

Análisis de expresión diferencial entre pólipos serrados (SPs) y tejido cecal circundante en el modelo transgénico HBUS utilizando DESeq2

Presentado por: Sofía Betancur Álvarez

Introducción.

El estudio realizado por Bongers et al. (2014) investigó los mecanismos que determinan la localización específica de los pólipos serrados (SPs) en el intestino, los cuales son lesiones con potencial de transformación a cáncer colorrectal por medio de la “vía serrada” (Carballal et al., 2013), empleando un modelo transgénico de ratón denominado HBUS, especie *Mus musculus*. Estos animales expresan el ligando HB-EGF a lo largo del tracto gastrointestinal, lo que activa la vía EGFR/MAPK, una vía de señalización celular, pero particularmente, desarrollan pólipos únicamente en el ciego, la primera parte del intestino grueso. El propósito principal del estudio fue entender cómo interactúan tres factores: la microbiota intestinal, las alteraciones genéticas y la respuesta inflamatoria, en la formación local de neoplasias intestinales. Mediante técnicas como RNA-seq, análisis de microbiota, ensayos funcionales e intervenciones como tratamiento antibiótico, los autores demostraron que el microbioma específico del ciego es indispensable para la aparición de los SPs y que la invasión bacteriana induce una respuesta inflamatoria e inmunitaria fundamental para su crecimiento.

En cuanto al experimento asociado al tema, el artículo incluye un análisis de transcriptómica comparativo mediante RNA-seq entre tejido de pólipos serrados y el tejido cecal (estructura del ciego) adyacente no afectado, con el objetivo de identificar genes diferencialmente expresados asociados a inflamación, defensa antimicrobiana y señalización EGFR, la cual controla la proliferación, supervivencia y migración de la célula. En el diseño experimental aplicado, cada muestra de pólipo serrado (SP) se compara directamente con su tejido cecal circundante del mismo individuo, por tanto, la variable biológica principal es el estado del tejido, con dos niveles: SP (pólipo serrado) y tejido circundante no afectado.

En el dataset de Expression Atlas utilizado para este proyecto se incluyen seis muestras, tres provienen de pólipos serrados (SP) y tres del tejido cecal normal adyacente. Esto corresponde a tres réplicas biológicas por condición, lo cual permite realizar este análisis con DESeq2. Las condiciones definidas fueron “SP vs tejido circundante”, es decir, se evaluaron los genes cuya expresión aumenta o disminuye específicamente en los pólipos y, finalmente, en el artículo, se reportaron 404 genes sobreexpresados y 272 genes subexpresados en los SPs.

Resultados.

Ahora bien, para evaluar los cambios transcriptómicos entre los pólipos serrados (SPs) y el tejido cecal circundante como se indica en el artículo, se realizó un análisis de expresión diferencial utilizando DESeq2. El volcano plot obtenido se indica en la figura 1 y resume estos resultados, facilitando la visualización de los genes con mayor magnitud de cambio y significancia estadística. Los puntos en color rojo representan genes que cumplen simultáneamente con los criterios de significancia de valor-p ajustado < 0.05 y $|\log_2FC| > 1$,

donde el valor-p ajustado (padj) es el valor p corregido para múltiples pruebas y $|\log_2FC| > 1$ es la magnitud de cambio equivalente a una sobreexpresión mayor a 2 veces, lo cual es considerado biológicamente relevante en un análisis de RNA-seq (Love et al., 2014). Por otro lado, los puntos azules corresponden a genes significativos solo por valor-p y los verdes hacen referencia a genes que superan únicamente el umbral de \log_2FC , es decir, tienen una magnitud de cambio significativa.

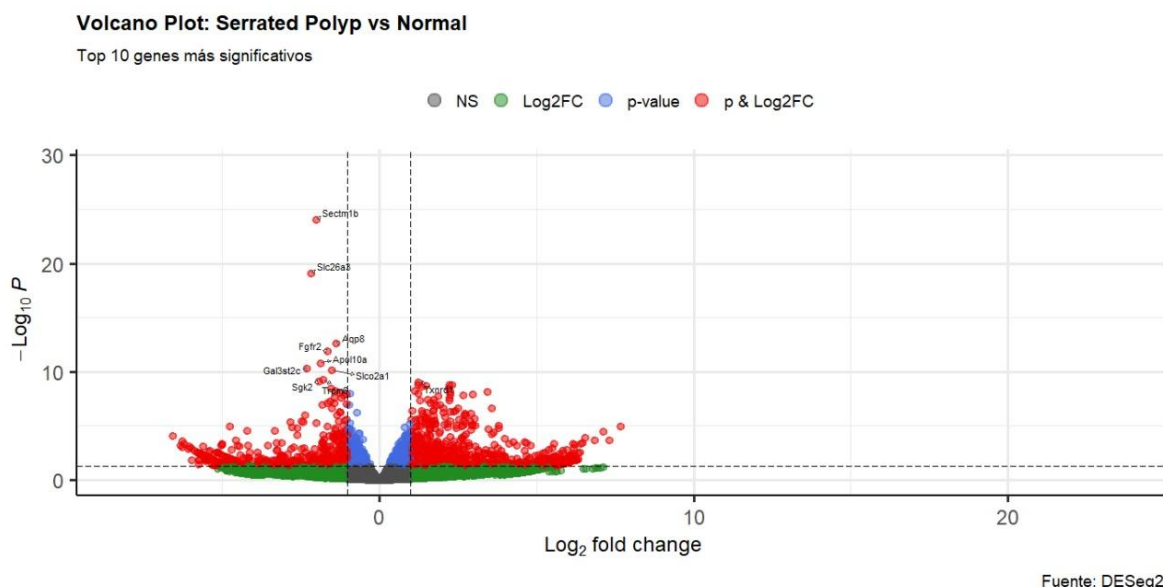


Figura 1. Volcano plot graficado en R studio.

En el volcano plot se observa que hay mayor cantidad de puntos rojos, reflejando que la mayoría de los cambios en los genes de SPs son altamente significativos y consistentes entre réplicas. Dentro de los genes más destacados se encuentran *Aqp8*, *Fgfr2*, *Gal3st2*, *Sectm1b*, *Sgk2*, *Slc26a8*, *Slco2a1*, *Trpm6* y *Txnrd1*, los cuales fueron etiquetados como los 10 genes más significativos de acuerdo con los valores de padj. Estos genes identificados se encuentran la tabla 1, junto con los valores correspondientes a los resultados del DESeq2 incluyendo los valores de \log_2 FoldChange, p-value y padj.

baseMean	log2FoldChange	lfcSE	stat	pvalue	padj	Gene.ID	Gene.Name	log10p	color
27.304.183.836.623	-216.231.886.394.837	0.237097967396679	-911.993.842.752.216	7,52E-07	5,77E-02	ENSMUSG000000001225	Slc26a3	191.239.726.505.795	Down
17.366.486.399.818	-193.011.652.138.563	0.313134561704766	-616.385.655.699.487	7,10E+04	1,21E+08	ENSMUSG000000017868	Sgk2	914.877.676.673.511	Down
18.816.883.981.499	124.018.104.793.765	0.202418889976961	612.680.490.481.301	8,97E+04	1,38E+08	ENSMUSG000000020250	Txnrd1	904.739.497.053.692	Up
300.013.802.282.576	-177.725.236.350.918	0.286917950848267	-619.428.780.337.644	5,85E+04	1,12E+08	ENSMUSG000000024727	Trpm6	923.247.933.719.997	Down
233.303.125.009.853	-13.665.721.790.517	0.186497276779534	-732.757.176.217.205	2,34E+01	1,20E+05	ENSMUSG000000030762	Aqp8	126.301.161.695.188	Down
544.318.354.723.076	-165.187.599.674.121	0.232801332542032	-709.564.665.590.119	1,29E+02	4,94E+05	ENSMUSG000000030849	Fgfr2	118.902.593.686.276	Down
453.218.414.092.752	-151.537.940.107.512	0.232271737669824	-652.416.611.800.289	6,84E+03	1,50E+07	ENSMUSG000000032548	Slco2a1	101.650.652.488.647	Down
804.693.060.932.281	-199.476.348.646.973	0.194071345848811	-102.785.059.677.163	8,81E-11	1,35E-06	ENSMUSG000000039364	Sectm1b	24.055.110.990.298	Down
208.319.858.035.574	-187.412.509.037.806	0.277919896809923	-674.340.020.952.089	1,55E+03	4,75E+06	ENSMUSG000000050982	Apol10a	108.104.472.106.552	Down
979.360.600.296.644	-230.161.916.249.167	0.350615657929214	-656.450.763.233.267	5,22E+03	1,34E+07	ENSMUSG000000073608	Gal3st2c	102.822.873.132.188	Down

Tabla 1. Subtabla de los resultados de DESeq2 con los 10 genes con mayor significancia.

Los genes identificados están principalmente involucrados en procesos de transporte transmembranal, regulación epitelial, señalización celular y mantenimiento de la homeostasis, especialmente en tejidos del tracto gastrointestinal. Genes como *Aqp8*, *Trpm6*,

Slco2a1 y *Slc26a8* participan directamente en el movimiento de agua, iones y metabolitos, los cuales son fundamentales para el equilibrio osmótico y la absorción de minerales (Yang et al., 1997; Ohana et al., 2011). Por otro lado, genes como *Fgfr2* y *Sectm1b* regulan vías de señalización relacionadas con proliferación celular, diferenciación, y activación inmunitaria (Turner & Grose, 2010). Además de su función individual, el comportamiento conjunto de estos genes sugiere alteraciones características del microambiente de los pólipos serrados. La mayoría de ellos se encuentran subexpresados (down) en los SPs, lo cual indica una posible pérdida de funciones epiteliales normales, especialmente aquellas relacionadas con el transporte iónico y la integridad de la superficie de la mucosa.

No obstante, *Txnrd1* aparece sobreexpresado (up), un resultado coherente con su papel antioxidante y la adaptación al estrés celular. Este gen suele inducirse en situaciones de inflamación crónica, estrés oxidativo o proliferación acelerada, condiciones típicas en microambientes tumorales tempranos (Arnér, 2009). El hecho de que haya una mayor cantidad de genes subexpresados (down) que sobreexpresados (up) sugiere una pérdida generalizada de funciones epiteliales normales en los pólipos serrados. Esto es consistente con los cambios tisulares, la disfunción de transporte y el estado de inflamación descrito en el artículo, donde la microbiota del ciego induce un microambiente que reduce la expresión de genes asociados con homeostasis epitelial.

Al comparar los diez genes más significativos obtenidos con DESeq2 con los valores reportados en Expression Atlas (tabla 2), se observó una coincidencia completa en la dirección del cambio para todos los genes analizados. Aunque los valores de log2FoldChange (log2FC), los cuales indican cuánto cambia la expresión de un gen entre dos condiciones, difieren en magnitud entre DESeq2 y Expression Atlas, es un resultado esperable debido a las diferencias en métodos de normalización y modelos estadísticos, y específicamente, la coincidencia en la dirección del cambio indica una alta reproducibilidad biológica entre ambos datasets. La consistencia observada en este top 10 confirma que los genes más relevantes del análisis están regulados de manera similar en la base de datos pública y en el análisis independiente realizado en R.

Gene.Name	Gene.ID.DESeq	log2FoldChange	padj	Gene.ID.Expression Atlas	foldChange_atlas	log2FC_atlas	direction_atlas	direction_DESeq	coincide_direccion
Aqp8	ENSMUSG00000030762	-13.665.721.790.517	1,20E+05	ENSMUSG00000030762	-1.3	-1.3	Down	Down	Si
Fgfr2	ENSMUSG00000030849	-165.187.599.674.121	4,94E+05	ENSMUSG00000030849	-1.6	-1.6	Down	Down	Si
Gal3st2c	ENSMUSG00000073608	-230.161.916.249.167	1,34E+07	ENSMUSG00000073608	-2.3	-2.3	Down	Down	Si
Sectm1b	ENSMUSG00000039364	-199.476.348.646.973	1,35E-06	ENSMUSG00000039364	-1.9	-1.9	Down	Down	Si
Sgk2	ENSMUSG00000017868	-193.011.652.138.563	1,21E+08	ENSMUSG00000017868	-1.7	-1.7	Down	Down	Si
Slc26a3	ENSMUSG00000001225	-216.231.886.394.837	5,77E-02	ENSMUSG00000001225	-2.1	-2.1	Down	Down	Si
Slco2a1	ENSMUSG00000032548	-151.537.940.107.512	1,50E+07	ENSMUSG00000032548	-1.5	-1.5	Down	Down	Si
Trpm6	ENSMUSG00000024727	-177.725.236.350.918	1,12E+08	ENSMUSG00000024727	-1.7	-1.7	Down	Down	Si
Txnrd1	ENSMUSG00000020250	124.018.104.793.765	1,38E+08	ENSMUSG00000020250	1.2	1.2	Up	Up	Si

Tabla 2. Comparación de los 10 genes identificados anteriormente con la base de datos de Expression Atlas.

En total se evaluaron 9.584 genes que estaban presentes en ambos conjuntos de datos, donde 9.564 genes presentaron la misma dirección del cambio (up o down), mientras que solo 20 genes mostraron discrepancias; esto indica que el análisis realizado reproduce de manera muy precisa los resultados oficiales del experimento. Por otro lado, en DESeq2 se identificaron 359 genes significativos ($p_{adj} < 0.05$), mientras que 9.198 genes no alcanzaron el umbral de

significancia. Por su parte, Expression Atlas reporta únicamente los genes significativos generados en el análisis original del experimento, razón por la cual los 9.584 genes disponibles en su archivo aparecen marcados como significativos. Finalmente, tanto en DESeq2 como en Expression Atlas se observó un predominio de genes reprimidos (Down), con 4.880 genes en DESeq2 y 4.884 en Atlas. De igual manera, el número de genes sobreexpresados (up) fue muy similar entre ambos análisis, 4.704 en DESeq2 y 4.700 en Atlas.

Conclusión.

El análisis de expresión diferencial realizado con DESeq2 permitió reproducir de manera consistente los resultados de transcriptómica descritos por Bongers et al. (2014), confirmando que los pólipos serrados del modelo HBUS presentan una reprogramación génica asociada a inflamación, estrés epitelial y pérdida de funciones intestinales normales. La predominancia de genes subexpresados indica que el epitelio del pólipo presenta una disminución significativa en procesos esenciales como transporte iónico, absorción de nutrientes y mantenimiento de la barrera mucosa. Al mismo tiempo, la sobreexpresión de genes como *Txnrd1* refleja la activación de mecanismos compensatorios frente al daño y al estrés oxidativo generado por la interacción con la microbiota.

La comparación con Expression Atlas mostró una gran coincidencia en la dirección del cambio, tanto en los diez genes más significativos como en el conjunto global de genes detectados. Esto valida no solo contraste SP vs tejido circundante, sino también la calidad del procesamiento bioinformático realizado en este proyecto. Por otro lado, desde un punto de vista biológico, los resultados refuerzan la idea de que los pólipos serrados no solo dependen de alteraciones genéticas como la activación de EGFR/MAPK, sino también de la interacción estrecha con la microbiota local, la cual genera un microambiente inflamatorio, aumentando posiblemente la carcinogénesis colorrectal.

Bibliografía.

- Arnér, E. S. J. (2009). Focus on mammalian thioredoxin reductases — Important selenoproteins with versatile functions. *Biochimica et Biophysica Acta (BBA) - General Subjects*, 1790(6), 495–526. <https://doi.org/10.1016/j.bbagen.2009.01.014>
- Bongers, G., Pacer, M. E., Geraldino, T. H., Chen, L., He, Z., Hashimoto, D., Furtado, G. C., Ochando, J., Kelley, K. A., Clemente, J. C., Merad, M., van Bakel, H., & Lira, S. A. (2014). *Interplay of host microbiota, genetic perturbations, and inflammation promotes local development of intestinal neoplasms in mice*. [The Journal of Experimental Medicine], 211(3). <https://doi.org/10.1084/jem.20131587>
- Carballal, S., Moreira, L., & Balaguer, F. (2013). *Pólipos serrados y síndrome de poliposis serrada*. *Cirugía Española*, 91(3), 141–148. [10.1016/j.ciresp.2012.12.001](https://doi.org/10.1016/j.ciresp.2012.12.001)

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15, 550.
<https://doi.org/10.1186/s13059-014-0550-8>

Ohana, E., Yang, D., Shcheynikov, N., & Muallem, S. (2009). Diverse transport modes by the solute carrier 26 family of anion transporters. *The Journal of Physiology*, 587(10), 2179–2185. [10.1113/jphysiol.2008.164863](https://doi.org/10.1113/jphysiol.2008.164863)

Turner, N., & Grose, R. (2010). Fibroblast growth factor signalling: From development to cancer. *Nature Reviews Cancer*, 10, 116–129. [Fibroblast growth factor signalling: from development to cancer | Nature Reviews Cancer](https://doi.org/10.1038/nrnc2009)

Yang, B., & Verkman, A. S. (1997). Water and glycerol permeabilities of aquaporins 1–5 and MIP determined quantitatively by expression of epitope-tagged constructs in *Xenopus* oocytes. *Journal of Biological Chemistry*, 272(26), 16140–16146.
<https://doi.org/10.1074/jbc.272.26.16140>