

Análisis de microarrays

https://github.com/IagoLast/ADO_PEC_1

Iago Lastra Rodríguez

Resumen

En este trabajo se realiza un análisis de microarrays utilizando los archivos .CEL originales de un estudio publicado en el año 2019 que investiga efectos secundarios de los inhibidores de TNF. Para realizar el análisis se ha verificado la validez de los datos, se han filtrado aquellos que proporcionan poca información y se han aplicado diferentes técnicas para comprobar que genes presentan una variabilidad estadísticamente significativa. Finalmente se han analizado estos genes para intentar descubrir las vías afectadas por el tratamiento.

Índice

1. Objetivos	2
2. Materiales y Métodos	2
3. Procedimiento de trabajo	3
3.1. Normalización	3
3.2. Filtrado	4
3.3. Selección de genes	5
3.4. Interpretación biológica de los resultados	6
4. Resultados	6
4.1. Keratinization	6
4.2. Chylomicron remodeling	6
4.3. Assembly of active LPL and LIPC lipase complexes	6
4.4. Plasma lipoprotein remodeling	6
4.5. Pyruvate metabolism	6
4.6. Metabolism of fat-soluble vitamins	7
5. Discusión	7
6. Apéndice	7

1. Objetivos

Aunque los inhibidores de factores de necrosis tumoral **TNF** son **utilizados en el tratamiento de enfermedades inflamatorias crónicas** no existe demasiada información acerca de cómo pueden afectar estos tratamientos al funcionamiento normal del sistema nervioso central.

En este trabajo se analizarán Microarrays de ARN para buscar diferencias estadísticamente significativas entre muestras sin tratar (WT) y muestras sometidas a tratamientos de inhibición de TNF.

2. Materiales y Métodos

Este trabajo se basa en estudio de comparación de grupos (class comparison) donde se han tomado muestras correspondientes al día 13.5 de la fase embrionaria (E13.5) al séptimo día de vida (P7) y en adultos de 2 y 4 meses de vida (A2 y A4 respectivamente) de un grupo de control (WT) de ratones **C57BL/6** y un segundo grupo de ratones tratados (TNF^{-/-}).

Los microarrays utilizados son del modelo GeneChip Mouse Gene 1.0 ST Array de Affymetrix que según su especificación contienen aproximadamente 25 sondas (probes) diseñadas para cubrir 28,853 genes bien conocidos y anotados.



Figura 1: GeneChip Mouse Gene 1.0 ST

3. Procedimiento de trabajo

En un primer paso se analizaron gráficamente los archivos .CEL originales buscando posibles errores en los datos. Aunque tanto el histograma como el boxplot mostraron datos de intensidades bastante uniformes se realizó una comprobación adicional utilizando el paquete `arrayQualityMetrics` para verificar que los datos no contenían errores.

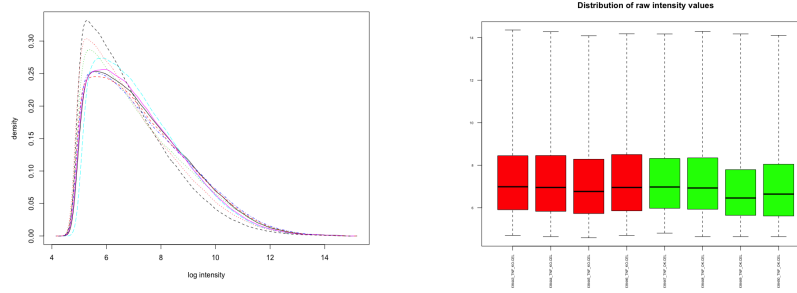


Figura 2: Distribución de intensidades en los datos originales.

3.1. Normalización

Para poder realizar un análisis de la expresión diferencial de los datos es necesario transformar los datos para que sean comparables entre sí.

Esta transformación se ha realizado utilizando el algoritmo [Robust Multichip Analysis \(RMA\)](#) que a grandes rasgos corrige el ruido de fondo, normaliza los datos y realiza una estimación final de la intensidad.

Una vez obtenidos los datos normalizados se repite el control de calidad sobre los mismos.

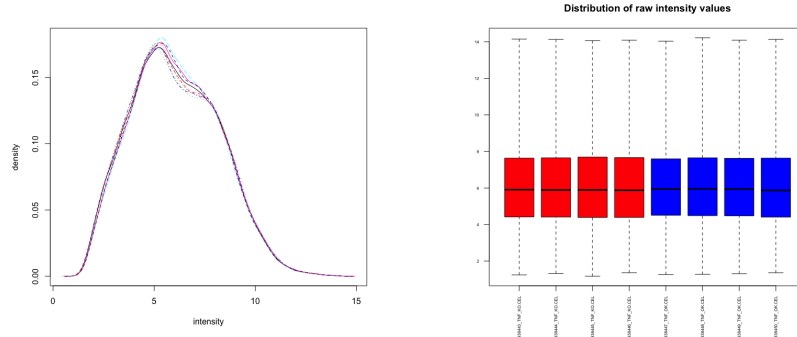


Figura 3: Distribución de intensidades en los datos normalizados.

3.2. Filtrado

Antes de empezar el análisis es interesante eliminar los genes cuya variabilidad puede ser consecuencia de un ruido aleatorio y aquellos de los que no se dispone de anotaciones.

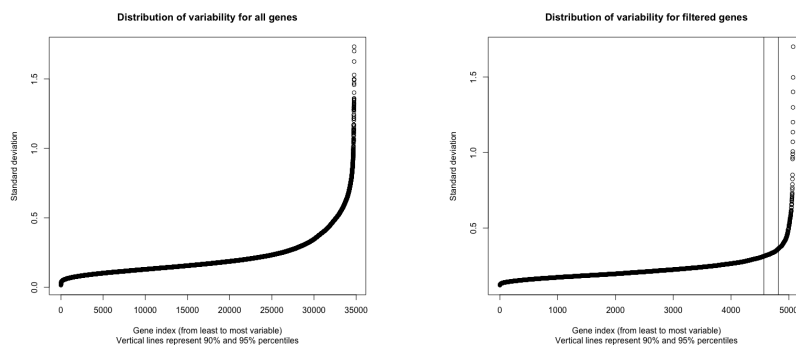


Figura 4: Variabilidad de los genes antes y después del filtrado.

Para ello se ha utilizado la función `nsFilter` paquete `geneFilter` para eliminar 1594 genes duplicados, 15225 con una variabilidad irrelevante y 12866 de los que actualmente no se disponen anotaciones

numDupsRemoved	1594
numLowVar	15225
numRemoved.ENTREZID	12866

3.3. Selección de genes

Para escoger los genes se han utilizado dos aproximaciones.

Por un lado se ha realizado un t-test (`rowttest`) y por otro se ha utilizado el método de Smyth (`limma`) visto en prácticas previas.

Debido al grán número de genes procesados se ha optado por aplicar una corrección sobre el p-valor. Dado que estamos dispuestos a asumir falsos positivos a cambio de maximizar los genes candidatos el método seleccionado es el de Benjamini & Hochber. Los genes seleccionados mediante rowtest y limma respectivamente han sido:

SYMBOL	BH
9430060I03Rik	0.0431307
Gm10787	0.0431307
Gm10024	0.0431307
Gm11696	0.0431307
Dnmt3aos	0.0491421
Gm10782	0.0431307
BC025933	0.0431307
Gm10536	0.0431307
Gm10532	0.0431307
Gm10857	0.0431307
Gm10804	0.0491421
Gm10714	0.0431307
Oog3	0.0431307
Gm10369	0.0431307
Gm10445	0.0491421
Gm10610	0.0431307
Fam129c	0.0431307
Gm10655	0.0431307

PROBE_ID	SYMBOL	adj.P.Val
10423836	Cthrc1	0.0311910
10418205	Plac9b	0.0311910
10566326	Trim12a	0.0311910
10471675	Glo1	0.0311910
10398432	Mir377	0.0311910
10572130	Lpl	0.0361161
10412394	Nnt	0.0410250

3.4. Interpretación biológica de los resultados

Con esto se ha podido realizar un análisis enrich dando como resultado las siguientes vías:

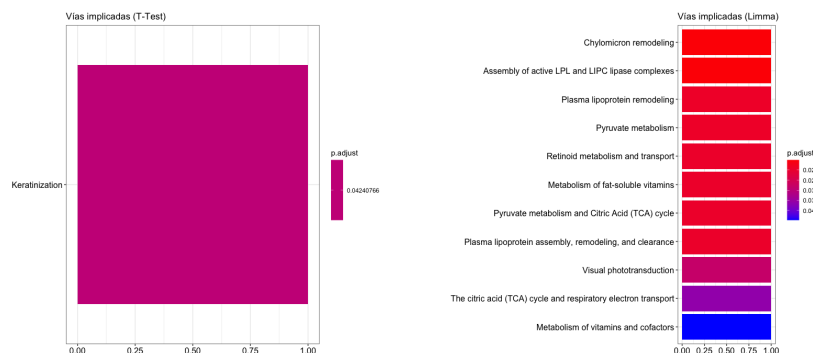


Figura 5: Vías con mayor significancia entre los grupos estudiados.

4. Resultados

4.1. Keratinization

Es interesante porque hay [un estudio previo](#) que relaciona los anti-TNF con los keratinocitos.

4.2. Chylomicron remodeling

Vemos que la primera vía implicada es [chylomicron remodeling](#).

<https://www.sciencedirect.com/topics/medicine-and-dentistry/chylomicron>

(Se relaciona con TNF y LPS)

4.3. Assembly of active LPL and LIPC lipase complexes

<http://scielo.isciii.es/pdf/nh/v27n6/01articuloespecial01.pdf> <https://www.ncbi.nlm.nih.gov/pubmed/3495531>

4.4. Plasma lipoprotein remodeling

<https://www.ncbi.nlm.nih.gov/pubmed/8572227>

4.5. Pyruvate metabolism

<https://www.ncbi.nlm.nih.gov/pubmed/29358703> <https://www.ncbi.nlm.nih.gov/pubmed/9450646>

##Retinoid metabolism and transport

4.6. Metabolism of fat-soluble vitamins

##Pyruvate metabolism and cytric acid (TCA) cycle

5. Discusión

- Los dos métodos dan vias diferentes que a la vez son diferentes a las del propio paper.

6. Apéndice

Estudio: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE134178>

ID__DATA: GSE134178

BioProject: <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA554146>