

PREGUNTA 1:

Existen varias debilidades del concepto de significación estadística en general, entre ellas:

La significación estadística solo indica la probabilidad de obtener los resultados observados si la hipótesis nula es verdadera. Sin embargo, esto no necesariamente significa que los resultados sean importantes o relevantes en la práctica.

La significación estadística puede ser influenciada por el tamaño de la muestra, lo que significa que un resultado puede ser significativo estadísticamente pero no tener importancia práctica en una muestra pequeña.

La significación estadística no tiene en cuenta la complejidad del modelo y puede ser engañosa en modelos complejos con múltiples variables.

La elección de la hipótesis nula puede influir en la significación estadística y puede ser subjetiva.

La significación estadística no tiene en cuenta la reproducibilidad de los resultados y puede ser influenciada por factores como la selección de variables y la elección del modelo.

En general, la significación estadística puede ser útil para evaluar la probabilidad de obtener los resultados observados si la hipótesis nula es verdadera, pero debe ser interpretada con precaución y en conjunto con otras medidas de importancia práctica y relevancia científica.

Los p-valores son una medida de la probabilidad de obtener los resultados observados si la hipótesis nula es verdadera. Sin embargo, los p-valores no indican la importancia práctica de los resultados y no necesariamente significan que los resultados sean importantes o relevantes en la práctica.

Por ejemplo, un resultado puede ser estadísticamente significativo con un p-valor muy bajo, pero puede tener una importancia práctica muy limitada. Por otro lado, un resultado puede no ser estadísticamente significativo con un p-valor alto, pero puede tener una gran importancia práctica.

Además, los p-valores no tienen en cuenta el tamaño del efecto, que es una medida de la magnitud de la relación entre las variables. Un resultado puede ser estadísticamente significativo con un p-valor bajo, pero tener un tamaño de efecto muy pequeño, lo que significa que la relación entre las variables es débil o insignificante en la práctica.

Por lo tanto, es importante tener en cuenta la importancia práctica de los resultados al interpretar los p-valores y utilizar otras medidas de importancia práctica y relevancia científica, como el tamaño del efecto y la validez externa

del modelo. Además, es importante considerar el contexto y la aplicación práctica de los resultados al tomar decisiones basadas en los resultados de un estudio.

REFERENCIAS:

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA's statement on p-values: context, process, and purpose. *The American Statistician*, 70(2), 129-133. <https://doi.org/10.1080/00031305.2016.1154108>

McShane, B. B., Gal, D., Gelman, A., Robert, C., & Tackett, J. L. (2019). Abandon statistical significance. *arXiv preprint arXiv:1902.06909*. <https://arxiv.org/abs/1902.06909>

Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., & Altman, D. G. (2016). Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *European Journal of Epidemiology*, 31(4), 337-350. <https://doi.org/10.1007/s10654-016-0149-3>

Nuzzo, R. (2014). Scientific method: statistical errors. *Nature*, 506(7487), 150-152. <https://doi.org/10.1038/506150a>

Gelman, A., & Loken, E. (2014). The statistical crisis in science. *American Scientist*, 102(6), 460-465. <https://doi.org/10.1511/2014.111.460>

Pregunta 1.1

La corrección de Bonferroni es una alternativa comúnmente utilizada a los p-valores para corregir los resultados en estudios de ómicas y reducir la probabilidad de encontrar resultados significativos por casualidad. La corrección de Bonferroni ajusta los p-valores para tener en cuenta el número de comparaciones realizadas y establece un umbral de significación más estricto para reducir la probabilidad de errores tipo I.

La corrección de Bonferroni es una técnica simple y fácil de aplicar que puede ser útil en estudios de ómicas con un número limitado de comparaciones. Sin embargo, en estudios de ómicas con un gran número de comparaciones, la corrección de Bonferroni puede ser demasiado conservadora y reducir la potencia del estudio.

Además, la corrección de Bonferroni no tiene en cuenta la correlación entre las variables, lo que puede ser un problema en estudios de ómicas donde las variables están altamente correlacionadas. En estos casos, la corrección de Bonferroni puede ser demasiado restrictiva y reducir la capacidad del estudio para detectar relaciones significativas entre las variables.

Por lo tanto, es importante considerar otras alternativas, como la corrección por métodos menos conservadores, como la corrección de Holm-Bonferroni o la corrección de Benjamini-Hochberg, que tienen en cuenta la correlación entre las variables y pueden ser más adecuados para estudios de ómicas. Además, es importante utilizar otras técnicas, como la validación cruzada y la selección de variables, para mejorar la fiabilidad de la selección de genes en estudios de ómicas. En general, es importante utilizar técnicas adecuadas para reducir la probabilidad de encontrar resultados significativos por casualidad y mejorar la fiabilidad de la selección de genes en estudios de ómicas.

Perneger, T. V. (1998). What's wrong with Bonferroni adjustments. BMJ, 316(7139), 1236-1238. <https://doi.org/10.1136/bmj.316.7139.1236>

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society: Series B (Methodological), 57(1), 289-300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>

Pregunta 1.3

| Muestras | Sano A | Sano B | Sano C | Sano D | Endo A | Endo B | Endo C | Endo D |
|----------|--------|--------|--------|--------|--------|--------|--------|--------|
| Sano A1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Sano A2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Sano A3 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Sano A4 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Sano A5 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Sano B1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| Sano B2 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| Sano B3 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| Sano B4 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| Sano B5 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| Sano C1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| Sano C2 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| Sano C3 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| Sano C4 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| Sano C5 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| Sano D1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| Sano D2 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| Sano D3 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| Sano D4 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| Sano D5 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| Endo A1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |

| | | | | | | | | |
|---------|---|---|---|---|---|---|---|---|
| Endo A2 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| Endo A3 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| Endo A4 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| Endo A5 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| Endo A6 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| Endo B1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| Endo B2 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| Endo B3 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| Endo B4 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| Endo B5 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| Endo B6 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| Endo C1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| Endo C2 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| Endo C3 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| Endo C4 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| Endo C5 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| Endo C6 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| Endo D1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| Endo D2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| Endo D3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| Endo D4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| Endo D5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| Endo D6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |

Teniendo en cuenta que los contrastes son los siguientes:

$\beta_1 = \alpha_1 - \alpha_5$ Diferencia entre la población celular A entre sano y enfermo.

$\beta_2 = \alpha_2 - \alpha_6$ Diferencia entre la población celular B entre sano y enfermo.

$\beta_3 = \alpha_3 - \alpha_7$ Diferencia entre la población celular C entre sano y enfermo.

$\beta_4 = \alpha_4 - \alpha_8$ Diferencia entre la población celular D entre sano y enfermo.

$\beta_5 = \alpha_3 - \alpha_4$ Diferencia entre la población celular C y D entre sanos.

$\beta_6 = \alpha_7 - \alpha_8$ Diferencia entre la población celular C y D entre enfermos.

Hemos creado la siguiente matriz de contrastes:

| | Contraste | α_1 | α_2 | α_3 | α_4 | α_5 | α_6 | α_7 | α_8 |
|---|-----------|------------|------------|------------|------------|------------|------------|------------|------------|
| 1 | β_1 | 1 | 0 | 0 | 0 | -1 | 0 | 0 | 0 |
| 2 | β_2 | 0 | 1 | 0 | 0 | 0 | -1 | 0 | 0 |
| 3 | β_3 | 0 | 0 | 1 | 0 | 0 | 0 | -1 | 0 |
| 4 | β_4 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | -1 |
| 5 | β_5 | 0 | 0 | 1 | -1 | 0 | 0 | 0 | 0 |
| 6 | β_6 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | -1 |

También he hecho la del número 2:

```
> design
      APO BT LT
GSM26804 0 0 1
GSM26867 0 0 1
GSM26868 0 0 1
GSM26869 0 0 1
GSM26870 0 0 1
GSM26871 0 1 0
GSM26872 0 0 1
GSM26873 0 0 1
GSM26874 0 0 1
GSM26875 0 0 1
GSM26876 0 0 1
GSM26877 0 0 1
GSM26878 1 0 0
GSM26879 0 0 1
GSM26880 0 1 0
GSM26881 0 0 1
GSM26882 0 1 0
GSM26883 1 0 0
GSM26884 0 1 0
GSM26885 0 0 1
GSM26886 1 0 0
GSM26887 1 0 0
GSM26888 0 1 0
GSM26889 0 1 0
GSM26890 0 0 1
[ reached getoption("max.print") -- omitted 24 rows ]
attr(,"assign")
[1] 1 1 1
attr(,"contrasts")
attr(,"contrasts")$group
[1] "contr.treatment"

> cont.matrix
      Contrasts
Levels APO-BT APO-LT BT-LT
      APO      1      1      0
      BT     -1      0      1
      LT      0     -1     -1
```

PREGUNTA 2

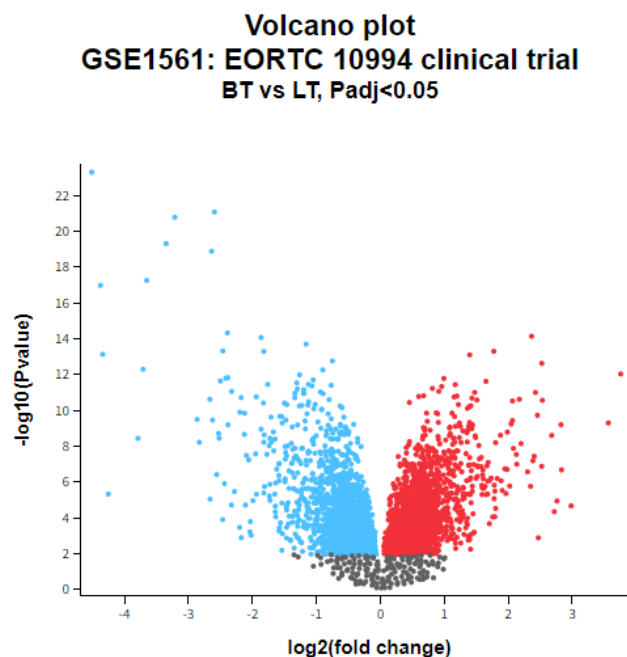
PARTE 1

```
> anotaciones<- AnnotationDbi::select (mogene10sttranscriptcluster.db,
keys=rownames(data), columns=c("ENTREZID","SYMBOL","GENENAME", "ENSEMB
L"))
> anotaciones_df <- as.data.frame(anotaciones[1:10,])
# Guardar el dataframe como una tabla CSV
> write.csv(anotaciones_df, file = "anotaciones.csv", row.names = FALS
E)
```

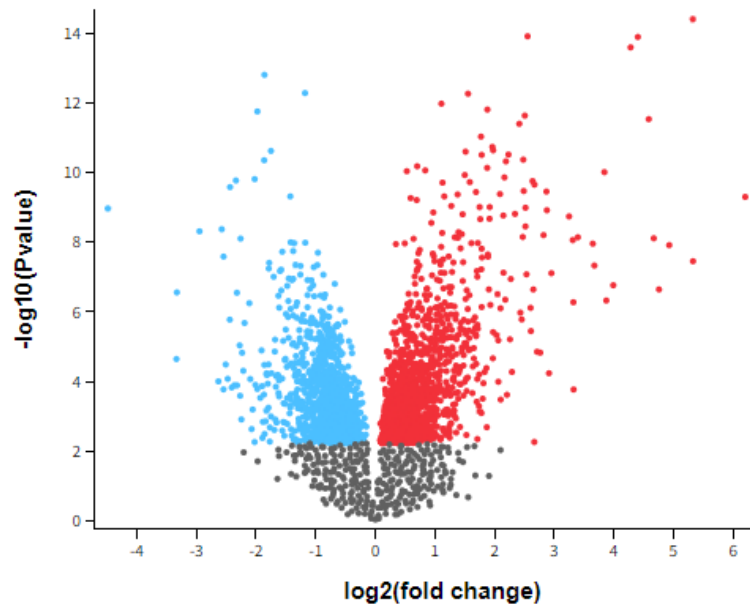
| PROBEID (double) | ENTREZID (double) | SYMBOL (character) | GENENAME (character) | ENSEMBL (character) |
|---------------------|----------------------|-----------------------|--|------------------------|
| 10470175 | 227627 | Obp2a | odorant binding protein 2A | ENSMUSG000000062061 |
| 10351443 | 110648 | Lmx1a | LIM homeobox transcription factor 1 alpha | ENSMUSG000000026686 |
| 10403796 | 218038 | Amph | amphiphysin | ENSMUSG000000021314 |
| 10522388 | 231290 | Slc10a4 | solute carrier family 10 (sodium/bile acid cotransporter fami... | ENSMUSG000000029219 |
| 10469358 | 17533 | Mrc1 | mannose receptor, C type 1 | ENSMUSG000000026712 |
| 10531869 | 26414 | Mapk10 | mitogen-activated protein kinase 10 | ENSMUSG000000046709 |
| 10400926 | 104001 | Rtn1 | reticulon 1 | ENSMUSG000000021087 |
| 10499189 | 80891 | Fcrls | Fc receptor-like 5, scavenger receptor | ENSMUSG000000015852 |
| 10474524 | 258022 | Or4f62 | olfactory receptor family 4 subfamily F member 62 | ENSMUSG000000049758 |
| 10474524 | 258440 | Or4f47 | olfactory receptor family 4 subfamily F member 47 | ENSMUSG000000050776 |

PARTE 2

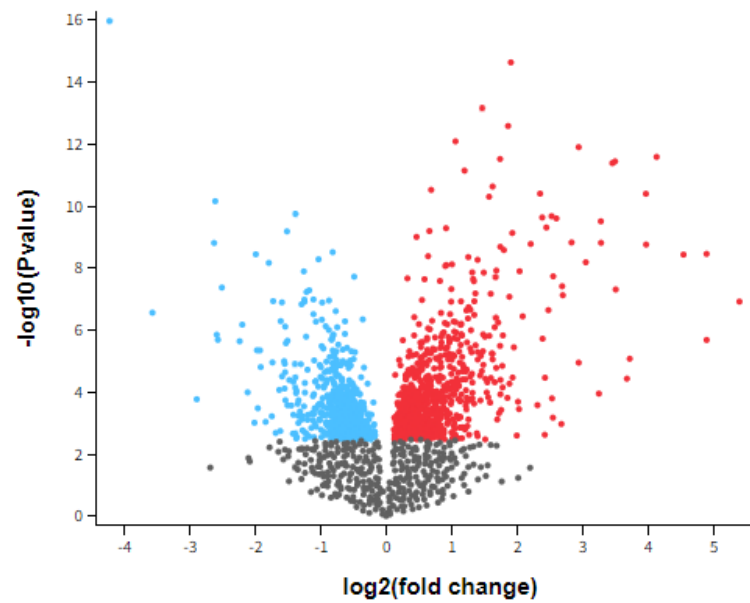
Utilizando GEO2R he hecho los volcano plots. Podemos observar la diferencia en los P-valores, y en gris abajo los de menos significancia.



Volcano plot
GSE1561: EORTC 10994 clinical trial
APO vs BT, $P_{adj} < 0.05$

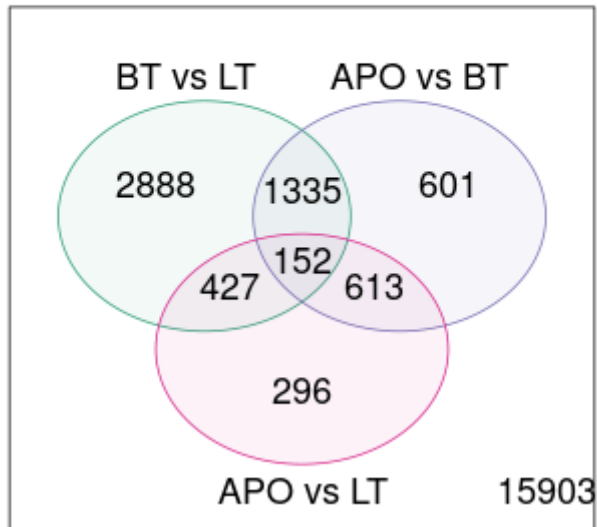


Volcano plot
GSE1561: EORTC 10994 clinical trial
APO vs LT, $P_{adj} < 0.05$

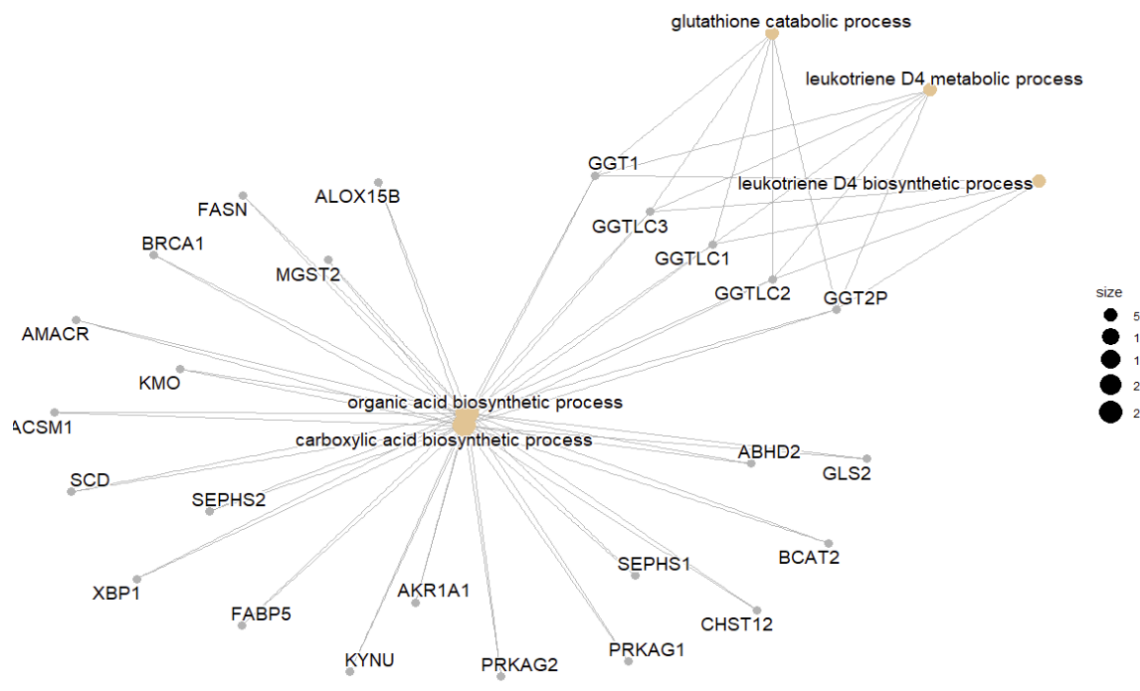


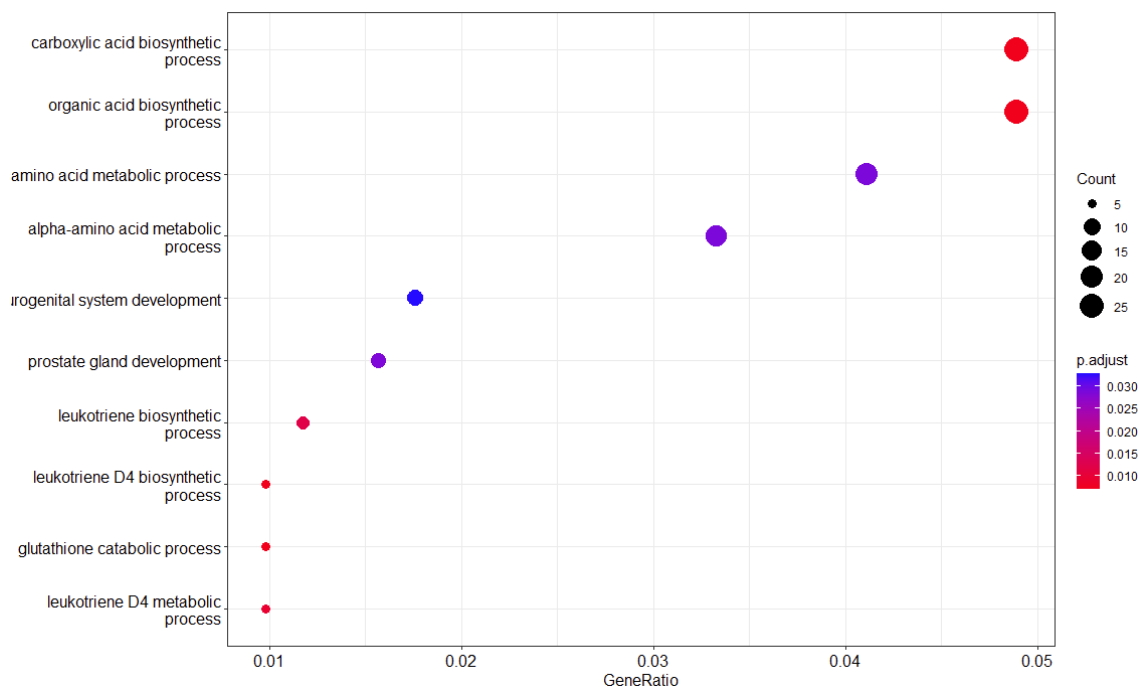
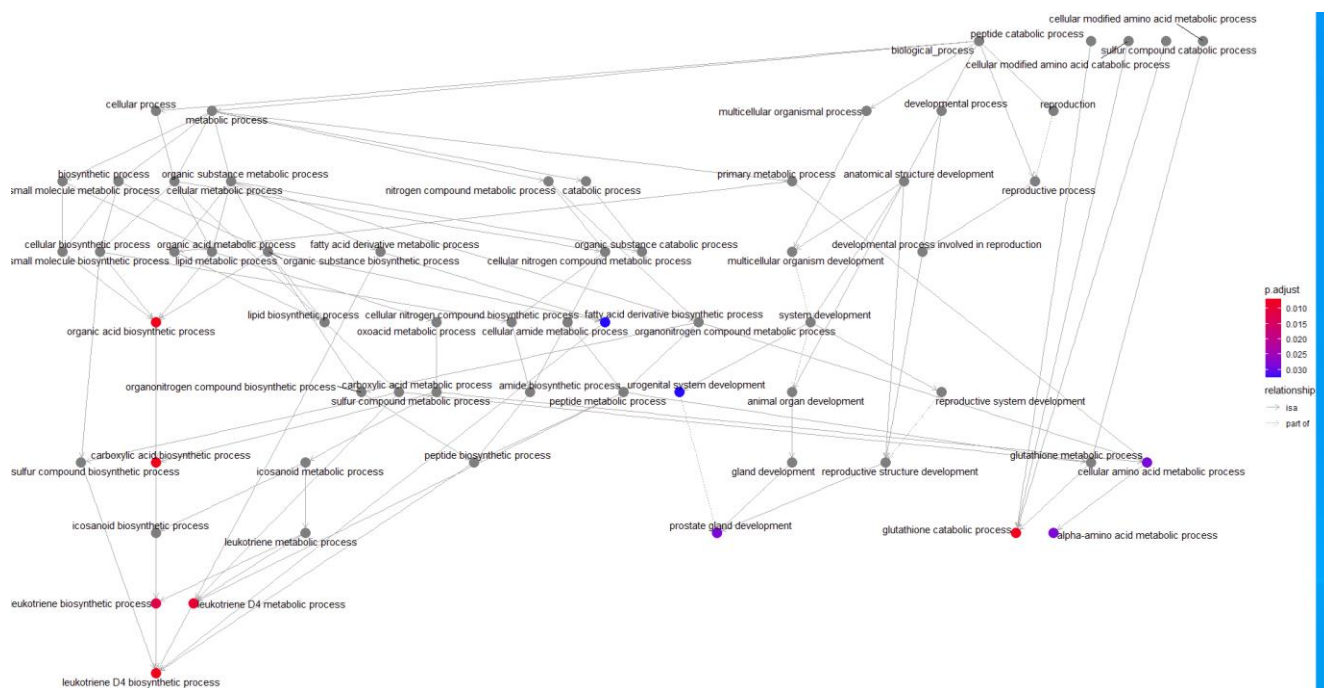
Venn Diagram

GSE1561: limma, Padj<0.05



Haciendo un análisis ORA con clusterProfiler en R y con los datos del caso 2, hemos obtenido los siguientes gráficos y podemos ver su interacción.





PREGUNTA 3

Tengo el problema de que no me encuentra el archivo. El repositorio fue clonado con éxito.

```
> targetsDF <-read.csv(file=file.path(dataDir,"targets.csv"), header =
TRUE, sep=";")
Error in file(file, "rt") : cannot open the connection
In addition: Warning message:
In file(file, "rt") :
```

cannot open file 'D:/Internship/dades/targets.csv': No such file or directory

No he conseguido que me lea el archivo. El archivo esta clonado en esa dirección y funciona cuando lo importo, pero corriendo ese código no me lo lee.