

Hernandez_Esteban_ADO_PEC1

Esteban Hernández Maldonado

1 5 2020

Contents

1	Resumen	1
2	Objetivos	1
3	Materiales y Métodos	2
3.1	Materiales	2
3.2	Métodos	2
4	Resultados	19
5	Discusión	20
6	Conclusión	20
7	Apéndice	20
7.1	Comentarios sobre la práctica	20
7.2	Código R	21
	Referencias	21

Repositorio de Github: https://github.com/ehm2411/PEC1_ADO

1 Resumen

El siguiente estudio de microarrays es la primera práctica de evaluación continua (PEC) de la asignatura **Análisis de datos ómicos** que se imparte en la Universitat Oberta de Catalunya (UOC). Para la realización de esta práctica vamos a seguir como guía principal el caso resuelto **Omics_Data_Analysis-Case_Study_1-Microarrays**, pero teniendo en cuenta lo aprendido a través de los materiales del curso y en los foros de la asignatura. El estudio a tratar se titula **Expression data from mice normal Macrophage and multiple myeloma tumor-associated macrophages (TAMs)**

2 Objetivos

Los macrófagos asociados a tumores (en inglés, TAM) son importantes células promotoras de tumores. Sin embargo, los mecanismos subyacentes en la forma en que el tumor y su microambiente reprograman estas células siguen siendo esquivos. Aquí informamos que los lípidos juegan un papel crucial en la generación de TAM en el microambiente tumoral (TME). Los macrófagos de los tejidos tumorales humanos y murinos se enriquecieron con lípidos debido al aumento de la absorción de lípidos por los macrófagos. Los TAM expresaron niveles elevados del receptor del captador CD36, lípidos acumulados y utilizaron oxidación de ácidos grasos (FAO) en lugar de glucólisis para obtener energía. Los altos niveles de la FAO promovieron la fosforilación oxidativa mitocondrial, la producción de especies reactivas de oxígeno, la fosforilación de JAK1

y la desfosforilación de SHP1, lo que condujo a la activación y transcripción de genes STAT6 que regulan la generación y función de TAM. Estos procesos fueron críticos para la polarización y actividad de TAM, tanto in vitro como in vivo. En resumen, destacamos la importancia del metabolismo de los lípidos en la diferenciación y la función de los TAM de los tumores en el TME.

3 Materiales y Métodos

3.1 Materiales

3.1.1 Software

Para esta práctica se ha utilizado el siguiente software:

- R version 3.6.3 (2020-02-29)
- RStudio Version 1.2.5033
- Bioconductor Version 3.10
- Sistema Operativo: Windows 10

3.1.2 Datos

Los datos se han extraído de un estudio con microarrays sobre datos de expresión de células macrófagos normales y asociadas a tumor. El estudio fue subido a la web del *Gene Expression Omnibus* con el numero de acceso **GSE143025** y puede ser accedido en : <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE143025>

Tenemos un total de 12 muestras con dos factores CD36 (Wild/OK) y tipo de macrófago (Asociado a tumor -TAM- o normal -NOR-). Las muestras están equitativamente repartidas en las 4 combinaciones posibles de estos dos factores.

3.2 Métodos

3.2.1 Instalación a carga de los paquetes de R

Se han instalado librerías genéricas de R y además otras específicas de Bioconductor.

```
> if (!require("ggplot2"))      install.packages("ggplot2")
> if (!require("ggrepel"))      install.packages("ggrepel")
> if (!require("kableExtra"))   install.packages("kableExtra")

> if (!requireNamespace("BiocManager", quietly = TRUE))
+   install.packages("BiocManager")
> #BiocManager::install()
>
> # Bioconductor packages
> if (!require("Biobase"))      BiocManager::install("Biobase")
> if (!require("oligo"))        BiocManager::install("oligo")
> if (!require("pd.mogene.2.0.st")) BiocManager::install("pd.mogene.2.0.st")
> if (!require("arrayQualityMetrics")) BiocManager::install("arrayQualityMetrics")
> #BiocManager::install("lima")
>
> library(Biobase)
> library(oligo)
> library(pd.mogene.2.0.st)
> library(arrayQualityMetrics)
```

Table 1: Contenido del fichero targets usado en este análisis

FileName	Group	CD36	TYPE	ShortName
GSM4249056_su-p18001.CEL	WT.NOR	WT	NOR	WT.NOR.1
GSM4249057_su-p18002.CEL	WT.NOR	WT	NOR	WT.NOR.2
GSM4249058_su-p18003.CEL	WT.NOR	WT	NOR	WT.NOR.3
GSM4249059_su-p18004.CEL	WT.TAM	WT	TAM	WT.TAM.1
GSM4249060_su-p18005.CEL	WT.TAM	WT	TAM	WT.TAM.2
GSM4249061_su-p18006.CEL	WT.TAM	WT	TAM	WT.TAM.3
GSM4249062_su-p18007.CEL	KO.NOR	KO	NOR	KO.NOR.1
GSM4249063_su-p18008.CEL	KO.NOR	KO	NOR	KO.NOR.2
GSM4249064_su-p18009.CEL	KO.NOR	KO	NOR	KO.NOR.3
GSM4249065_su-p18010.CEL	KO.TAM	KO	TAM	KO.TAM.1
GSM4249066_su-p18011.CEL	KO.TAM	KO	TAM	KO.TAM.2
GSM4249067_su-p18012.CEL	KO.TAM	KO	TAM	KO.TAM.3

3.2.2 Preparación del entorno

Se han creado las subcarpetas “data” y “results” desde Windows Explorer. En “data” se guardaron los 12 ficheros .CEL descargados de la página del estudio en GEO así como el fichero targets que fue creado de manera manual.

3.2.3 Preparar los datos para el análisis

Primero hemos creado manualmente el fichero *targets*. En él se puede ver como hay tres muestras para cada una de las cuatro combinaciones posibles de los dos factores.

```
> targets <- read.csv2("./data/targets.csv", header = TRUE, sep = ",")
> knitr::kable(targets, booktabs = TRUE, caption = 'Contenido del fichero targets usado en este análisis')
```

3.2.4 Leer los ficheros CEL

Ahora leemos los ficheros CEL y asociamos el fichero targets

```
> my.targets@data$ShortName->rownames(pData(rawData))
> colnames(rawData) <-rownames(pData(rawData))
> rownames(pData(rawData))

[1] "WT.NOR.1" "WT.NOR.2" "WT.NOR.3" "WT.TAM.1" "WT.TAM.2" "WT.TAM.3"
[7] "KO.NOR.1" "KO.NOR.2" "KO.NOR.3" "KO.TAM.1" "KO.TAM.2" "KO.TAM.3"
```

3.2.5 Control de calidad del rawdata

Una vez cargado el rawdata procedemos, con su control de calidad mediante la librería *arrayQualityMetrics*. Los ficheros resultantes están en la carpeta *arrayQualityMetrics report for rawData*.

Podemos observar que sólo hay presencia de outliers según Boxplots con la muestra 11, y tenemos dos outliers en las muestras 11 y 12 según los MA plots.

El análisis de componentes principales nos muestra que las dos componentes principales explican algo más del 70% del modelo (PC1: 48.9% , PC2: 23.2%)

```
> plotPCA3(exprs(rawData), labels = targets$ShortName, factor = targets$Group,
+          title="Raw data", scale = FALSE, size = 3,
+          colores = c("red", "blue", "green", "yellow"))
```

	array	sampleNames	*1	*2	*3	Group	Genotype	Temperature	ShortName
<input type="checkbox"/>	1	WT.RT.1				WT.RT	WT	RT	WT.RT.1
<input type="checkbox"/>	2	WT.RT.2				WT.RT	WT	RT	WT.RT.2
<input type="checkbox"/>	3	WT.RT.3				WT.RT	WT	RT	WT.RT.3
<input type="checkbox"/>	4	KO.RT.1				KO.RT	KO	RT	KO.RT.1
<input type="checkbox"/>	5	KO.RT.2				KO.RT	KO	RT	KO.RT.2
<input type="checkbox"/>	6	KO.RT.3				KO.RT	KO	RT	KO.RT.3
<input type="checkbox"/>	7	WT.COLD.1				WT.COLD	WT	COLD	WT.COLD.1
<input type="checkbox"/>	8	WT.COLD.2		x		WT.COLD	WT	COLD	WT.COLD.2
<input type="checkbox"/>	9	WT.COLD.3			x	WT.COLD	WT	COLD	WT.COLD.3
<input type="checkbox"/>	10	KO.COLD.1				KO.COLD	KO	COLD	KO.COLD.1
<input type="checkbox"/>	11	KO.COLD.2		x		KO.COLD	KO	COLD	KO.COLD.2
<input type="checkbox"/>	12	KO.COLD.3				KO.COLD	KO	COLD	KO.COLD.3

Figure 1: Aspecto del resumen, en el fichero index.html, producido por el paquete arrayQualityMetrics sobre los datos crudos

De la misma manera, se puede visualizar la distribución de la intensidad usando boxplots. La figura 3 muestra un boxplot múltiple describiendo las distribución de las intensidades a lo largo de todas las muestras. Observaremos una pequeña variación en las mismas, pero está dentro de lo normal para datos sin normalizar.

3.2.6 Normalización de los datos

Se utiliza el método RMA para normalizar los datos.

3.2.7 Control de calidad de los datos normalizados

Volvemos aplicar los mismo controles de calidad para los datos normalizados.

```
> plotPCA3(exprs(eset_rma), labels = targets$ShortName, factor = targets$Group,
+          title="Datos Normalizados", scale = FALSE, size = 3,
+          colores = c("red", "blue", "green", "yellow"))

> boxplot(eset_rma, cex.axis=0.5, las=2, which="all",
+          col = c(rep("red", 3), rep("blue", 3), rep("green", 3), rep("yellow", 3)),
+          main="Boxplot de intensidad: Datos normalizados")
```

3.2.8 Detección de los genes más variables

Vamos a ver que genes son los que más varían. En 6 podemos ver que los genes más variables son aquellos con una desviación típica por evema de 90-95% de todas las desviaciones típicas.

```
> sds <- apply (exprs(eset_rma), 1, sd)
> sds0<- sort(sds)
> plot(1:length(sds0), sds0, main="Distribución de la variabilidad para todos los genes",
+      sub="Las líneas verticales representan los percentiles 90% y 95%",
+      xlab="Índice del Gen (de menos a más diferente)", ylab="Desviación típica")
> abline(v=length(sds)*c(0.9,0.95))
```

3.2.9 Filtraje de los genes menos variables

Vamos a continuar con el proceso de filtraje y para ello vamos a utilizar la funcion *nsFilter* del paquete *genefilter*.

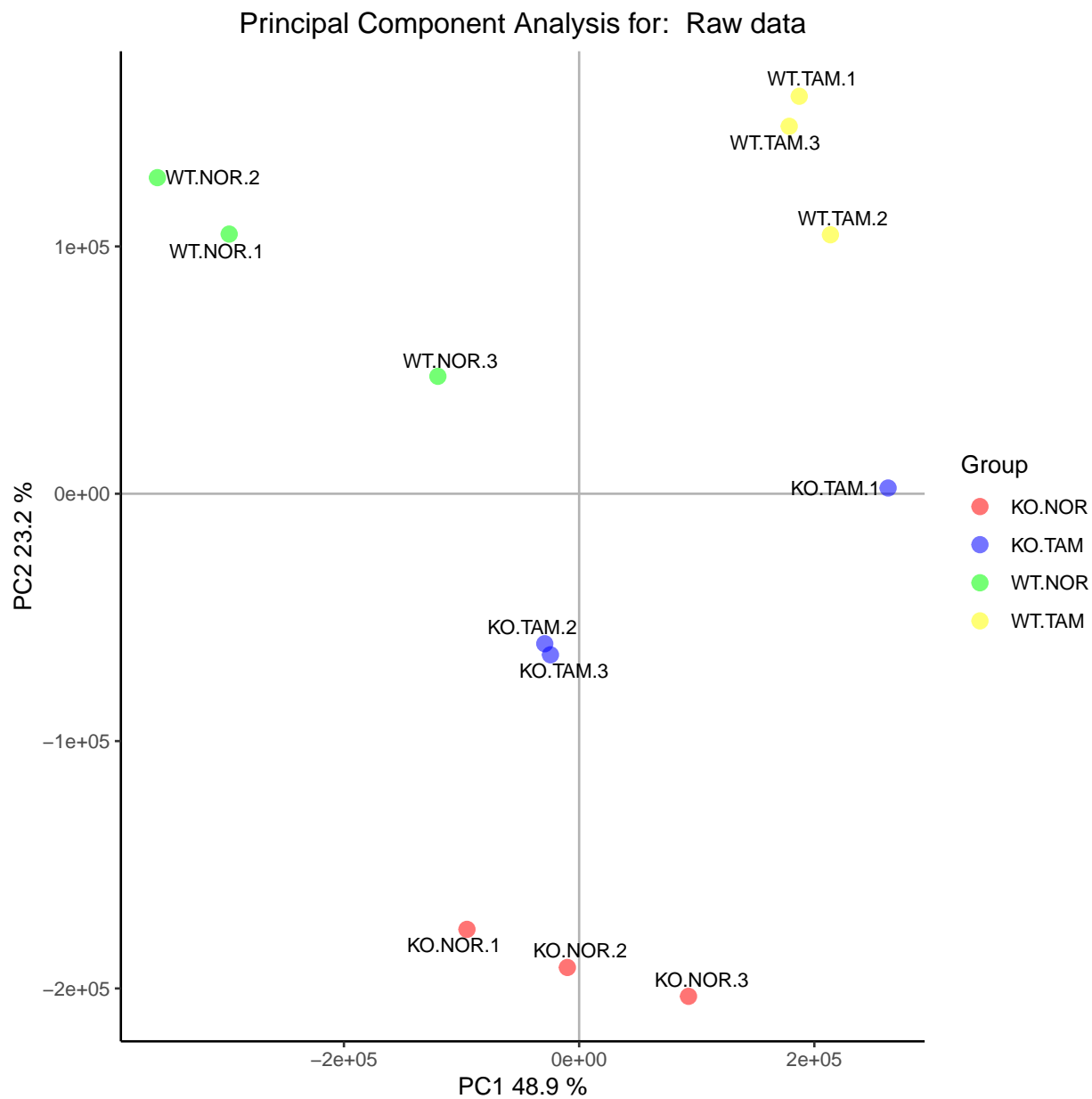


Figure 2: Visualización de las dos componentes principales para rawdata

Distribucion de los valores de intensidad

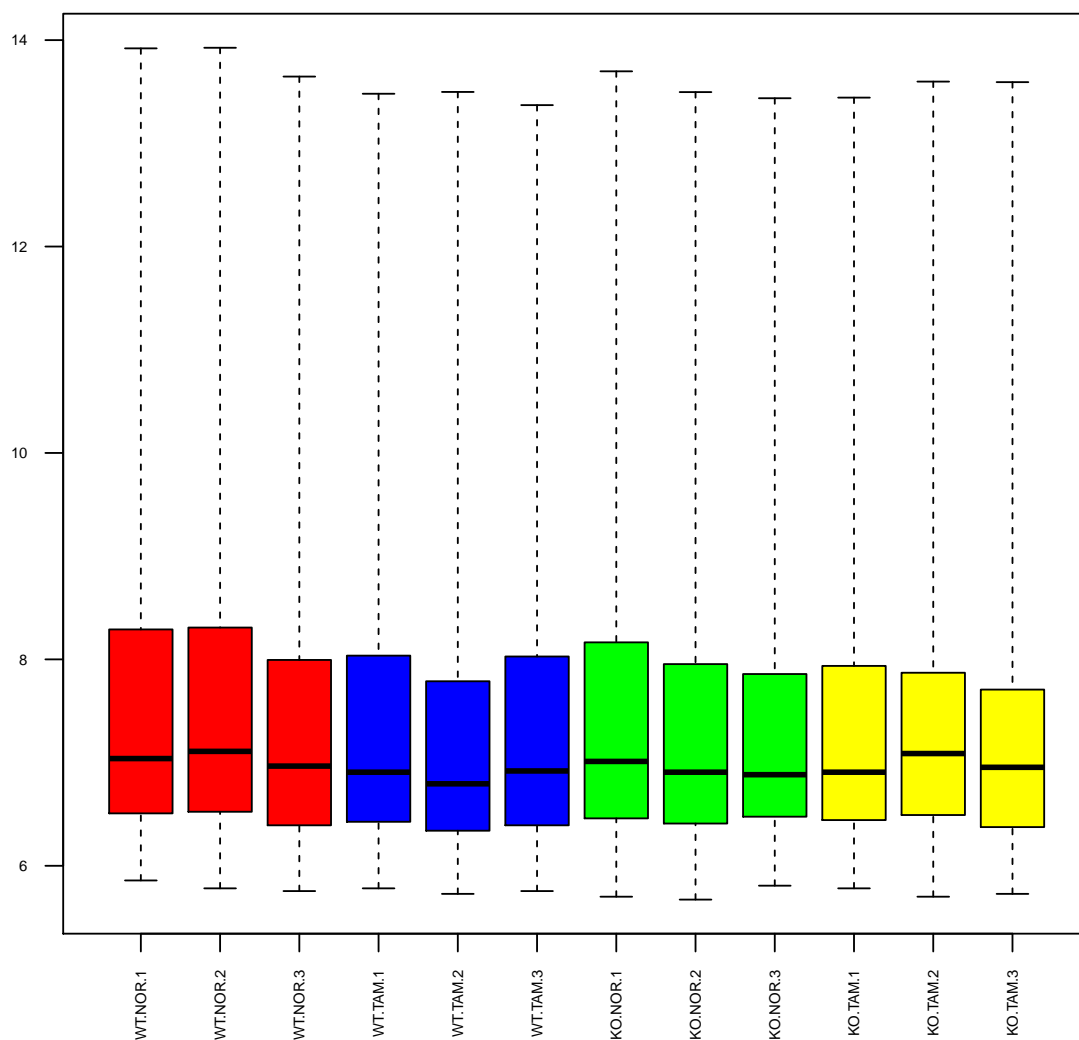


Figure 3: Boxplot para intensidades de arrays (Raw Data)

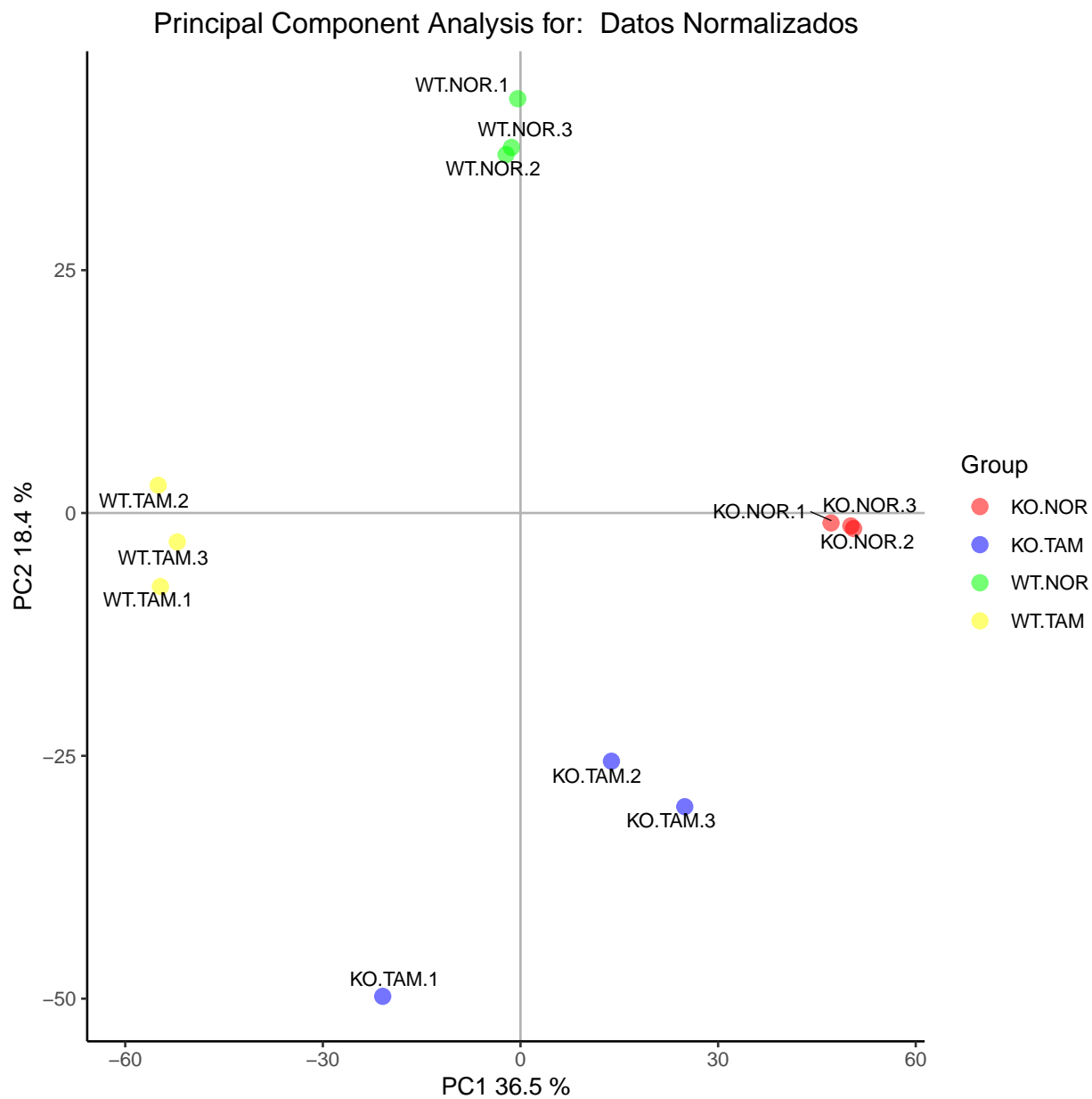


Figure 4: Visualización de las dos componentes principales de los datos normalizados

Boxplot de intensidad: Datos normalizados

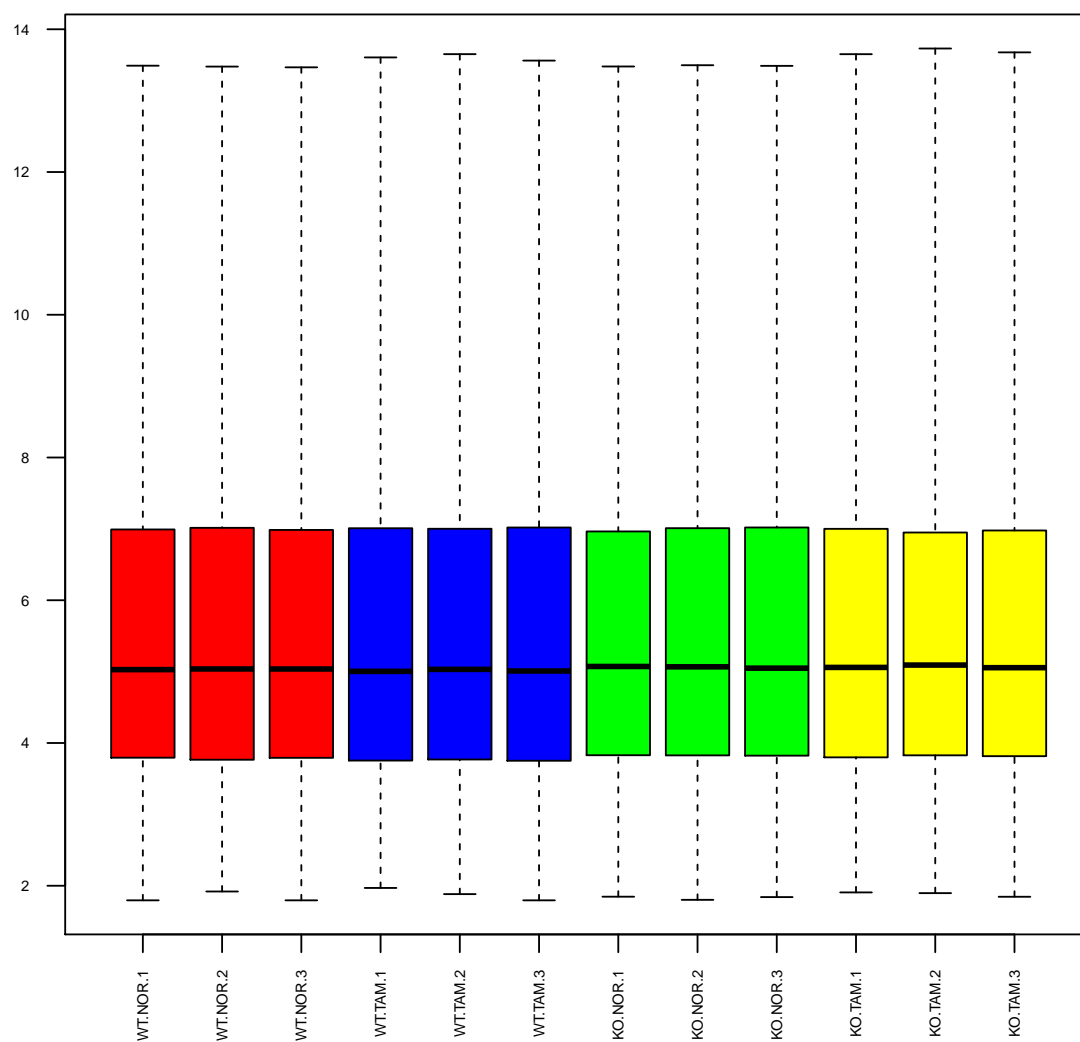


Figure 5: Distribución de las intensidades para los datos normalizados

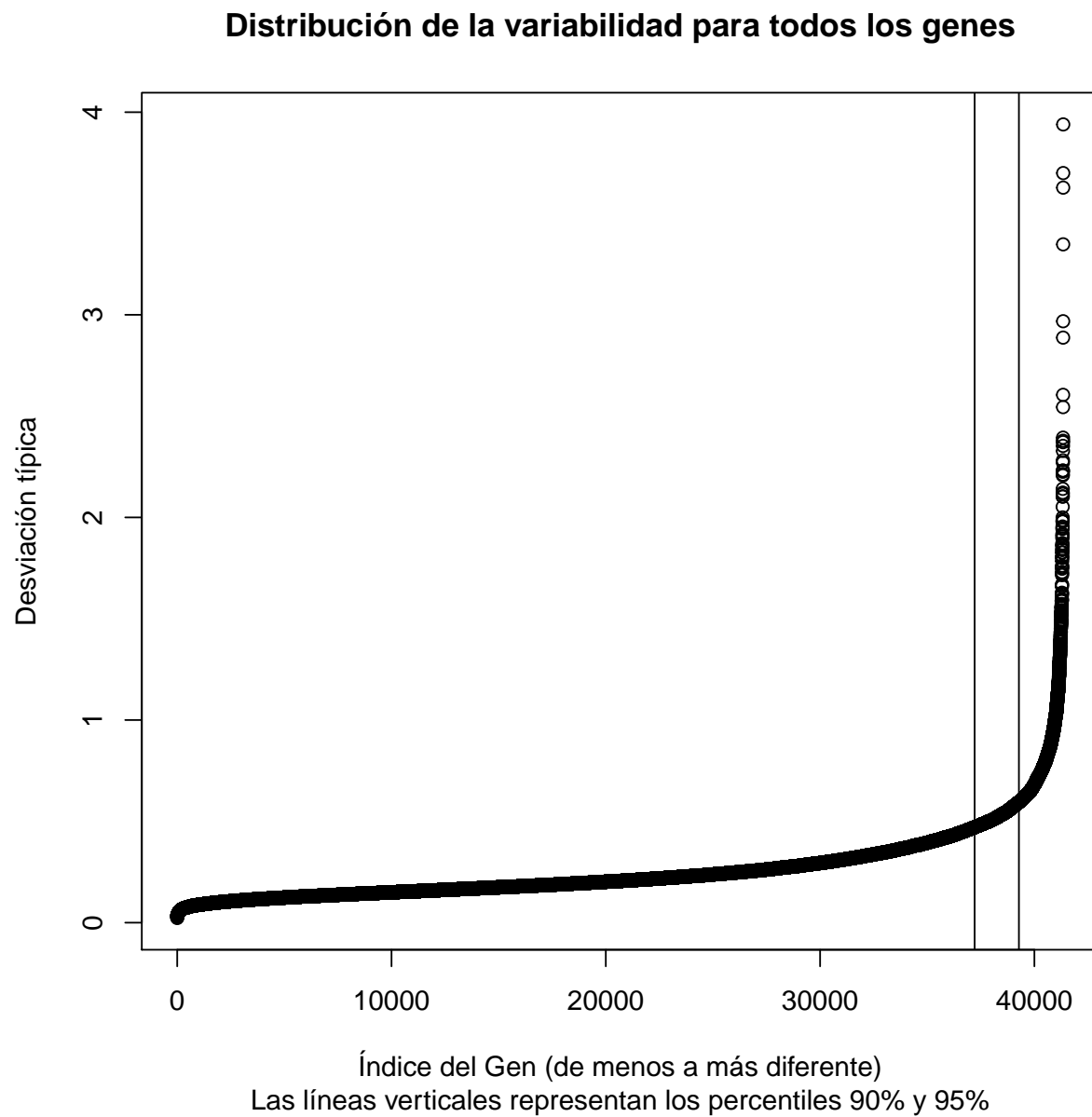


Figure 6: Valores de la desviación típica a lo largo de todas las muestras para todos los genes ordenados de menor a mayor

```

> library(genefilter)
> library(mogene21sttranscriptcluster.db)
> annotation(eset_rma) <- "mogene21sttranscriptcluster.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")

```

3.2.10 Guardando datos normalizados y filtrados

Guardamos los ficheros obtenidos hasta ahora.

```

> 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")

```

3.2.11 Matriz de diseño

Vamos a crear la matriz de diseño. Recordemos que tenemos 2 factores de dos niveles cada uno que se pueden agrupar en un factor de 4 niveles.

```

> library(limma)
> designMat<- model.matrix(~0+Group, pData(eset_filtered))
> colnames(designMat) <- c("KO.NOR", "KO.TAM", "WT.NOR", "WT.TAM")
> print(designMat)

```

	KO.NOR	KO.TAM	WT.NOR	WT.TAM
WT.NOR.1	0	0	1	0
WT.NOR.2	0	0	1	0
WT.NOR.3	0	0	1	0
WT.TAM.1	0	0	0	1
WT.TAM.2	0	0	0	1
WT.TAM.3	0	0	0	1
KO.NOR.1	1	0	0	0
KO.NOR.2	1	0	0	0
KO.NOR.3	1	0	0	0
KO.TAM.1	0	1	0	0
KO.TAM.2	0	1	0	0
KO.TAM.3	0	1	0	0

```

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

```

3.2.12 Definiendo comparaciones con la matriz de contrastes

Vamos a hacer las comparaciones de los genotipos KT vs OK en macrófagos normales y en macrófagos TAM. También haremos una comparación global.

```

> cont.matrix <- makeContrasts (KOvsWT.NOR = KO.NOR-WT.NOR,
+                               KOvsWT.TAM = KO.TAM-WT.TAM,
+                               INT = (KO.NOR-WT.NOR) - (KO.TAM-WT.TAM),
+                               levels=designMat)
> print(cont.matrix)

```

Levels	Contrasts		
	KOvsWT.NOR	KOvsWT.TAM	INT
KO.NOR	1	0	1
KO.TAM	0	1	-1
WT.NOR	-1	0	-1
WT.TAM	0	-1	1

3.2.13 Estimación del modelo y selección de genes

Con los datos filtrados y con la matriz de diseño, creamos un modelo utilizando la función `lmFit`.

3.2.14 Lista de genes diferencialmente expresados

Para la comparación 1 (KOvsWT.NOR): Genes que cambian su expresión entre KO and WT in macrofagos normales:

```
> topTab_KOvsWT.NOR <- topTable (fit.main, number=nrow(fit.main), coef="KOvsWT.NOR", adjust="fdr")
> head(topTab_KOvsWT.NOR)
```

	logFC	AveExpr	t	P.Value	adj.P.Val	B
17550478	-7.663604	8.534005	-51.12151	1.944216e-14	1.164780e-10	19.64538
17546797	6.543300	6.688255	39.26458	3.496972e-13	1.047518e-09	18.33158
17546834	5.582473	7.506259	34.43146	1.468473e-12	2.369739e-09	17.52773
17546287	4.986325	5.744844	34.19691	1.582200e-12	2.369739e-09	17.48317
17546316	5.801854	6.541594	30.24331	6.036925e-12	7.233443e-09	16.63741
17344873	-3.678150	6.498091	-15.88617	6.173373e-09	6.164113e-06	11.06259

Para la comparación 2 (KOvsWT.TAM): Genes que cambian su expresión entre KO and WT en macrofagos tipo TAM:

```
> topTab_KOvsWT.TAM <- topTable (fit.main, number=nrow(fit.main), coef="KOvsWT.TAM", adjust="fdr")
> head(topTab_KOvsWT.TAM)
```

	logFC	AveExpr	t	P.Value	adj.P.Val	B
17550478	-7.407465	8.534005	-49.41289	2.822195e-14	1.690777e-10	19.375446
17546797	6.256074	6.688255	37.54101	5.712073e-13	1.711051e-09	17.975457
17546834	5.444533	7.506259	33.58068	1.929393e-12	3.852999e-09	17.283077
17546287	4.735994	5.744844	32.48010	2.774842e-12	4.156019e-09	17.062272
17546316	5.530511	6.541594	28.82888	1.016555e-11	1.218036e-08	16.221451
17462748	2.528795	7.144847	13.50090	3.411977e-08	2.955096e-05	9.450324

Para la comparación 3 (INT): Genes que se comportan diferente entre la comparación 1 y la comparación 2:

```
> topTab_INT <- topTable (fit.main, number=nrow(fit.main), coef="INT", adjust="fdr")
> head(topTab_INT)
```

	logFC	AveExpr	t	P.Value	adj.P.Val	B
17437969	-2.245740	5.547429	-9.374475	1.397977e-06	0.008090792	5.451538
17408024	-1.164196	11.161725	-8.380991	4.169406e-06	0.008090792	4.535215
17344873	-2.696009	6.498091	-8.233720	4.943542e-06	0.008090792	4.389079
17404230	-2.818322	8.735236	-8.157803	5.401964e-06	0.008090792	4.312651
17304817	1.047081	9.093600	7.832350	7.956305e-06	0.009533245	3.976347
17462748	-1.967643	7.144847	-7.428146	1.308323e-05	0.011795159	3.538440

La primera columna de cada `topTable` contiene el ID de Affymetrix para cada muestra. El siguiente paso es averiguar que gen corresponde con cada ID. Ese proceso se llama *anotación*.

Table 2: Anotaciones añadidas a los resultados "topTable" para la comparación "KOvsWT.TAM"

PROBEID	SYMBOL	ENTREZID	GENENAME
17210953	St18	240690	suppression of tumorigenicity 18
17210984	Pcmt1d	319263	protein-L-isoaspartate (D-aspartate) O-methyltransferase domain containing 1
17211405	Gsta3	14859	glutathione S-transferase, alpha 3
17211498	Lmbrd1	68421	LMBR1 domain containing 1
17211587	Dst	13518	dystonin

3.2.15 Anotación de genes

Hemos obtenido tres tablas de anotaciones, se pueden consultar en la carpeta *results* de este proyecto con los nombres *topAnnotated_KOvsWT_NOR.csv*, *topAnnotated_KOvsWT_TAM.csv* y *topAnnotated_INT.csv*. No se muestra su contenido en este informe debido a lo grande de su tamaño, pero mostramos por los primeros registros de "KOvsWT.TAM".

	PROBEID	SYMBOL	ENTREZID		GENENAME
1	17210953	St18	240690		suppression of tumorigenicity 18
2	17210984	Pcmt1d	319263		protein-L-isoaspartate (D-aspartate) O-methyltransferase domain containing 1
3	17211405	Gsta3	14859		glutathione S-transferase, alpha 3
4	17211498	Lmbrd1	68421		LMBR1 domain containing 1
5	17211587	Dst	13518		dystonin

3.2.16 Visualizando expresión diferencial

Vamos a utilizar los "Volcano Plots", la figura 7 muestra un volcano plot para la comparación entre KO and WT. los nombres de los genes 4son mostrados en el gráfico.

3.2.17 Comparaciones múltiples

Vamos a utilizar las funciones `decideTests` y `VennDiagram` del paquete `limma`.

Ontenemos tantas columnas como comparaciones y tantas filas como genes. La significancia es indicada con los valores +1 y -1 mientras que 0 indica no significancia.

```
> sum.res.rows<-apply(abs(res),1,sum)
> res.selected<-res[sum.res.rows!=0,]
> print(summary(res))
```

	KOvsWT.NOR	KOvsWT.TAM	INT
Down	451	576	40
NotSig	5465	5270	5928
Up	75	145	23

Podemos visualizar estos datos con un diagrama de Venn.

```
> vennDiagram (res.selected[,1:3], cex=0.9)
> title("Genes en común entre las tres comparaciones\n Genes seleccionados con FDR < 0.1 and logFC > 1")
```

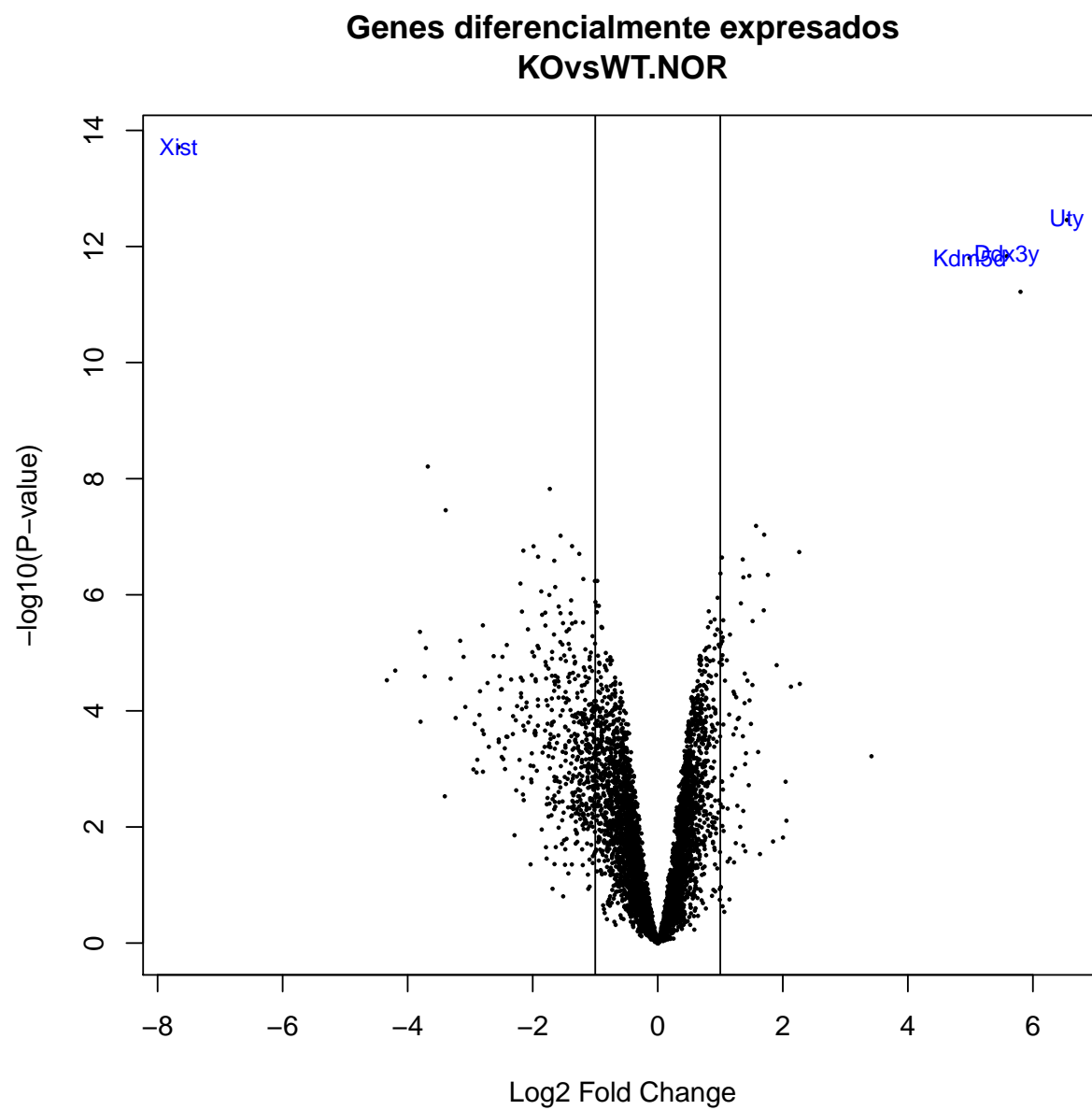


Figure 7: Volcano plot para la comparación entre KO y WT. El nombre de los 4 están mostrados en la gráfica

Genes en común entre las tres comparaciones
Genes seleccionados con $FDR < 0.1$ and $\log FC > 1$

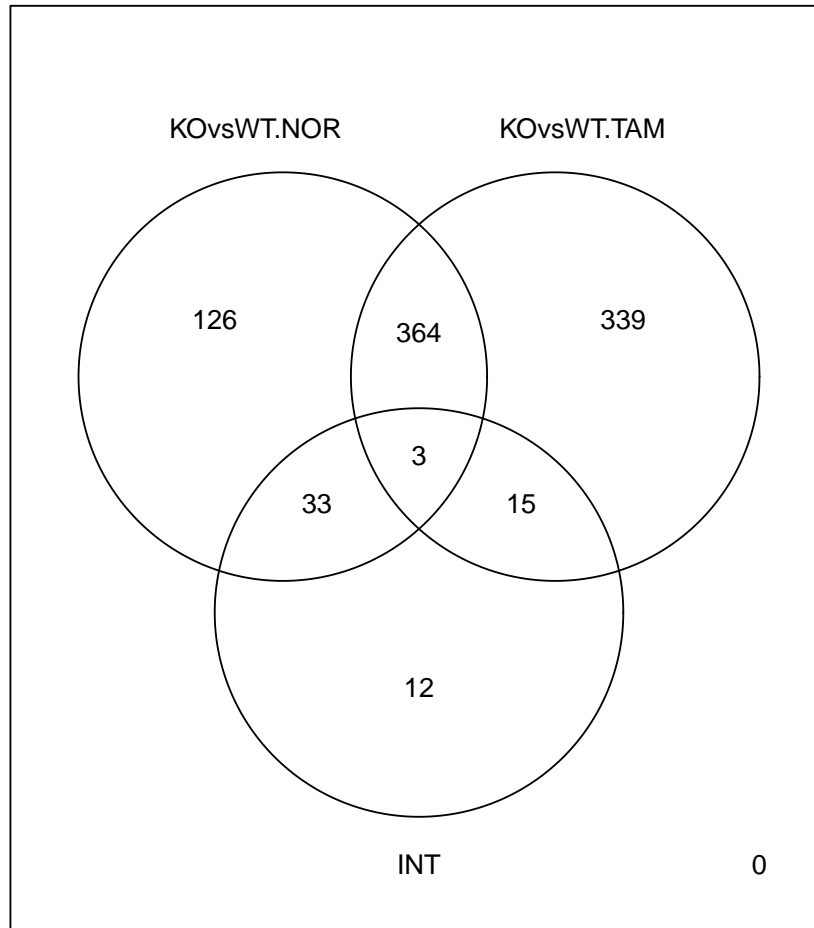


Figure 8: Diagrama de Venn mostrando los genes in común entre las tres comparaciones

3.2.18 Heatmaps

Los genes seleccionados como diferencialmente expresados puede ser visualizados mediante un Heatmap.

```
> probesInHeatmap <- rownames(res.selected)
> HMdata <- exprs(eset_filtered)[rownames(exprs(eset_filtered)) %in% probesInHeatmap,]
>
> geneSymbols <- select(mogene21sttranscriptcluster.db, rownames(HMdata), c("SYMBOL"))
> SYMBOLS<- geneSymbols$SYMBOL
> rownames(HMdata) <- SYMBOLS
> write.csv(HMdata, file = file.path("./results/data4Heatmap.csv"))
```

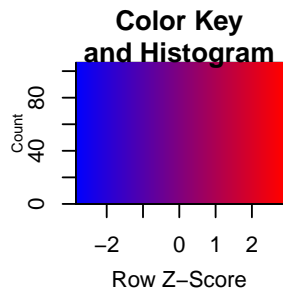
Con los datos seleccionados un heatmap puede ser generado con o sin agrupar genes o/y muestras.

La figura 9 muestra un heatmap producido por todos los genes seleccionados con el mismo criterio ($FDR < 0.1$ and $\log FC > 1$).

```
> my_palette <- colorRampPalette(c("blue", "red"))(n = 299)
> library(gplots)
>
> heatmap.2(HMdata,
+           Rowv = FALSE,
+           Colv = FALSE,
+           main = "Diferencia de genes expresada \n FDR < 0,1, logFC >=1",
+           scale = "row",
+           col = my_palette,
+           sepcolor = "white",
+           sepwidth = c(0.05,0.05),
+           cexRow = 0.5,
+           cexCol = 0.9,
+           key = TRUE,
+           keysize = 1.5,
+           density.info = "histogram",
+           ColSideColors = c(rep("red",3),rep("blue",3), rep("green",3), rep("yellow",3)),
+           tracecol = NULL,
+           dendrogram = "none",
+           srtCol = 30)
```

Figure 10 muestra un heatmap producido por todos los genes seleccionados con el mismo criterio ($FDR < 0.1$ and $\log FC > 1$) donde los genes y muestras son forzados a agruparse por similitud de fila y columna.

```
> heatmap.2(HMdata,
+           Rowv = TRUE,
+           Colv = TRUE,
+           dendrogram = "both",
+           main = "Diferencia de genes expresada \n FDR < 0,1, logFC >=1",
+           scale = "row",
+           col = my_palette,
+           sepcolor = "white",
+           sepwidth = c(0.05,0.05),
+           cexRow = 0.5,
+           cexCol = 0.9,
+           key = TRUE,
+           keysize = 1.5,
+           density.info = "histogram",
+           ColSideColors = c(rep("red",3),rep("blue",3), rep("green",3), rep("yellow",3)),
+           tracecol = NULL,
```



**Diferencia de genes expresada
FDR < 0,1, logFC >=1**

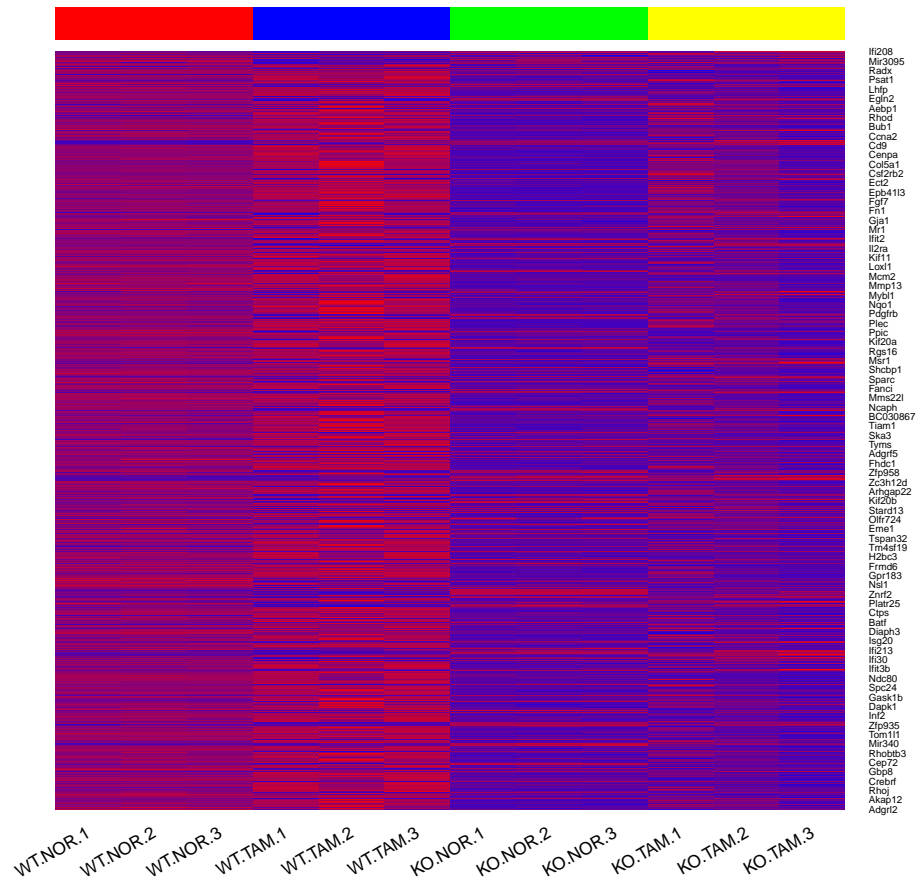


Figure 9: Heatmap para los datos sin ningún grupo

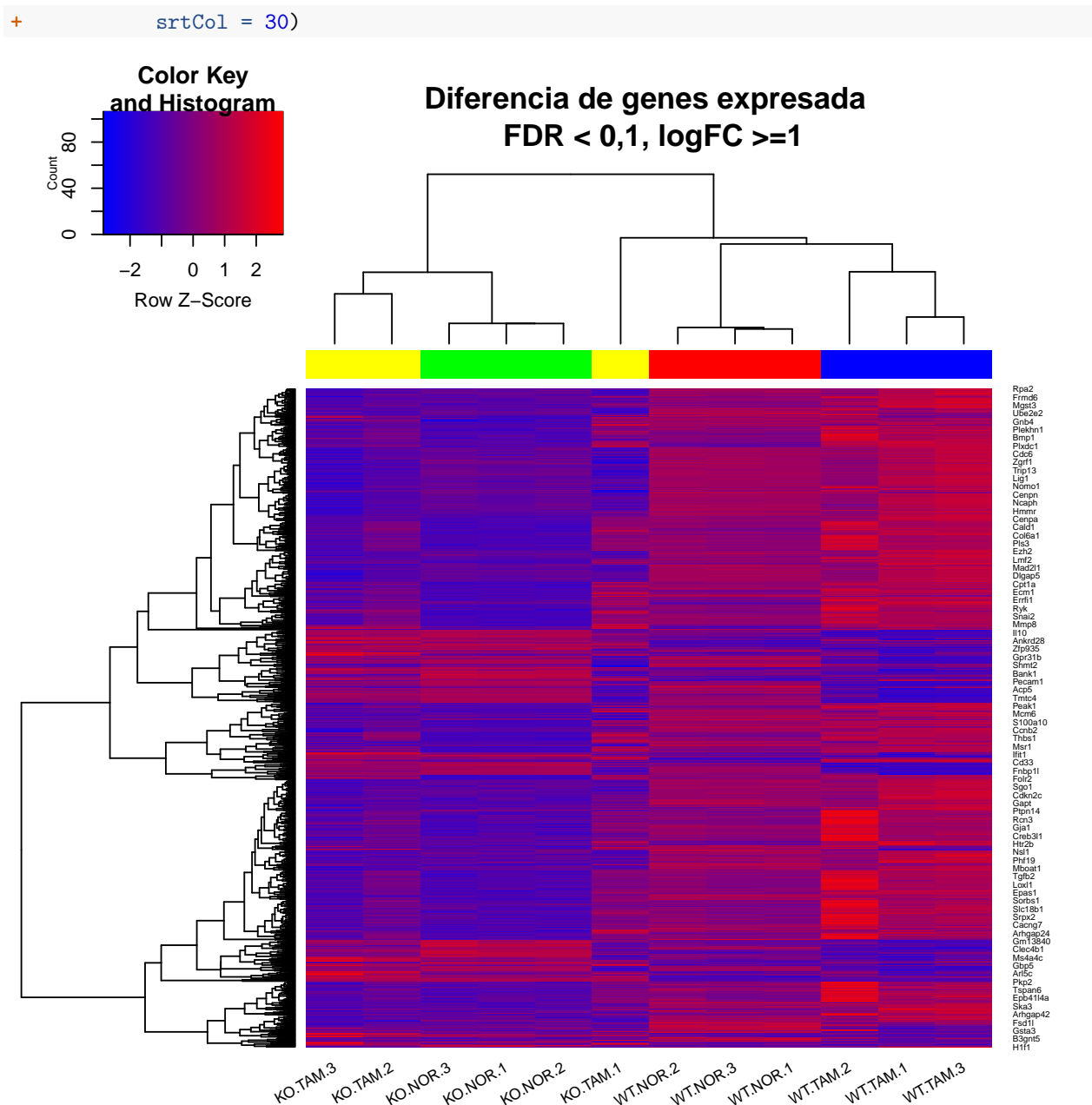


Figure 10: Heatmap for expression data grouping genes (rows) and samples (columns) by their similarity

3.2.19 Significacia biológica

Hay muchas variantes para hacer este tipo de análisis, ver Khatri, Sirota, and Butte (2012), pero se usará aquí el analisis implementado por el paquete **ReactomePA**. The analysis se realiza con las anotaciones de ReactomePA <https://reactome.org/>.

KOvsWT.NOR	KOvsWT.TAM	INT
3128	3295	209

The Biological significance analysis will be applied only to the first two lists. Sometimes yet another decomposition is applied so that up and downregulated genes are separately analyzed. This will not be done here because there is no clear biological argument to proceed so in all cases.

```
> library(ReactomePA)
>
> listOfData <- listOfSelected[1:2]
> comparisonsNames <- names(listOfData)
> universe <- mapped_genes
>
> for (i in 1:length(listOfData)){
+   genesIn <- listOfData[[i]]
+   comparison <- comparisonsNames[i]
+   enrich.result <- enrichPathway(gene = genesIn,
+                                   pvalueCutoff = 0.05,
+                                   readable = T,
+                                   pAdjustMethod = "BH",
+                                   organism = "mouse",
+                                   universe = universe)
+
+   cat("#####")
+   cat("\nComparison: ", comparison, "\n")
+   print(head(enrich.result))
+
+   if (length(rownames(enrich.result@result)) != 0) {
+     write.csv(as.data.frame(enrich.result),
+               file = paste0("./results/", "ReactomePA.Results.", comparison, ".csv"),
+               row.names = FALSE)
+
+     pdf(file=paste0("./results/", "ReactomePABarplot.", comparison, ".pdf"))
+     print(barplot(enrich.result, showCategory = 15, font.size = 4,
+                   title = paste0("Reactome Pathway Analisis para ", comparison, ". Barplot")))
+     dev.off()
+
+     pdf(file = paste0("./results/", "ReactomePACnetplot.", comparison, ".pdf"))
+     print(cnetplot(enrich.result, categorySize = "geneNum", showCategory = 15,
+                    vertex.label.cex = 0.75))
+     dev.off()
+   }
+ }
```

#####

Comparison: KOvsWT.NOR

	ID	Description	GeneRatio
R-MMU-69278	R-MMU-69278	Cell Cycle, Mitotic	187/1540
R-MMU-69620	R-MMU-69620	Cell Cycle Checkpoints	108/1540
R-MMU-68962	R-MMU-68962	Activation of the pre-replicative complex	25/1540
R-MMU-2500257	R-MMU-2500257	Resolution of Sister Chromatid Cohesion	56/1540
R-MMU-68877	R-MMU-68877	Mitotic Prometaphase	75/1540
R-MMU-68886	R-MMU-68886	M Phase	123/1540
	BgRatio	pvalue	p.adjust
R-MMU-69278	499/8772	6.705560e-28	6.886610e-25
R-MMU-69620	281/8772	2.777925e-17	1.426464e-14
R-MMU-68962	30/8772	6.291948e-15	2.153943e-12
R-MMU-2500257	118/8772	5.126354e-14	1.316191e-11

```
R-MMU-68877 189/8772 3.475500e-13 7.138677e-11 5.933959e-11
R-MMU-68886 381/8772 6.807855e-13 1.165278e-10 9.686264e-11
```

```
R-MMU-69278 Tubb6/Hist1h2an/Ube2c/Mcm10/Cdk1/Cenpt/Lmnbl/Ccnbl/Ercc6l/Cdc6/Cdc45/Gtse1/Dna2/Tpx2/Cdk6
R-MMU-69620
R-MMU-68962
R-MMU-2500257
R-MMU-68877
R-MMU-68886
```

```
Count
R-MMU-69278 187
R-MMU-69620 108
R-MMU-68962 25
R-MMU-2500257 56
R-MMU-68877 75
R-MMU-68886 123
```

```
#####
```

```
Comparison: KOvsWT.TAM
```

	ID	Description	GeneRatio	
R-MMU-69278	R-MMU-69278	Cell Cycle, Mitotic	198/1618	
R-MMU-68877	R-MMU-68877	Mitotic Prometaphase	83/1618	
R-MMU-69620	R-MMU-69620	Cell Cycle Checkpoints	108/1618	
R-MMU-2500257	R-MMU-2500257	Resolution of Sister Chromatid Cohesion	59/1618	
R-MMU-5663220	R-MMU-5663220	RHO GTPases Activate Formins	63/1618	
R-MMU-194315	R-MMU-194315	Signaling by Rho GTPases	143/1618	
	BgRatio	pvalue	p.adjust	qvalue
R-MMU-69278	499/8772	2.036051e-30	2.093061e-27	1.720999e-27
R-MMU-68877	189/8772	3.081912e-16	1.584103e-13	1.302513e-13
R-MMU-69620	281/8772	1.077482e-15	3.692172e-13	3.035853e-13
R-MMU-2500257	118/8772	5.345366e-15	1.373759e-12	1.129560e-12
R-MMU-5663220	132/8772	1.073387e-14	2.206885e-12	1.814590e-12
R-MMU-194315	428/8772	2.226691e-14	3.815064e-12	3.136900e-12

```
R-MMU-69278 Ccne2/Fbxo5/Cdc45/Cdk1/Pole/Mcm6/Zwilch/Mcm7/Kntc1/Cdc6/Hist1h2an/Mcm5/H2afx/Lig1/Spc24/T
R-MMU-68877
R-MMU-69620
R-MMU-2500257
R-MMU-5663220
R-MMU-194315
```

```
Count
R-MMU-69278 198
R-MMU-68877 83
R-MMU-69620 108
R-MMU-2500257 59
R-MMU-5663220 63
R-MMU-194315 143
```

4 Resultados

```
> listOfFiles <- dir("./results/")
> knitr::kable(
+   listOfFiles, booktabs = TRUE,
```

Table 3: List of files generated in the analysis

List_of_Files
data4Heatmap.csv
normalized.Data.csv
normalized.Data.Rda
normalized.Filtered.Data.csv
QCDir.Norm
ReactomePA.Results.KOvsWT.NOR.csv
ReactomePA.Results.KOvsWT.TAM.csv
ReactomePABarplot.KOvsWT.NOR.pdf
ReactomePABarplot.KOvsWT.TAM.pdf
ReactomePAcnetplot.KOvsWT.NOR.pdf
ReactomePAcnetplot.KOvsWT.TAM.pdf
topAnnotated_INT.csv
topAnnotated_KOvsWT_NOR.csv
topAnnotated_KOvsWT_TAM.csv

```
+ caption = 'List of files generated in the analysis',
+ col.names="List_of_Files"
+ )
```

5 Discusión

N/A

6 Conclusión

Este estudio destaca el papel del metabolismo de los lípidos en la diferenciación y función de los TAM y sugiere enfocarse en la oxidación de los ácidos grasos TAM como una modalidad terapéutica potencial para los cánceres humanos.

7 Apéndice

7.1 Comentarios sobre la práctica

La realización de esta práctica ha llevado más tiempo del pensado por mí inicialmente. Había querido utilizar el paquete `geoQuery` para descargar los ficheros CEL y montar el fichero targets de manera automática, pero errores de ejecución que no pude solventar, me hicieron decantar por la manera manual observada en los casos de estudio.

Por otro lado, para la significación biológica, hubiese deseado utilizar el paquete `cluterProfiler`, pero tampoco pude hacer frente a los problemas encontrado en la ejecución del código.

Ha sido una práctica dura en la que, intentando comprender los conceptos biológicos intrínsecos (de los que carezco), he tenido que dedicar mucho tiempo. Además, el estudio elegido para esta práctica ha sido el tercero, por lo que previamente había empleado tiempo en dos estudios que por las dificultades encontradas tuvieron que ser descartados.

7.2 Código R

Se puede consultar el código fuente de ese informe en el documento `*hernandez_esteban_ADO_PEC1.rmd`, localizado en el directorio principal de este proyecto.

Referencias

Khatri, Purvesh, Marina Sirota, and Atul J. Butte. 2012. “Ten Years of Pathway Analysis: Current Approaches and Outstanding Challenges.” *PLOS Computational Biology* 8 (2): e1002375. <https://doi.org/10.1371/journal.pcbi.1002375>.