

Anàlisi de Dades Òmiques PAC1

Oriol Nualart Mundó

25/4/2020

Enllaç al repositori de GitHub associat a la PAC: <https://github.com/onualart/liver.git>

Títol de l'estudi: Expression data from liver of *Cyclocarya paliurus* triterpenic acids treated high-fat-diet-induced mice (Wang Y, Yin Z 2019) (GSE125045)

CONTINGUTS

- 1. Abstract
 - 2. Objectius
 - 3. Materials i mètodes
 - 3.1. Disseny de l'experiment
 - 3.2. Procediment d'anàlisi de les dades
 - * 3.2.1. Preparació de les dades
 - * 3.2.2. Control de qualitat de les dades crues
 - * 3.2.3. Normalització
 - * 3.2.4. Control de qualitat de les dades normalitzades
 - * 3.2.5. Filtratge no específic
 - * 3.2.6. Identificació de gens diferencialment expressats
 - * 3.2.7. Anotació dels resultats
 - * 3.2.8. Comparació entre comparacions
 - * 3.2.9. Anàlisi de significació biològica
 - 4. Resultats
 - 5. Discussió
 - Referències
-

1. Abstract

Es sap que els àcids triterpenics de la planta *Cyclocarya paliurus* (ATCP) tenen efectes beneficiosos en pacients amb esteatosi hepàtica no alcohòlica (EHNA), paliant l'estrès oxidatiu i la disfunció mitocondrial associats a aquesta malaltia. Aquest estudi utilitza microarrays per estudiar, en teixit hepàtic d'un model de ratolí, l'expressió gènica subjacent als símptomes de l'EHNA i al tractament amb ATCP.

2. Objectius

Tot i que la informació que podem obtenir a partir de la pàgina de GEO no especifica els objectius concrets (l'estudi no està publicat), sabem que busca fer un perfil d'expressió. Podem marcar uns objectius lògics.

- Identificar gens que s'expressen diferencialment entre els grups de ratolins estudiats.
 - Per ampliar el coneixement general dels processos subjacents als símptomes de la malaltia, al comparar grups de ratolins no tractats.
 - Per estudiar l'efecte del tractament amb ATCP, al comparar ratolins sense tractament amb ratolins tractats.
 - A partir d'aquests gens expressats diferencialment, identificar processos i *pathways* alterats en aquestes comparacions, i veure quina relació poden tenir, ja sigui directa o indirectament, amb els coneixements previs que tenim de la malaltia i del tractament.
-

3. Materials i mètodes

3.1. Disseny de l'experiment

D'acord amb la informació disponible, 15 ratolins van ser repartits aleatoriament en tres grups (5 per grup) corresponents als tres nivells d'un factor:

- Grup de control normal (grup NC): va rebre una dieta estàndard.
- Grup de dieta alta en greixos (grup HFD): va rebre una dieta alta en greixos (sals biliars sòdiques 0.5%, colesterol 2%, rovell d'ou 10%, llard 10%, dieta estàndard 77.5%; aproximadament 4016 kcal/kg, 30% de calories de greix).
- Grup tractat amb ATCP (grup CPT): va ser administrat amb ATCP (160 mg/kg) diàriament durant 8 setmanes després de rebre la dieta alta en greixos durant 12 setmanes.

Els ratolins van ser sacrificats i se'n van extreure mostres de fetge per ser analitzades.

Els microarrays utilitzats per fer l'anàlisi són del model Affymetrix Mouse Gene 2.0 ST Array (GPL16570). Es tracta d'arrays d'oligonucleòtids d'un color, amb sondes corresponents a fragments d'una selecció de gens de ratolí.

3.2. Procediment d'anàlisi de les dades

El procediment detallat amb tot el codi de R utilitzat, juntament amb l'*output* dels fragments de codi, es pot trobar a l'arxiu "pipeline.pdf" al repositori de GitHub, generat a partir de l'arxiu de markdown "pipeline.Rmd". El codi pressuposa tenir instal·lats tots els paquets de R necessaris, inclosos els paquets d'anotacions *mogene20sttranscriptcluster.db* i *org.Mm.eg.db*.

El codi utilitzat està basat principalment en el que apareix al *Case Study 1* dels materials de l'assignatura (https://github.com/ASPteaching/Omics_Data_Analysis-Case_Study_1-Microarrays).

3.2.1. Preparació de les dades

Abans que res, el primer que hem fet és descarregar els arxius .CEL de la pàgina de GEO i preparar l'arxiu *targets*. Aquests arxius els hem posat al subdirectori *data* del repositori.

A l'arxiu *targets* hem organitzat les dades, d'acord amb el disseny experimental descrit, en un factor (*Diet*) amb tres nivells (NC, HFD i CPT). Les 15 mostres les hem assignat als diferents nivells d'acord amb la informació disponible.

A partir d'aquí hem utilitzat R per fer el tractament de les dades.

Primer hem carregat l'arxiu *targets*, i les dades crues contingudes als arxius .CEL les hem passat a l'objecte *rawData*.

3.2.2. Control de qualitat de les dades crues

Per fer el control de qualitat de les dades crues hem aplicat la funció *arrayQualityMetrics* del paquet amb el mateix nom a les dades de *rawData*. Aquesta funció ens genera un informe molt complet, que inclou diversos gràfics que permeten fer una primera visualització de les dades, i també detecta possibles *outliers* segons diferents criteris.

L'informe complet es pot consultar a l'arxiu “index.html” de la carpeta “qcRaw” del repositori.

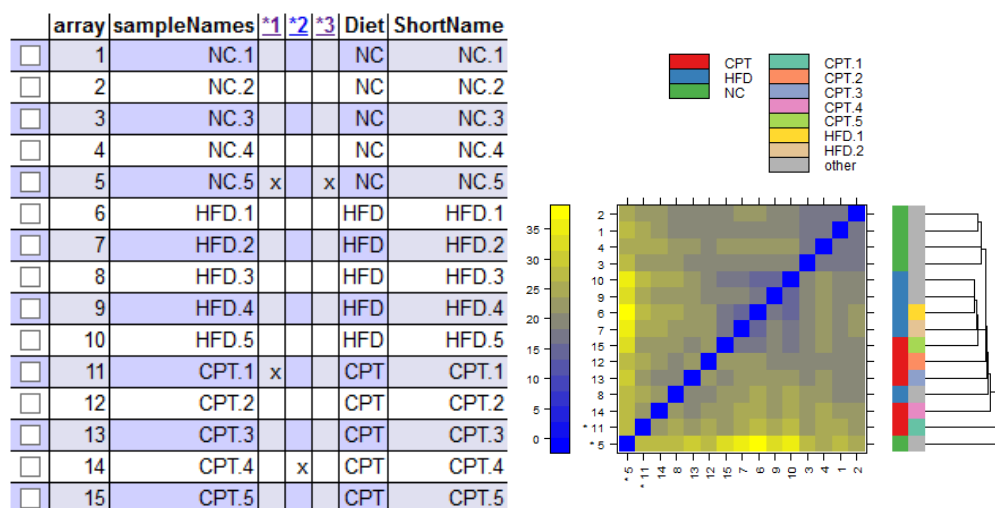


Figure 1: Taula resum del control de qualitat de les dades crues i heatmap corresponent

Al resum del control de qualitat comprovem que es detecten alguns *outliers*, especialment a la mostra NC.5. Al *heatmap*, per exemple, es veu clarament que aquesta mostra (corresponent a la primera columna/última fila) es diferencia força de la resta.

Tenint en compte que es tracta d'una mostra del grup control, en el qual no esperem ni volem que hi hagi grans variacions, optem per considerar-la poc fiable i no utilitzar-la en els anàlisis. Per aquesta raó, creem una versió modificada de la carpeta de dades (“moddata”) sense la mostra NC.5 i amb un arxiu *targets* modificat.

Un cop tretes les dades de la mostra, repetim el control de qualitat. L'informe en aquest cas és a la carpeta “qcMod”.

Observem que ara no hi ha cap mostra amb més d'un marcador d'*outliers*. El *heatmap* és ara més uniforme, només la mostra CPT.1 és detectada com a *outlier*, però no es visualitza una diferència tan destacada.

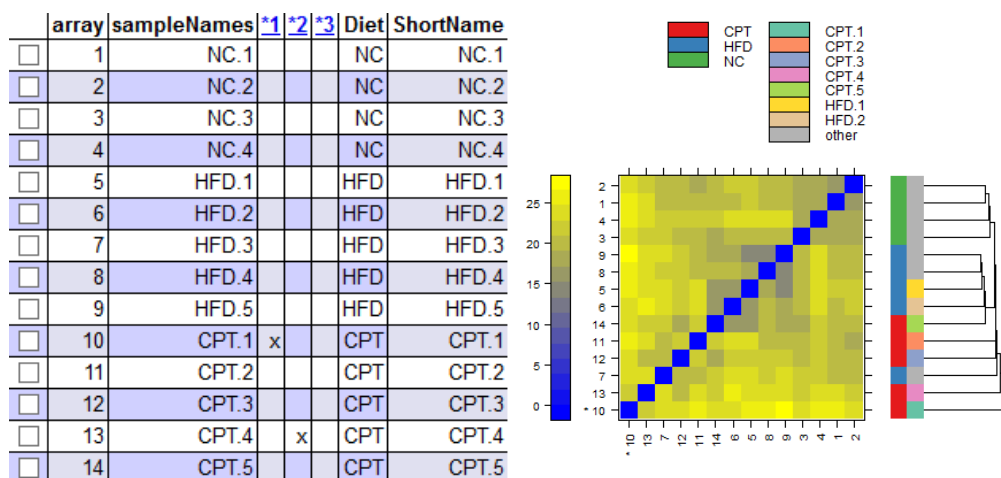


Figure 2: Taula resum del control de qualitat de les dades modificades i heatmap

3.2.3. Normalització

Per poder treballar amb les dades dels microarrays que tenim a *rawData* necessitem abans fer una sèrie d'ajustos. Hem de corregir el soroll de fons, normalitzar les dades i “sumaritzar-les” (convertir els valors de les diferents sondes de cada *probeset* en un únic valor).

Utilitzem el mètode Robust Multichip Analysis (RMA) d'Irizarry et al. (2003) per fer aquests processos. Aquest mètode, fàcil d'aplicar amb R mitjançant la funció *rma* del paquet *oligo*, engloba tots els passos necessaris.

Un cop realitzat el procés, les dades normalitzades queden guardades a *normData*, un *ExpressionSet* força més reduït després de la “sumarització”.

3.2.4. Control de qualitat de les dades normalitzades

Repetim el control de qualitat que hem fet amb la funció *arrayQualityMetrics*, aquesta vegada sobre les dades normalitzades i sumaritzades. L'informe ens apareix a la carpeta “qcNorm”.

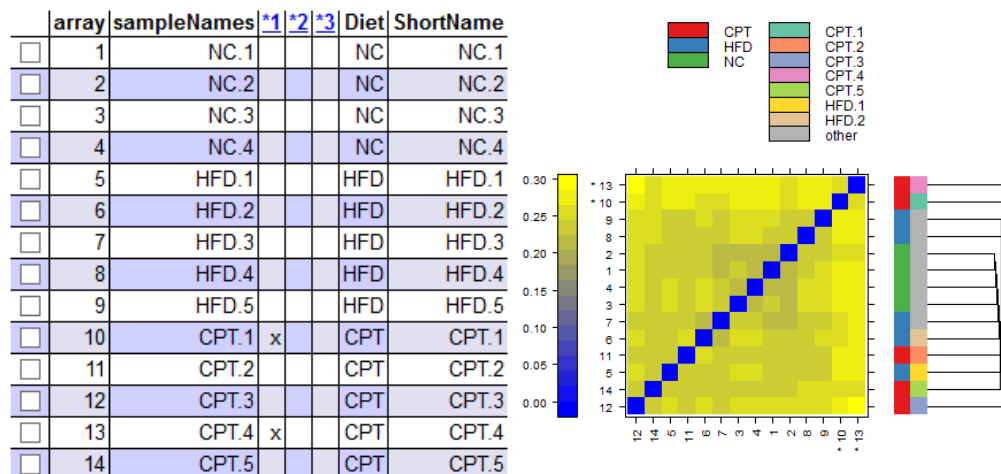


Figure 3: Taula resum del control de qualitat de les dades normalitzades i heatmap

En aquest cas detectem només *outliers* en les distàncies entre arrays. Al no haver-hi cap mostra amb més d'un marcador, donem per bones totes les mostres.

Al *heatmap*, efectivament, observem diferència entre les mostres marcades com a *outliers* (les dues primeres files / últimes columnes) i la resta, però visualment no sembla una diferència especialment gran.

3.2.5. Filtratge no específic

Per filtrar els gens menys variables del nostre *ExpressionSet*, apliquem la funció *nsFilter* del paquet *genefilter*.

En aquest cas el llindar de variabilitat que hem posat és baix, de 0,5, ja que hem optat per ser conservadors amb les dades. Això vol dir que dels gens del nostre conjunt de dades que podríem utilitzar per fer els anàlisis, ens quedem amb la meitat que mostra major variabilitat i descartem l'altra meitat.

El paquet *mogene20sttranscriptcluster.db*, que conté anotacions corresponents als gens de l'array, ens proporciona els identificadors dels gens, cosa que ens permet deixar fora també els gens que no tinguin identificador.

Aquests són els gens eliminats en el procés de filtratge:

```
## $numDupsRemoved
## [1] 671
##
## $numLowVar
## [1] 11982
##
## $numRemoved.ENTREZID
## [1] 16710
```

El resultat del filtratge queda guardat a l'*ExpressionSet* *fltData*. També guardem les dades normalitzades i les dades filtrades en arxius a la carpeta “results” del repositori (“normData.csv” i “fltData.csv”).

3.2.6. Identificació de gens diferencialment expressats

Per identificar els gens diferencialment expressats, primer hem de definir el model i crear les matrius de disseny i de contrastos.

El nostre és un model lineal amb un factor,

$$Y = X\alpha + \epsilon$$

on X és la matriu de disseny.

Els contrastos vénen definits com

$$\beta = C\alpha$$

on C és la matriu de contrastos.

La funció *model.matrix* del paquet *limma* (Smyth, G. K. 2005) ens genera fàcilment la matriu de disseny a partir de les nostres dades filtrades.

```
##          CPT HFD NC
## NC.1      0   0  1
## NC.2      0   0  1
## NC.3      0   0  1
## NC.4      0   0  1
## HFD.1     0   1  0
## HFD.2     0   1  0
## HFD.3     0   1  0
## HFD.4     0   1  0
## HFD.5     0   1  0
## CPT.1     1   0  0
## CPT.2     1   0  0
## CPT.3     1   0  0
## CPT.4     1   0  0
## CPT.5     1   0  0
## attr("assign")
## [1] 1 1 1
## attr("contrasts")
## attr("contrasts")$Diet
## [1] "contr.treatment"
```

De cara a decidir quins contrastos farem per analitzar les dades, ens fixem en l'interès que puguin tenir de cara a obtenir coneixements rellevants.

Partint dels tres grups de ratolins estudiats, hi ha tres comparacions possibles: les resultants de combinar per parelles els tres grups.

- D'entrada, la comparació entre NC (dieta estàndard) i HFD (dieta grassa) ens donarà informació sobre els processos subjacents als símptomes estudiats.
- La comparació entre el grup HFD (dieta grassa) i el grup CPT (dieta grassa i tractament) ens dóna informació sobre el tractament.
- El contrast entre NC (dieta estàndard) i CPT (dieta grassa i tractament) per sí sol no ens dóna una informació útil, ja que en el fons està estudiant dos “factors” a la vegada, la dieta i el tractament (encara que al dissenyar l'estudi els haguem considerat com a parts d'un únic factor). Tanmateix, pot ser interessant comparar dos contrastos: el contrast entre NC i HFD amb el contrast entre NC i CPT. La diferència entre aquests dos contrastos també ens donarà informació rellevant sobre l'efecte del tractament.

Per tant, inclourem a l'anàlisi els tres contrastos possibles.

Amb els contrastos decidits, generem la matriu amb la funció *makeContrasts* del paquet *limma*.

```
##          Contrasts
## Levels HFDvsCPT NCvsHFD NCvsCPT
##      CPT          1          0          1
##      HFD         -1          1          0
##      NC           0         -1         -1
```

Estimem el model amb la funció *lmFit* del paquet *limma*: en cada grup definit a la matriu de disseny, calculem els paràmetres del model per a cada gen, a partir dels valors d'expressió de les diferents mostres del grup. La funció *lmfit* utilitza per defecte el mètode de mínims quadrats per fer aquest càlcul.

Amb els paràmetres del model calculats i amb la matriu de contrastos, estimem els diferents contrastos amb la funció *contrasts.fit*.

Utilitzant la funció *eBayes*, calculem per a cada gen i comparació una sèrie d'estadístics que ens ajudaran a decidir quins gens podem considerar que estan diferencialment expressats i quins no. El mètode combina una

estimació de la variabilitat basada en tota la matriu amb estimacions individuals per cada valor (Smyth, G. K. 2004).

Obtenim per a cada comparació un llistat de gens associat a una sèrie de valors, entre ells el *fold-change* i el p-valor per a aquell gen i comparació. Per controlar el número de falsos positius deguts a l'elevat número de tests, aquesta funció també calcula p-valors ajustats mitjançant el mètode de Benjamini i Hochberg (1995).

Els resultats queden guardats a l'objecte *fit.main*, i també els guardem a l'arxiu "liver.fit.main.Rda" a la carpeta de resultats.

Per fer una primera visualització d'algunes de les dades obtingudes, explorem amb la funció *topTable* els gens amb un p-valor més petit, dintre de les diferents comparacions. La funció, aplicada a *fit.main*, genera un *data frame* amb les dades de la comparació en qüestió ordenades per p-valor. Com a resultat obtenim els *data frames* *top_HFDvsCPT*, *top_NCvsHFD* i *top_NCvsCP*.

HFDvsCPT:

##	logFC	AveExpr	t	P.Value	adj.P.Val	B
## 17335540	0.9994833	5.082067	7.319599	1.569727e-06	0.01880847	4.793782
## 17364642	0.8786221	7.263510	6.269726	1.041219e-05	0.04481886	3.285387
## 17411955	-1.5712617	2.671452	-6.229835	1.122155e-05	0.04481886	3.224270
## 17246967	1.2477222	7.255852	5.909740	2.062156e-05	0.05603154	2.723798
## 17260644	-0.8431216	5.115747	-5.732672	2.904490e-05	0.05603154	2.439352
## 17481725	0.8290971	4.589389	5.690607	3.152622e-05	0.05603154	2.370993

NCvsHFD

##	logFC	AveExpr	t	P.Value	adj.P.Val	B
## 17468759	-0.972245	2.589143	-9.402244	5.620902e-08	0.0006217612	8.051459
## 17254395	-1.409242	5.008661	-8.992759	1.037825e-07	0.0006217612	7.550835
## 17269521	-1.835453	6.236747	-7.838127	6.512347e-07	0.0015196324	6.007204
## 17273348	-2.156673	6.020434	-7.831958	6.579512e-07	0.0015196324	5.998408
## 17362717	-1.577277	7.702808	-7.811196	6.810921e-07	0.0015196324	5.968761
## 17342642	-1.996605	6.308200	-7.661105	8.759012e-07	0.0015196324	5.752406

NCvsCPT:

##	logFC	AveExpr	t	P.Value	adj.P.Val	B
## 17362717	-2.128860	7.702808	-10.542817	1.120887e-08	0.0001132889	9.530148
## 17468759	-1.050836	2.589143	-10.162270	1.890984e-08	0.0001132889	9.107213
## 17254395	-1.525307	5.008661	-9.733403	3.470289e-08	0.0001386034	8.608845
## 17435934	-1.150220	9.633207	-9.181977	7.798893e-08	0.0001979705	7.932493
## 17482897	-1.932815	4.681236	-9.143625	8.261161e-08	0.0001979705	7.883909
## 17399266	-2.061990	7.776123	-9.018480	9.980473e-08	0.0001993100	7.723951

A destacar d'aquests primers resultats, els valors força elevats dels p-valors ajustats en el contrast HFDvsCPT, cosa que indicaria que els dos grups es diferencien poc.

3.2.7. Anotació dels resultats

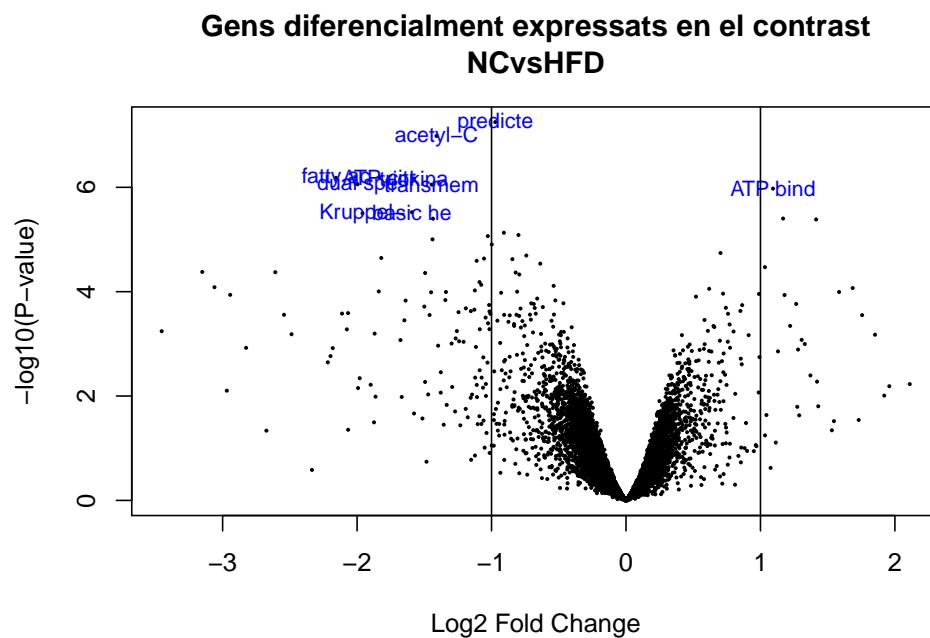
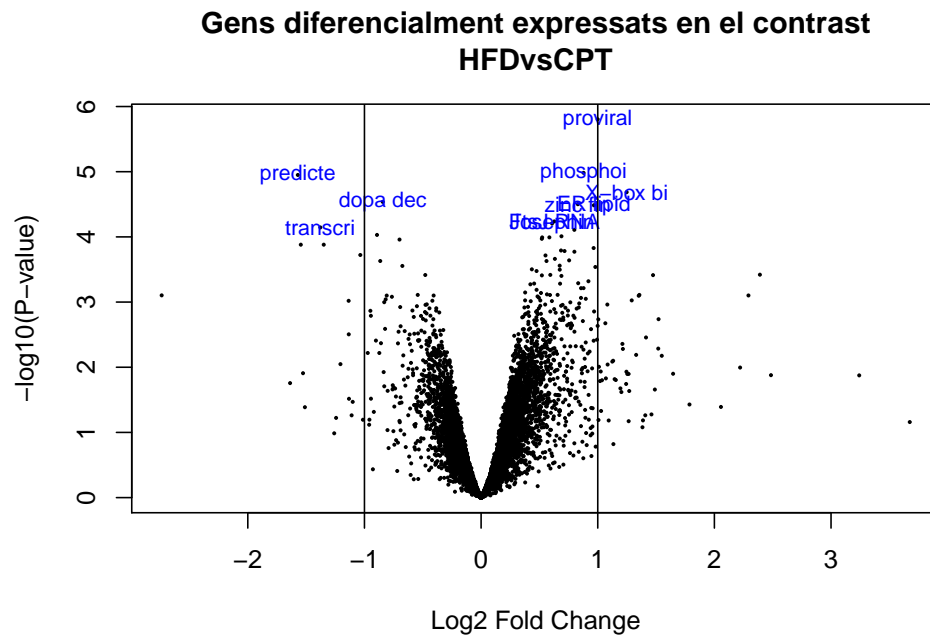
Basant-nos en la funció *annotatedTopTable* que apareix al *Case_Study_1* dels materials de l'assignatura, generem les anotacions per als tres conjunts de gens diferencialment expressats que hem obtingut.

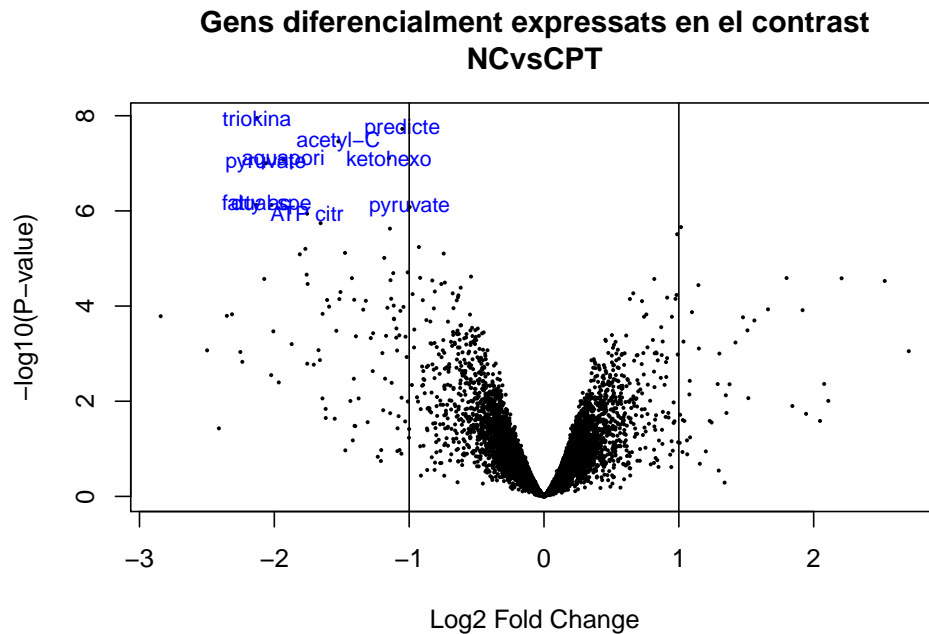
El que fa la funció és associar els gens que tenim en els *data frames* obtinguts en els últims passos amb les anotacions disponibles per a aquests gens, que treiem del paquet d'anotacions corresponent al microarray utilitzat (*mogene20sttranscriptcluster.db*). Concretament, agafa els paràmetres "SYMBOL", "ENTREZID" i "GENENAME", que ens seran útils en els següents passos del nostre procediment d'anàlisi.

Els *data frames* obtinguts, que anomenem *annot_HFDvsCPT*, *annot_NCvsHFD* i *annot_NCvsCPT*, els guardem també en arxius .csv a la carpeta de resultats.

Per tenir una nova visualització de les dades un cop anotades, generem volcano plots per les tres comparacions, incloent els noms dels gens amb un p-valor més baix.

El volcano plot ens dóna una idea de com es distribueixen els gens pel que fa al *fold-change*, que ens indica les diferències d'expressió, i pel que fa al p-valor, que ens indica el grau de significació estadística d'aquestes diferències. Els punts que queden més amunt i més allunyats del centre en l'eix d'abcises són els que més ens interessin.





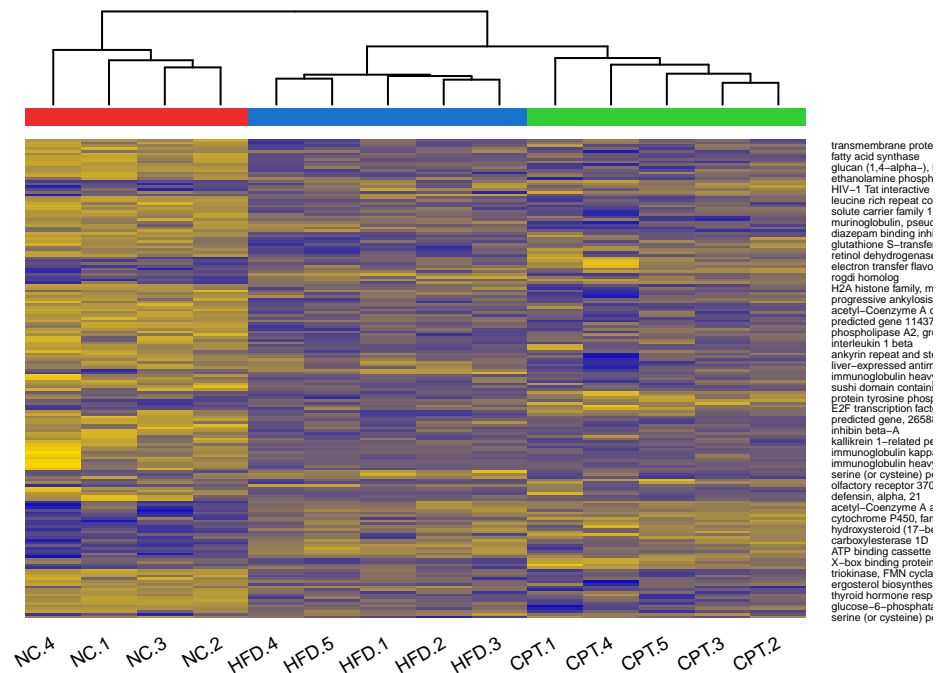
Els volcano plots ens confirmen la impressió que el contrast HFDvsCPT mostra poca diferència entre els dos grups, havent-hi pocs gens amb un p-valor baix i un *fold-change* destacable.

També crida l'atenció el fet que la majoria de gens que apareixen destacats en els contrastos entre el grup de dieta estàndard (NC) i un grup amb dieta alta en greixos (HFD i CPT), ho fan amb un *fold-change* negatiu, és a dir que la majoria d'aquests gens quedarien inhibits per l'excés de greixos en la dieta.

Passem a seleccionar els gens diferencialment expressats en alguna de les comparacions. La funció *decideTests* ho fa, a partir de les estimacions dels nostres contrastos que tenim a *fit.main*, agafant els gens que tenen un p-valor per sota d'un determinat número i un valor absolut del Log2 del *fold-change* per sobre d'un altre número. En el nostre cas hem utilitzat un límit de 0,1 per al p-valor i de 0.75 per al Log2 del *fold-change*.

Podem fer una ullada general al nivell d'expressió dels gens seleccionats amb un *heatmap*.

Gens diferencialment expressats FDR < 0,1, logFC >=0.75



Hem indicat que les mostres s'ajuntin per semblança entre files i entre columnes.

Observem que els tres grups experimentals s'han separat correctament, i és especialment visible la diferència entre el nivell d'expressió dels gens del grup NC i el dels altres dos grups.

3.2.8. Comparació entre comparacions

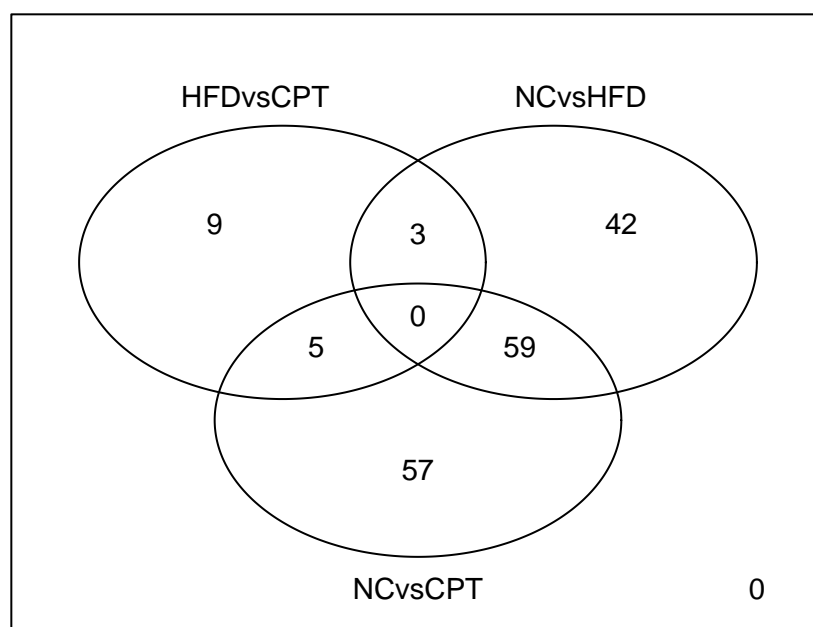
Mirem quants gens diferencialment expressats hem obtingut amb cada contrast segons la selecció que hem fet.

##	HFDvsCPT	NCvsHFD	NCvsCPT
## Down	8	81	86
## NotSig	11965	11878	11861
## Up	9	23	35

Com ja ens anunciaven les visualitzacions preliminars, la comparació entre els grups HFD i CPT és clarament la que detecta menys gens diferencialment expressats (i el grup NC és per tant el que més es diferencia dels altres dos).

Fem un diagrama de Venn per veure quants gens coincideixen a expressar-se diferencialment entre les comparacions.

Gens diferencialment expressats segons cada comparació.



Veiem que els dos contrastos amb major número de gens diferencialment expressats, NCvsHFD i NCvsCPT, comparteixen un número important de gens. Això té sentit si ens fixem que els dos contrastos comparen el grup de dieta estàndard amb un grup de dieta alta en greixos (amb o sense tractament), i que aquesta diferència en la dieta sembla ser, pel que hem vist fins ara, la principal font de contrast entre els grups.

3.2.9. Anàlisi de significació biològica

Per fer l'anàlisi de significació biològica, primer hem de generar el llistat d'identificadors *Entrez* dels gens.

A partir dels *data frames* *top_HFDvsCPT*, *top_NCvsHFD* i *top_NCvsCP*, agafem de cada comparació els gens amb valors més significatius d'expressió diferencial. Amb la funció *select* fem un llistat amb els seus respectius identificadors *Entrez*, trets del paquet d'anotacions *mogene20sttranscriptcluster.db*, corresponent al model de microarray utilitzat.

Per tal d'obtenir uns llistats extensos, hem optat per fer servir un llindar de p-valor ajustat poc restrictiu, de 0,25.

Podem fer-nos una idea de la mida d'aquests llistats mirant quants gens inclouen.

```
## HFDvsCPT  NCvsHFD  NCvsCPT
##      506      456      647
```

Després generem un nou llistat amb tots els gens de ratolí coneguts que tenen almenys una anotació de Gene Ontology o de KEGG. La funció *mappedkeys* del paquet *AnnotationDbi* s'encarrega de seleccionar-los a partir del paquet d'anotacions del genoma complet de ratolí: *org.Mm.eg.db*.

Finalment, a partir de les llistes que hem generat, apliquem l'anàlisi de significació biològica mitjançant les funcions *enrichGO* i *enrichKEGG* del paquet *clusterProfiler*. Aquestes anàlisis consisteixen a comparar les

anotacions dels gens seleccionats mitjançant tests de sobrerrepresentació, per buscar-hi termes GO i *pathways* en comú.

Els resultats de les anàlisis són, per a cada contrast, una sèrie de termes GO i de *pathways* amb estadístiques associades. Al fer l'anàlisi guardem aquestes dades en arxius .csv i generem alguns gràfics que també guardem en arxius.

4. Resultats

Com a resultat de l'anàlisi hem obtingut una sèrie de dades per a cada comparació entre grups:

- Un llistat de gens diferencialment expressats, associats a diversos estadístics que ens indiquen el seu diferencial d'expressió i el nivell de significació estadística d'aquest diferencial. Aquests llistats han quedat guardats als arxius “annot_XXXXXXXX.csv”.
- Un llistat de termes GO relacionats amb aquests gens juntament amb alguns estadístics associats. Aquests llistats estan guardats als arxius “enrichGO.Results.XXXXXXXXX.csv”.
- Un llistat de *pathways* relacionats amb aquests gens juntament amb alguns estadístics associats. Aquests llistats estan guardats als arxius “enrichKEGG.Results.XXXXXXXXX.csv”.

Per a cada llistat de termes GO i de *pathways* també hem generat a mode de resum dos gràfics en arxius separats: un gràfic de barres mostrant el número de gens associats als principals termes, i un gràfic de ret d'aquests termes.

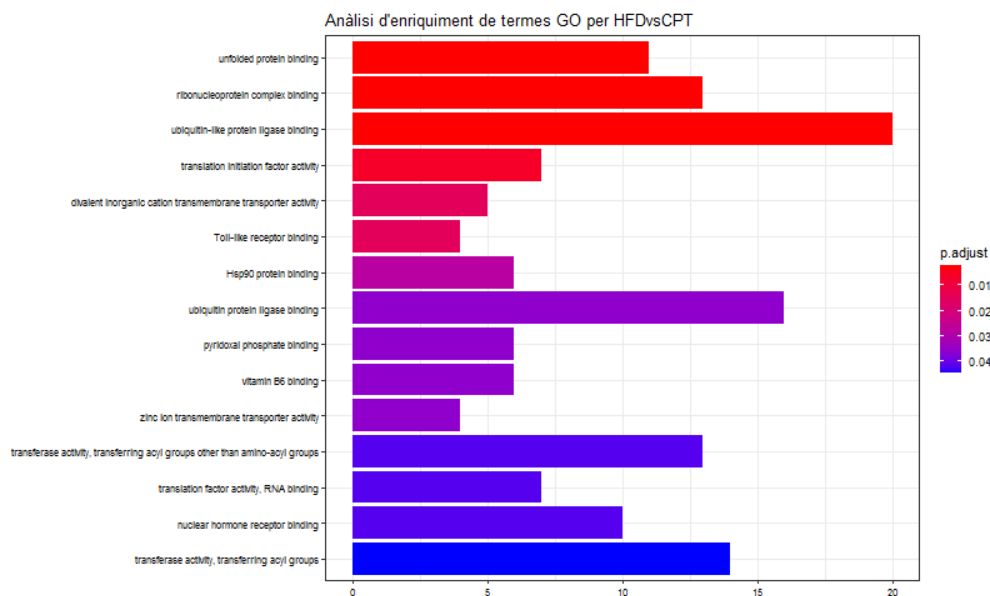


Figure 4: Gràfic de barres de termes GO per HFD vs CPT

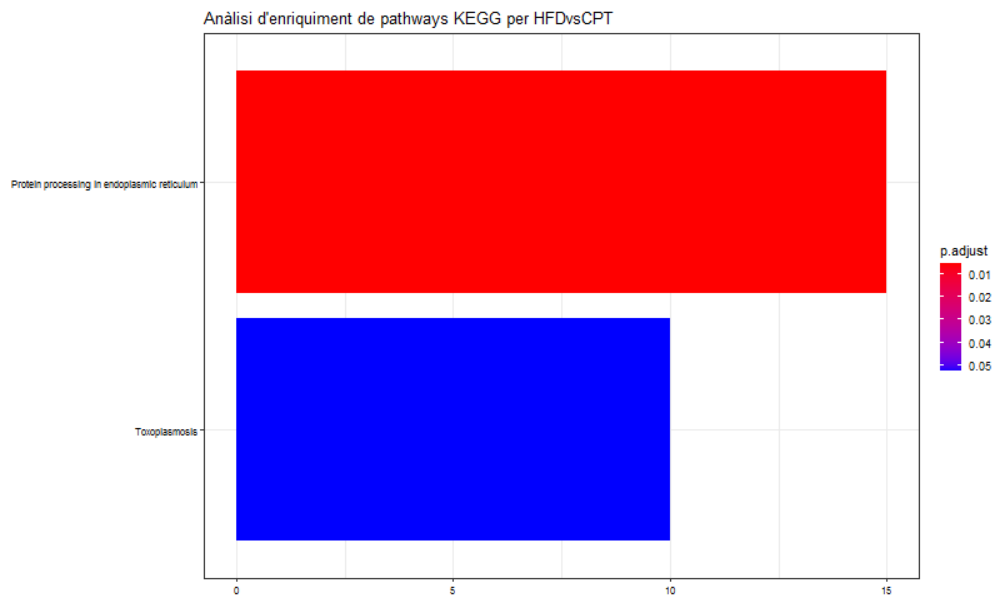


Figure 5: Gràfic de barres de pathways per HFD vs CPT

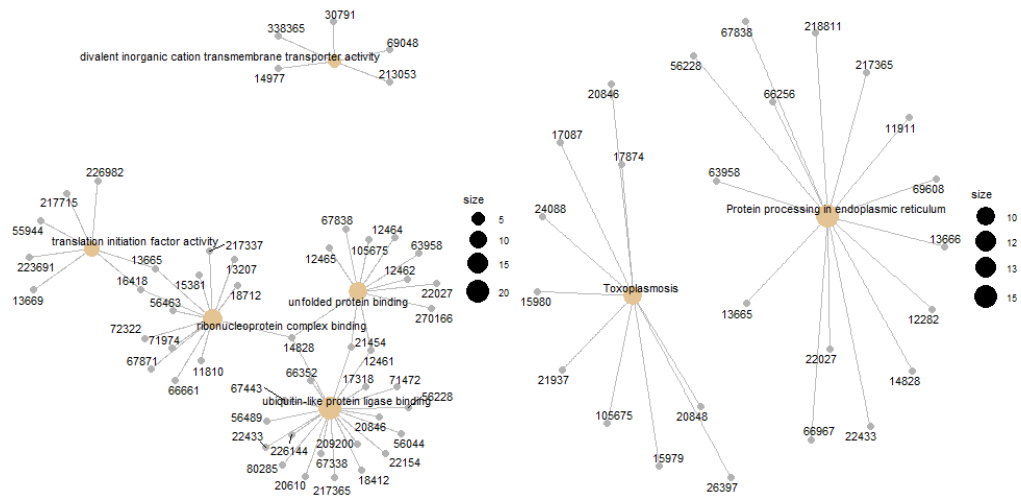


Figure 6: Gràfics de ret de termes GO i de pathways per HFD vs CPT

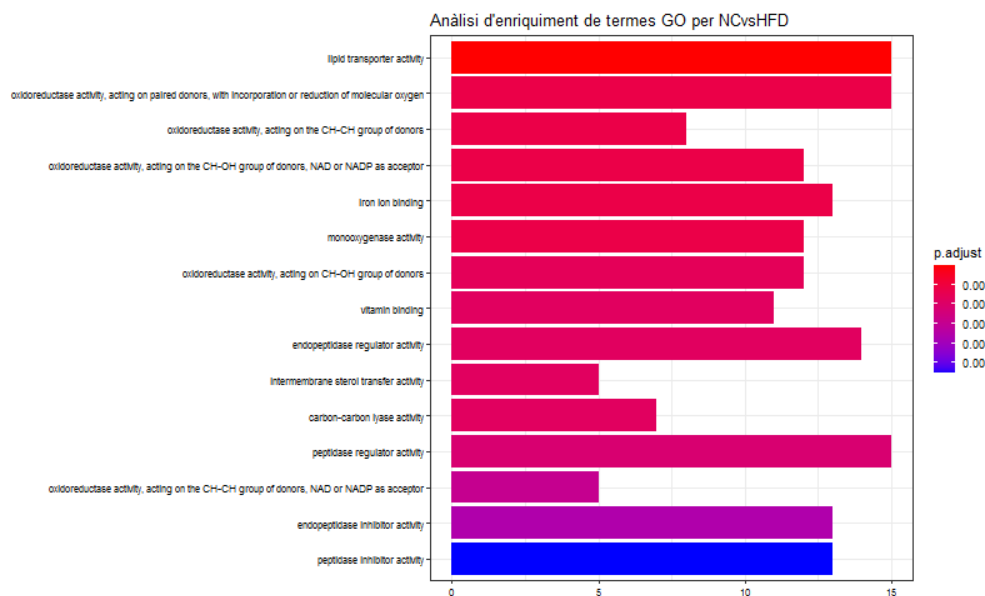


Figure 7: Gràfic de barres de termes GO per NC vs HFD

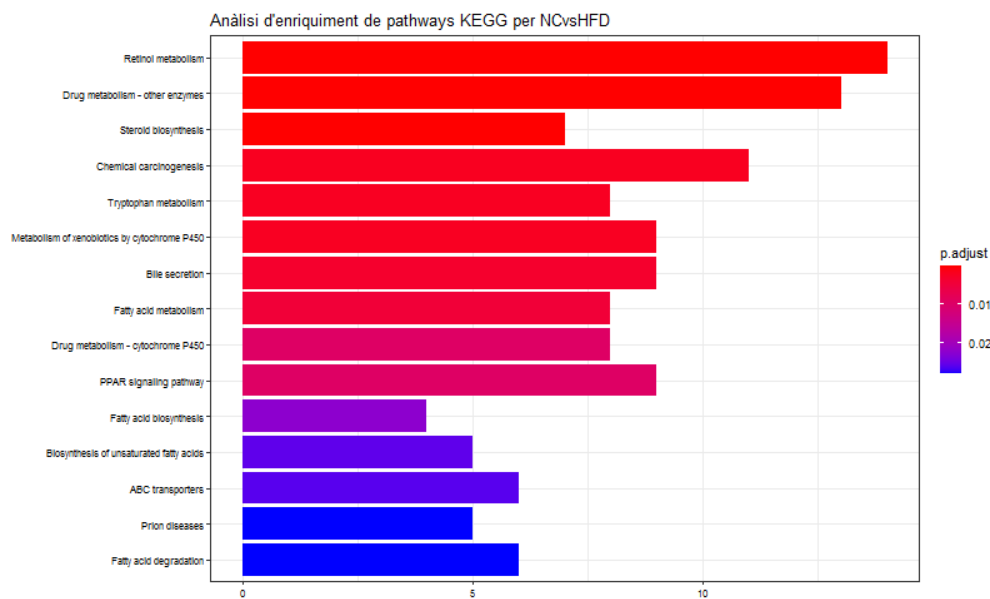


Figure 8: Gràfic de barres de pathways per NC vs HFD

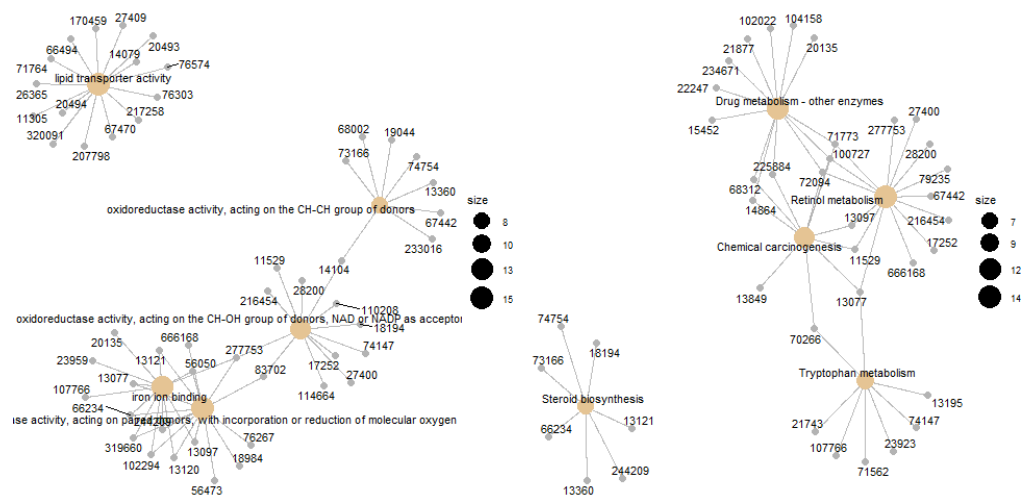


Figure 9: Gràfics de ret de termes GO i de pathways per NC vs HFD

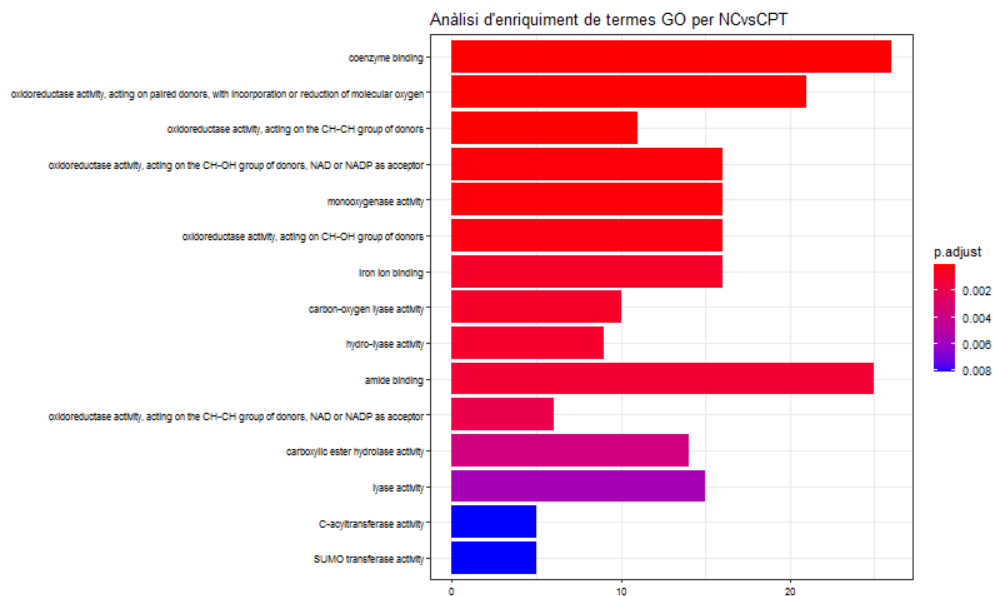


Figure 10: Gràfic de barres de termes GO per NC vs CPT

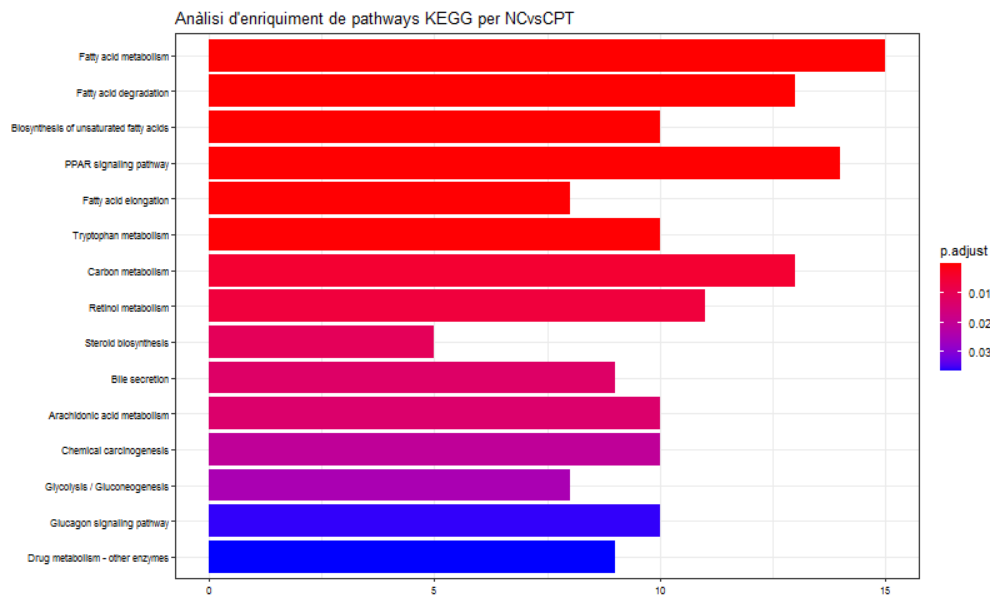


Figure 11: Gràfic de barres de pathways per NC vs CPT

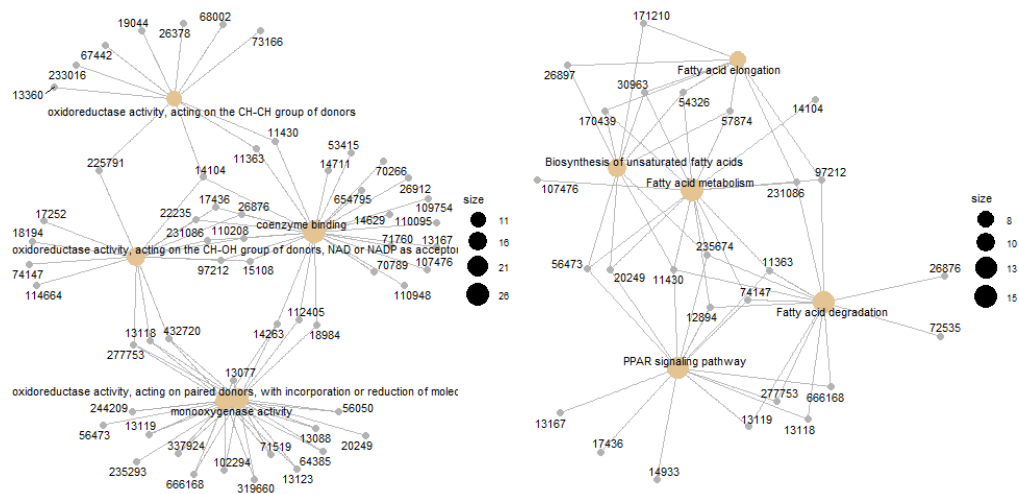


Figure 12: Gràfics de ret de termes GO i de pathways per NC vs CPT

5. Discussió

Un aspecte que sens dubte s'ha tingut en compte al dissenyar l'estudi, tot i que no és específic d'un estudi de *microarrays*, són les limitacions del model d'organisme utilitzat. Aquest simula la malaltia d'interès, l'esteatosi hepàtica no alcohòlica, a partir de diferències de dieta entre grups de ratolins, i sembla obvi que tant per l'organisme estudiat com per la manera de provocar els símptomes ha de tenir limitacions.

Pel que fa a la interpretació dels resultats obtinguts, sense pretendre extreure'n conclusions concretes, sí que podem apuntar alguns comentaris.

Els termes GO i *pathways* detectats amb la comparació NCvsHFD semblen tenir bastant en comú amb el tipus de processos que esperem, ja que inclouen termes relacionats amb el metabolisme dels greixos i amb processos oxidatius. Té sentit que sigui així ja que els ratolins comparats en aquest contrast (ratolins control contra ratolins amb dieta grassa) esperem que es diferenciïn precisament en la presència de símptomes com l'acumulació de greixos o l'estrès oxidatiu.

Tanmateix, com hem vist en els volcano plots, la majoria dels gens diferencialment expressats en aquest contrast ho són per inhibició (amb l'excés de greixos a la dieta, el gen reduiria la seva expressió), i això d'entrada resulta poc intuïtiu. Caldria veure quina és exactament la relació que poden tenir aquests gens amb les diferències de dieta.

El resultat és semblant en el cas del contrast NCvsCPT. També té sentit que sigui així perquè de nou estem comparant ratolins control amb ratolins amb dieta grassa, només que aquests últims a més reben tractament.

Seria interessant fer una comparació detallada entre els resultats dels dos contrastos i veure quina relació tenen les diferències entre ells amb el fet que el grup CPT rebi tractament. Per exemple, en el contrast que inclou el grup de tractament observem un increment en el nombre de gens relacionats amb processos oxidatius, cosa que podria indicar que l'ATCP efectivament tindria un efecte regulador d'aquests processos.

Pel que fa HFDvsCPT, observem com a termes destacats uns de diferents de la resta de contrastos, i a més aquests termes tenen un nivell més baix de significació estadística. Aquesta discrepància probablement s'explica perquè en aquelles comparacions la principal font de contrast eren les diferències de dieta (estàndard contra alta en greixos). Al comparar dos grups amb dieta alta en greixos, aquesta font de contrast desapareix, quedant només l'efecte del tractament, comparativament més reduït.

Referències

Benjamini, Yoav, and Yoel Hochberg. 1995. *Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing*. <http://www.jstor.org/stable/2346101><http://www.jstor.org/><http://www.jstor.org/action/showPublisher?publisherCode=black>.

Irizarry, Rafael A., Bridget Hobbs, Francois Collin, Yasmin D. Beazer-Barclay, Kristen J. Antonellis, Uwe Scherf, and Terence P. Speed. 2003. *Exploration, Normalization, and Summaries of High Density Oligonucleotide Array Probe Level Data*. <https://doi.org/10.1093/biostatistics/4.2.249>.

Smyth, G. K. 2004. *Linear Models and Empirical Bayes Methods for Assessing Differential Expression in Microarray Experiments*. <https://doi.org/10.2202/1544-6115.1027>.

———. 2005. *Limma: Linear Models for Microarray Data*. <https://doi.org/10.1007>.

Wang Y, Yin Z. 2019. *Expression Data from Liver of Cyclocarya paliurus Triterpenic Acids Treated High-Fat-Diet-Induced Mice*. Nanjing, China: Department of Natural Medicinal Chemistry & State Key Laboratory of Natural Medicines, China Pharmaceutical University. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE125045>.