

PEC 2 – Análisis de datos ómicos

Yeison Santamaría Alza

Tabla de contenido:

1. Resumen
2. Objetivos
3. Métodos
 - 3.1 Software empleado
 - 3.2 Datos
 - 3.3 Extracción de datos
 - 3.4 Normalización
 - 3.5 Identificación de genes diferencialmente expresados
 - 3.6 Genes comunes
 - 3.7 Anotación de resultados
 - 3.8 Análisis de significancia biológica
4. Resultados
 - 4.1 Extracción de resultados
 - 4.2 Normalización
 - 4.3 Identificación de genes diferencialmente expresados
 - 4.3.1 NIT vs SFI
 - 4.3.2 SFI vs ELI
 - 4.3.3 NIT vs ELI
 - 4.4 Genes comunes
 - 4.5 Anotación de resultados
 - 4.6 Análisis de significancia biológica
5. Discusión
6. Conclusiones
7. Bibliografía

1. Resumen

La glándula tiroides se encuentra involucrada en una gran cantidad de procesos metabólicos, sin embargo, su disfunción mediada por mecanismo inflamatorios puede llevar a patologías importantes. Se sabe que estas patologías inflamatorias autoinmunes están mediadas por linfocitos autorreactivos y que la extensión del infiltrado por estas células tiene implicaciones pronósticas en los pacientes que presentan estas patologías. Por lo anterior se quiere evaluar la presencia de expresión diferencial de genes en tejido tiroideo, según la extensión del infiltrado. Se encontraron genes con expresión diferencial que se encuentran involucrados en procesos como ciclo celular, señalización intracelular y generación de regiones variables de cadenas livianas de las inmunoglobulinas.

2. Objetivos

Con la práctica se busca realizar análisis de expresión diferencial de datos de ultrasecuenciación, en este caso, la expresión diferencial de genes en muestras de tejido tiroideo en tres grupos: NIT, SFI y ELI.

3. Métodos

3.1. Software empleado

El análisis se realizó usando el software RStudio, con lenguaje de R, en la versión 4.0 para Windows. Para toda la ejecución del proceso se empleó RMarkdown. El gestor de librerías usado más fue Bioconductor. Las librerías ejecutadas fueron limma, edgeR, VennDiagram, biomaRt y topGO

3.2. Datos:

Se nos suministró una base de datos provenientes del repositorio GTEx. En esta base de datos, se encontraban datos respecto a RNAseq de tejidos tiroideos y clasificados en tres grupos.

3.3. Extracción de los datos

Se realizó extracción manual de los datos. Se cargaron los datos en RStudio y luego se separaron en tres grupos según la extensión del infiltrado. Luego se extrajeron 10 muestras aleatorias de cada grupo y se unificaron en una base. Luego se eliminaron los datos faltantes y los que pudieran interferir con el análisis

3.4. Normalización

Se realizó normalización tipo TMM. Primero se normalizaron los datos con función de la librería twweDEseq. Luego se graficó en un MA plot.

3.5. Identificación de genes diferencialmente expresados

Se utilizó la librería edgeR, con lo que se determinaron los genes diferencialmente expresados. Estos datos fueron graficados en un MDS plot para evaluar la distancia entre los genes, un BCV plot para determinar el coeficiente de variabilidad biológica y un smear plot para evaluar el número de genes regulados al alta y a la baja

3.6. Genes comunes

Se realizó evaluación con diagrama de Venn para evaluar los genes que estaban comunes en las tres diferentes comparaciones

3.7. Anotación de resultados

Se determinó el nombre de los genes que tenían expresión diferencial usando la librería biomaRt.

3.8. Análisis de significancia biológica

Se usó la librería topGO, con lo que se encontraron los procesos relacionados con los genes expresados diferencialmente, también se empleó el test de Fisher para identificar aquellas anotaciones con $p < 0,01$

4. Resultados

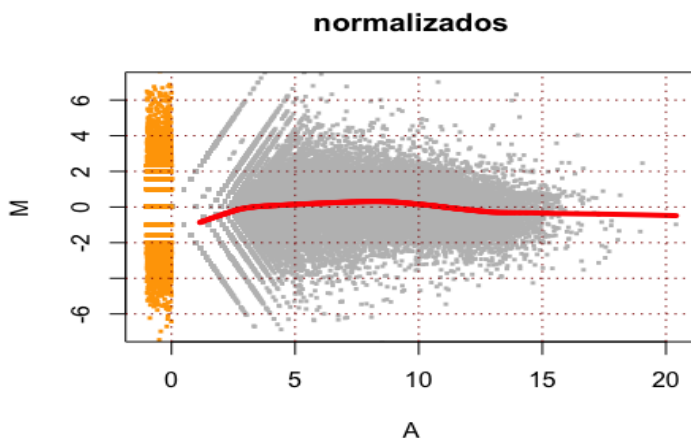
4.1. Extracción de datos

En total se extrajeron 30 muestras, con 56202 datos. Luego de retirar los datos faltantes y los que podían interferir con el análisis se obtuvieron 30 muestras con 46787 datos.

4.2. Normalización

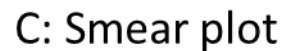
Se ejecutó la normalización y se comprobó gráficamente con un MA plot. En la gráfica 1 se muestra el MA plot, donde se puede identificar poca variabilidad y logaritmo del radio está cercano al cero (línea roja)

Gráfica 1: MA plot luego de normalizar los datos



4.3.1. NIT vs SFI

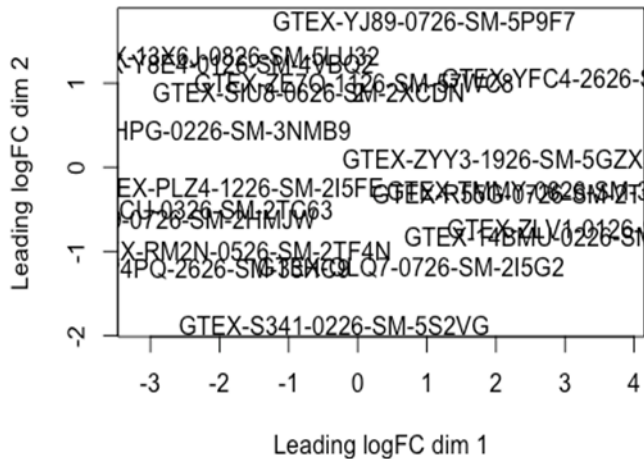
Gráfica 2: Expresión diferencial de genes entre NIT y SFI



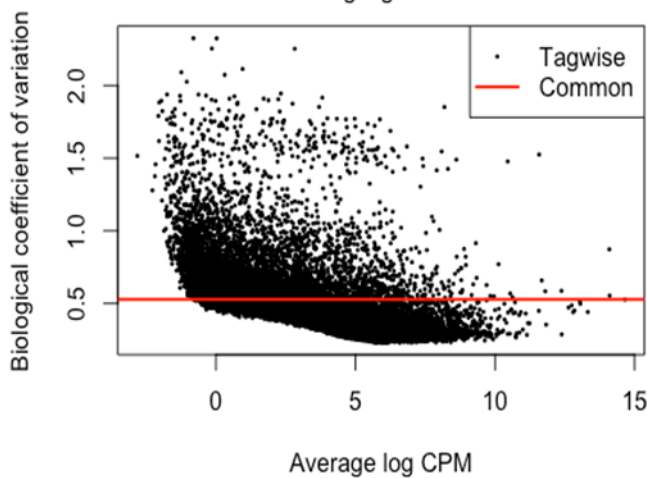
4.3.2. SFI vs ELI

Se encontraron 389 genes con expresión diferencial, de los cuales 50 tuvieron baja expresión y 339 alta expresión. En la gráfica 3 se muestran el MDS plot, BCV plot y smear plot de esta comparación

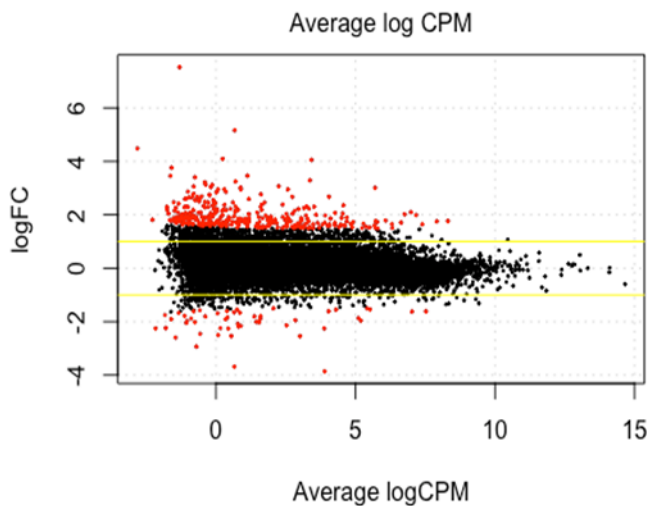
Gráfica 3: Expresión diferencial de genes entre SFI y ELI



A: MDS plot



B: BCV plot



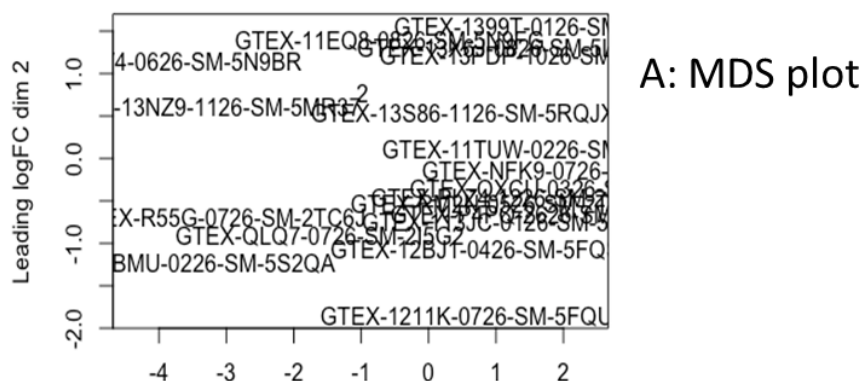
C: Smear plot

4.3.3. NIT vs ELI

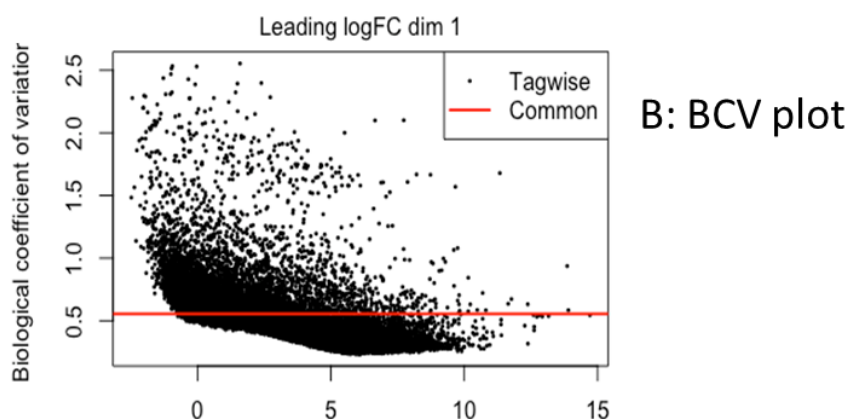
Al comparar los grupos NIT y ELI, se encontraron 201 genes con expresión diferencial. De estos 201, 154 tuvieron baja expresión y 47 tuvieron alta expresión. En la gráfica 4 se muestran el MDS plot, BCV plot y smear plot de esta comparación.

Llamativamente, la cantidad de genes diferencialmente expresados y su distribución entre alta y baja expresión, es la misma a la encontrada en la comparación NIT vs SFI

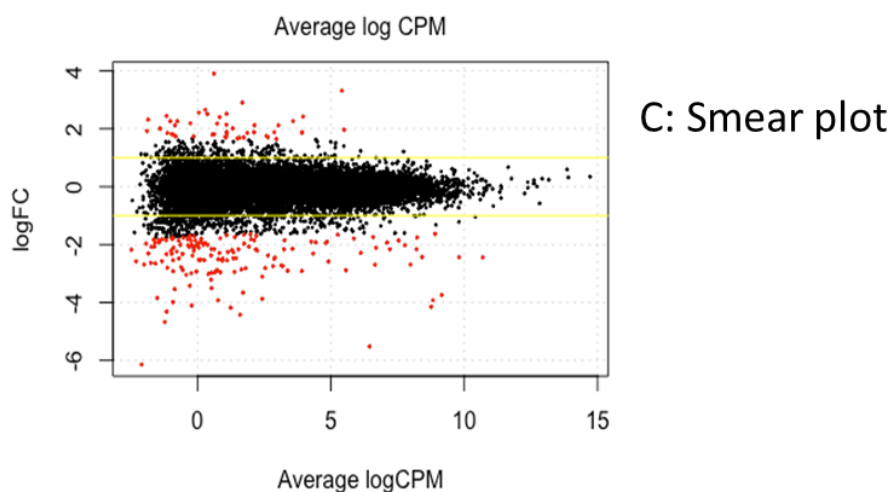
Gráfica 4: Expresión diferencial de genes entre NIT y ELI



A: MDS plot



B: BCV plot



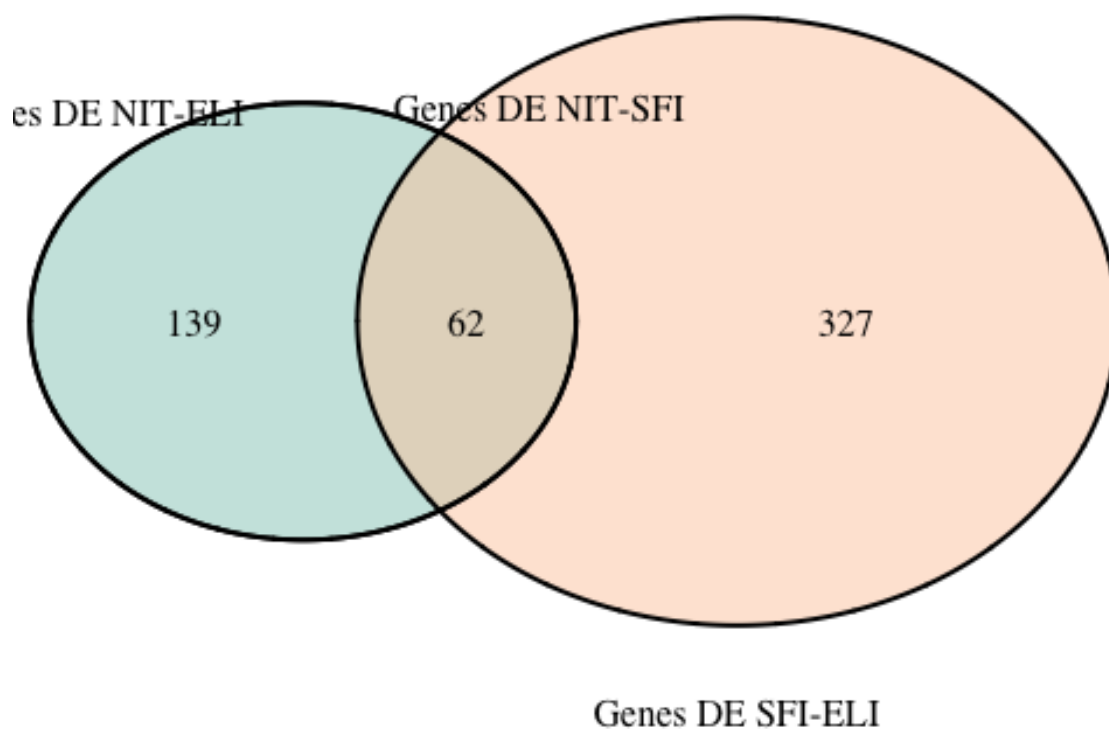
C: Smear plot

4.4. Genes comunes

Al realizar el diagrama de Venn, ocurrió un problema. Debido a que, en el análisis de expresión diferencial de genes, la comparación entre NIT vs SFI y la comparación entre NIT y ELI arrojaron la misma cantidad de genes, se omitió la comparación entre SFI vs ELI.

En la gráfica 5 se muestra el diagrama de Venn, en el que se muestra que al comparar las comparaciones de NIT vs ELI y NIT vs SFI, se encontraron 62 genes comunes.

Gráfica 5: Diagrama de Venn



4.5. Anotación de resultados

En el análisis de anotación de resultados se encontraron 10 genes:

```
## [1] "CRNN"      "SNORD14E"  "IGLV4-60"  "IGLV4-3"   "MTND4P24"  "CICP27"
## [7] "IGKV2D-28" "MTND5P11"  "LINC02562" "BNIP3P13"
```

Algunos de los genes encontrados realizan las siguientes funciones

CRNN: asociado a cáncer o proteína de choque térmico

IGLV4-69, IGLV4-3: región variable de la cadena Lambda de las inmunoglobulinas

MTND4P24, MTND5P11: Ubiquinona

IGKV2D-28: Región variable de cadena Kappa de las inmunoglobulinas

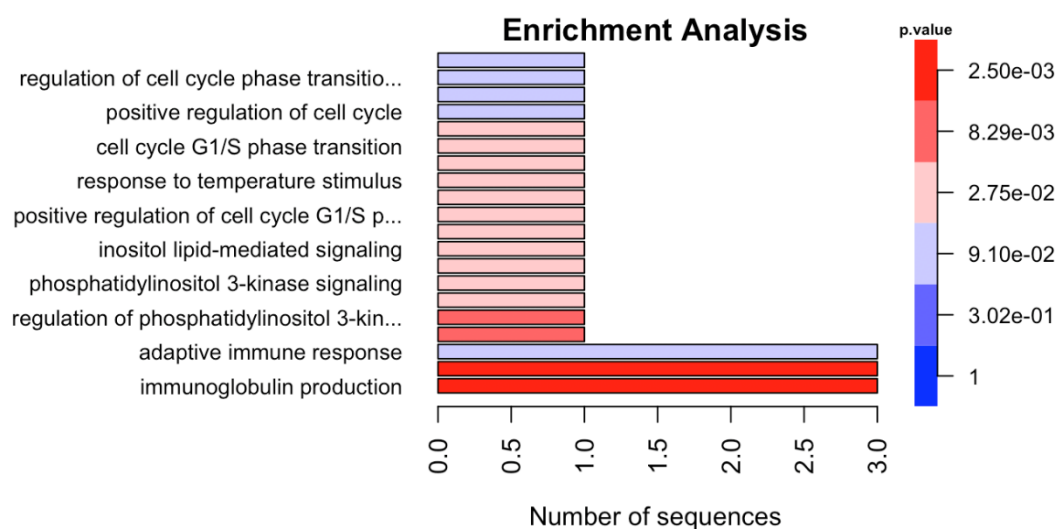
BNIP3P13: Interacción con adenovirus

4.6. Análisis de significancia biológica

Se encontraron 4438 anotaciones, de las cuales se analizaron 683, que pertenecían a 4 genes diferencialmente expresados. Al determinar la significancia estadística se encontraron 2 anotaciones con valor de $p < 0,01$ en el test de Fisher.

Finalmente se realizó un plot de enriquecimiento, donde se muestran los procesos relacionados con los genes que mostraron una expresión diferencial (gráfica 6)

Gráfica 6: Plot de enriquecimiento



5. Discusión

La glándula tiroides, tiene funciones muy importantes en el metabolismo humano. Esta glándula es propensa a tener enfermedades, algunas de ellas son de características inflamatorias, donde sus principales exponentes son la tiroiditis de Hashimoto y la enfermedad de Graves(1,2).

Estas enfermedades se caracterizan por la producción de autoanticuerpos por parte de linfocitos B autorreactivos, que generan ataque a las células tiroideas, generando dos espectros diferentes de la enfermedad hipo e hipertiroidismo(3,4).

Es sabido que el infiltrado linfocitario encontrado en la histopatología tiene relación con el pronóstico de la enfermedad, por lo que se quiere saber si la expresión diferencial de genes en las diferentes extensiones de infiltrado puede tener algún efecto clínico, a la vez que la expresión diferencial podría tener un efecto en la extensión del infiltrado inflamatorio(5).

En el análisis realizado se encontraron genes diferencialmente expresados y en la significancia biológica llama la atención la presencia de genes involucrados en el ciclo celular. Estos genes podrían estar en relación con la probabilidad de progresión a neoplasia de células tiroideas (6).

También se encontraron genes involucrados en procesos de señalización intracelular. Estas vías están relacionadas con los procesos fisiopatológicos, producción de citoquinas o quimioquinas y persistencia celular que tienen efecto inflamatorio importante en las enfermedades autoinmunes (7).

Finalmente se encontraron genes relacionados con la formación de las regiones variables de las cadenas livianas de las inmunoglobulinas. Esto puede ser importante, porque puede ser la causa de la autorreactividad de los linfocitos B, al generar regiones variables que potencialmente puedan reconocer como extraños, algunos antígenos encontrados en la glándula tiroides (8).

6. Conclusión

Al realizar el análisis se encontraron genes que tienen diferente expresión según la extensión del infiltrado linfocitario en el tejido tiroideo. Estos genes están involucrados en procesos de ciclo celular, señalización intracelular y producción de las regiones variables de las inmunoglobulinas.

En el repositorio github encontrado en el siguiente link, se puede encontrar el Rmarkdown para la reproducibilidad de los resultados.

Link: <https://github.com/yeison1807/PEC2>

7. Bibliografía

1. Smith TJ, Hegedüs L. Graves' disease [Internet]. Longo DL, editor. Vol. 375, New England Journal of Medicine. Massachusetts Medical Society; 2016 [cited 2020 Jun 14]. p. 1552–65.

Available from: <http://www.nejm.org/doi/10.1056/NEJMra1510030>

2. Caturegli P, De Remigis A, Rose NR. Hashimoto thyroiditis: Clinical and diagnostic criteria. Vol. 13, Autoimmunity Reviews. Elsevier; 2014. p. 391–7.
3. McIver B, Morris JC. The pathogenesis of graves' disease. Endocrinol Metab Clin North Am. 1998 Mar 1;27(1):73–89.
4. Rapoport B. Pathophysiology of Hashimoto's Thyroiditis and Hypothyroidism. Annu Rev Med. 1991 Feb;42(1):91–6.
5. Ajjan RA, Weetman AP. The Pathogenesis of Hashimoto's Thyroiditis: Further Developments in our Understanding. Horm Metab Res. 2015 Apr 16;47(10):702–10.
6. Iliadou PK, Effraimidis G, Konstantinos M, Grigorios P, Mitsakis P, Patakiouta F, et al. Chronic lymphocytic thyroiditis is associated with invasive characteristics of differentiated thyroid carcinoma in children and adolescents. Eur J Endocrinol. 2015 Dec 1;173(6):827–33.
7. Pinzón CE, Serrano ML, Sanabria MC. Role of phosphatidylinositol 3-kinase pathway (PI3K/Akt) in humans.
8. Chardès T, Chapal N, Bresson D, Bès C, Giudicelli V, Lefranc MP, et al. The human anti-thyroid peroxidase autoantibody repertoire in Graves' and Hashimoto's autoimmune thyroid diseases. Vol. 54, Immunogenetics. 2002. p. 141–57.