

PEC1 Informe

Beatriz Pardo Montenegro

26/4/2020

Contents

Contents.....	1
1.Abstract.....	2
2.Objetivos	2
3.Materiales y métodos.....	2
3.1 Naturaleza de los datos.....	2
3.2 Tipo de experimento.....	2
3.3 Diseño general.....	2
3.4 Materiales	2
3.5 Análisis de datos.....	3
3.5.1 Datos y grupos	3
3.5.2 Control de calidad de los datos crudos	3
3.5.3 Normalización.....	5
3.5.4 Control de calidad de los datos normalizados.....	6
3.5.5 Identificación de genes diferencialmente expresados.....	8
3.5.6 Anotación de los resultados	10
4.Resultados.....	11
5.Discusión.....	12
6.Apéndice	12
7.Referencias bibliográficas.....	18

URL GitHub: https://github.com/bpardom/AO_PEC1.git

1. Abstract

La progresión metastásica en pacientes con cáncer de mama se asocia a la resistencia a terapias como la quimioterapia. Estudiamos las respuestas genéticas de células metastásicas de cáncer de mama tratadas con Paclitaxel y observamos una actividad elevada de JNK así como de la expresión de SPP1 o TNC. Su inhibición puede ser una futura estrategia terapéutica en el cáncer de mama metastásico Insua-Rodríguez (2018).

2. Objetivos

Sabiendo que un gran problema en el tratamiento de las metástasis de cáncer de mama es la resistencia de estas a la quimioterapia, el objetivo de este estudio es estudiar qué genes y procesos intervienen en la resistencia a la quimioterapia de las células metastásicas para buscar así nuevas dianas terapéuticas.

3. Materiales y métodos

3.1 Naturaleza de los datos

Entro en la base de datos GEO, tal y como se sugiere en el enunciado de la PEC, y selecciono el estudio cuya referencia es GSE98238¹.

3.2 Tipo de experimento

El tipo de experimento que planteo es de Comparación de grupos o class comparison. El objetivo de los estudios comparativos es determinar si los perfiles de expresión génica difieren entre grupos previamente identificados, así como seleccionar genes diferencialmente expresados². En mi experimento comparo 2 grupos, uno tratado con DMSO y otro con Paclitaxel.

3.3 Diseño general

Mi estudio lo componen 2 grupos, el primero son 3 muestras de la línea celular MDA231-LM2 tratadas con un vehículo (DMSO) y el segundo grupo son 3 muestras de la misma línea celular tratadas con Paclitaxel 5 nM durante 48 horas.

3.4 Materiales

Cultivo celular: línea celular MDA231-LM2, es una de las más utilizadas para el estudio experimental in vitro del cáncer de mama hormono-independiente. Estas células se

¹ <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE98238>

² http://materials.cv.uoc.edu/daisy/Materials/PID_00192730/pdf/PID_00192743.pdf

aislaron en 1973 a partir de una muestra de derrame pleural de una paciente con cáncer de mama que falleció en Houston, estas células presentan un crecimiento extraordinariamente rápido en medios de cultivo poco enriquecidos.

Array: Affymetrix GeneChip® Human Genome U133 Plus 2.0 microarray.

3.5 Análisis de datos

3.5.1 Datos y grupos

Tal y como se puede ver en el diseño del estudio, tengo 6 muestras divididas en 2 grupos. El primero de ellos son 3 muestras de células tratadas con DMSO y el segundo grupo son 3 muestras de células tratadas con Paclitaxel. En mi estudio se utiliza el array de Affymetrix que pertenece al grupo de los de un color. El resultado de escanear la imagen de este tipo de arrays es un archivo de extensión .CEL. Accedo a la web y me descargo los 6 archivos .CEL, uno por cada muestra³.

3.5.2 Control de calidad de los datos crudos

A través de GEOquery, sin necesidad de descargarse ningún archivo, se pueden cargar los datos de cualquier estudio. Esto es lo primero que hago pero al hacer el Control de calidad de los datos crudos a partir del expressionset puedo comprobar que o bien los datos han sido tratados previamente o que los datos son comparables sin necesidad de normalizar.

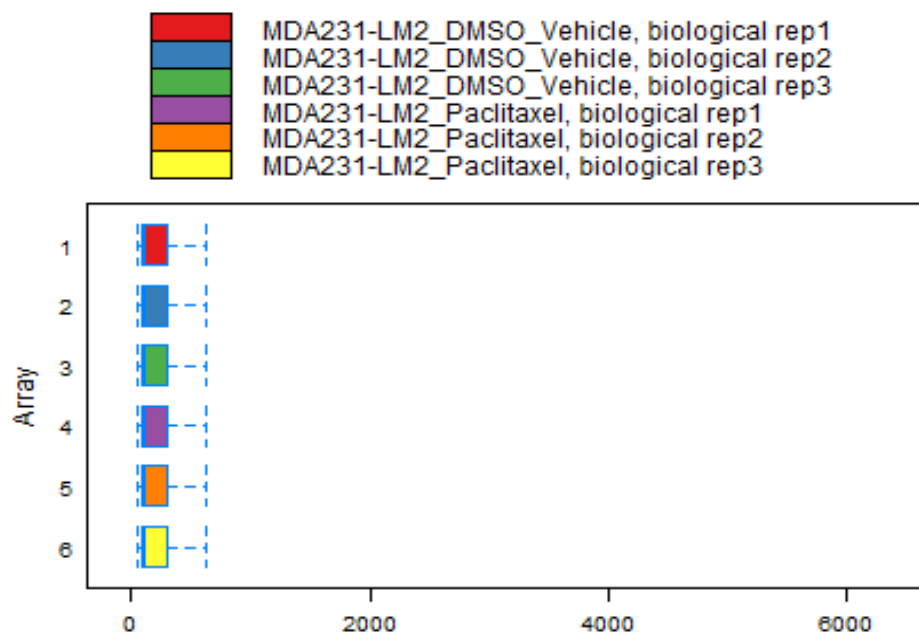


Figura1: Boxplot aplicando control de calidad de datos cargados con GEOquery

³ <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE98238>

Cargo los datos de los archivos CEL descargados previamente. Instalo el paquete affy y con la función ReadAffy leo los archivos. Realizo sobre estos datos el control de calidad de los datos crudos con la función arrayQualityMetrics. Se genera un informe en el directorio elegido, en él puedo consultar el boxplot pero en este caso el boxplot lo genero yo con la función boxplot.

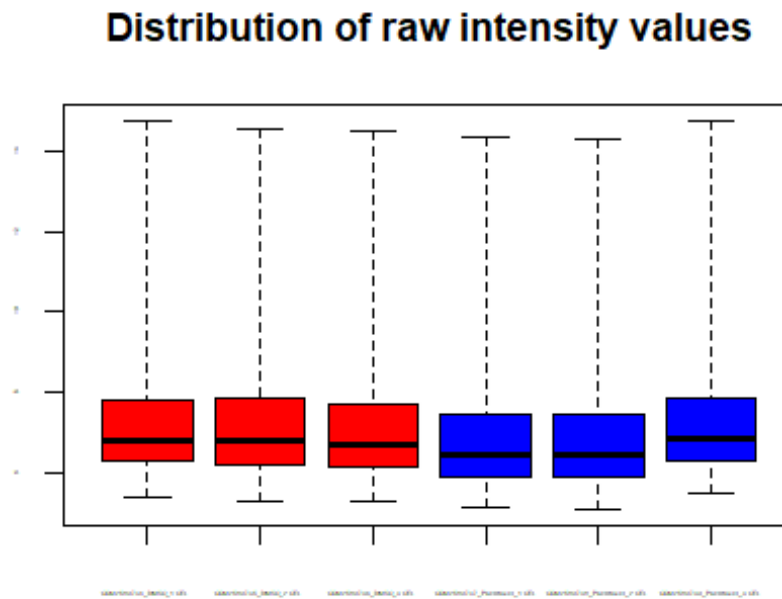


Figura2: Boxplot de datos cargados desde archivos CEL

Se observa un mínimo desplazamiento de unos arrays respecto a otros pero el resultado es realmente bueno sin necesidad de rechazar ninguna muestra. Los 3 primeros, en rojo, son los tratados con DMSO y en azul, los tratados con Paclitaxel.

En el archivo index.html que se genera tras realizar el control de calidad, en el resumen inicial aparece el análisis de los datos evaluando varios criterios de calidad. En este primero se evalúan 6 criterios y solamente en el array 2 y en el 5 aparece una cruz en alguno de los criterios. Con sólo una cruz no es necesario descartar dichos arrays.

	array	sampleNames	*1	*2	*3	*4	*5	*6	sample	ScanDate
<input type="checkbox"/>	1	GSM2589734_DMSO_1.CEL							1	2016-04-06T14:12:58Z
<input checked="" type="checkbox"/>	2	GSM2589735_DMSO_2.CEL			x				2	2016-04-06T11:31:14Z
<input type="checkbox"/>	3	GSM2589736_DMSO_3.CEL							3	2016-04-06T11:42:27Z
<input type="checkbox"/>	4	GSM2589737_Paclitaxel_1.CEL							4	2016-04-06T13:40:49Z
<input checked="" type="checkbox"/>	5	GSM2589738_Paclitaxel_2.CEL					x		5	2016-04-06T12:27:24Z
<input type="checkbox"/>	6	GSM2589739_Paclitaxel_3.CEL							6	2016-04-06T13:12:36Z

Figura3: Tabla resumen de index

En el gráfico de análisis de componentes principales generado por arrayQualityMetrics con estos datos sin normalizar, detecto que los arrays 2 y 3, ambos de las líneas celulares tratadas con DMSO y los arrays 4 y 5, de las tratadas con Paclitaxel, se agrupan de forma natural. Sin embargo en el array 1 y el 6 no se ve proximidad en este gráfico.

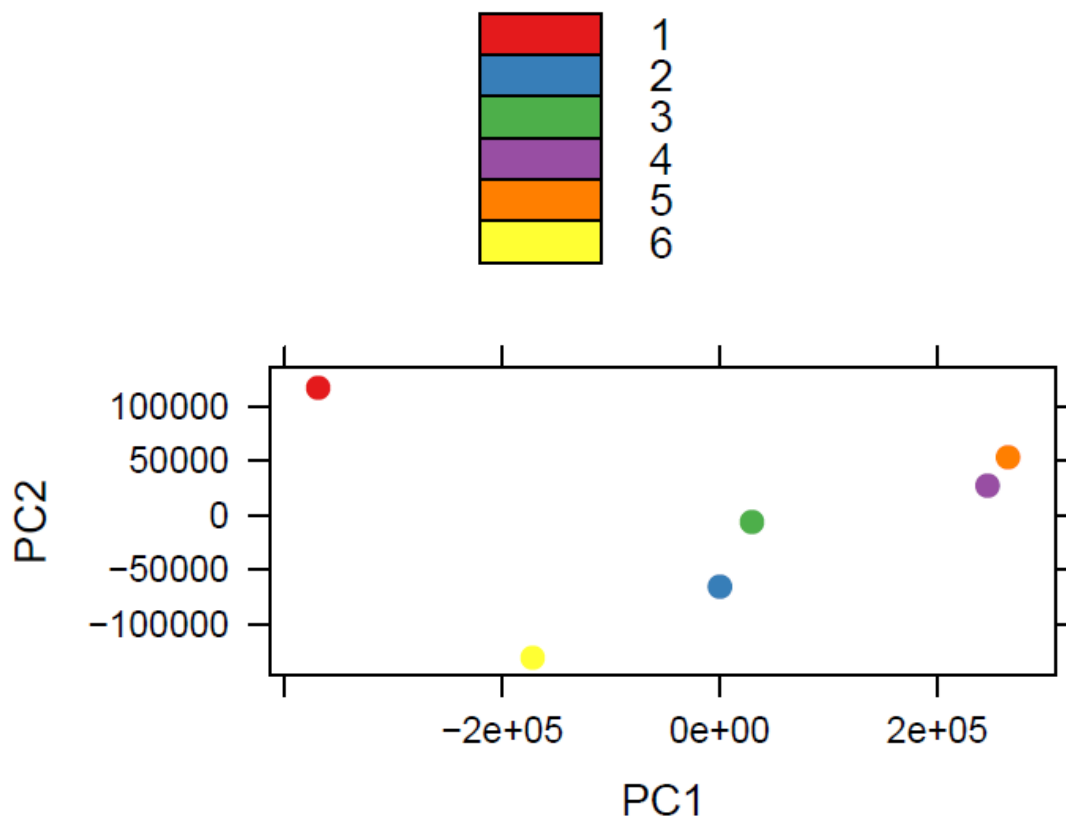


Figura4:PCA de index

3.5.3 Normalización

Una vez que acepto que la calidad de los datos es aceptable, paso a la normalización. Muchas son las ventajas de este paso del análisis de datos, las dos principales: nos permite eliminar el ruido de fondo y nos permite hacer comparables todos los valores del estudio. Uno de los problemas de la normalización es que podemos detectar sesgos, ya que todos los

datos se han igualado y no podemos eliminarlos, por lo que el análisis de calidad previo es importante.

Hay distintas formas de normalizar, al tratarse de un array de Affymetrix voy a utilizar el método RMA. Es un método basado en la modelización de las intensidades de las sondas que, en vez de basarse en las distintas sondas de un gen dentro de un mismo array, se basa en los distintos valores de la misma sonda entre todos los arrays disponibles⁴.

```
affyceles <- rma(affyceset)

## Background correcting
## Normalizing
## Calculating Expression
```

3.5.4 Control de calidad de los datos normalizados

Este es un paso opcional en el pipeline del análisis del enunciado de la PEC1. Lo realizo ya que se pueden observar muy fácilmente los efectos de la normalización de los datos. En el resumen del archivo index.html puedo comprobar que ya no aparece ninguna cruz en ninguno de los 3 criterios de calidad que aplica, esto quiere decir que los datos son comparables.

	array	sampleNames	*1	*2	*3	sample	ScanDate
<input type="checkbox"/>	1	GSM2589734_DMSO_1.CEL				1	2016-04-06T14:12:58Z
<input type="checkbox"/>	2	GSM2589735_DMSO_2.CEL				2	2016-04-06T11:31:14Z
<input type="checkbox"/>	3	GSM2589736_DMSO_3.CEL				3	2016-04-06T11:42:27Z
<input type="checkbox"/>	4	GSM2589737_Paclitaxel_1.CEL				4	2016-04-06T13:40:49Z
<input type="checkbox"/>	5	GSM2589738_Paclitaxel_2.CEL				5	2016-04-06T12:27:24Z
<input type="checkbox"/>	6	GSM2589739_Paclitaxel_3.CEL				6	2016-04-06T13:12:36Z

Figura5: Tabla resumen de index datos normalizados

En el análisis de componentes principales podemos observar como se agrupan los arrays del grupo tratado con DMSO y los arrays del grupo tratado con Paclitaxel.

⁴ http://materials.cv.uoc.edu/daisy/Materials/PID_00192730/pdf/PID_00192743.pdf

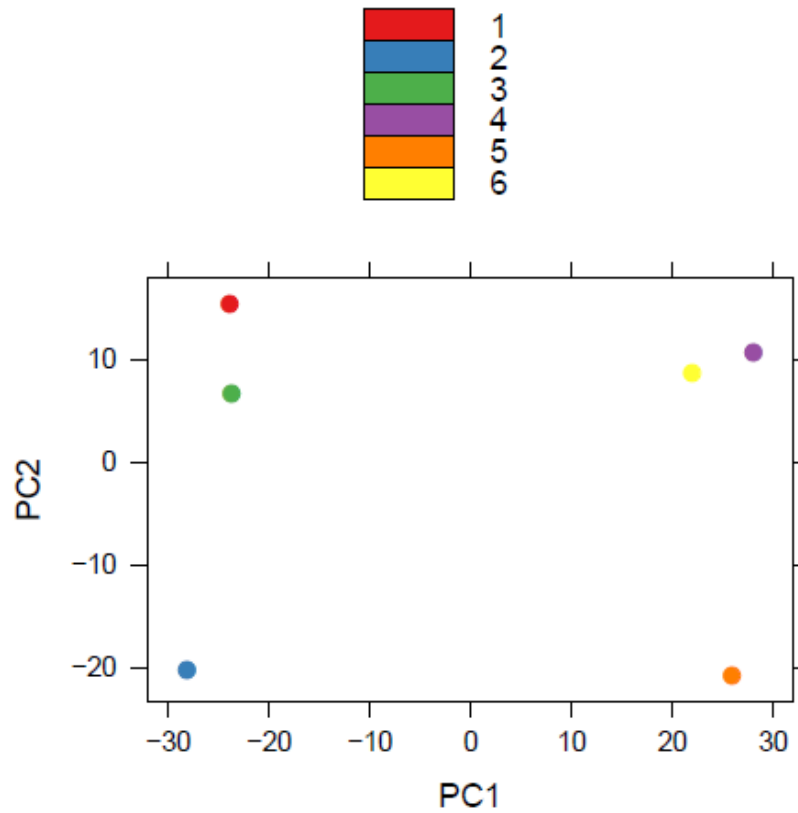


Figura6: PCA datos normalizados

Realizo también el boxplot donde se puede ver claramente que los valores están en una escala en donde se pueden comparar.

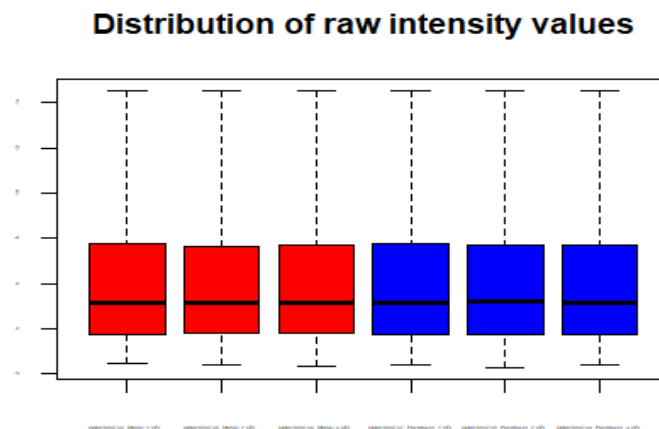


Figura7: Boxplot de datos normalizados

Los arrays de Affymetrix contienen millones de sondas por lo que no pueden examinarse a simple vista. A pesar de ello se puede obtener por ejemplo un MPlot de un canal. La única forma de definir M (el log ratio) es comparar entre cada array o bien con otros arrays o bien con un array de referencia creado, por ejemplo tomando gen a gen la mediana de todas las expresiones⁵.

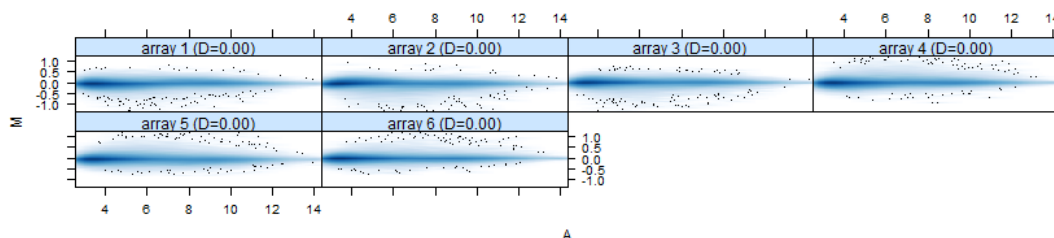


Figura8: MPlot de un canal

3.5.5 Identificación de genes diferencialmente expresados

Realizo el análisis basado en modelos lineales. Para la aplicación de dicho modelo lineal, lo primero que hago es definir la matriz de diseño.

```
##                               DMSO Paclitaxel
## GSM2589734_DMSO_1.CEL        1          0
## GSM2589735_DMSO_2.CEL        1          0
## GSM2589736_DMSO_3.CEL        1          0
## GSM2589737_Paclitaxel_1.CEL   0          1
## GSM2589738_Paclitaxel_2.CEL   0          1
## GSM2589739_Paclitaxel_3.CEL   0          1
```

Con el modelo lineal definido a través de una matriz de diseño, pueden formularse las preguntas de interés como contrastes, es decir, comparaciones entre los parámetros del modelo. Para ello creo la matriz de contrastes donde en mi estudio se compara que la línea celular haya sido tratada con DMSO o con Paclitaxel.

```
##                               Contrasts
## Levels      Paclitaxel - DMSO
## DMSO                -1
## Paclitaxel           1
```

⁵ http://materials.cv.uoc.edu/daisy/Materials/PID_00192730/pdf/PID_00192743.pdf

Una vez creadas ambas matrices paso a estimar el modelo donde puedo comparar la expresión de los genes según las muestras hayan sido tratadas con DMSO o Paclitaxel. Podré observar si hay genes diferencialmente expresados.

El análisis proporciona los estadísticos de test habituales como Fold-change o p-valores ajustados, que se utilizan para ordenar los genes de más a menos diferencialmente expresados.

En los arrays de Affymetrix sabemos que cada valor no corresponde a la expresión de un gen sino de una sonda y hay varios valores (sondas) por cada gen.

Fijando el lfc en 1 salen 257 valores sobreexpresados en los tratados con Paclitaxel respecto a los tratados con DMSO.

```
##      Paclitaxel - DMSO
## Down                0
## NotSig             54418
## Up                 257
```

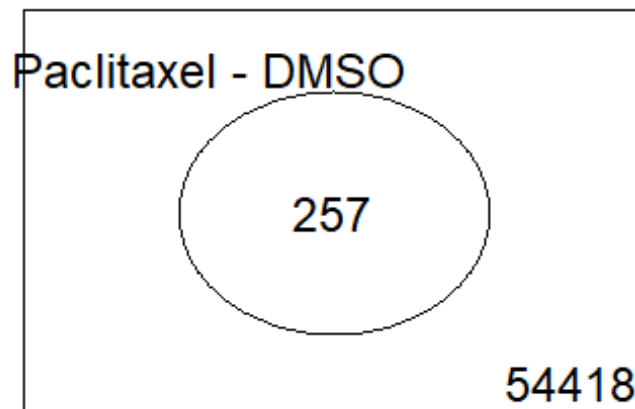


Figura9: Diagrama genes diferencialmente expresados

Pruebo a bajar a 0,5 el valor del lfc, así observo 273 genes down regulados y 1323 genes up regulados.

```
##      Paclitaxel - DMSO
## Down                273
```

```
## NotSig          53079
## Up              1323
```

3.5.6 Anotación de los resultados

La identificación de los genes seleccionados será más sencilla si le asigno el nombre y el símbolo del gen. A este proceso se le llama anotación, utilizo la tabla generada por la función `toptable` y me descargo la librería específica del array utilizado en mi estudio (Affymetrix GeneChip® Human Genome U133 Plus 2.0 microarray). Haciendo esto se asocian los identificadores que aparecen en la tabla, con características como el Símbolo del gen, el identificador del gen (EntrezID) o la descripción del gen.

```
connombres <- annotatedTopTable(toptarmacont, anotPackage="hgu133plus2.db")
## 'select()' returned 1:many mapping between keys and columns
```

Ordeno de menor a mayor p valor ajustado y muestro los 4 con menor p valor ajustado.

```
##          PROBEID SYMBOL ENTREZID
## 1132    1553736_at ZFC3H1    196441
## 24144  213183_s_at CDKN1C     1028
## 19786   209189_at  FOS       2353
## 31152  219534_x_at CDKN1C     1028
##
##                                     GENENAME      logFC
AveExpr
## 1132                                zinc finger C3H1-type containing 2.074206
5.248166
## 24144                                cyclin dependent kinase inhibitor 1C 2.138604
5.402596
## 19786 Fos proto-oncogene, AP-1 transcription factor subunit 2.424858
8.289016
## 31152                                cyclin dependent kinase inhibitor 1C 2.294521
6.456400
##          t      P.Value    adj.P.Val      B
## 1132  33.23183 4.055458e-10 2.217322e-05 12.09450
## 24144  27.64489 1.843481e-09 5.039617e-05 11.26561
## 19786  26.18148 2.880934e-09 5.250502e-05 10.99185
## 31152  23.95620 5.966146e-09 5.436650e-05 10.51727
```

Puede obtenerse una visualización de la expresión diferencial global utilizando gráficos volcano plot. Estas gráficas muestran si hay muchos o pocos genes diferenciados, en mi caso como son muchos, en lugar de ponerles el nombre los resalto en otro color.

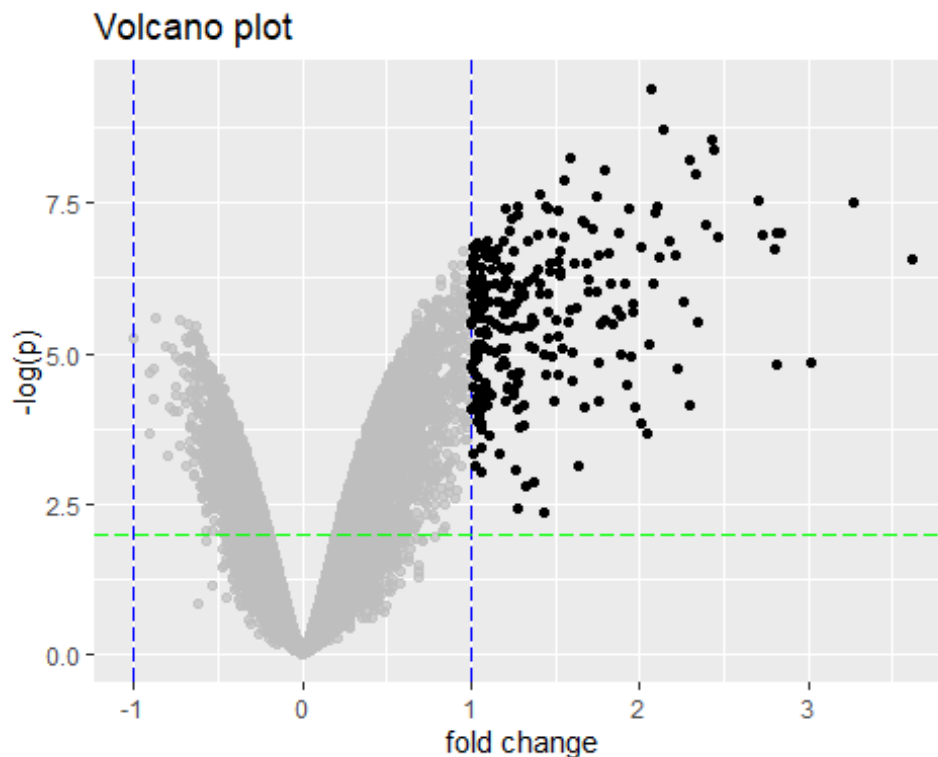


Figura10: Volcano plot genes diferencialmente expresados

Selecciono aquellos con $\log FC > 2$ y con el menos logaritmo del p valor mayor a 1. Los ordenos de menor a mayor p valor ajustado y saco el resultado de los 4 primeros.

##	PROBEID	SYMBOL	ENTREZID	GENENAME	logFC	
##	11573	201566_x_at	ID2	3398	inhibitor of DNA binding 2	2.042212
##	14581	204470_at	CXCL1	2919	C-X-C motif chemokine ligand 1	2.008540
##	20413	209774_x_at	CXCL2	2920	C-X-C motif chemokine ligand 2	2.302282
##	25546	214455_at	H2BC4	8347	H2B clustered histone 4	2.218867
##	AveExpr	t	P.Value	adj.P.Val	B	
##	11573	5.712298	6.258288	2.094538e-04	0.008647600	0.9709774
##	14581	8.489757	6.625965	1.401215e-04	0.006943269	1.3895860
##	20413	9.168898	7.280843	7.114518e-05	0.004737045	2.0920575
##	25546	6.677119	8.759285	1.802476e-05	0.002215687	3.4968726

4.Resultados

Se observan un número bastante elevado de genes que incrementan su expresión entre las células tratadas con Paclitaxel. Estos genes pueden estar implicados en los mecanismos de resistencia a quimioterapia que desarrollan las células metastásicas de cáncer de mama. Su inhibición por lo tanto podría ser una nueva estrategia terapéutica. En el artículo publicado con los datos de mi experimento se concluye que la quimioterapia produce una actividad elevada de JNK que incrementa la actividad en la matriz extracelular (ECM), cicatrización de heridas y una red de células madre en las células cancerosas lo que conduce a una

menor eficacia terapéutica. El tratamiento con fármacos quimioterapéuticos induce la actividad de JNK en las células de cáncer de mama, lo que refuerza la producción de SPP1 y TNC que promueven las metástasis pulmonares M (2020). Con lo que la inhibición tanto de JNK como de la expresión de SPP1 o TNC hace que los tumores mamarios sean más sensibles a la quimioterapia, pudiendo ser esta una futura estrategia en el tratamiento del cáncer de mama metastásico Insua-Rodríguez (2018).

5. Discusión

El análisis de microarrays presentado en este informe es un análisis sencillo cuyo fin principal era aprender las herramientas disponibles para el análisis de datos. Un inconveniente es el número de muestras, solamente 6 con lo que el estudio tendrá muy poca potencia estadística. Es muy importante la elección de un correcto tamaño muestral. Si queremos ganar precisión podremos recurrir a la replicación o repetición de un experimento de forma idéntica en un número determinado de unidades. Para futuros experimentos sería interesante la aleatorización, es decir, la asignación de todos los factores al azar a las unidades experimentales. Con ello se consigue disminuir el efecto de los factores no controlados por el experimentador en el diseño experimental y que podrían influir en los resultados. Otra herramienta con la que podríamos mejorar los resultados en futuros estudios es el bloqueo o control local que consiste en agrupar las unidades experimentales de forma que la variabilidad dentro de los grupos sea inferior a la variabilidad de todas las unidades antes de agrupar.

6. Apéndice

Tal y cómo se ha hablado en repetidas ocasiones, la reproducibilidad del estudio es fundamental a la hora de trabajar como bioinformáticos, con lo que creo un repositorio en Github con todo lo relativo al proyecto de forma que se pueda clonar en otro ordenador y reproducir mi trabajo⁶.

URL(puesta también al inicio del informe): https://github.com/bpardom/AO_PEC1.git

Pongo a continuación el código de R utilizado para la realización del análisis. También está disponible en el documento RMD disponible en repositorio GitHub:

#Cargo las librerías que utilizo para la realización del análisis

```
library(BiocManager)
library(GEOquery)
library(affy)
library(arrayQualityMetrics)
library(limma)
library(hgu133plus2.db)
```

⁶ <https://cfss.uchicago.edu/setup/git-with-rstudio/>

```

library(ggplot2)
library(gghighlight)

#Cargo Los datos utilizando GEOquery

library(BiocManager)
library(GEOquery)
Mi_gse <- getGEO("GSE98238", destdir = "C:/Bea/Master/Datos omicos Bea/PEC1",
                GSEMatrix = TRUE)

## Found 1 file(s)

## GSE98238_series_matrix.txt.gz

## Using locally cached version: C:/Bea/Master/Datos omicos
Bea/PEC1/GSE98238_series_matrix.txt.gz

## Parsed with column specification:
## cols(
##   ID_REF = col_character(),
##   GSM2589734 = col_double(),
##   GSM2589735 = col_double(),
##   GSM2589736 = col_double(),
##   GSM2589737 = col_double(),
##   GSM2589738 = col_double(),
##   GSM2589739 = col_double()
## )

## Using locally cached version of GPL570 found here:
## C:/Bea/Master/Datos omicos Bea/PEC1/GPL570.soft

#Control de calidad de datos cargados con GEOquery

arrayQualityMetrics(Mi_gse$GSE98238_series_matrix.txt.gz,
                    outdir = "C:/Bea/Master/Datos omicos Bea/PEC1/arrayQM",
                    intgroup =
colnames(Mi_gse$GSE98238_series_matrix.txt.gz@phenoData),
                    reporttitle = "arrayQualityMetrics",
                    do.logtransform = FALSE, force = TRUE)

## The report will be written into directory 'C:/Bea/Master/Datos omicos
Bea/PEC1/arrayQM'.

## (loaded the KernSmooth namespace)

#Cargo Los archivos CEL

archivos_cel <- list.celfiles("C:/Bea/Master/Datos omicos Bea/PEC1/Data/CEL")
affycel <- ReadAffy(celfile.path = "C:/Bea/Master/Datos omicos
Bea/PEC1/Data")

```

```

#Control de calidad de datos contenidos en archivos CEL

arrayQualityMetrics(affycel,
                     outdir = "C:/Bea/Master/Datos omicos
Bea/PEC1/arrayQMcel",
                     intgroup = colnames(affycel@phenoData@data),
                     reporttitle = "affycelquality",
                     do.logtransform = FALSE, force = TRUE)

## The report will be written into directory 'C:/Bea/Master/Datos omicos
Bea/PEC1/arrayQMcel'.

#Normalización RMA

affyceles <- rma(affycel)

## Background correcting
## Normalizing
## Calculating Expression

#Control de calidad de datos normalizados

arrayQualityMetrics(affyceles,
                     outdir = "C:/Bea/Master/Datos omicos
Bea/PEC1/arrayQMceles",
                     intgroup = colnames(affyceles@phenoData),
                     reporttitle = "affycelesquality",
                     do.logtransform = FALSE, force = TRUE)

## The report will be written into directory 'C:/Bea/Master/Datos omicos
Bea/PEC1/arrayQMceles'.

#Creación de matriz de diseño

madis <- cbind(DMSO = c(1,1,1,0,0,0),
               Paclitaxel = c(0,0,0,1,1,1))
rownames(madis) <- rownames(affyceles@phenoData)

#Creación de matriz de contraste

macont <- makeContrasts(Paclitaxel - DMSO, levels= madis)

#Estimación del modelo lineal

lineartar <- lmFit(affyceles, madis)

tarmacont <- contrasts.fit(lineartar, macont)
tarmacont <- eBayes(tarmacont)
toptarmacont <- topTable(tarmacont, number=nrow(tarmacont),
                        coef = "Paclitaxel - DMSO" , adjust="fdr")
head(toptarmacont, n = 2)

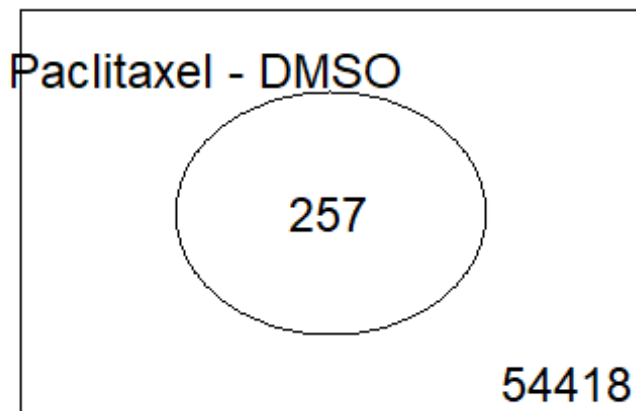
```

```
##           logFC AveExpr      t      P.Value    adj.P.Val      B
## 1553736_at  2.074206 5.248166 33.23183 4.055458e-10 2.217322e-05 12.09450
## 213183_s_at 2.138604 5.402596 27.64489 1.843481e-09 5.039617e-05 11.26561

resultados<-decideTests(tarmacont, method="separate", adjust.method="fdr",
p.value=0.1, lfc=1)
sumabs<-apply(abs(resultados),1,sum)
rescero<-resultados[sumabs!=0,]
print(summary(resultados))

##           Paclitaxel - DMSO
## Down                      0
## NotSig                   54418
## Up                       257

vennDiagram(resultados)
```



```
resultados2<-decideTests(tarmacont, method="separate", adjust.method="fdr",
p.value=0.1, lfc=0.5)
sumabs<-apply(abs(resultados2),1,sum)
rescero<-resultados2[sumabs!=0,]
print(summary(resultados2))

##           Paclitaxel - DMSO
## Down                      273
## NotSig                   53079
## Up                       1323
```

#Anotación de genes

```
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")
  return(annotatedTopTab)
}

connombres <- annotatedTopTable(toptarmacont, anotPackage="hgu133plus2.db")

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

Topgenedif = connombres [order(connombres$adj.P.Val,decreasing = FALSE), ]
head(Topgenedif, n = 4)

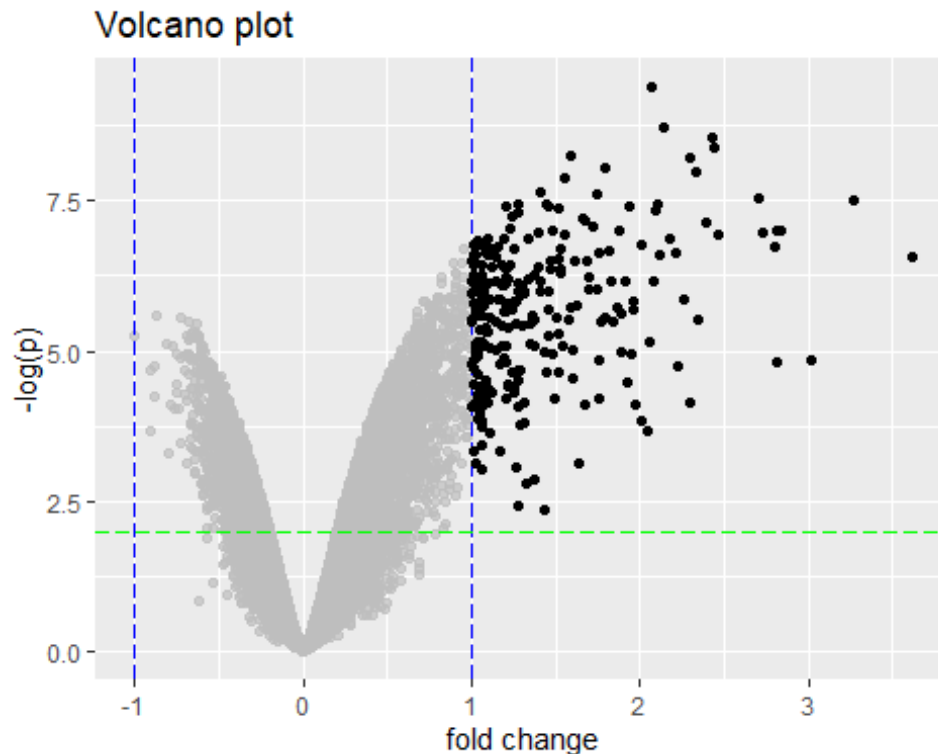
##          PROBEID SYMBOL ENTREZID
## 1132   1553736_at ZFC3H1   196441
## 24144 213183_s_at CDKN1C    1028
## 19786  209189_at  FOS      2353
## 31152 219534_x_at CDKN1C    1028
##                                     GENENAME      logFC
AveExpr
## 1132                                zinc finger C3H1-type containing 2.074206
5.248166
## 24144                                cyclin dependent kinase inhibitor 1C 2.138604
5.402596
## 19786 Fos proto-oncogene, AP-1 transcription factor subunit 2.424858
8.289016
## 31152                                cyclin dependent kinase inhibitor 1C 2.294521
6.456400
##          t      P.Value    adj.P.Val      B
## 1132 33.23183 4.055458e-10 2.217322e-05 12.09450
## 24144 27.64489 1.843481e-09 5.039617e-05 11.26561
## 19786 26.18148 2.880934e-09 5.250502e-05 10.99185
## 31152 23.95620 5.966146e-09 5.436650e-05 10.51727
```

#Elaboración Volcano plot

```
attach(connombres)
ggplot(data = connombres, aes(x = logFC, y = -log10(P.Value))) +
  geom_point()+
  geom_vline(xintercept = -1,colour="blue", linetype = "longdash")+
  geom_vline(xintercept = 1,colour="blue", linetype = "longdash")+
  geom_hline(yintercept = -log10(0.01),colour="green", linetype =
"longdash")+
```



```
xlab("fold change")+
ylab("-log(p)")+
labs(title = "Volcano plot") +
gghighlight(-log10(P.Value) > 2 & abs(logFC) > 1)
```



```
detach(connombres)

attach(connombres)
genesdif<-subset(connombres, logFC > 2 & abs(-log10(P.Value)) > 1)
detach(connombres)
genes = genesdif [order(genesdif$adj.P.Val,decreasing = TRUE), ]
head(genes,4)
```

##	PROBEID	SYMBOL	ENTREZID	GENENAME	logFC
##	11573	201566_x_at	ID2	3398 inhibitor of DNA binding 2	2.042212
##	14581	204470_at	CXCL1	2919 C-X-C motif chemokine ligand 1	2.008540
##	20413	209774_x_at	CXCL2	2920 C-X-C motif chemokine ligand 2	2.302282
##	25546	214455_at	H2BC4	8347 H2B clustered histone 4	2.218867

```
## AveExpr t P.Value adj.P.Val B
## 11573 5.712298 6.258288 2.094538e-04 0.008647600 0.9709774
## 14581 8.489757 6.625965 1.401215e-04 0.006943269 1.3895860
## 20413 9.168898 7.280843 7.114518e-05 0.004737045 2.0920575
## 25546 6.677119 8.759285 1.802476e-05 0.002215687 3.4968726
```

7. Referencias bibliográficas

Insua-Rodríguez, Jacob. 2018. "Stress Signaling in Breast Cancer Cells Induces Matrix Components That Promote Chemoresistant Metastasis." *EMBO Mol Med* online.

M, Pein. 2020. "Metastasis-Initiating Cells Induce and Exploit a Fibroblast Niche to Fuel Malignant Colonization of the Lungs." *Nat Commun* online.