

banuls_marc_ADO_PEC2

Marc Bañuls Tornero

27/5/2020

Contents

Abstract	1
Objetivos	1
Materiales y Métodos	2
Obtención de datos previos al análisis y materiales	2
Métodos	2
Resultados	3
Control de calidad	3
Genes diferencialmente expresados	4
Gráficas de genes diferencialmente expresados por grupos	4
Análisis de significación biológica	4
Resumen de archivos obtenidos	14
Discusión	15
Conclusión	15
Apéndice	15
Referencias	15

Todos los archivos y código creados en este informe se encuentran en un repositorio Github. Para acceder a él visite la página web:

https://github.com/MarcBanuls/OMICAS_PEC2_UOC.git

Abstract

Se han obtenido datos de expresión génica procedentes de un análisis de tiroides, los cuales están subdivididos según el tipo de infiltración (No infiltrados, Infiltración focal Pequeña y infiltración linfoide extensa). A partir de estos datos se pretende realizar un análisis de la expresión génica en los distintos tipos de infiltración y un análisis de expresión diferencial entre los grupos. La conclusión más relevante de los resultados del análisis de expresión diferencial obtenidos indica que a mayor nivel de infiltración, mayor respuesta inmunitaria se genera en el tejido, en este caso la tiroides.

Objetivos

En este estudio el objetivo principal consiste realizar un análisis de expresión diferencial entre los tres tipos de infiltración de los datos procedentes del análisis de tiroides. Para ello se pretende obtener primero los genes

diferencialmente expresados entre cada grupo, observar la cantidad de genes diferencialmente expresados entre los tres grupos y finalmente realizar un análisis de significación biológica, tanto de las rutas metabólicas en las que participan (KEGG) como de su función a partir de su término GO (Gene Ontology).

Materiales y Métodos

Obtención de datos previos al análisis y materiales

Métodos

El método utilizado para este estudio está basado principalmente en los manuales de cada paquete indicado en su respectiva página de bioconductor, y los pipelines de ejemplo suministrados durante el curso (Pla and Gonzalo (2020)) y la página web “R-bloggers” (Turner (2015))

Selección de los datos

Los datos a utilizar en este análisis proceden de un estudio obtenido del repositorio GTEx. Concretamente, el archivo obtenido contiene datos de expresión de un análisis de tiroides, donde se han comparado tres tipos de infiltración. Este archivo contiene 292 muestras pertenecientes a tres grupos: - Not infiltrated tissues (NIT): 236 samples - Small focal infiltrates (SFI): 42 samples - Extensive lymphoid infiltrates (ELI): 14 samples

Para el análisis de estos datos se requiere de la selección de 30 muestras aleatorias totales, siendo 10 de cada grupo mencionado. Para ello se separan por grupos todas las muestras, obteniendo las variables NIT, SFI i ELI con sus respectivas muestras. Posteriormente se seleccionan 10 muestras aleatorias de cada grupo utilizando la función sample. También se incluye un seed concreto (123) por si se quiere replicar el estudio con las mismas muestras.

Se unen y ordenan todas las muestras en una nueva tabla de datos, se filtra la tabla de conteos para que sólo contenga las muestras aleatoriamente seleccionadas y a partir de los archivos obtenidos se crea un objeto `DeSeqDataSet` mediante el paquete `DESeq2` (Love, Huber, and Anders (2014)).

Preprocesado de los datos: filtraje y normalización

Filtraje

Se realiza un filtrado de las líneas que no dan ningún tipo de información (las que no tienen conteos en ninguna muestra). De esta manera se facilitan los posteriores procesos y análisis.

Normalización de datos

El análisis se realiza automáticamente al crear objetos `DESeq`, por lo que no es requerido ninguna normalización adicional.

Identificación de genes diferencialmente expresados

Utilizando la función `DESeq` se ajustan los parámetros de los conteos.

Para identificar los genes diferencialmente expresados entre las distintas condiciones se realizan 3 comparaciones: SFIvsNIT, ELIvsNIT y ELIvsSFI.

Anotación de los resultados

Para realizar las anotaciones se utiliza el paquete de anotaciones `AnnotationDbi` (Pagès et al. (2019)). Después de añadir las anotaciones para cada tabla de comparaciones se pueden ordenar por el p valor, para así encontrar en los primeros lugares los genes más significativos. Finalmente se pueden guardar las tablas anotadas y ordenadas para posteriores observaciones.

Es una buena idea utilizar además el paquete `ReportingTools` (Huntley et al. (n.d.)) para crear una página web con la tabla de datos y una interfaz intuitiva en la que realizar búsquedas.

Análisis de significación biológica (“Gene Enrichment Analysis”)

KEGG pathway

Para realizar el análisis de enriquecimiento de los genes se realiza un análisis de las rutas metabólicas. Para realizar este análisis las anotaciones de las rutas del KEGG son de gran ayuda. En este estudio se utilizan los paquetes `gage` (Luo et al. (2009)) y `gageData`(Luo (2019)).

Gene Ontology (GO)

De la misma manera se realiza con el mismo paquete que en el análisis de las rutas metabólicas diferencialmente expresada un análisis con los términos GO más expresados diferencialmente.

Resultados

Control de calidad

Utilizando un gráfico de cajas se puede observar que los datos están normalizados, aunque existe una significativa presencia de outliers en todos los grupos.

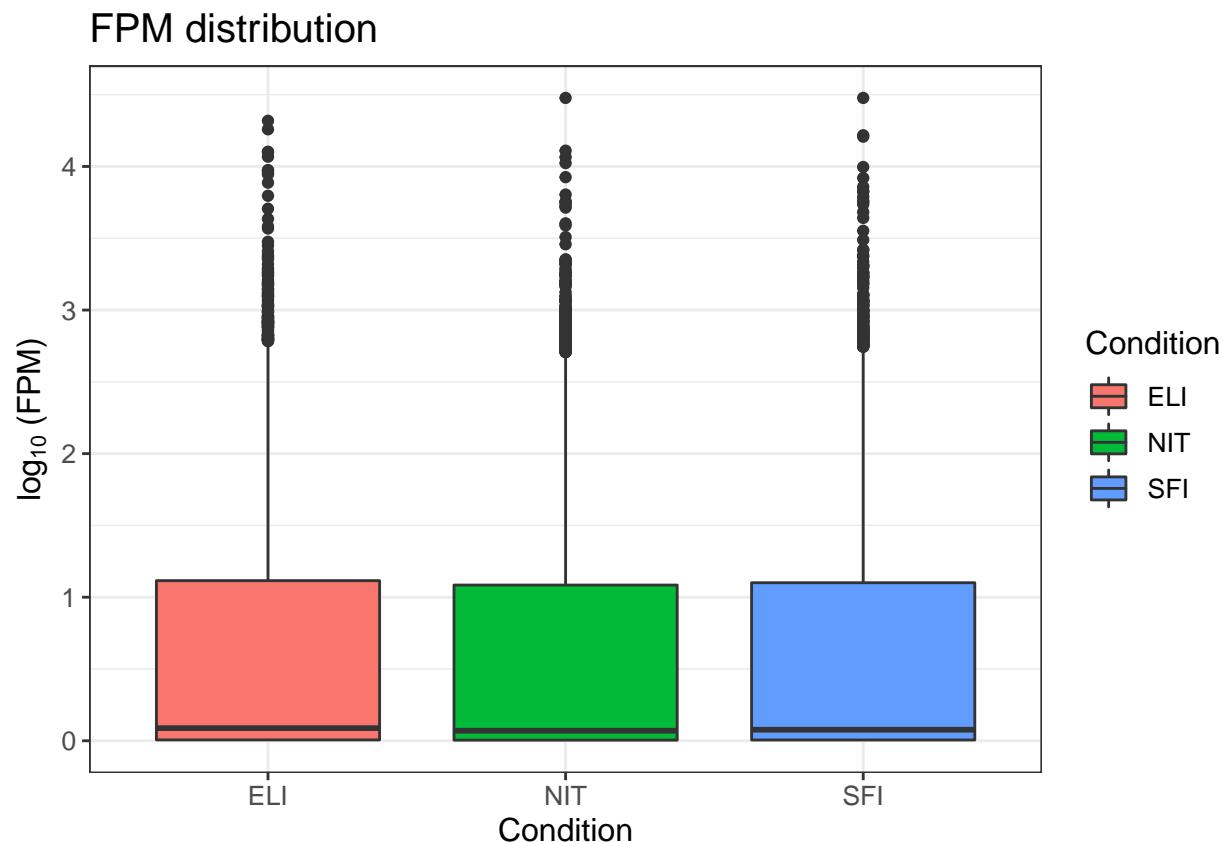


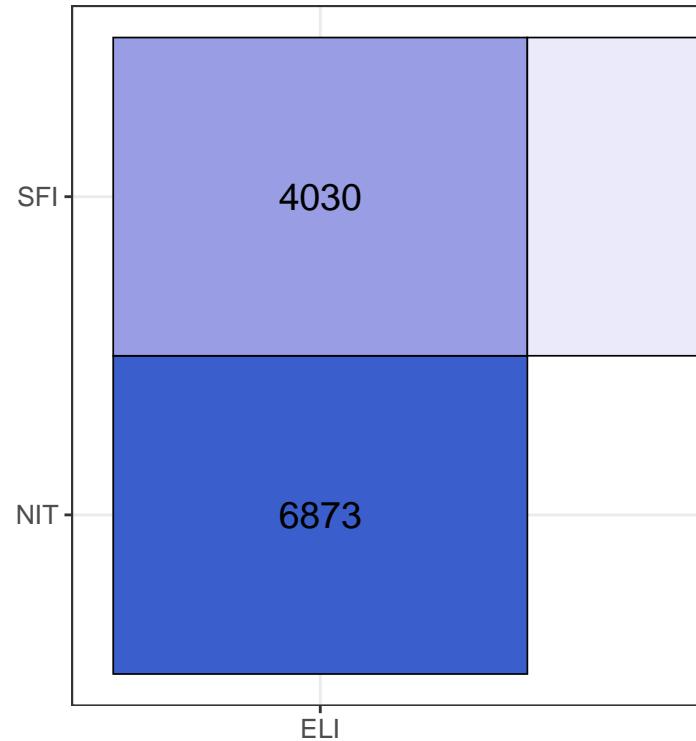
Figure 1: Figura 1. Gráfico de cajas

Debido a la elevada cantidad de conteos se puede considerar que algunos de éstos se desvén de la distribución sin afectar significativamente al análisis global. Para realizar el gráfico se utiliza el paquete `viðger` (Monier et al. (2019)).

Genes diferencialmente expresados

Se puede realizar una matriz que identifica la cantidad de genes diferencialmente expresados entre las distintas

Significant transcripts at $\alpha = 0.1$



condiciones, utilizando también funciones del paquete `vidger`:

De esta manera se puede visualizar rápidamente que entre las condiciones SFI y NIT es donde hay una menor cantidad de genes diferencialmente expresados, mientras que entre las condiciones NIT y ELI se encuentra la mayor cantidad de genes diferencialmente expresados (con un nivel de significación del 0.1).

Gráficas de genes diferencialmente expresados por grupos

Gráficos MA

Utilizando gráficos MA del paquete `vidger` se puede visualizar la cantidad de genes diferencialmente expresados entre cada par de condiciones:

Gráficos de volcán

De la misma manera que en los gráficos MA, mediante gráficos de volcán se pueden visualizar los genes diferencialmente expresados, utilizando el paquete `vidger`:

Con estos gráficos no se identifican qué genes son los que se encuentran diferencialmente expresados, pero da una idea de la cantidad de genes con una expresión significativa entre las distintas condiciones.

Análisis de significación biológica

KEGG

Se pueden visualizar las rutas metabólicas más diferencialmente expresadas en cada comparación (se mostrarán aquí las 6 rutas más significativas para evitar saturar el informe de datos):

SFI vs NIT

MA Matrix

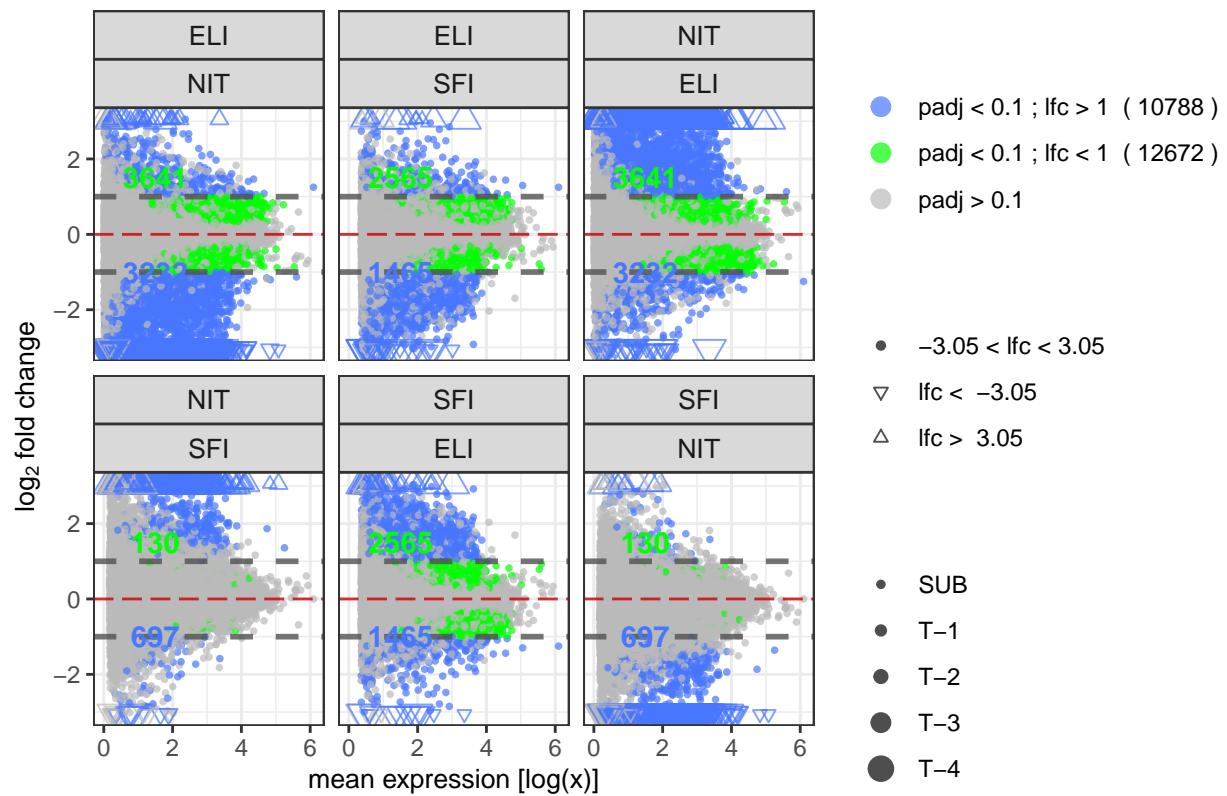


Figure 2: Figura 3. Matriz MA por grupos por grupos apareados

Volcano Matrix

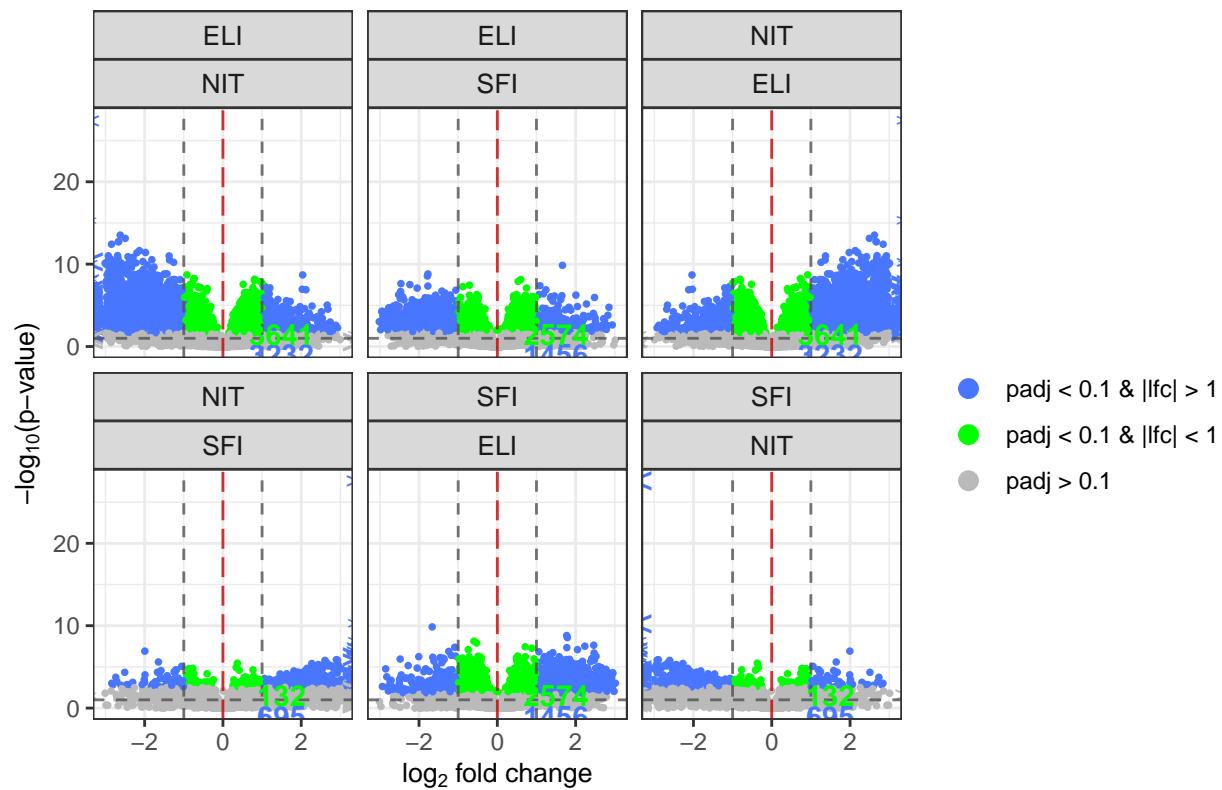


Figure 3: Figura 4. Gráficos de volcán por grupos apareados

```

## $greater
##
##          p.geomean stat.mean
## hsa04640 Hematopoietic cell lineage           1.415487e-05 4.319651
## hsa04660 T cell receptor signaling pathway     7.752574e-04 3.207032
## hsa04672 Intestinal immune network for IgA production 8.766567e-04 3.246071
## hsa04662 B cell receptor signaling pathway      1.713048e-03 2.979391
## hsa04514 Cell adhesion molecules (CAMs)        1.787950e-03 2.941650
## hsa04380 Osteoclast differentiation            3.840850e-03 2.690343
##
##          p.val      q.val
## hsa04640 Hematopoietic cell lineage           1.415487e-05 0.002321399
## hsa04660 T cell receptor signaling pathway     7.752574e-04 0.047923902
## hsa04672 Intestinal immune network for IgA production 8.766567e-04 0.047923902
## hsa04662 B cell receptor signaling pathway      1.713048e-03 0.058644758
## hsa04514 Cell adhesion molecules (CAMs)        1.787950e-03 0.058644758
## hsa04380 Osteoclast differentiation            3.840850e-03 0.092828263
##
##          set.size      exp1
## hsa04640 Hematopoietic cell lineage           84 1.415487e-05
## hsa04660 T cell receptor signaling pathway     107 7.752574e-04
## hsa04672 Intestinal immune network for IgA production 45 8.766567e-04
## hsa04662 B cell receptor signaling pathway      74 1.713048e-03
## hsa04514 Cell adhesion molecules (CAMs)        127 1.787950e-03
## hsa04380 Osteoclast differentiation            125 3.840850e-03
##
## $less
##
##          p.geomean stat.mean      p.val
## hsa04972 Pancreatic secretion                0.004422032 -2.647395 0.004422032
## hsa04974 Protein digestion and absorption    0.006141145 -2.534235 0.006141145
## hsa03010 Ribosome                          0.028710422 -1.922861 0.028710422
## hsa00190 Oxidative phosphorylation          0.029476458 -1.901875 0.029476458
## hsa04142 Lysosome                           0.041305297 -1.747308 0.041305297
## hsa00480 Glutathione metabolism            0.051150052 -1.651185 0.051150052
##
##          q.val set.size      exp1
## hsa04972 Pancreatic secretion                0.5035739 98 0.004422032
## hsa04974 Protein digestion and absorption    0.5035739 80 0.006141145
## hsa03010 Ribosome                          0.6200958 88 0.028710422
## hsa00190 Oxidative phosphorylation          0.6200958 132 0.029476458
## hsa04142 Lysosome                           0.6200958 121 0.041305297
## hsa00480 Glutathione metabolism            0.6200958 49 0.051150052
##
## $stats
##
##          stat.mean      exp1
## hsa04640 Hematopoietic cell lineage           4.319651 4.319651
## hsa04660 T cell receptor signaling pathway     3.207032 3.207032
## hsa04672 Intestinal immune network for IgA production 3.246071 3.246071
## hsa04662 B cell receptor signaling pathway      2.979391 2.979391
## hsa04514 Cell adhesion molecules (CAMs)        2.941650 2.941650
## hsa04380 Osteoclast differentiation            2.690343 2.690343

```

Con estas rutas metabólicas se refleja que el grupo SFI tiene una mayor respuesta inmunológica que el grupo NIT, ya que las rutas metabólicas diferencialmente expresadas están relacionadas con el sistema inmune principalmente. En cambio, el grupo NIT indica una mayor actividad pancreática, de ribosomas y fosforilación oxidativa (entre otros) que el grupo SFI. Razonando las expresiones diferenciales entre los dos grupos, el hecho de que haya una leve infiltración provoca una mayor actividad inmune, teniendo cierta lógica.

ELI vs NIT

```

## $greater
##
##          p.geomean stat.mean
## hsa04640 Hematopoietic cell lineage      3.221728e-06 4.699853
## hsa04672 Intestinal immune network for IgA production 3.811187e-06 4.838449
## hsa04660 T cell receptor signaling pathway    3.796501e-05 4.042546
## hsa04650 Natural killer cell mediated cytotoxicity 1.019715e-04 3.770863
## hsa04380 Osteoclast differentiation        1.808471e-04 3.616729
## hsa04612 Antigen processing and presentation 2.302682e-04 3.591423
##
##          p.val      q.val
## hsa04640 Hematopoietic cell lineage      3.221728e-06 0.0003125173
## hsa04672 Intestinal immune network for IgA production 3.811187e-06 0.0003125173
## hsa04660 T cell receptor signaling pathway    3.796501e-05 0.0020754204
## hsa04650 Natural killer cell mediated cytotoxicity 1.019715e-04 0.0041808320
## hsa04380 Osteoclast differentiation        1.808471e-04 0.0054136150
## hsa04612 Antigen processing and presentation 2.302682e-04 0.0054136150
##
##          set.size     exp1
## hsa04640 Hematopoietic cell lineage      84 3.221728e-06
## hsa04672 Intestinal immune network for IgA production 45 3.811187e-06
## hsa04660 T cell receptor signaling pathway    107 3.796501e-05
## hsa04650 Natural killer cell mediated cytotoxicity 127 1.019715e-04
## hsa04380 Osteoclast differentiation        125 1.808471e-04
## hsa04612 Antigen processing and presentation 68 2.302682e-04
##
## $less
##
##          p.geomean stat.mean
## hsa00190 Oxidative phosphorylation 0.0000995722 -3.801202
## hsa04141 Protein processing in endoplasmic reticulum 0.0013385075 -3.032475
## hsa03013 RNA transport            0.0020524339 -2.907986
## hsa04360 Axon guidance           0.0021274513 -2.887903
## hsa04974 Protein digestion and absorption 0.0031861091 -2.765283
## hsa04962 Vasopressin-regulated water reabsorption 0.0033694381 -2.806799
##
##          p.val      q.val
## hsa00190 Oxidative phosphorylation 0.0000995722 0.01632984
## hsa04141 Protein processing in endoplasmic reticulum 0.0013385075 0.08722550
## hsa03013 RNA transport            0.0020524339 0.08722550
## hsa04360 Axon guidance           0.0021274513 0.08722550
## hsa04974 Protein digestion and absorption 0.0031861091 0.09209797
## hsa04962 Vasopressin-regulated water reabsorption 0.0033694381 0.09209797
##
##          set.size     exp1
## hsa00190 Oxidative phosphorylation 132 0.0000995722
## hsa04141 Protein processing in endoplasmic reticulum 164 0.0013385075
## hsa03013 RNA transport            146 0.0020524339
## hsa04360 Axon guidance           128 0.0021274513
## hsa04974 Protein digestion and absorption 80 0.0031861091
## hsa04962 Vasopressin-regulated water reabsorption 44 0.0033694381
##
## $stats
##
##          stat.mean     exp1
## hsa04640 Hematopoietic cell lineage      4.699853 4.699853
## hsa04672 Intestinal immune network for IgA production 4.838449 4.838449
## hsa04660 T cell receptor signaling pathway    4.042546 4.042546
## hsa04650 Natural killer cell mediated cytotoxicity 3.770863 3.770863
## hsa04380 Osteoclast differentiation        3.616729 3.616729
## hsa04612 Antigen processing and presentation 3.591423 3.591423

```

En la comparación de estos grupos, se observa de la misma manera que entre SFI y NIT que ELI tiene una mayor expresión de las rutas metabólicas basadas en la actividad inmunológica que el grupo NIT. El grupo NIT sin embargo tiene una expresión diferencial de unas rutas metabólicas distintas a las encontradas al realizar la comparación con SFI. Concretamente se observa que las rutas metabólicas más diferencialmente expresadas se basan en la fosforilación oxidativa, procesamiento de proteínas en el retículo plasmático y transporte de RNA entre otras, no teniendo una relación concreta entre estas rutas (como sí pasa con las rutas metabólicas relacionadas con la actividad inmune en SFI y ELI).

ELI vs SFI

```
## $greater
##                                     p.geomean stat.mean
## hsa04672 Intestinal immune network for IgA production 0.0000110955 4.491794
## hsa04650 Natural killer cell mediated cytotoxicity    0.0001066766 3.757404
## hsa04660 T cell receptor signaling pathway            0.0005856524 3.290480
## hsa04062 Chemokine signaling pathway                  0.0012492786 3.045569
## hsa04640 Hematopoietic cell lineage                 0.0016607011 2.982760
## hsa04380 Osteoclast differentiation                  0.0031062782 2.760058
##
##                                     p.val      q.val
## hsa04672 Intestinal immune network for IgA production 0.0000110955 0.001819663
## hsa04650 Natural killer cell mediated cytotoxicity    0.0001066766 0.008747478
## hsa04660 T cell receptor signaling pathway            0.0005856524 0.032015664
## hsa04062 Chemokine signaling pathway                  0.0012492786 0.051220421
## hsa04640 Hematopoietic cell lineage                 0.0016607011 0.054470996
## hsa04380 Osteoclast differentiation                  0.0031062782 0.084904939
##
##                                     set.size     exp1
## hsa04672 Intestinal immune network for IgA production      45 0.0000110955
## hsa04650 Natural killer cell mediated cytotoxicity        127 0.0001066766
## hsa04660 T cell receptor signaling pathway                107 0.0005856524
## hsa04062 Chemokine signaling pathway                      176 0.0012492786
## hsa04640 Hematopoietic cell lineage                     84 0.0016607011
## hsa04380 Osteoclast differentiation                     125 0.0031062782
##
## $less
##                                     p.geomean stat.mean
## hsa04141 Protein processing in endoplasmic reticulum 3.913968e-05 -4.017211
## hsa00190 Oxidative phosphorylation                   7.907076e-04 -3.207852
## hsa04530 Tight junction                           2.478353e-03 -2.841004
## hsa04360 Axon guidance                            2.564922e-03 -2.828858
## hsa04510 Focal adhesion                           5.301581e-03 -2.568838
## hsa04962 Vasopressin-regulated water reabsorption 5.699114e-03 -2.606346
##
##                                     p.val      q.val
## hsa04141 Protein processing in endoplasmic reticulum 3.913968e-05 0.006418907
## hsa00190 Oxidative phosphorylation                   7.907076e-04 0.064838022
## hsa04530 Tight junction                           2.478353e-03 0.105161798
## hsa04360 Axon guidance                            2.564922e-03 0.105161798
## hsa04510 Focal adhesion                           5.301581e-03 0.155775782
## hsa04962 Vasopressin-regulated water reabsorption 5.699114e-03 0.155775782
##
##                                     set.size     exp1
## hsa04141 Protein processing in endoplasmic reticulum 164 3.913968e-05
## hsa00190 Oxidative phosphorylation                   132 7.907076e-04
## hsa04530 Tight junction                           130 2.478353e-03
## hsa04360 Axon guidance                            128 2.564922e-03
## hsa04510 Focal adhesion                           200 5.301581e-03
## hsa04962 Vasopressin-regulated water reabsorption 44 5.699114e-03
```

```

## $stats
##          stat.mean     exp1
## hsa04672 Intestinal immune network for IgA production 4.491794 4.491794
## hsa04650 Natural killer cell mediated cytotoxicity   3.757404 3.757404
## hsa04660 T cell receptor signaling pathway           3.290480 3.290480
## hsa04062 Chemokine signaling pathway                 3.045569 3.045569
## hsa04640 Hematopoietic cell lineage                2.982760 2.982760
## hsa04380 Osteoclast differentiation                 2.760058 2.760058

```

En la comparación entre los grupos ELI y SFI se observa que ELI tiene una mayor expresión que SFI en las rutas metabólicas relacionadas con el sistema inmune principalmente, mientras que el grupo SFI tiene una mayor expresión en otros grupos aparentemente no relacionados entre ellos. Al haber observado que tanto el grupo SFI como ELI tienen una sobre-expresión de las rutas metabólicas del sistema inmune en comparación a NIT, se puede suponer que ELI es el grupo con una mayor activación de las rutas metabólicas relacionadas con el sistema inmune, seguido por el grupo SFI y finalmente el que menos tiene expresado estas rutas metabólicas es NIT.

Cabe comentar que en los archivos de la carpeta *results* de este informe se encuentran gráficos de cada ruta metabólica más diferencialmente expresada, dividida en cada comparación de grupos.

GO

También se puede realizar un análisis de los términos GO de los genes más diferencialmente expresados en cada comparación (mostrando los 6 términos GO más significativos en cada comparación):

SFI VS NIT

```

## $greater
##          p.geomean stat.mean
## GO:0046649 lymphocyte activation      3.468925e-11 6.601856
## GO:0042110 T cell activation        1.951530e-08 5.560992
## GO:0050778 positive regulation of immune response 2.132264e-08 5.541939
## GO:0051249 regulation of lymphocyte activation 5.030896e-08 5.396384
## GO:0002694 regulation of leukocyte activation 1.325565e-07 5.202306
## GO:0050865 regulation of cell activation    2.030033e-07 5.115156
##          p.val      q.val
## GO:0046649 lymphocyte activation      3.468925e-11 1.506207e-07
## GO:0042110 T cell activation        1.951530e-08 3.086097e-05
## GO:0050778 positive regulation of immune response 2.132264e-08 3.086097e-05
## GO:0051249 regulation of lymphocyte activation 5.030896e-08 5.461037e-05
## GO:0002694 regulation of leukocyte activation 1.325565e-07 1.151121e-04
## GO:0050865 regulation of cell activation    2.030033e-07 1.469067e-04
##          set.size     exp1
## GO:0046649 lymphocyte activation      459 3.468925e-11
## GO:0042110 T cell activation        330 1.951530e-08
## GO:0050778 positive regulation of immune response 347 2.132264e-08
## GO:0051249 regulation of lymphocyte activation 285 5.030896e-08
## GO:0002694 regulation of leukocyte activation 325 1.325565e-07
## GO:0050865 regulation of cell activation    350 2.030033e-07
##          p.geomean stat.mean      p.val
## GO:0006643 membrane lipid metabolic process 0.001704745 -2.961937 0.001704745
## GO:0006665 sphingolipid metabolic process   0.003213514 -2.760320 0.003213514
## GO:0060271 cilium morphogenesis            0.003240581 -2.767393 0.003240581

```

```

## GO:0006790 sulfur compound metabolic process 0.003476630 -2.715827 0.003476630
## GO:0019080 viral genome expression 0.003598297 -2.722091 0.003598297
## GO:0019083 viral transcription 0.003598297 -2.722091 0.003598297
##
## q.val set.size exp1
## GO:0006643 membrane lipid metabolic process 0.8200427 160 0.001704745
## GO:0006665 sphingolipid metabolic process 0.8200427 120 0.003213514
## GO:0060271 cilium morphogenesis 0.8200427 107 0.003240581
## GO:0006790 sulfur compound metabolic process 0.8200427 229 0.003476630
## GO:0019080 viral genome expression 0.8200427 149 0.003598297
## GO:0019083 viral transcription 0.8200427 149 0.003598297
##
## $stats
## stat.mean exp1
## GO:0046649 lymphocyte activation 6.601856 6.601856
## GO:0042110 T cell activation 5.560992 5.560992
## GO:0050778 positive regulation of immune response 5.541939 5.541939
## GO:0051249 regulation of lymphocyte activation 5.396384 5.396384
## GO:0002694 regulation of leukocyte activation 5.202306 5.202306
## GO:0050865 regulation of cell activation 5.115156 5.115156

```

Los términos GO mayormente expresados en SFI respecto a NIT son una vez más los relacionados al sistema inmune o respuesta inmunitaria, mientras que los términos GO más expresados en NIT respecto a SFI tienen una menor significancia y están más esparcidos en distintas funciones.

ELI vs NIT

```

## $greater
## p.geomean stat.mean
## GO:0046649 lymphocyte activation 2.829325e-17 8.562309
## GO:0042110 T cell activation 2.914244e-14 7.696423
## GO:0050778 positive regulation of immune response 5.057536e-13 7.278024
## GO:0051249 regulation of lymphocyte activation 6.634670e-13 7.276220
## GO:0002694 regulation of leukocyte activation 9.699365e-13 7.193750
## GO:0050865 regulation of cell activation 1.839063e-12 7.085940
##
## p.val q.val
## GO:0046649 lymphocyte activation 2.829325e-17 1.228493e-13
## GO:0042110 T cell activation 2.914244e-14 6.326823e-11
## GO:0050778 positive regulation of immune response 5.057536e-13 7.201934e-10
## GO:0051249 regulation of lymphocyte activation 6.634670e-13 7.201934e-10
## GO:0002694 regulation of leukocyte activation 9.699365e-13 8.422929e-10
## GO:0050865 regulation of cell activation 1.839063e-12 1.330869e-09
##
## set.size exp1
## GO:0046649 lymphocyte activation 459 2.829325e-17
## GO:0042110 T cell activation 330 2.914244e-14
## GO:0050778 positive regulation of immune response 347 5.057536e-13
## GO:0051249 regulation of lymphocyte activation 285 6.634670e-13
## GO:0002694 regulation of leukocyte activation 325 9.699365e-13
## GO:0050865 regulation of cell activation 350 1.839063e-12
##
## $less
## p.geomean
## GO:0006091 generation of precursor metabolites and energy 1.695730e-07
## GO:0006412 translation 1.314267e-05
## GO:0015980 energy derivation by oxidation of organic compounds 2.501966e-05
## GO:0034660 ncRNA metabolic process 4.598629e-05

```

```

## GO:0008380 RNA splicing                                4.787775e-05
## GO:0045333 cellular respiration                      4.835338e-05
##
## GO:0006091 generation of precursor metabolites and energy -5.147740
## GO:0006412 translation                                 -4.232000
## GO:0015980 energy derivation by oxidation of organic compounds -4.088652
## GO:0034660 ncRNA metabolic process                  -3.957916
## GO:0008380 RNA splicing                                -3.944739
## GO:0045333 cellular respiration                      -3.976227
##
## GO:0006091 generation of precursor metabolites and energy      p.val 1.695730e-07
## GO:0006412 translation                                     1.314267e-05
## GO:0015980 energy derivation by oxidation of organic compounds 2.501966e-05
## GO:0034660 ncRNA metabolic process                     4.598629e-05
## GO:0008380 RNA splicing                                4.787775e-05
## GO:0045333 cellular respiration                      4.835338e-05
##
## GO:0006091 generation of precursor metabolites and energy      q.val 0.0007362859
## GO:0006412 translation                                     0.0285327298
## GO:0015980 energy derivation by oxidation of organic compounds 0.0294471564
## GO:0034660 ncRNA metabolic process                     0.0294471564
## GO:0008380 RNA splicing                                0.0294471564
## GO:0045333 cellular respiration                      0.0294471564
##
## GO:0006091 generation of precursor metabolites and energy      set.size 450
## GO:0006412 translation                                     458
## GO:0015980 energy derivation by oxidation of organic compounds 329
## GO:0034660 ncRNA metabolic process                     288
## GO:0008380 RNA splicing                                319
## GO:0045333 cellular respiration                      162
##
## GO:0006091 generation of precursor metabolites and energy      exp1 1.695730e-07
## GO:0006412 translation                                     1.314267e-05
## GO:0015980 energy derivation by oxidation of organic compounds 2.501966e-05
## GO:0034660 ncRNA metabolic process                     4.598629e-05
## GO:0008380 RNA splicing                                4.787775e-05
## GO:0045333 cellular respiration                      4.835338e-05
##
## $stats
##
## GO:0046649 lymphocyte activation                         stat.mean 8.562309   exp1 8.562309
## GO:0042110 T cell activation                           7.696423 7.696423
## GO:0050778 positive regulation of immune response    7.278024 7.278024
## GO:0051249 regulation of lymphocyte activation       7.276220 7.276220
## GO:0002694 regulation of leukocyte activation        7.193750 7.193750
## GO:0050865 regulation of cell activation              7.085940 7.085940

```

Una vez más los términos GO más diferencialmente expresados respecto a NIT son los relacionados a la respuesta inmune, mientras que los términos GO más diferencialmente expresados en NIT respecto a ELI varían en mayor medida, no teniendo una relación alta entre las distintas funciones diferencialmente expresadas.

ELI vs SFI

```
## $greater
```

```

##                                     p.geomean stat.mean
## GO:0046649 lymphocyte activation      1.611132e-12 7.063762
## GO:0042110 T cell activation        3.445207e-11 6.633110
## GO:0051249 regulation of lymphocyte activation 1.477585e-09 6.030734
## GO:0002694 regulation of leukocyte activation 1.607314e-09 6.004025
## GO:0050778 positive regulation of immune response 3.015546e-09 5.889863
## GO:0050865 regulation of cell activation    3.171439e-09 5.880793
##
##                                     p.val      q.val
## GO:0046649 lymphocyte activation      1.611132e-12 6.995537e-09
## GO:0042110 T cell activation        3.445207e-11 7.479544e-08
## GO:0051249 regulation of lymphocyte activation 1.477585e-09 1.744740e-06
## GO:0002694 regulation of leukocyte activation 1.607314e-09 1.744740e-06
## GO:0050778 positive regulation of immune response 3.015546e-09 2.295065e-06
## GO:0050865 regulation of cell activation    3.171439e-09 2.295065e-06
##
##                                     set.size     exp1
## GO:0046649 lymphocyte activation      459 1.611132e-12
## GO:0042110 T cell activation        330 3.445207e-11
## GO:0051249 regulation of lymphocyte activation 285 1.477585e-09
## GO:0002694 regulation of leukocyte activation 325 1.607314e-09
## GO:0050778 positive regulation of immune response 347 3.015546e-09
## GO:0050865 regulation of cell activation    350 3.171439e-09
##
## $less
##
##                                     p.geomean
## GO:0006091 generation of precursor metabolites and energy 1.521018e-07
## GO:0007507 heart development          2.134669e-06
## GO:0015980 energy derivation by oxidation of organic compounds 6.344998e-06
## GO:0061024 membrane organization       8.393360e-06
## GO:0016044 cellular membrane organization 1.581567e-05
## GO:0006412 translation                 3.448689e-05
##
##                                     stat.mean
## GO:0006091 generation of precursor metabolites and energy -5.165191
## GO:0007507 heart development          -4.635165
## GO:0015980 energy derivation by oxidation of organic compounds -4.404676
## GO:0061024 membrane organization       -4.331581
## GO:0016044 cellular membrane organization -4.186872
## GO:0006412 translation                 -4.005706
##
##                                     p.val
## GO:0006091 generation of precursor metabolites and energy 1.521018e-07
## GO:0007507 heart development          2.134669e-06
## GO:0015980 energy derivation by oxidation of organic compounds 6.344998e-06
## GO:0061024 membrane organization       8.393360e-06
## GO:0016044 cellular membrane organization 1.581567e-05
## GO:0006412 translation                 3.448689e-05
##
##                                     q.val
## GO:0006091 generation of precursor metabolites and energy 0.000660426
## GO:0007507 heart development          0.004634365
## GO:0015980 energy derivation by oxidation of organic compounds 0.009110992
## GO:0061024 membrane organization       0.009110992
## GO:0016044 cellular membrane organization 0.013734330
## GO:0006412 translation                 0.024957010
##
##                                     set.size
## GO:0006091 generation of precursor metabolites and energy 450
## GO:0007507 heart development          384

```

```

## GO:0015980 energy derivation by oxidation of organic compounds      329
## GO:0061024 membrane organization                                428
## GO:0016044 cellular membrane organization                         423
## GO:0006412 translation                                         458
##                                                               exp1
## GO:0006091 generation of precursor metabolites and energy      1.521018e-07
## GO:0007507 heart development                                    2.134669e-06
## GO:0015980 energy derivation by oxidation of organic compounds 6.344998e-06
## GO:0061024 membrane organization                           8.393360e-06
## GO:0016044 cellular membrane organization                     1.581567e-05
## GO:0006412 translation                                         3.448689e-05
##
## $stats
##                               stat.mean     exp1
## GO:0046649 lymphocyte activation      7.063762 7.063762
## GO:0042110 T cell activation        6.633110 6.633110
## GO:0051249 regulation of lymphocyte activation    6.030734 6.030734
## GO:0002694 regulation of leukocyte activation    6.004025 6.004025
## GO:0050778 positive regulation of immune response  5.889863 5.889863
## GO:0050865 regulation of cell activation       5.880793 5.880793

```

Finalmente Los términos GO de ELI más diferencialmente expresados respecto a SFI son también las funciones relacionadas al sistema inmunitario, implicando de nuevo que la mayor respuesta inmune de los tres grupos es ELI, seguido por SFI teniendo como menor respuesta inmune el grupo NIT.

Resumen de archivos obtenidos

Para información más detallada se recomienda investigar entre los distintos archivos que se han obtenido durante el estudio, los cuales se encuentran separados según la comparación. Como ejemplo se deja una tabla de archivos obtenidos en la comparación entre la condición ELI y NIT.

Table 1: Lista de archivos obtenidos en el análisis

<u>Lista_de_archivos</u>
csslib
GOdown_elis_nit.csv
GOup_elis_nit.csv
hsa04380 Osteoclast differentiation.png
hsa04380 Osteoclast differentiation.xml
hsa04640 Hematopoietic cell lineage.png
hsa04640 Hematopoietic cell lineage.xml
hsa04650 Natural killer cell mediated cytotoxicity.png
hsa04650 Natural killer cell mediated cytotoxicity.xml
hsa04660 T cell receptor signaling pathway.png
hsa04660 T cell receptor signaling pathway.xml
hsa04672 Intestinal immune network for IgA production.png
hsa04672 Intestinal immune network for IgA production.xml
jslib
report_elis_nit.html
results_elis_nit.csv

Discusión

Debido a la baja cantidad de muestras del grupo ELI, se han tenido que elegir tan solo 10 muestras aleatorias de cada grupo. Con esta cantidad de muestras (un total de 30, 10 por grupo) los resultados obtenidos tienen una menor precisión y fiabilidad respecto a qué genes están más o menos diferencialmente expresados entre grupos. Además, se han detectado varias rutas metabólicas y términos GO definidos como NA, indicando que a la hora de realizar anotaciones no se ha encontrado el código ENSEMBL pertinente para la respectiva ruta metabólica o término GO. Esto implica que si se encuentra una ruta o término GO de interés no anotada, se tiene que realizar la búsqueda del código pertinente en la base de datos ENSEMBL manualmente.

Conclusión

Aunque una mayor cantidad de muestras para realizar el análisis de expresión diferencial, se han obtenido resultados concluyentes respecto a la expresión de términos GO y rutas metabólicas más o menos expresadas entre grupos. La conclusión más clara es que a mayor infiltración en la tiroides mayor respuesta inmunitaria existe en el tejido. Para obtener más conclusiones relevantes se puede observar la página web generada por el paquete **ReportingTools**.

Apéndice

Todo el código utilizado se encuentra en el archivo “.rmd” en el repositorio github de la PEC2.

Referencias

- Huntley, Melanie A., Jessica L. Larson, Christina Chaivorapol, Gabriel Becker, Michael Lawrence, Jason A. Hackney, and Joshua S. Kaminker. n.d. “ReportingTools: An Automated Result Processing and Presentation Toolkit for High Throughput Genomic Analyses.” *Bioinformatics*. <https://doi.org/10.1093/bioinformatics/btt551>.
- Love, Michael I., Wolfgang Huber, and Simon Anders. 2014. “Moderated Estimation of Fold Change and Dispersion for Rna-Seq Data with Deseq2.” *Genome Biology* 15 (12): 550. <https://doi.org/10.1186/s13059-014-0550-8>.
- Luo, Weijun. 2019. *GageData: Auxillary Data for Gage Package*.
- Luo, Weijun, Friedman, Michael, Shedden, Kerby, Hankenson, Kurt, Woolf, and Peter. 2009. “GAGE: Generally Applicable Gene Set Enrichment for Pathway Analysis.” *BMC Bioinformatics* 10: 161.
- Monier, Brandon, Adam McDermaid, Jing Zhao, and Qin Ma. 2019. *Vidger: Create Rapid Visualizations of Rnaseq Data in R*.
- Pagès, Hervé, Marc Carlson, Seth Falcon, and Nianhua Li. 2019. *AnnotationDbi: Manipulation of Sqlite-Based Annotations in Bioconductor*.
- Pla, Alex Sanchez, and Ricardo Gonzalo. 2020. “RNASEQ Pipeline - Bioconductor.”
- Turner, Stephen. 2015. “Tutorial: RNA-Seq Differential Expression & Pathway Analysis with Sailfish, Deseq2, Gage, and Pathview.”