

# Proyecto Final

## Análisis de Expresión Diferencial del set de datos SRP107565

### Introducción

El cáncer de mama es una patología que afecta a mujeres de todo el mundo. Ocupa el quinto lugar de muertes en mujeres a nivel mundial (1). En México, en 2020, el cáncer de mama fue la enfermedad más frecuente (15.3%) y la primera causa de muerte en mujeres, con una tasa de incidencia de 40.5 y de mortalidad de 10.6 por cada 100,000 casos (1). Todos los tipos de cáncer se caracterizan por una proliferación descontrolada y por la migración de células cancerígenas a otros tejidos, el cáncer de mama no es la excepción. Las células cancerígenas presentan una regulación de ciclo celular aberrante, esta mala regulación permite que las células pueden formar tumores debido al desarrollo y proliferación descontrolada de la misma (7).

El cáncer de mama tiene una clasificación molecular que permite la correlación con el riesgo de caída de la enfermedad. Existen 4 tipos: Luminal A, luminal B (HER2 negativo o positivo), HER-2 y triple negativos. Los tumores luminales tienen receptores de estrógeno, los HER2 tienen expresión de HER2 (receptor 2 del factor de crecimiento epidérmico humano) y de receptores de estrógeno y los triple negativos no expresan ni receptores hormonales ni HER2. El subtipo Luminal A es el de mejor pronóstico y el triple negativo es el de peor pronóstico (2).

Para el tratamiento del cáncer de mama se tiene una variedad de opciones que permiten atacar de distintas formas a cualquiera de los cuatro tipos de cáncer. Para el cáncer de mama uno de los tratamientos más comunes es la inhibición de ciclinas. Las ciclinas son una familia de proteínas que regulan el ciclo celular. Las ciclinas son reguladoras de las cinasas dependientes de ciclina (CDKs, por sus siglas en español), éstas permiten la progresión del ciclo celular cuando son activadas por las ciclinas (4). La inhibición de ciclinas busca interferir con la progresión anormal del ciclo celular de las células cancerígenas. De los tratamientos más aceptados para el tratamiento de cáncer de mama están palbociclib, ribiciclib y abemaciclib para ciertos tipos de cáncer, en muchas ocasiones para los ER positivos (3).

En el presente proyecto se busca hacer un análisis de expresión diferencial para buscar aquellos genes a los que afecta el tratamiento de abemaciclib en la línea celular MCF7. MCF7 es una línea celular usada comúnmente para el estudio de cáncer de mama. Esta línea celular expresa adenocarcinoma invasivo ductal de mujer de 69 años, es estrógeno positivo y HER2 negativo (5, 6). Del artículo "Multiomics Profiling Establishes the Polypharmacology of FDA-Approved CDK4/6 Inhibitors and the Potential for Differential Clinical Activity" obtenemos el dataset que queremos analizar. El dataset tiene el objetivo de hacer una comparación de tres inhibidores de CDK4/6 para cáncer de mama en etapas

avanzadas. Para esto se hace un perfil transcripcional entre siete líneas celulares de cáncer de mama después de 6 y 24 horas de exposición a concentraciones de 0.3 a 3.0 micromolares. Para tener una comparación menos general utilizamos los datos de la línea celular MCF7 con el tratamiento de abemaciclin en tres concentraciones; 0.3, 1.0 y 3.0 micromolar por 24 horas.

## Hipótesis

Esperamos ver que los genes que pertenecen a la vía de las CDKs tengan una expresión diferencial o aquellos que tengan un papel importante en el ciclo celular.

## Métodos

Los análisis se hicieron con RStudio, el código y workflow se puede ver en:

[https://github.com/dianadg159/proyecto\\_transcriptomica](https://github.com/dianadg159/proyecto_transcriptomica)

Para los datos se obtuvieron de recount3 en:

<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE99116>

El análisis de expresión diferencial se hizo con DAVID:

<https://david.ncifcrf.gov/tools.jsp>

## Resultados

Tomando solo en cuenta las muestras a las que les aplicaron abemaciclib por 24hrs en la línea celular MCF7, nos quedamos con doce muestras. Para visualizar las diferencias entre las muestras utilizamos un *multidimensional plot*. En la figura 1 podemos ver que las muestras más diferentes son aquellas con las que se trataron con la concentración más alta de abemaciclib. Para seguir visualizar mejor aquellas muestras que se separan más del resto realizamos un PCA (figura 2), donde vemos agrupadas a las muestras control y un poco más dispersas, pero cercanas las muestras tratadas con palbociclib y con ribociclib en todas sus concentraciones. Las muestras de abemaciclib de 0.3 micromolar se juntan con aquellas de palbociclib y ribocilcin a 3.0 micromolar. Las muestras que nos interesan más son las que se alejan, aquellas están tratadas con abemaciclib a 3.0 y 1.0 micromolar. Esto nos dice que abemaciclib a altas concentraciones tiene un efecto más fuerte que ribociclib y palbociclib en la línea celular MCF7 a 24h de exposición del tratamiento.

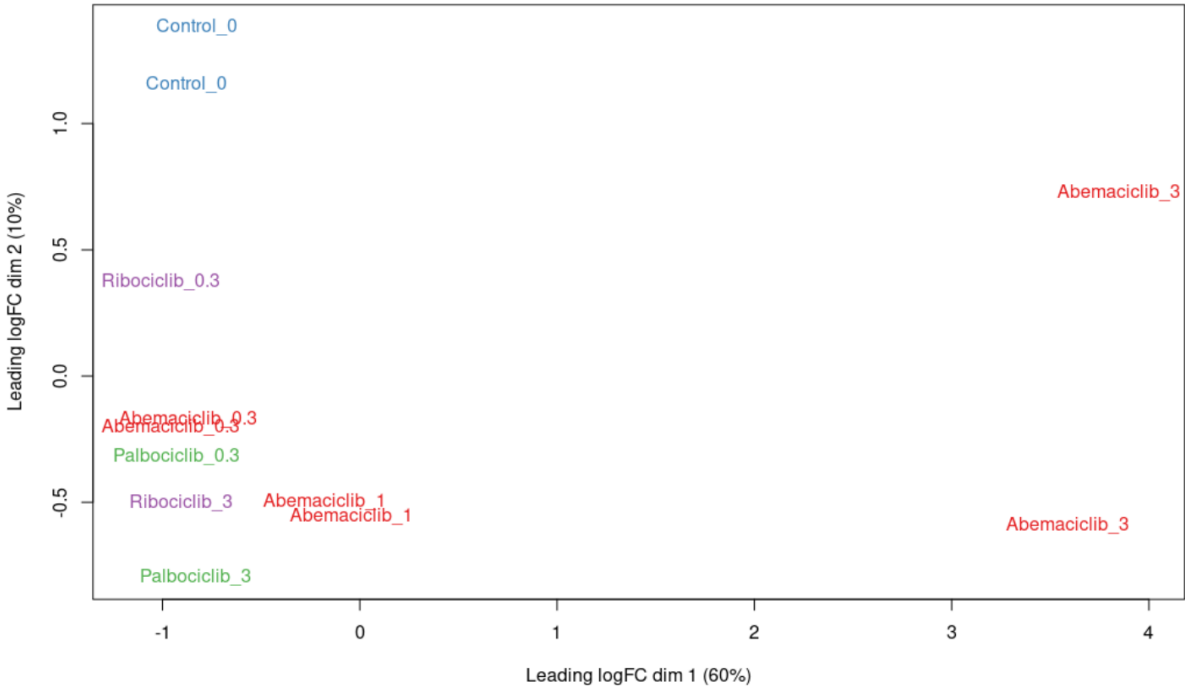


Figura 1. MDS plot

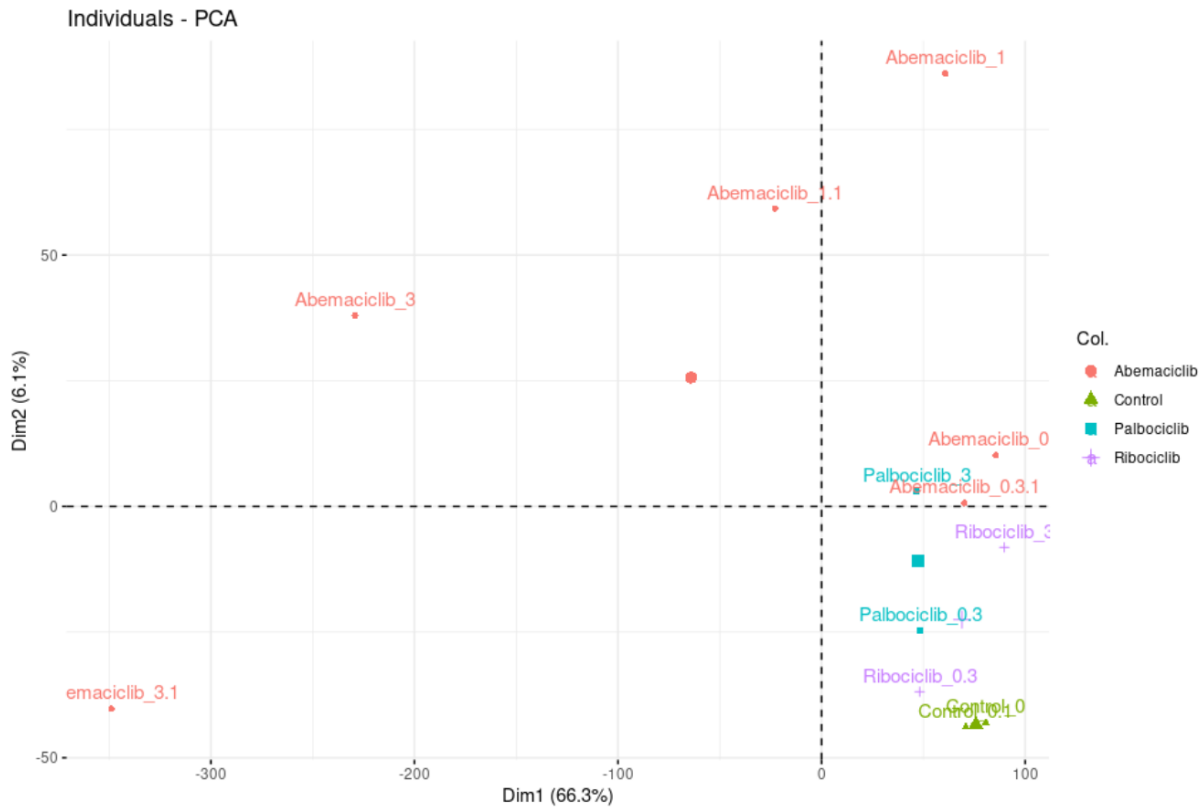


Figura 2. PCA

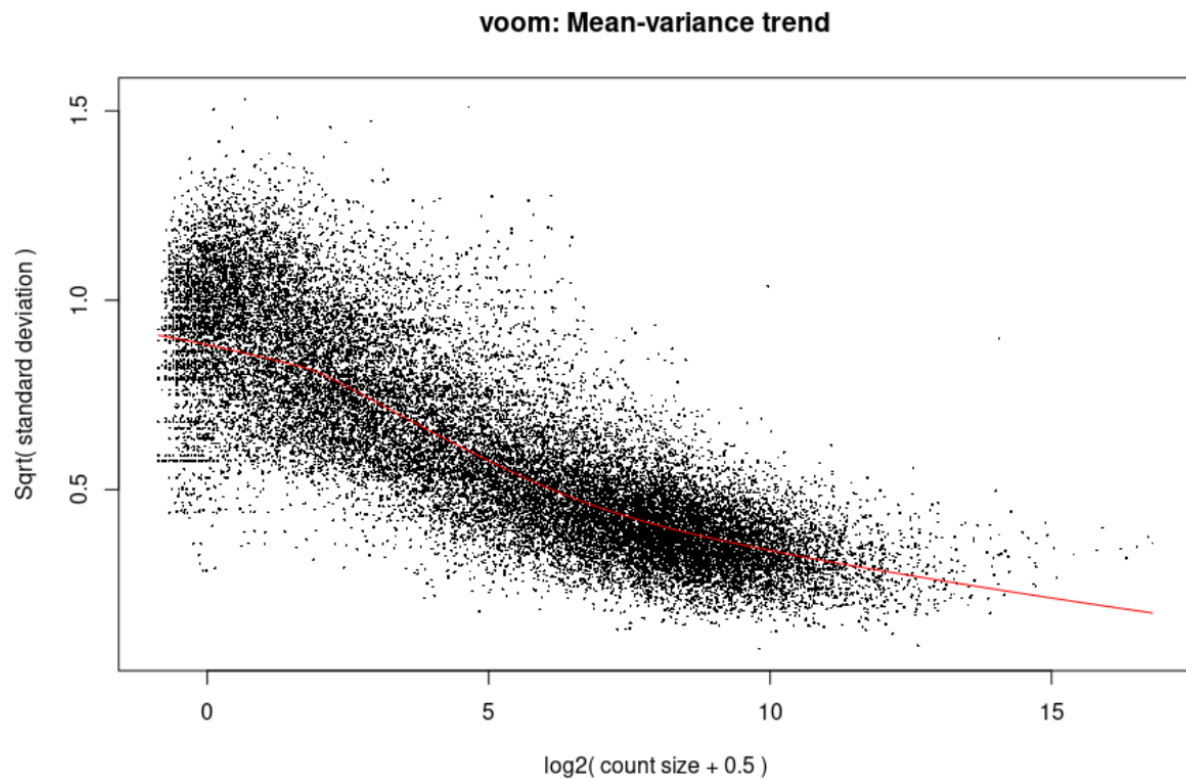


Figura 3. Modelo con voom.

Después de hacer una transformación de los datos a un modelo lineal obtenemos la figura 3. Este modelo nos quita la variación de las cuentas por gen en cada una de las muestras. Se modela la varianza promedio contra los genes y luego se interpola la varianza promedio para predecir las varianzas de las observaciones individuales. La figura tres sigue el modelo de las varianzas como se espera (8).

Una vez que los datos se ajustaron a un modelo lineal graficamos los genes estadísticamente más significativos y los graficamos con base a su  $p$ -valor  $< 0.05$  contra su cambio transformado al logaritmo (figura 4). En esta gráfica vemos los 20 genes diferencialmente expresados que pueden ser significativos para el análisis. De los veinte genes, aquellos que se encuentran en la parte superior derecha son aquellos que podrían ser los más significativos (MT-CO2, KRT8, KRT18, ACTB y ACTG1).

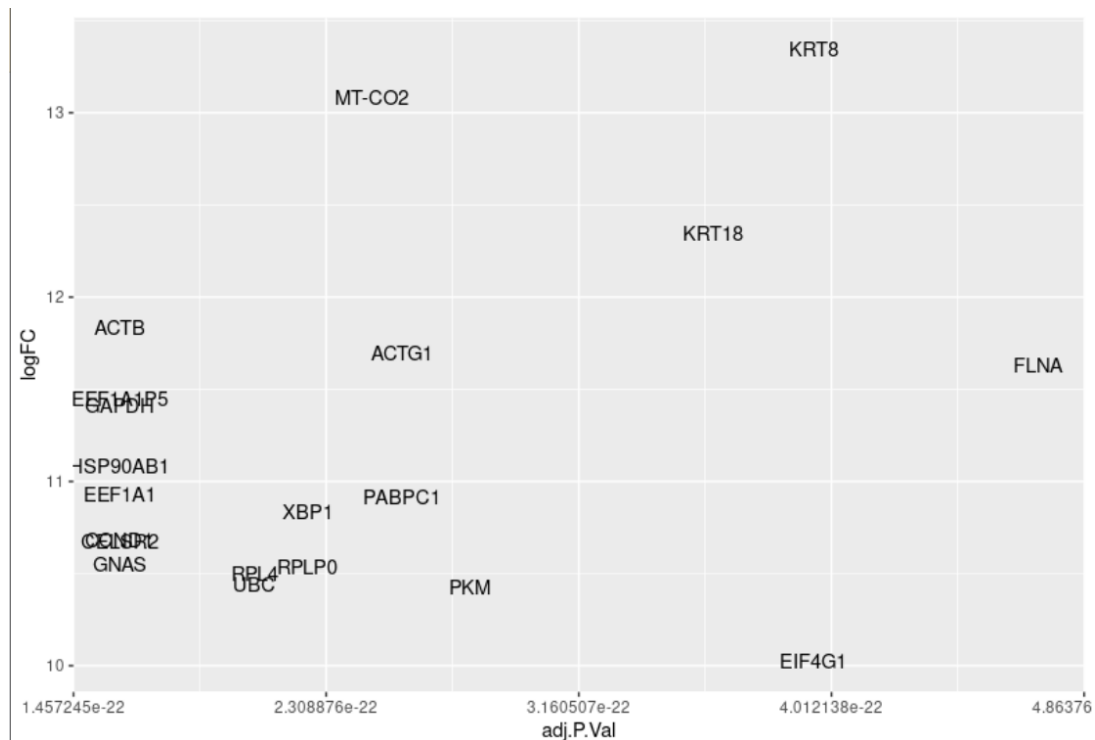


Figura 4. 20 DEGs

Para ver más gráficamente estos genes vemos el heatmap de la figura 5. En el heatmap aparecen todas las condiciones a las que se sometieron las doce muestras y los primeros 50 DEGs con  $p\text{-valor} < 6.364881e-22$ . Es importante ver el heatmap contra las demás condiciones para poder comparar los DEGs en cada una de las muestras y ver cuales son las condiciones que más afectaron a las líneas celulares. Podemos apreciar que, efectivamente como vimos en las figuras 1 y 2, las muestras con abemaciclib fueron aquellas que tuvieron un mayor grado de expresión diferencial que aquellas muestras tratadas con palbociclib o ribociclib. También evidenciamos que la concentración del tratamiento juega un papel muy importante al momento de medir cualitativamente la expresión diferencial. Vemos que KRT8 es de los genes más diferencialmente expresados en la línea celular MCF7 y que tiene un  $p\text{-valor}$  significativo, pero la expresión es similar a la del control. Genes como SENP3-EIF4A1, EIF4A1, RP11-371A22.1 y MT-RNR2 son algunos de los más diferencialmente expresados. Dentro del heatmap vemos a los 20 DEGs pero estos no necesariamente fueron los que tienen niveles de expresión más o menos alto, es decir que pudieron haber genes más diferencialmente expresados pero con un  $p\text{-valor}$  menos significativo pero aún así menor a 0.05.

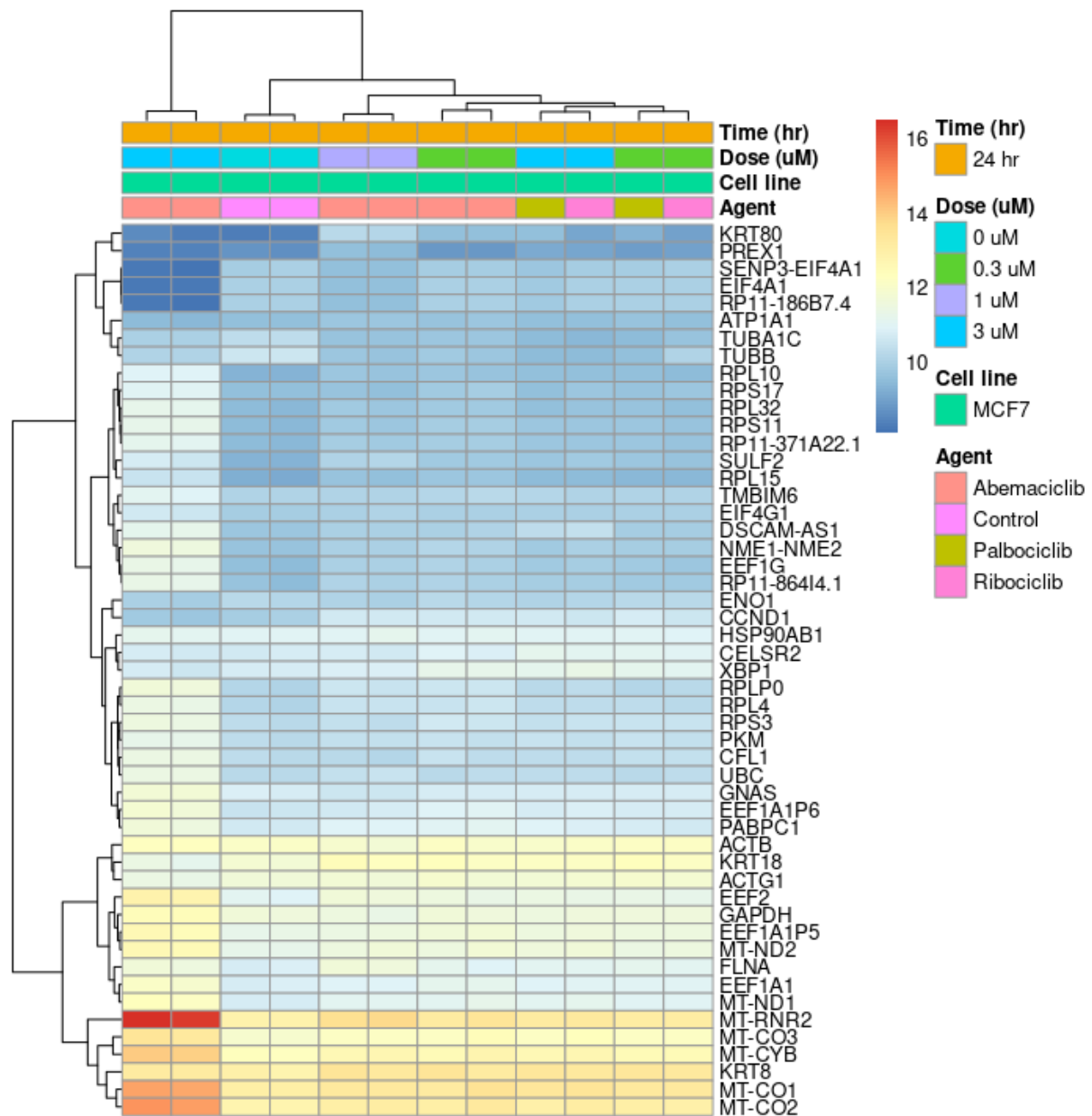


Figura 5. Heatmap de los 50 DEGs

Al hacer el análisis de Gene Ontology vemos que hay diversas funciones celulares que se vieron afectadas a lo largo del experimento. Se usaron los veinte genes diferencialmente expresados con un valor  $p$  menor a 0.05 más aquellos que se visualizaron en el heatmap. De estos genes las diez primeras anotaciones fueron las que se ven en la figura.

En la tabla apreciamos que la mayoría de los genes se ven envueltos en procesos de unión de RNA, unión de proteínas quinasas y unión de proteínas idénticas.

Al hacer el análisis por cluster vemos que tenemos 16 clusters con el que tiene una mayor *enrichment score* de 5.53. De este primer cluster las funciones biológicas tienen que ver con: metilación, unión isopeptídica, acetilación, conjugación Ubl, exoma extracelular, fosfoproteínas y unión de proteínas.

Como vemos en la figura 8, de los genes que participan en todas las funciones del cluster está KRT8, RPL4 y PKM.

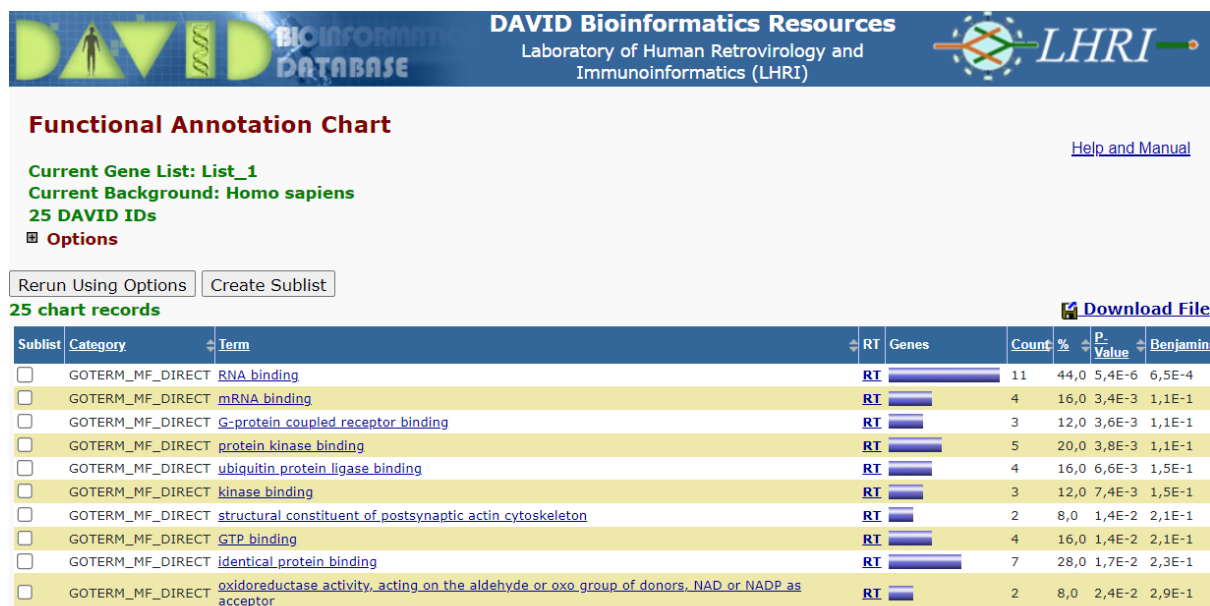


Figura 6. Tabla de anotaciones de GEO

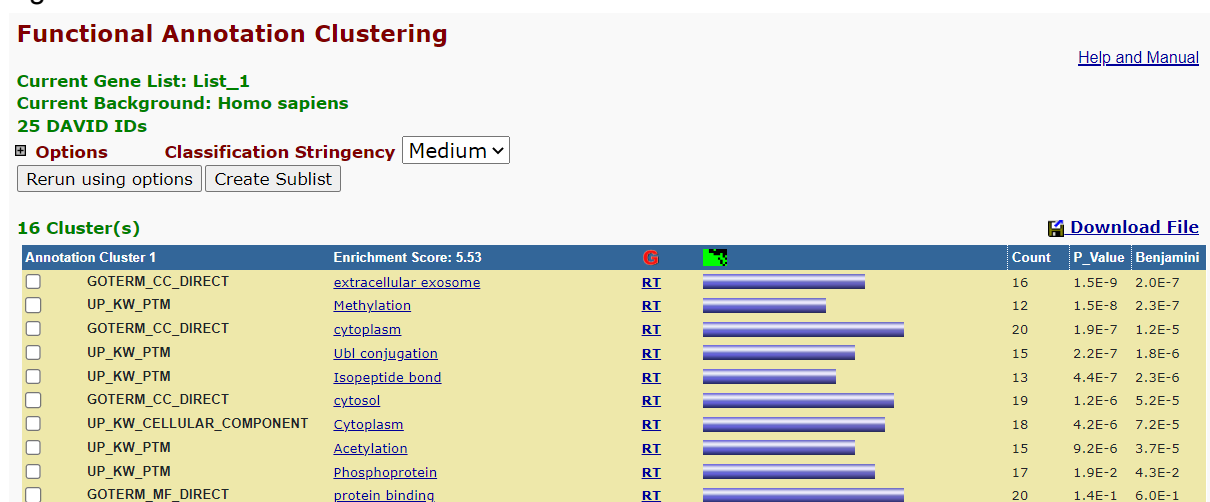


Figura 7. Tabla de Clustering de anotaciones

## Conclusiones

KRT8 (keratin 8) es un gen que pertenece al tipo I y II de keratinas y se heteropolimeriza para formar filamentos intermedios en el citoplasma de células epiteliales de una sola capa. Este gen juega un rol muy importante en mantener la integridad estructural de la transducción de señales en el ambiente celular. RPL4 es una proteína ribosomal que se encuentra en la subunidad 60s del ribosoma. La proteína ribosomal L4 (RPL4, por sus siglas en inglés) ha sido asociada a mantener la estabilidad nuclear. PKM (Piruvato quinasa M1/2) es un gen que codifica a una proteína de la glucólisis. Ésta ha sido demostrada que interactúa con la hormona tiroidea y puede afectar los efectos metabólicos de las hormonas tiroideas.

Podemos decir entonces que abemaciclib tiene actividad antitumoral, afectando funciones celulares importantes como la estabilidad citoplasmática y nuclear. En este reporte, el gen más afectados con el tratamiento de abemaciclib a 3 microMolar fue KRT8.

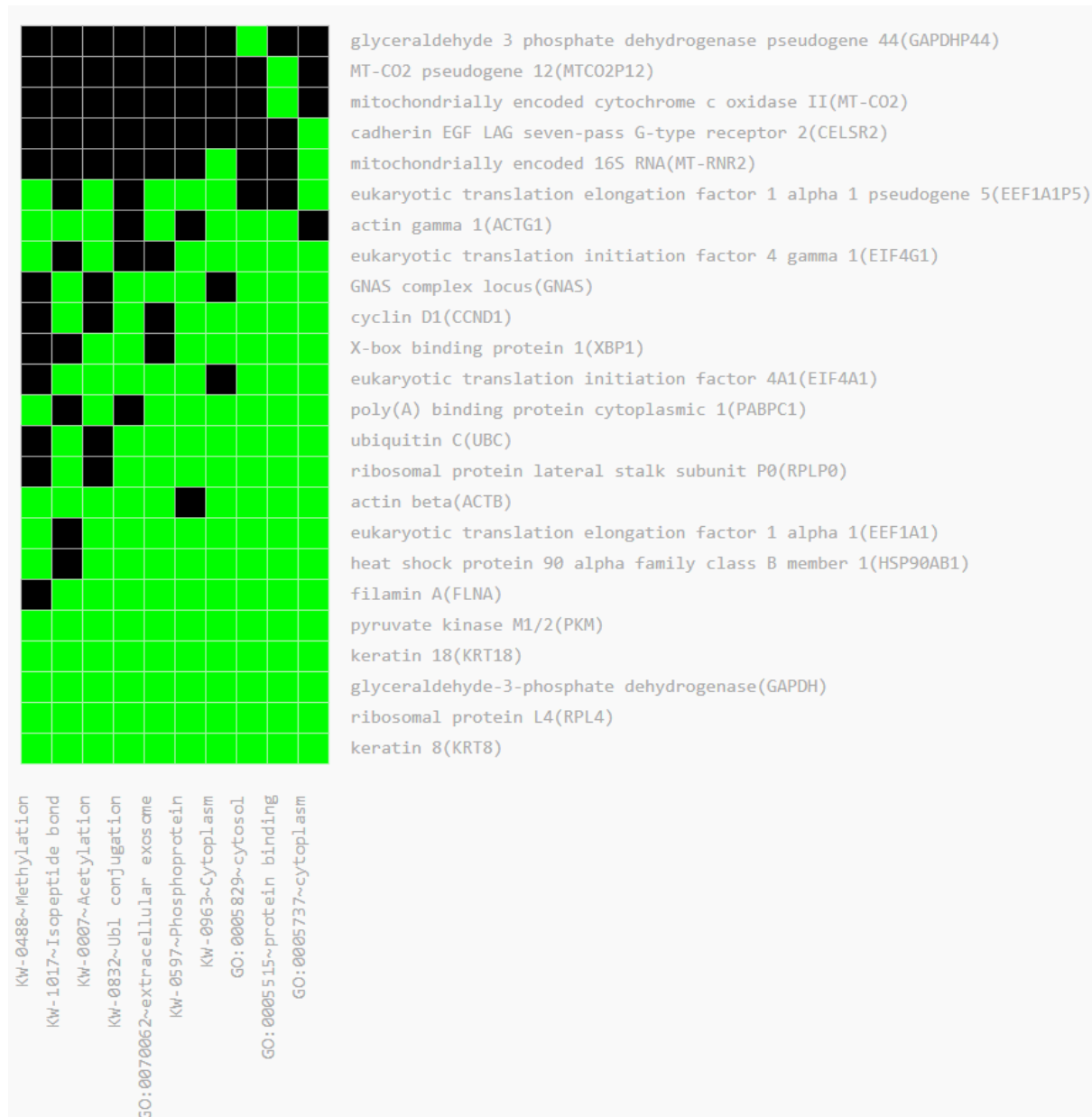


Figura 8. Primer cluster con *enrichment score* de 5.53

## Bibliografía

1. Gobierno de México, del. (2023). *Epidemiología del cáncer de mama*.  
 Gob.mx.  
<https://www.gob.mx/imss/articulos/epidemiologia-del-cancer-de-mama-318014>
2. Santaballa,A., (2023, January 24). *Cáncer de mama - SEOM: Sociedad Española de Oncología Médica* © 2019. Seom.org.



<https://seom.org/125-Informaci%C3%B3n%20al%20P%C3%ABlico%20-%20Patolog%C3%ADas/cancer-de-mama>

3. Poratti, M., & Marzaro, G. (2019). *Third-generation CDK inhibitors: A review on the synthesis and binding modes of Palbociclib, Ribociclib and Abemaciclib*. 172, 143–153. <https://doi.org/10.1016/j.ejmech.2019.03.064>
4. Spring, L., Wander, S. A., Lajos Pusztai, Moy, B., Turner, N. J., & Aditya Bardia. (2020). *Cyclin-dependent kinase 4 and 6 inhibitors for hormone receptor-positive breast cancer: past, present, and future*. 395(10226), 817–827. [https://doi.org/10.1016/s0140-6736\(20\)30165-3](https://doi.org/10.1016/s0140-6736(20)30165-3)
5. Kostya Ostrikov, Cheng, H., Bai, Z., & Li, J. (2017). *Breast Cancer Cell Line Classification and Its Relevance with Breast Tumor Subtyping*. 8(16), 3131–3141. <https://doi.org/10.7150/jca.18457>
6. ATCC. MCF7. <https://www.atcc.org/products/htb-22>. Recuperado el 27 de mayo del 2023.
7. Matthews, H. K., Bertoli, C., & Robertus. (2021). *Cell cycle control in cancer*. 23(1), 74–88. <https://doi.org/10.1038/s41580-021-00404-3>
8. Law, C. W., Chen, Y., Shi, W., & Smyth, G. K. (2014). *voom: precision weights unlock linear model analysis tools for RNA-seq read counts*. 15(2), R29–R29. <https://doi.org/10.1186/gb-2014-15-2-r29>
9. Hafner, M., Mills, C., Subramanian, K., Chen, C., Chung, M., Boswell, S. A., Everley, R. A., Liu, C., Walmsley, C. S., Juric, D., & Sorger, P. K. (2019). *Multomics Profiling Establishes the Polypharmacology of FDA-Approved CDK4/6 Inhibitors and the Potential for Differential Clinical Activity*. 26(8), 1067–1080.e8. <https://doi.org/10.1016/j.chembiol.2019.05.005>

10. Tolaney, S. M., Muralidhar Beeram, J. Thaddeus Beck, Conlin, A., E. Claire Dees, Puhalla, S., Rexer, B. N., Burris, H. A., Jhaveri, K., Helsten, T., Becerra, C., Kalinsky, K., Moore, K. N., Manuel, A. M., Lithio, A., Price, G. D., Chapman, S. C., Litchfield, L. M., & Matthew Bidwell Goetz. (2022). *Abemaciclib in Combination With Endocrine Therapy for Patients With Hormone Receptor-Positive, HER2-Negative Metastatic Breast Cancer: A Phase 1b Study*. 11. <https://doi.org/10.3389/fonc.2021.810023>