
transcripts_under_notch1

Victor Ulises Plascencia Pérez

Abstract

En este reporte se describen resultados de un análisis de expresión diferencial utilizando datos de expresión de células sometidas a una activación continua del receptor Notch1 y células cuyo receptor Notch1 no fue activado.

Objetivo

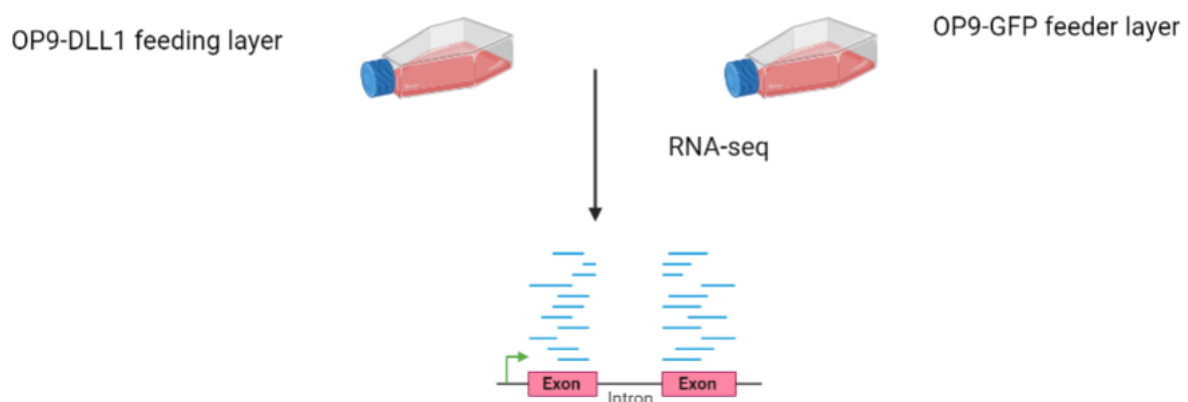
El objetivo de este estudio fue obtener genes diferencialmente expresados en una condición donde el receptor Notch1 se encontraba continuamente unido a ligandos (estado activo).

Justificación

Entre las principales mutaciones que ocurren en ciertos tipos de cáncer, particularmente en la leucemia linfoblástica aguda de células T, están aquellas que tienen lugar sobre el gen que codifica para el receptor membranar NOTCH1. Muchas de estas mutaciones tienen un efecto activador en NOTCH1. La activación continua de NOTCH1 está asociada al desarrollo de fenotipos cancerígenos en ciertos tipos de células, por lo que conocer cómo cambia el transcriptoma en estas células es fundamental para comprender cómo mutaciones en NOTCH1 pueden inducir fenotipos cancerígenos.

Diseño general

Cultivos celulares de células CD34+ fueron crecidos sobre capas de células que expresaban el ligando de NOTCH1, DLL1 o proteína verde fluorescente (control negativo). Posteriormente, se realizó la extracción del RNA y su secuenciación.



Análisis de expresión diferencial

Este análisis de expresión diferencial se realizó utilizando Bioconductor.

Para el análisis de expresión diferencial se utilizaron datos obtenidos de Recount3, con el identificador SRP048604. Las lecturas fueron reescaladas utilizando funciones del paquete Recount3. Se realizaron análisis exploratorios del objeto RangeSummarizedExperiment (RSE), así como la determinación de algunos parámetros de calidad de las lecturas, tales como la proporción de lecturas asignadas a genes y la distribución de los valores de lecturas. Los datos anteriores fueron utilizados para descartar lecturas consideradas insignificantes. Posteriormente los datos fueron normalizados, un procedimiento que tiene como objetivo eliminar efectos técnicos sistemáticos (Robinson, M.D., et al (2010). Finalmente, se generó un modelo estadístico adecuado y se realizó el análisis de expresión diferencial utilizando el paquete Limma.

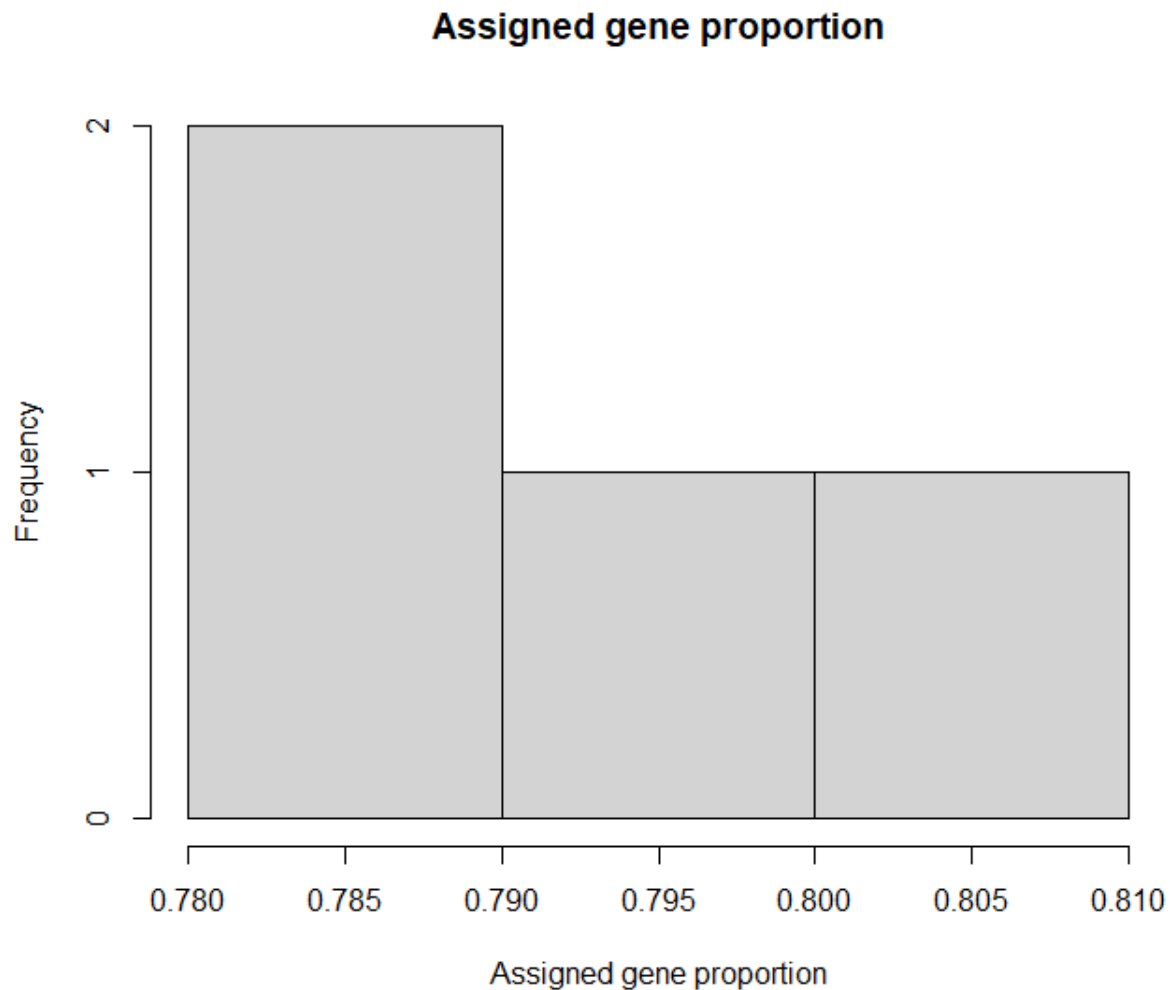
Resultados

Análisis exploratorios

Los análisis exploratorios del objeto permitieron comprobar que sólo había dos réplicas de lecturas para cada condición, lo que verdaderamente podría impactar en la obtención de resultados.

Parámetros de calidad de las lecturas en el RSE previo al análisis

La determinación de la proporción de lecturas asignadas a genes permitió corroborar que las lecturas eran representativas del transcriptoma del modelo usado. Este parámetro se obtuvo mediante la relación del atributo con las cuentas totales asignadas a genes y el atributo con las cuentas totales. En la figura 1 se observa un histograma con la distribución de este parámetro.



Asi mismo, los valores obtenidos por las cuentas mostraron la siguiente distribución de valores.

Minimo	1er quartil	Mediana	Media	3er quartil	Máximo
0.0	0.0	1.5	2095.0	197.0	1355387.2

Limpieza de los datos

La limpieza de los datos se fundamentó en los parámetros de calidad anteriores. Se eliminaron genes con una expresión por debajo de 0.1, quedando 61.89 % de genes restantes.

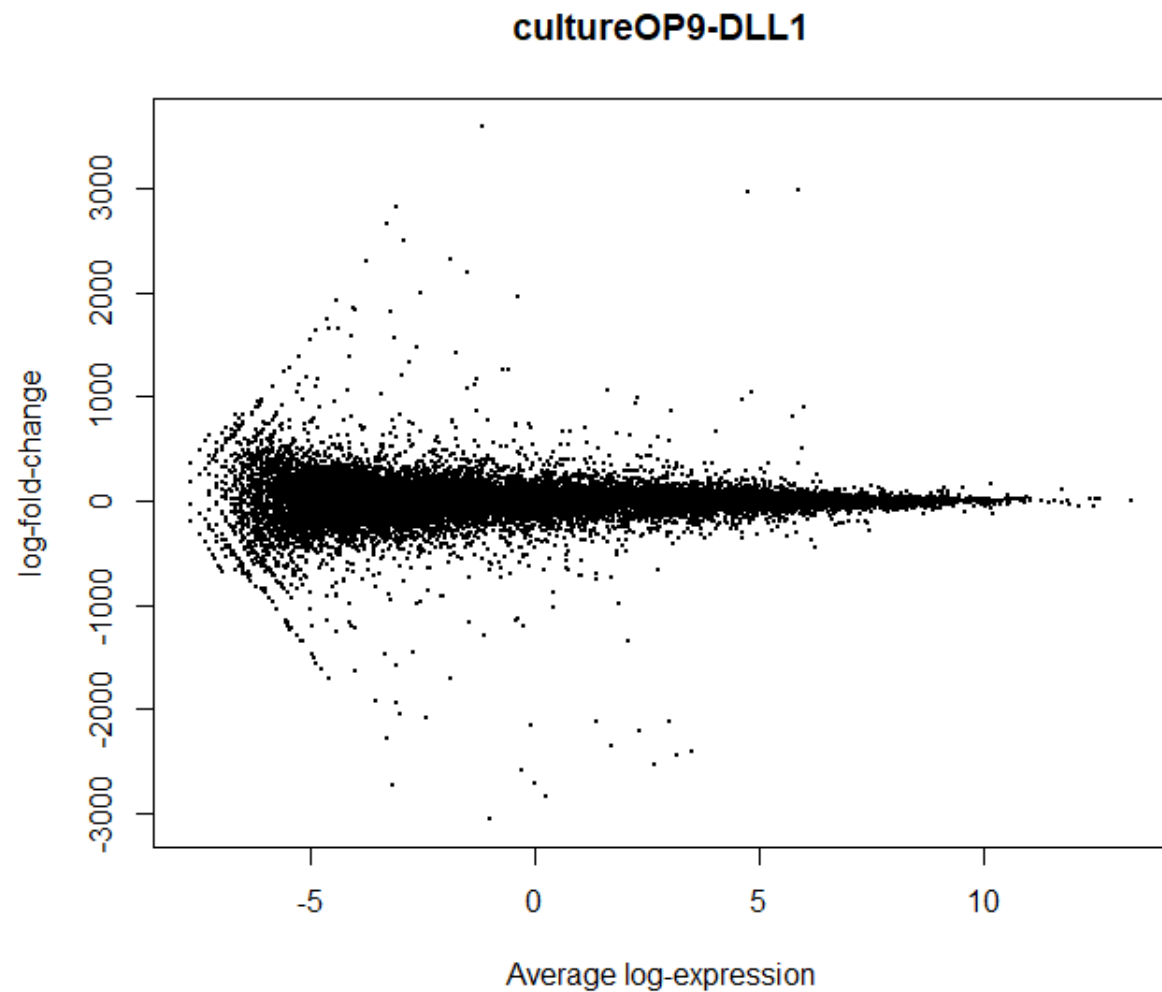
Análisis de expresión diferencial

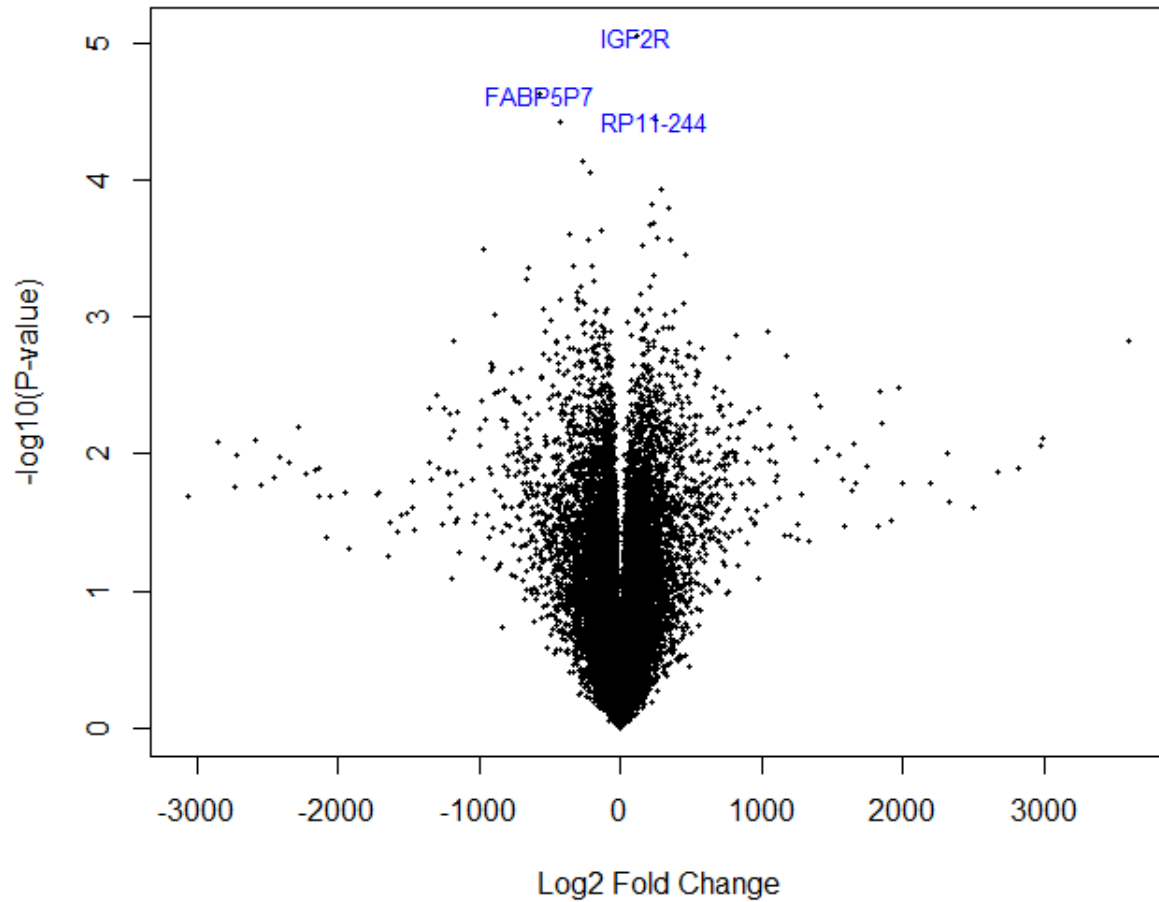
Tras la limpieza y normalización de los datos, se procedió a establecer un modelo estadístico para la comparación de datos. Al tener una estructura muy sencilla (comparación tratamiento - control), no fue necesario utilizar visualizaciones gráficas como las que pudiera proporcionar el paquete de Bioconductor ExploreModelMatrix.

Finalmente, se realizó el análisis de expresión diferencial y se seleccionó aquellos genes con un valor P ajustado inferior a 0.05. Sin embargo, no se obtuvieron genes cuyo valor estuviera por debajo de este umbral, lo cual significa que la tasa de descubrimiento de falsos esta por encima del 5% y confiar en que los genes con un logFC (logaritmo de cambio duplicacion) alto son en realidad genes

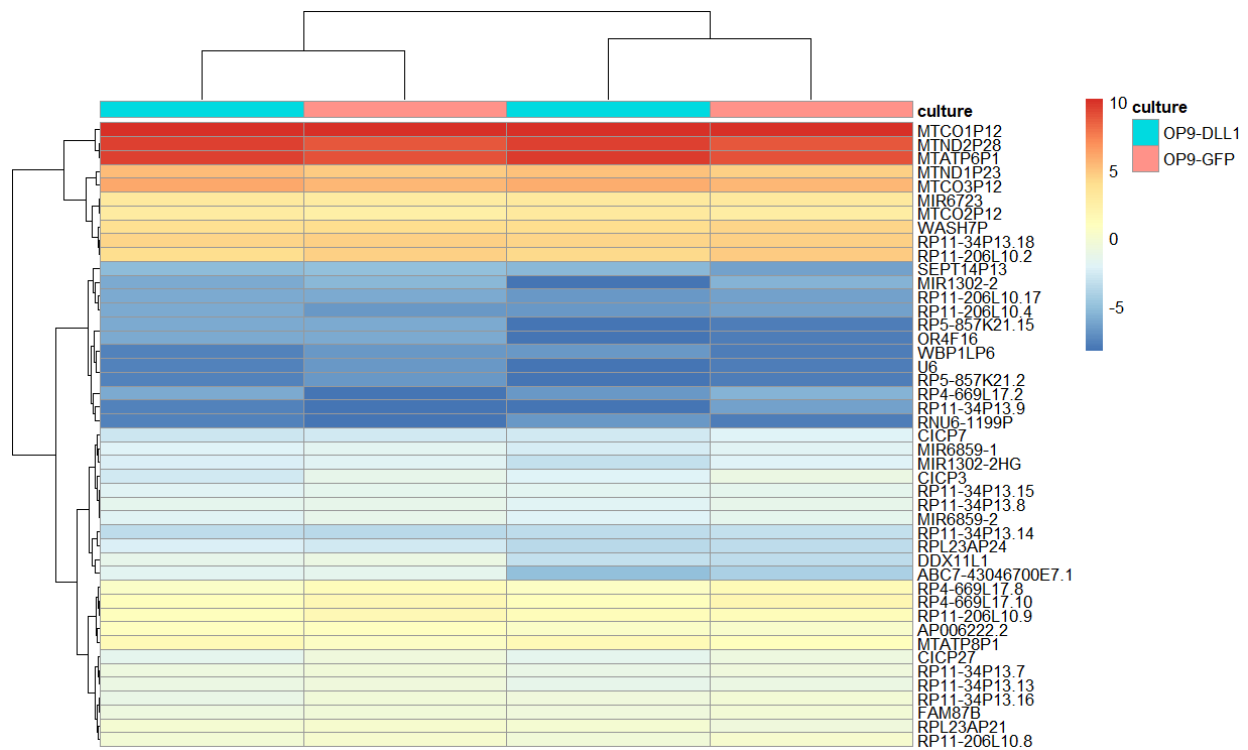
diferencialmente expresados (DEGs) es arriesgado.

En las siguientes figuras podemos observar gráficas MA y Volcán, donde se observa que efectivamente hay genes con logFC con valores suficientemente grandes (Fig 2.), así como con valores P mayores a 0.05 (Fig 3.) como para ser considerados DEGs.





Finalmente, los genes con el valor p ajustado más alto fueron utilizados para genera un mapa de calor (Fig 4.). Sin embargo, como se podría esperar, muchos de estos genes no muestran cambios significativos de expresión en las dos condiciones del experimento, pues su criterio de selección no está basado en el logFC o el pValue. Por lo anterior, no se plantea en este reporte que los genes mostrados en el siguiente mapa de calor sean DEGs.



Conclusiones y discusiones

Los resultados obtenidos del presente análisis de expresión diferencial no permitieron identificar con veracidad genes diferencialmente expresados, primordialmente por obtener valores p ajustados demasiado altos. Uno de los factores que influyeron en ello fue la cantidad de replicas disponibles para cada condición, las cuales son insuficientes para obtener datos estadísticamente significativos. Según SCHURCH N. et al (2016), se recomienda al menos una cantidad de 6 replicas por condición.

Dada la relevancia que podrían tener los resultados obtenidos de este análisis, sería recomendable repetir el experimento con una cantidad de replicas mayor a las recomendadas y ejecutar el análisis planteado en este reporte con esos nuevos datos.

Bibliografías

1. Zhao, Y., Li, MC., Konaté, M.M. et al. TPM, FPKM, or Normalized Counts? A Comparative Study of Quantification Measures for the Analysis of RNA-seq Data from the NCI Patient-Derived Models Repository. *J Transl Med* 19, 269 (2021). <https://doi.org/10.1186/s12967-021-02936-w>
2. Schurch, N. J., Schofield, P., Gierliński, M., Cole, C., Sherstnev, A., Singh, V., ... & Barton, G. J. (2016). How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use?. *Rna*, 22(6), 839-851.
3. Robinson, M.D., Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol* 11, R25 (2010). <https://doi.org/10.1186/gb-2010-11-3-r25>