ReactomePA** de Bioconductor con un FDR < 0.15 (menos restrictivo) para tratar de asegurar el listado de genes suficientes para poder llegar a coincidencias en vías metabólicas, mostando así los siguientes valores totales de genes para cada grupo (Figura 12):

```
Infec.E Infec.N      MEM
    2632    3023        4
```

```
#####
```

```
Comparison:  Infec.E
```

	ID	Description			
R-MMU-1483255	R-MMU-1483255	PI Metabolism			
R-MMU-168898	R-MMU-168898	Toll-like Receptor Cascades			
R-MMU-1660516	R-MMU-1660516	Synthesis of PIPs at the early endosome membrane			
R-MMU-449147	R-MMU-449147	Signaling by Interleukins			
R-MMU-168138	R-MMU-168138	Toll Like Receptor 9 (TLR9) Cascade			
R-MMU-5617833	R-MMU-5617833	Cilium Assembly			
	GeneRatio	BgRatio	pvalue	p.adjust	qvalue
R-MMU-1483255	27/1226	74/8772	1.003952e-06	0.0009828691	0.0008655124
R-MMU-168898	39/1226	133/8772	2.939510e-06	0.0013779552	0.0012134244
R-MMU-1660516	10/1226	15/8772	4.222539e-06	0.0013779552	0.0012134244
R-MMU-449147	63/1226	262/8772	6.640942e-06	0.0016253706	0.0014312978
R-MMU-168138	26/1226	85/8772	5.871189e-05	0.0107598123	0.0094750672
R-MMU-5617833	48/1226	198/8772	6.594369e-05	0.0107598123	0.0094750672

```
R-MMU-1483255
```

```
R-MMU-168898
```

```
R-MMU-1660516
```

```
R-MMU-449147  Il1a/Il23a/Nfkb2/Nfkb1/Nfkb1a/Il6/Nfkbib/Socs3/Il2ra/Mapkapk2/Syk/Rela/Irak3/Il20rb/Stat5b
```

```
R-MMU-168138
```

```
R-MMU-5617833  Ift140/Sd
```

	Count
R-MMU-1483255	27
R-MMU-168898	39
R-MMU-1660516	10
R-MMU-449147	63
R-MMU-168138	26
R-MMU-5617833	48

```
#####
```

```
Comparison:  Infec.N
```

	ID	Description			
R-MMU-69278	R-MMU-69278	Cell Cycle, Mitotic			
R-MMU-1483255	R-MMU-1483255	PI Metabolism			
R-MMU-1660516	R-MMU-1660516	Synthesis of PIPs at the early endosome membrane			
R-MMU-174417	R-MMU-174417	Telomere C-strand (Lagging Strand) Synthesis			
R-MMU-69190	R-MMU-69190	DNA strand elongation			
R-MMU-69242	R-MMU-69242	S Phase			
	GeneRatio	BgRatio	pvalue	p.adjust	qvalue
R-MMU-69278	122/1394	499/8772	2.311884e-07	0.0002323444	0.0002049059
R-MMU-1483255	29/1394	74/8772	1.081416e-06	0.0005434116	0.0004792381
R-MMU-1660516	10/1394	15/8772	1.385912e-05	0.0046428052	0.0040945190
R-MMU-174417	12/1394	22/8772	3.365045e-05	0.0067637400	0.0059649847
R-MMU-69190	12/1394	22/8772	3.365045e-05	0.0067637400	0.0059649847

R-MMU-69242 41/1394 142/8772 6.011900e-05 0.0100699319 0.0088807360

R-MMU-69278 Cdkn1b/Dna2/Cdc25b/Pds5b/Numa1/Optn/Vrk1/Sdccag8/Mad1l1/Rbl1/Pole2/Ccnd3/Cep192/Ncapd2/Rf

R-MMU-1483255

R-MMU-1660516

R-MMU-174417

R-MMU-69190

R-MMU-69242

	Count
R-MMU-69278	122
R-MMU-1483255	29
R-MMU-1660516	10
R-MMU-174417	12
R-MMU-69190	12
R-MMU-69242	41



Figure 12: Red obtenida del análisis Reactome enrichment de la lista obtenida de las comparaciones de Naive Experienced y MEM

Table 1: Primeras filas y columnas de los resultados de Reactome utilizando la comparación de Infec.E.csv

	Description	GeneRatio	BgRatio	pvalue
R-MMU-1483255	PI Metabolism	27/1226	74/8772	1.00395205521062e-06
R-MMU-168898	Toll-like Receptor Cascades	39/1226	133/8772	2.93951028706845e-06
R-MMU-1660516	Synthesis of PIPs at the early endosome membrane	10/1226	15/8772	4.22253895303042e-06
R-MMU-449147	Signaling by Interleukins	63/1226	262/8772	6.64094204174427e-06

Es interesante que la vía metabólica más relevante en los cambios genéticos sea la vía del metabolismo de los fosfolípidos de membrana seguida de la de cascadas de receptores Toll-like y de la de síntesis de fosfolípidos de la membrana de formación de endosomas, así como la de señalización de Interleukinas.

Por último se creó un archivo con las descripciones y contenidos de todos los archivos generados en este análisis (Tabla2).

Table 2: Lista de archivos generados en el análisis

List_of_Files
data4Heatmap.csv
normalized.Data.csv
normalized.Data.Rda
normalized.Filtered.Data.csv
QCDir.Norm
ReactomePA.Results.Infec.E.csv
ReactomePA.Results.Infec.N.csv
ReactomePABarplot.Infec.E.pdf
ReactomePABarplot.Infec.N.pdf
ReactomePAcnetplot.Infec.E.pdf
ReactomePAcnetplot.Infec.N.pdf
topAnnotated_Infec.E.csv
topAnnotated_Infec.N.csv
topAnnotated_MEM.csv

## Discusión

Los resultados de este análisis en paralelo a los acontecidos en el estudio objeto muestran numerosas similitudes como por ejemplo en la producción de glicoproteínas y aumento de la señalización inmune que concuerdan con la producción de fosfolípidos de membrana (catalogado como R-MMU-1483255 en este análisis). No obstante, hay numerosas diferencias también que han de ser resaltadas. En el estudio se ha correlacionado el fenotipo con el transcrito y eso le confiere una visión mucho más amplia de lo que se supone que pueda aportar este análisis por sí mismo. Además, es de extrañar que algunos de los genes que más repuntan en el estudio completo (Ccl22, Tnf, Marco..) no figuran en la toptable del presente análisis. Una supuesta hipótesis de por qué esto está sucediendo es quizá a los distintos protocolos de normalización que podrían haberse dado para analizar el estudio. Otro aspecto muy importante a tener en cuenta es que, tal y como hemos podido ver en la figura 7, exista una enorme variabilidad no contemplada, ya catalogada como residual y que pueda enmascarar estos datos. Esto sería compatible con la gran variabilidad y discordancia observada en el Array E.C.3 o N.C.3 que pudiera ser que se haya eliminado previamente a la realización de los análisis efectuados en el estudio original. En todo caso, en este análisis no se ha eliminado ningún Array, puesto que uno de los objetivos principales de este análisis es comparar los resultados obtenidos en el presente estudio con algunos de los del estudio de referencia. Por otra parte, un aspecto fundamental que debería haberse tenido en cuenta,

es una mayor profundización en los que se ha definido en la matriz de contrastes como el efecto memoria, que no es otra cosa que la diferencia entre los perfiles de expresión de los macrófagos estudiados una vez se haya extraído la variabilidad de sus respectivos controles. En éste análisis, al ser sólo dos genes que han sido diferencialmente expresados de una manera significativa, no se ha podido profundizar más en este aspecto.

## Conclusión

A modo de conclusiones finales, en este estudio de análisis de Arrays se ha podido: - Aprender a realizar un análisis completo de los datos contenidos de los Arrays. - Descubrir las diferencias de expresión génica en macrófagos alveolares de ratones infectados con SP3. - Hipotetizar que las diferencias entre los resultados obtenidos en el presente estudio con algunos de los del estudio de referencia puedan deberse al distinto manejo de los datos brutos.