

AdriaFernandez_ADO_PEC1

Adria_Fernandez

30/4/2020

Contents

1	Abstract	2
2	Objectius	2
3	Materials	2
3.1	Software	2
3.2	Dades	2
4	Mètodes	2
4.1	Pipeline	2
4.2	Procediment	3
5	Resultats:	12
6	Discusió:	12
7	Apèndix:	13

1 Abstract

Aquest informe esta basat en l'estudi de les dades de microarray extretes del Gene Expression Omnibus (GSE148501), en el qual s'estudiava l'efecte d'un oligo-flucoidan en la prevenció del desenvolupament de l'hepatocarcinogènesis (estimulada per una mutació en p53 amb presència de HBx). Partim d'unes dades de microarrays de Affymatrix sobre els fetges de peix zebra mutats (tots femelles i de mateixa edat) després del seguiment d'una dieta hipocalòrica on part del grup han estat tractats amb OF.

El codi d'execució, així com al resta de fitxers descarregats i generats es poden trobar a : [https://github.com/Adria-FerDi/ADO_PEC1]

2 Objectius

L'objectiu d'aquest exercici es el d'estudiar els efectes de OF en la prevenció de la hepatocarcinogènesis en Peix Zebra. Això sefektuarà mitjançant el re-anàlisi de microarrays d'expressió gènica de les dades de l'estudi. Comparant els nivells dels diferents condicions (Control-Dieta hipocalòrica-Dieta hipocalòrica+Oligo-fucoidan) per als dos tipus de transgènics (p53+/-).

3 Materials

3.1 Software

Aquest informe ha utilitzat [R]¹ Com a programari principal de treball, mitjançant l'interfaç [RStudio]². Algunes llibreries emprades no formen part del llistat de R, i s'ha recolzat molt en el paquet [Bioconductor]³ com a eina principal per a l'anàlisi de dades òmiques.

3.2 Dades

Les dades emprades són d'accés públic des de la base de dades [GEO]⁴.l'experiment, amb codi GSE148501, tracta amb 17 individus, organitzats en 3 factors:

- Mutació (p53 / p53-)
- Dieta (Control / Obesa)
- Tractament (cap / Oligo-fucoidan)

En l'estudi base s'han emprat microarrays Affymetrix [ZebGene-1_1-st] Zebrafish Gene 1.1 ST Array [transcript (gene) version]

4 Mètodes

4.1 Pipeline

- 1 Preparació del entron de treball i descàrrega de dades.
- 2 Identificar grups i classificar individus.

¹<https://cran.r-project.org/index.html>

²<https://www.rstudio.com/>

³<https://www.bioconductor.org/>

⁴<https://www.ncbi.nlm.nih.gov/geo/browse/?view=series>

- 3 Control de qualitat de les dades.
- 4 Normalització de les dades.
- 5 Control de qualitat de les dades normalitzades.
- 6 Filtratge no específic.
- 7 Control de qualitat de les dades filtrades.
- 8 Identificació de gens diferencialment expressats.
- 9 Anotació dels resultats
- 10 Comparació entre comparacions
- 11 Anàlisi de significancia biològica.

4.2 Procediment

4.2.1 Preparació de l'entorn de treball i descàrrega de dades:

Primer de tot generem les carpetes necessàries per a l'efectuació del treball. I posteriorment descarreguem les dades des de GEO amb la funció `getGEOSuppFiles()`. Aquest procés ha necessitat de la llibreria `GEOquery`.

4.2.2 Identificar grups i classificar individus:

A partir de les dades de GEO, generem un arxiu `targets.csv` amb la classificació dels arxius a cadascun dels grups, el qual té la següent forma (per mostrar-ho aquí s'han canviat els noms dels arxius per a una millor visualització):

##	FileName	Dieta	Tractament	Transgen	Grup
##	GSM4472532_C	X	control	none	HBx_Src p53+Ctrl
##	GSM4472533_C	X	control	none	HBx_Src p53+Ctrl
##	GSM4472534_C	X	control	none	HBx_Src p53+Ctrl
##	GSM4472535_Ob	X	obesity	none	HBx_Src p53+Obes
##	GSM4472536_Ob	X	obesity	none	HBx_Src p53+Obes
##	GSM4472537_Ob	X	obesity	none	HBx_Src p53+Obes
##	GSM4472538_ObOF	X	obesity	OF	HBx_Src p53+Obes_OF
##	GSM4472539_ObOF	X	obesity	OF	HBx_Src p53+Obes_OF
##	GSM4472540_ObOF	X	obesity	OF	HBx_Src p53+Obes_OF
##	GSM4472541_p53-C	X	control	none	HBx_Src_p53- p53-Ctrl
##	GSM4472542_p53-C	X	control	none	HBx_Src_p53- p53-Ctrl
##	GSM4472543_p53-Ob	X	obesity	none	HBx_Src_p53- p53-Obes
##	GSM4472544_p53-Ob	X	obesity	none	HBx_Src_p53- p53-Obes
##	GSM4472545_P53-Ob	X	obesity	none	HBx_Src_p53- p53-Obes
##	GSM4472546_P53-ObOF	X	obesity	OF	HBx_Src_p53- p53-Obes_OF
##	GSM4472547_P53-ObOF	X	obesity	OF	HBx_Src_p53- p53-Obes_OF
##	GSM4472548_P53-ObOF	X	obesity	OF	HBx_Src_p53- p53-Obes_OF

Seguidament bolquem el contingut dels arxius al nostre entorn R, mitjançant la funció `read.celfiles`, combinant la llista dels noms d'arxius i el nostre dataframe de `targets`, generem l'objecte `raw_data`. Aquest procés ha necessitat de les llibreries `oligo` i `Biobase`.

4.2.3 Control de qualitat de les dades:

En aquest punt evaluem la qualitat de les dades per a continuar amb l'estudi. * Primer de tot, mitjançant la funció `arrayQualityMetrics()` generem un conjunt de anàlisis de les dades. * Seguidament farem l'anàlisi gràfic mitjançant un boxplot (Figure 2) i un cluster jeràrquic (Figure 3).

Del resum de l'anàlisi (Figure 1) podem extreure que hi ha diverses mostres que tenen outliers i per tant no tenen una qualitat òptima. Del boxplot (Figure 2) podem extreure que hi ha relatives diferències entre les mostres, i que no semblen agrupar-se del tot segons els grups, tenint com exemple clar el cas de l'última mostra, que està força diferenciada. Això es corrobora amb el cluster (Figura 3) on s'observa que si més no la majoria de grups estan relativament agrupats, hi ha un cas visiblement allunyat. Aquest procés ha necessitat de les llibreries `arrayQualityMetrics`, `ggplot2` i `ggrepel`.

array	sampleNames	*1	*2	*3	FileName	Dieta	Tractament	Transgen
1	GSM4472532_C				GSM4472532_HBx_src_NOR-2	control	none	
2	GSM4472533_C				GSM4472533_HBx_src_NOR-3	control	none	HBx_Src
3	GSM4472534_C				GSM4472534_HBx_src_NOR-4	control	none	HBx_Src
4	GSM4472535_Ob				GSM4472535_HBx_src_DIO-1	obesity	none	HBx_Src
5	GSM4472536_Ob				GSM4472536_HBx_src_DIO-2	obesity	none	HBx_Src
6	GSM4472537_Ob				GSM4472537_HBx_src_DIO-4	obesity	none	HBx_Src
7	GSM4472538_ObOF				GSM4472538_HBx_src_DIO+OF-1	obesity	OF	HBx_Src
8	GSM4472539_ObOF				GSM4472539_HBx_src_DIO+OF-7	obesity	OF	HBx_Src
9	GSM4472540_ObOF				GSM4472540_HBx_src_DIO+OF-8	obesity	OF	HBx_Src
10	GSM4472541_p53-C				GSM4472541_HBx_src_p53-NOR-1	control	none	HBx_Src_p53-
11	GSM4472542_p53-C	x	x		GSM4472542_HBx_src_p53-NOR-2	control	none	HBx_Src_p53-
12	GSM4472543_p53-Ob		x		GSM4472543_HBx_src_p53-DIO-1	obesity	none	HBx_Src_p53-
13	GSM4472544_p53-Ob		x		GSM4472544_HBx_src_p53-DIO-2	obesity	none	HBx_Src_p53-
14	GSM4472545_p53-Ob		x		GSM4472545_HBx_src_p53-DIO-3	obesity	none	HBx_Src_p53-
15	GSM4472546_p53-ObOF				GSM4472546_HBx_src_p53-DIO+OF-2	obesity	OF	HBx_Src_p53-
16	GSM4472547_p53-ObOF				GSM4472547_HBx_src_p53-DIO+OF-3	obesity	OF	HBx_Src_p53-
17	GSM4472548_p53-ObOF	x	x		GSM4472548_HBx_src_p53-DIO+OF-6	obesity	OF	HBx_Src_p53-

Figure 1: Resum dels resultats obtinguts en l'arxiu `index.html` per al control de qualitat de les dades

```
library(ggplot2)
library(ggrepel)
boxplot(raw_data, cex.axis = 0.5, las=2, which="all", main="", col = c(rep("gold2", 3), rep("darkorange", 3), rep("darkblue", 11)))
```

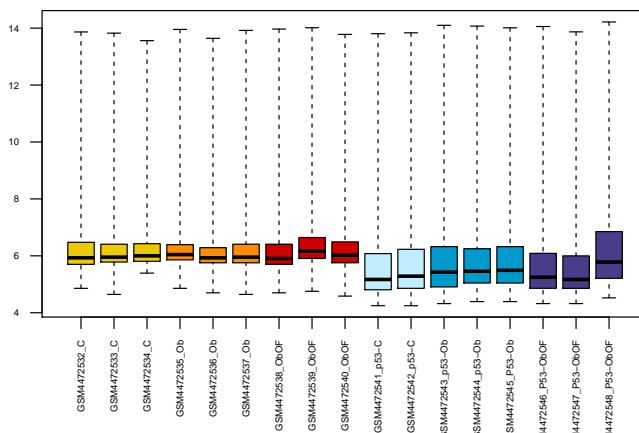


Figure 2: Boxplot d'intensitat dels arrays

```
plot(hclust(dist(t(exprs(raw_data)))))
```

4.2.4 Normalització de les dades:

Passem doncs, a la normalització de les dades, per veure si es corregeixen els errors detectats anteriorment. Amb aquets procés s'intenta que les mostres siguin més comparables entre si, intentant evadir la variabilitat deguda a araons tècniques o d'anàlisi. En aquets cas hem fet servir el mètode de Robust Multichip Analysis (RMA) mitjançant la funció `rma()`.

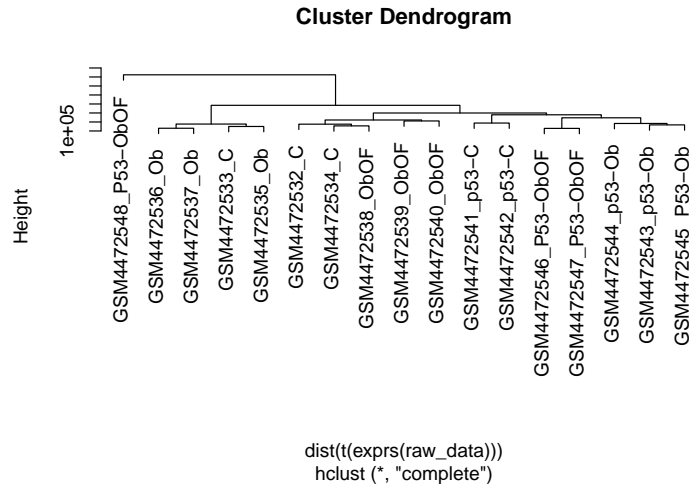


Figure 3: Cluster jeràrquic

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

4.2.5 Control de qualitat de les dades normalitzades:

Procedirem com en les dades en cru, i efectuarem els mateixos anàlisis de qualitat. En aquest cas, veiem com en el resum de l'anàlisi (Figure 4) encara queda algun grup amb outliers (que de fet no són els mateixos que tenim abans). En el cas del boxplot (Figure 5) si que s'aprecia una millora amb la normalització de les dades. Però el cluster (Figure 6) segueix tenint algunes incoherències d'agrupament, tot i que no tant abruptes com en el cas anterior.

array	sampleNames	*1	*2	*3	FileName	Dieta	Tractament	Transgen
1	GSM4472532_C				GSM4472532_HBx_src_NOR-2	control	none	
2	GSM4472533_C				GSM4472533_HBx_src_NOR-3	control	none	HBx_Src
3	GSM4472534_C				GSM4472534_HBx_src_NOR-4	control	none	HBx_Src
4	GSM4472535_Ob	x			GSM4472535_HBx_src_DIO-1	obesity	none	HBx_Src
5	GSM4472536_Ob		x		GSM4472536_HBx_src_DIO-2	obesity	none	HBx_Src
6	GSM4472537_Ob				GSM4472537_HBx_src_DIO-4	obesity	none	HBx_Src
7	GSM4472538_ObOF				GSM4472538_HBx_src_DIO+OF-1	obesity	OF	HBx_Src
8	GSM4472539_ObOF				GSM4472539_HBx_src_DIO+OF-7	obesity	OF	HBx_Src
9	GSM4472540_ObOF				GSM4472540_HBx_src_DIO+OF-8	obesity	OF	HBx_Src
10	GSM4472541_p53-C				GSM4472541_HBx_src_p53-NOR-1	control	none	HBx_Src_p53-
11	GSM4472542_p53-C				GSM4472542_HBx_src_p53-NOR-2	control	none	HBx_Src_p53-
12	GSM4472543_p53-Ob				GSM4472543_HBx_src_p53-DIO-1	obesity	none	HBx_Src_p53-
13	GSM4472544_p53-Ob				GSM4472544_HBx_src_p53-DIO-2	obesity	none	HBx_Src_p53-
14	GSM4472545_P53-Ob				GSM4472545_HBx_src_p53-DIO-3	obesity	none	HBx_Src_p53-
15	GSM4472546_P53-ObOF				GSM4472546_HBx_src_p53-DIO+OF-2	obesity	OF	HBx_Src_p53-
16	GSM4472547_P53-ObOF				GSM4472547_HBx_src_p53-DIO+OF-3	obesity	OF	HBx_Src_p53-
17	GSM4472548_P53-ObOF	x			GSM4472548_HBx_src_p53-DIO+OF-6	obesity	OF	HBx_Src_p53-

Figure 4: Resum dels resultats obtinguts en l'arxiu index.html per al control de qualitat de les dades normalitzades

4.2.6 Filtratge no específic:

Executarem un filtratge no específic de les dades, per evitar la sobrecomparació de gens, buscant com objectiu comapràr només aquells gens que estan evidentment més diferenciats en les diferents mostres.

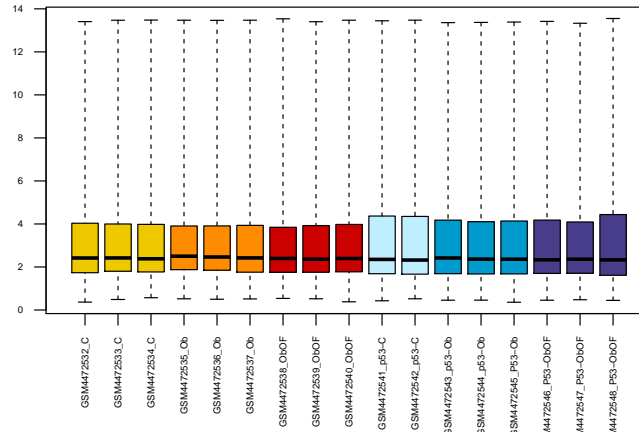


Figure 5: Boxplot d'intensitat dels arrays de les dades normalitzades

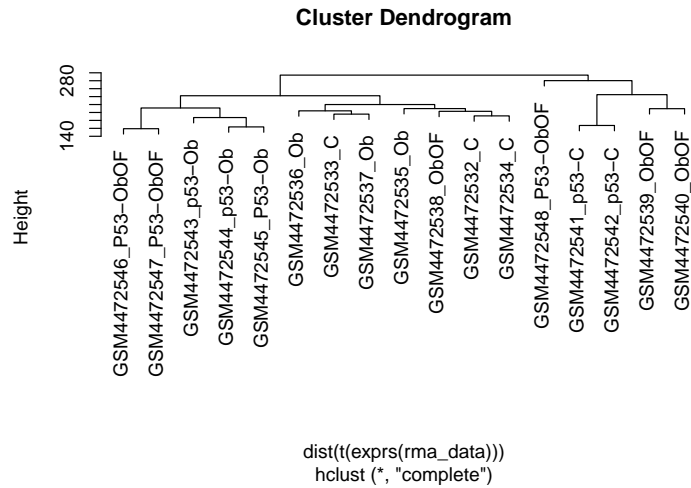


Figure 6: Cluster jeràrquic de les dades normalitzades

Primerament generem un gràfic on visualitzem la desviació standard dels gens (Figure 7) per poder decidir quins poden considerar-se diferencialment expressats. S'han afegit unes línies verticals en els percentils 90 i 95. Seguidament fem el filtrat de gens. Mitjançant la funció `nsFilter()` que permet filtrar gens marcant un umbral de variabilitat, tot junt amb el paquet d'anotacions pertinent que permet també, evitar aquells gens que no tenen un identificador associat. Aquest procés ha necessitat les llibreries `genefilter`, `DBI` i `pd.zebgene.1.1.st`

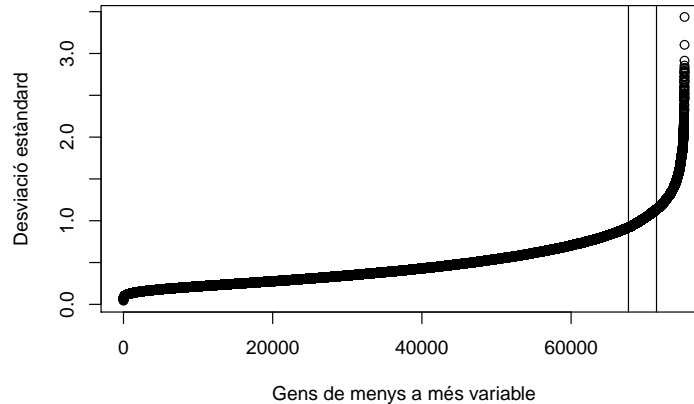


Figure 7: Distribució de la variabilitat dels gens

```
## [1] "Nombre de Gens de les dades normalitzades"
## [1] 75212
## [1] "Nombre de gens una vegada filtrats"
## [1] 22564
## [1] "Nombre de gens eliminats"
## $numLowVar
## [1] 52648
```

4.2.7 Control de qualitat de les dades filtrades:

De manera anàloga als casos anteriors, procedim amb el control de la qualitat de les dades. En el resum de l'anàlisi (Figure 8) observem que finalment no hi ha cap mostra amb outliers. En el cas del boxplot (Figure 9) es manté la millora anterior, no hi ha grans diferències d'entre els diferents grups de l'estudi. Però el cluster (Figure 10) segueix tenint algunes incoherències d'agrupament, tot i que si ens fixem en on es donen (alguns tractaments estan massa aprop dels controls), sembla que podrien ser resultats alentadors per l'estudi.

4.2.8 Identificació de gens diferencialment expressats:

La identificació de gens diferencialment expressats es basa en un conjunt de mètodes i test per comparar grups. El paquet `limma` està especialment dissenyat per a l'anàlisi de microarrays, amb les eines necessàries per a seleccionar els gens diferenciats.

4.2.8.1 Obtenció de les matrius:

array	sampleNames	1	2	3	FileName	Dieta	Tractament	Transgen
1	GSM4472532_C				GSM4472532_HBx_src_NOR-2	control	none	
2	GSM4472533_C				GSM4472533_HBx_src_NOR-3	control	none	HBx_Src
3	GSM4472534_C				GSM4472534_HBx_src_NOR-4	control	none	HBx_Src
4	GSM4472535_Ob				GSM4472535_HBx_src_DIO-1	obesity	none	HBx_Src
5	GSM4472536_Ob				GSM4472536_HBx_src_DIO-2	obesity	none	HBx_Src
6	GSM4472537_Ob				GSM4472537_HBx_src_DIO-4	obesity	none	HBx_Src
7	GSM4472538_ObOF				GSM4472538_HBx_src_DIO+OF-1	obesity	OF	HBx_Src
8	GSM4472539_ObOF				GSM4472539_HBx_src_DIO+OF-7	obesity	OF	HBx_Src
9	GSM4472540_ObOF				GSM4472540_HBx_src_DIO+OF-8	obesity	OF	HBx_Src
10	GSM4472541_p53-C				GSM4472541_HBx_src_p53-_NOR-1	control	none	HBx_Src_p53-
11	GSM4472542_p53-C				GSM4472542_HBx_src_p53-_NOR-2	control	none	HBx_Src_p53-
12	GSM4472543_p53-Ob				GSM4472543_HBx_src_p53-_DIO-1	obesity	none	HBx_Src_p53-
13	GSM4472544_p53-Ob				GSM4472544_HBx_src_p53-_DIO-2	obesity	none	HBx_Src_p53-
14	GSM4472545_P53-Ob				GSM4472545_HBx_src_p53-_DIO-3	obesity	none	HBx_Src_p53-
15	GSM4472546_P53-ObOF				GSM4472546_HBx_src_p53-_DIO+OF-2	obesity	OF	HBx_Src_p53-
16	GSM4472547_P53-ObOF				GSM4472547_HBx_src_p53-_DIO+OF-3	obesity	OF	HBx_Src_p53-
17	GSM4472548_P53-ObOF				GSM4472548_HBx_src_p53-_DIO+OF-6	obesity	OF	HBx_Src_p53-

Figure 8: Resum dels resultats obtinguts en l'arxiu index.html per al control de qualitat de les dades filtrades

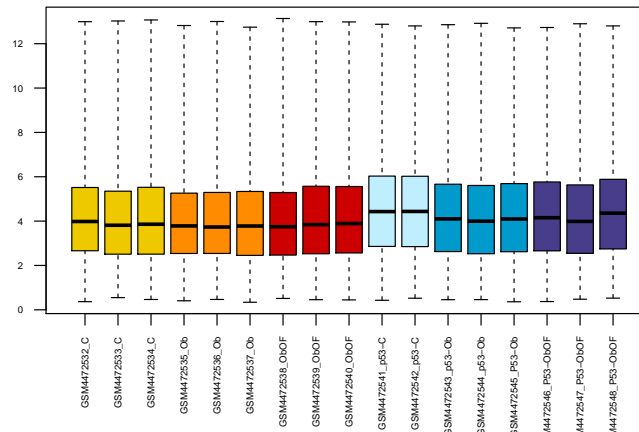


Figure 9: Boxplot d'intensitat dels arrays amb les dades filtrades

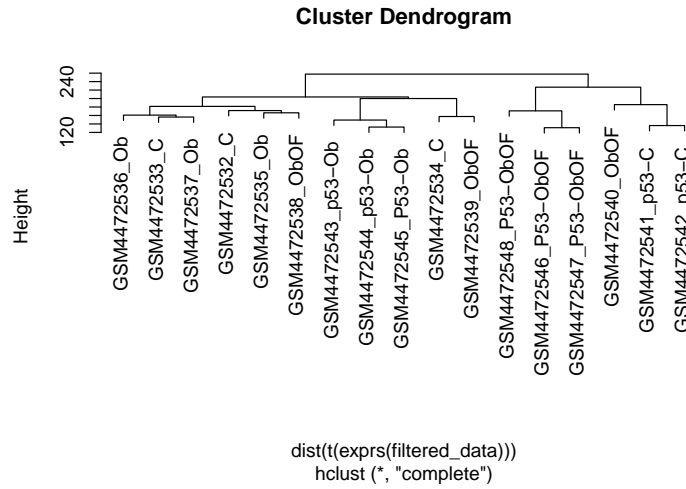


Figure 10: Cluster jeràrquic de les dades filtrades

Per executar els mètodes lineals necessitem crear una matriu de disseny on s'engloba a quin grup pertany cada mostra.

Amb aquesta matriu es poden fer múltiples comparacions, ja que tenim 6 grups, però els que estudiarem seran: control vs obesitat, control vs obesitat amb OF, i obesitat vs obesitat amb OF, duplicant-ho per a cada tipus de mutació p53. Aquest procés ha necessitat de la llibreria limma.

4.2.8.2 Estimació del model:

Amb les matrius definides podem passar a estimar les comparacions i realitzar els tests necessaris per obtenir els gens diferencialment expressats. Utilitzarem aquí del paquet limma, amb la funció `lmFit()`, basat en els models de Bayes, que retorna el Fold-Change i els p-valors ajustats.

```
fit<-lmFit(filtered_data,matriu_disseny)
fit.main<-contrasts.fit(fit,matriu_contrast)
fit.main<-eBayes(fit.main)
```

4.2.8.3 Llistat de gens

Mitjançant la funció `topTable()` de limma ordenem de manera ascendent els gens en funció de la p-value. Entenent que els primers gens seran els més diferenciats a nivell d'expressió.

```
Tab_CvsOb <- topTable(fit.main, number = nrow(fit.main),
                      coef="CvsOb", adjust="fdr")

Tab_CvsOb_OF <- topTable(fit.main, number = nrow(fit.main),
                         coef="CvsOb_OF", adjust="fdr")

Tab_ObvsOb_OF <- topTable(fit.main, number = nrow(fit.main),
                          coef="ObvsOb_OF", adjust="fdr")

Tab_p53_CvsOb <- topTable(fit.main, number = nrow(fit.main),
                          coef="p53_CvsOb", adjust="fdr")

Tab_p53_CvsOb_OF <- topTable(fit.main, number = nrow(fit.main),
                             coef="p53_CvsOb_OF", adjust="fdr")

Tab_p53_ObvsOb_OF <- topTable(fit.main, number = nrow(fit.main),
                              coef="p53_ObvsOb_OF", adjust="fdr")
```

```
head(Tab_CvsOb)
```

##		logFC	AveExpr	t	P.Value	adj.P.Val	B
##	12928077	3.520810	2.750083	7.876882	5.013657e-07	0.002339724	6.386836
##	12928995	3.520810	2.750083	7.876882	5.013657e-07	0.002339724	6.386836
##	12932403	3.851441	2.681263	7.857132	5.184639e-07	0.002339724	6.356864
##	12935111	3.851441	2.681263	7.857132	5.184639e-07	0.002339724	6.356864
##	12936297	3.851441	2.681263	7.857132	5.184639e-07	0.002339724	6.356864
##	13167184	-3.711221	4.344890	-7.411208	1.120619e-06	0.002534014	5.663938

```
head(Tab_CvsOb_OF)
```

##		logFC	AveExpr	t	P.Value	adj.P.Val	B
##	13296652	6.158737	3.758306	16.252911	1.113759e-11	2.513086e-07	14.963481
##	13243004	-6.150121	7.842997	-9.418999	4.246626e-08	3.433029e-04	8.583112
##	13250265	-6.001671	10.407449	-9.371278	4.564389e-08	3.433029e-04	8.520867

```
## 12968399 -4.429782 3.849711 -8.768154 1.162202e-07 6.555982e-04 7.706929
## 13047898 -3.589519 4.140295 -8.451629 1.931187e-07 8.715062e-04 7.258957
## 13282138 3.645191 7.557628 8.053456 3.723244e-07 1.400188e-03 6.674328
```

```
head(Tab_ObvsOb_OF)
```

```
##          logFC AveExpr          t          P.Value      adj.P.Val          B
## 13296652  5.927345 3.758306 17.488586 3.486062e-12 7.865950e-08 16.081120
## 13166611 -4.433912 6.054477 -9.023387 7.785002e-08 2.586945e-04 8.146308
## 13247288  4.497267 5.063768  8.658526 1.383777e-07 2.586945e-04 7.631354
## 13247193  4.495578 5.019007  8.483338 1.834387e-07 2.586945e-04 7.377329
## 13247199  4.495578 5.019007  8.483338 1.834387e-07 2.586945e-04 7.377329
## 13247218  4.495578 5.019007  8.483338 1.834387e-07 2.586945e-04 7.377329
```

```
head(Tab_p53_CvsOb)
```

```
##          logFC AveExpr          t          P.Value      adj.P.Val          B
## 13276618 -3.763201 4.991126 -7.458874 1.030697e-06 0.02325664 4.658840
## 12968666 -3.072267 4.914031 -6.838527 3.135419e-06 0.02475021 3.864871
## 12963242 -3.443605 4.040821 -6.812223 3.290668e-06 0.02475021 3.829701
## 12991992  2.968432 5.431177  6.514209 5.727164e-06 0.03230693 3.422553
## 12970583  3.064055 2.179415  5.872908 1.965988e-05 0.07807520 2.492008
## 13276616 -2.836278 6.188679 -5.773682 2.391056e-05 0.07807520 2.341427
```

```
head(Tab_p53_CvsOb_OF)
```

```
##          logFC AveExpr          t          P.Value      adj.P.Val          B
## 13166611 -4.784677 6.054477 -9.737223 2.641688e-08 0.0005960705 8.451248
## 12968399 -3.510477 3.849711 -7.768673 6.028627e-07 0.0053333934 5.942821
## 13108641  2.953243 2.879214  7.674138 7.091021e-07 0.0053333934 5.807342
## 12921773  3.292303 4.595212  7.269975 1.438363e-06 0.0054092056 5.211617
## 12937747  3.292303 4.595212  7.269975 1.438363e-06 0.0054092056 5.211617
## 12938855  3.292303 4.595212  7.269975 1.438363e-06 0.0054092056 5.211617
```

```
head(Tab_p53_ObvsOb_OF)
```

```
##          logFC AveExpr          t          P.Value      adj.P.Val          B
## 13166611 -4.136900 6.054477 -8.418942 2.036591e-07 0.004252635 6.994142
## 13270821  3.772431 5.733413  8.046098 3.769398e-07 0.004252635 6.472932
## 12976100 -4.505206 7.319717 -7.486204 9.825667e-07 0.007390212 5.649165
## 13232792 -3.188015 4.923736 -7.023717 2.237120e-06 0.011151862 4.930500
## 13074027  2.747108 4.536478  6.911913 2.741300e-06 0.011151862 4.751502
## 13114159 -3.176573 7.037997 -6.868936 2.965395e-06 0.011151862 4.682148
```

Observem que tots els p-values son inferiors a 0.05 i per tant són significatius, rebutjant les hipòtesis nul·les i per tant demostrant que hi ha diferències d'expressió de gens en les nostres mostres.

4.2.9 Anotació dels resultats:

A continuació afegim informació als gens seleccionats. Definirem una funció que combinarà la informació dels arrays generats amb la informació de les bases de dades per a poder interpretar millor els resultat, els quals es guardaran en arxius .csv diferenciats.

```
annotatedTable <- 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)
}

```

El codi R que executariem està a l'apèndix en forma de text.

La funció no funciona, he fet proves amb altres db de Peix Zebra i el problema no es soluciona, he intentat buscar bibliografia per extreure la informació que necessito d'una altra manera, però no ho he aconseguit. La funció si funcionava per altres estudis (i altres bases de dades). I he arribat a la conclusió de que aquesta base de dades té una estructura diferent.

4.2.10 Comparació entre comparacions:

En el punt actual de l'estudi es interessant comparar quins gens es troben en comú en les diferents comparacions. Utilitzant la funció `decideTest` generem un objecte de R que ens permetrà després comptar les repeticions en diferents comparacions:

A continuació un petit resum del nombre de sobreexpressions o sotaexpressions de les comparacions:

##	CvsOb	CvsOb_OF	ObvsOb_OF	p53_CvsOb	p53_CvsOb_OF	p53_ObvsOb_OF
## Down	2109	2659	2409	330	1532	3006
## NotSig	15370	15172	18527	21516	19588	17531
## Up	5085	4733	1628	718	1444	2027

I finalment els diagrames de Venn, mitjançant la funció `vennDiagram`, obtenim els diferents diagrames per a les mostres p53- (Figure 11), i per a les mostres p53+ (Figure 12). Amb tot això observem que en el cas de les p53+ tenim menys semblances entre els tres casos.

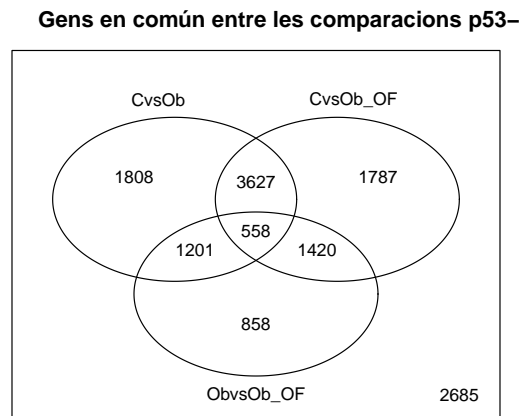


Figure 11: Gens en comú entre les comparacions p53+

4.2.11 Anàlisi de significancia biològica:

Amb el llistat de gens que hauriem de tenir en aquest punt, hauriem d'analitzar-los per veure a que son degudes aquestes diferències. Un biòleg o un gran entes en al materia faria un estudi ben exhaustiu, pero mitjançant

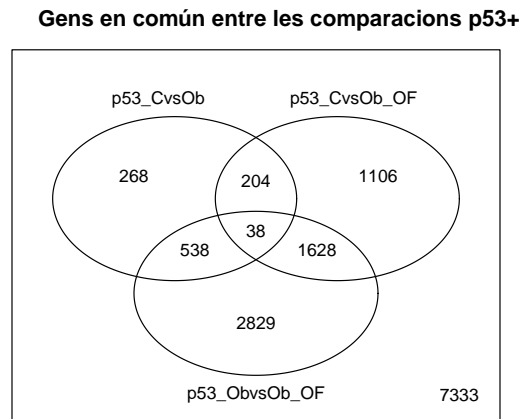


Figure 12: Gens en comú entre les comparacions p53+

paquets com Gene Enrichment Analysis podriem executar un conjunt de funcions que ens permetrien fer una anàlisi previ i prou complet.

Sense els noms dels gens aquest apartat és molt d'executar i no s'obtidrien els resultats desitjats.

5 Resultats:

Fins aquí tenim el re-anàlisi de les dades de l'estudi, no puc comparar els resultats obtinguts amb els de l'estudi original ja que aquest encara no està publicat. Del meu estudi s'extreu que si hi ha diferències entre els grups, però caldria veure la resta d'interessos de l'estudi per saber si es confirmen o es desmenteixen les seves hipòtesis, i amb quins matissos.

6 Discussió:

El meu treball m'ha suposat dificultats, probablement suscidades per la meua elecció de la font de base de dades, per diferents motius:

- Tot i que només hi havia 17 mostres, estaven agrupades en 6 grups diferents, que no vaig sapiguer caure en les dificultats afegides que em suposaria (i una vegada executat gran part del treball no em veia amb ànims de canviar les dades).
- Són dades recents, d'un article que encara no està publicat, i per tant tot i que no he tingut problemes a l'hora d'obtenir les dades, i classificar-les, no he pogut contrastar els meus resultats, entre d'altres.
- Sembla que el organisme que han fet servir per l'estudi no m'ha ajudat amb el tema de les bases de dades (tot i que en part també es culpa meua).

I tot això ha ajudat a que la part final del meu treball hagi quedat una mica incomplet.

7 Apèndix:

Aquest es el codi de R que em dona l'error en la secció del anotament:

```
library(zebrafish.db) anotatedTab_CvsOb<-annotatedTable(Tab_CvsOb,anotationsPackage="zebrafish.db")
write.csv(anotatedTab_CvsOb, file="/resultats/anotatedTab_CvsOb")

anotatedTab_CvsOb_OF<-annotatedTable(Tab_CvsOb_OF,anotationsPackage="zebrafish.db")
write.csv(anotatedTab_CvsOb_OF, file="/resultats/anotatedTab_CvsOb_OF")

anotatedTab_ObvsOb_OF<-annotatedTable(Tab_ObvsOb_OF,anotationsPackage="zebrafish.db")
write.csv(anotatedTab_ObvsOb_OF, file="/resultats/anotatedTab_ObvsOb_OF")

anotatedTab_p53_CvsOb<-annotatedTable(Tab_p53_CvsOb,anotationsPackage="zebrafish.db")
write.csv(anotatedTab_p53_CvsOb, file="/resultats/anotatedTab_p53_CvsOb")

anotatedTab_p53_CvsOb_OF<-annotatedTable(Tab_p53_CvsOb_OF,anotationsPackage="zebrafish.db")
write.csv(anotatedTab_p53_CvsOb_OF, file="/resultats/anotatedTab_p53_CvsOb_OF")

anotatedTab_p53_ObvsOb_OF<-annotatedTable(Tab_p53_ObvsOb_OF,anotationsPackage="zebrafish.db")
write.csv(anotatedTab_p53_ObvsOb_OF, file="/resultats/anotatedTab_p53_ObvsOb_OF")
```