

PONTIFICIA UNIVERSIDAD CATÓLICA DEL ECUADOR

FACULTAD DE INGENIERÍA

MAESTRÍA EN BIOLOGÍA COMPUTACIONAL

TRABAJO DE TITULACIÓN

“Análisis de enriquecimiento de genes asociados a la heterogeneidad en Glioblastoma
Multiforme”

AUTORA: Susana Isabel Balvoa Caguana

TUTORA: Dra. Doris Vela

Quito, 2024

DEDICATORIA

A mis padres María Caguana y Manuel Balvoa, a mi abuela María Tránsito Santos y a mis hermanos, por todo lo que representan en mi vida.

AGRADECIMIENTO

Expreso mi infinito agradecimiento a Dios, a mis queridos padres y hermanos, por su amor y apoyo incondicional, en cada día. Gracias por ser mi fortaleza.

Un enorme agradecimiento a mi tutora Doris Vela, por su gran apoyo incondicional en el desarrollo de esta investigación.

Agradezco a mis profesores y compañeros de maestría, por las grandes enseñanzas y aprendizajes en este campo de la Biología Computacional.

ÍNDICE GENERAL

1. RESUMEN.....	9
2. Abstract	11
3. INTRODUCCIÓN	13
3.1. Justificación.....	13
3.2. Planteamiento del problema	14
3.3. Objetivos	15
3.3.1. Objetivo General	15
3.3.2. Objetivos Específicos	16
3.4. Alcance.....	16
3.5. Revisión de la literatura.....	16
3.5.1. Generalidades del Glioblastoma Multiforme (GBM)	16
3.5.2. Heterogeneidad Genómica y Subclases Moleculares en GBM.....	17
3.5.3. Generalidades y ventajas de ARN de Célula Única (scRNA-seq).....	18
3.5.4. Análisis de Expresión Diferencial (DE).....	20
3.5.5. Análisis de DE Machine Learning	20
3.5.6. Resultados del Estudio de Aelaga et al. (2023).....	21
3.5.7. Análisis Funcional.....	22
3.5.8. Metodologías de Análisis Funcional.....	23
3.5.9. Enriquecimiento de Genes	24
4. METODOLOGÍA	26
4.1. Definición de la lista de genes de interés	26
4.1.1. Selección la base de datos de anotaciones funcionales.....	27

4.1.2. Aplicación de una prueba estadístico de enriquecimiento	27
4.1.3. Interpretación de los términos enriquecidos significativos	27
5. RESULTADOS Y DISCUSIÓN.....	28
6. CONCLUSIONES	58
7. RECOMENDACIONES	59
8. BIBLIOGRAFÍA.....	60
9. ANEXOS.....	64

ÍNDICE DE FIGURAS

Figura 1. Grafo Dirigido Acíclico, Fuente: (López, et al.,2021).	25
Figura 2. Pipeline de Análisis de Enriquecimiento de Genes	28
Figura 4. Análisis de enriquecimiento en la base de datos KEGG, Escenario 1.....	34
Figura 5. Términos KEGG enriquecidos, Escenario 1.....	35
Figura 6. Rutas metabólicas KEGG, Escenario 1	37
Figura 7. Enriquecimiento de genes GO, Escenario 2.	39
Figura 8. Análisis de enriquecimiento en la base de datos KEGG, Escenario 2.....	42
Figura 9. Rutas metabólicas KEGG, Escenario 2.	44
Figura 10. Enriquecimiento de genes GO, Escenario 3	45
Figura 11. Términos KEGG enriquecidos, Escenario 3.....	48
Figura 12. Rutas metabólicas KEGG, Escenario 3.	50
Figura 13. Enriquecimiento de genes GO, Escenario 4	51
Figura 14. Análisis de enriquecimiento en la base de datos KEGG, Escenario 4.....	54
Figura 15. Términos KEGG enriquecidos, Escenario 4.....	55
Figura 16. Rutas metabólicas KEGG, Escenario 4.	57

INDICE DE TABLAS

Tabla 1. Parámetros de salida de un análisis de enriquecimiento.	29
Tabla 2. Listado de genes para el Escenario 1	30
Tabla 3. Listado de genes para el Escenario 2	39
Tabla 4. Listado de genes para el Escenario 3	45
Tabla 5. Listado de Genes para el Escenario 4.....	52

ÍNDICE DE ANEXOS

ANEXO 1: Genes relevantes obtenidas del agrupamiento del modelo LRST por Arteaga-Arteaga et al. 2023 para el escenario 1.	64
ANEXO 2: Dataframe de la salida del análisis de enriquecimiento KEGG y GO	65

1. RESUMEN

El Glioblastoma Multiforme (GBM) es un tumor cerebral agresivo y común en adultos, caracterizado por su heterogeneidad genómica y mal pronóstico, con una supervivencia media de aproximadamente un año tras el diagnóstico. La clasificación del GBM incluye variantes primarias y secundarias, diferenciadas por su origen y características moleculares. Los modelos Machine Learning han revolucionado el análisis de expresión diferencial en estos tumores sin embargo requieren de análisis complementarios y validaciones con estudios biológicos para la comprensión de funciones biológicas, procesos celulares y vías metabólicas implicadas. Este enfoque integral facilitaría la identificación de nuevos objetivos terapéuticos y la mejora de la precisión de las estrategias de tratamiento en la era de la Inteligencia artificial. En este estudio se realizó el análisis de enriquecimiento de genes para interpretar los mecanismos biológicos subyacentes a la heterogeneidad del GBM. Utilizando los resultados del análisis de expresión diferencial del estudio de Arteaga-Arteaga et al. (2023), para ello se determinó una lista de genes de interés que muestran expresión diferencial en GBM. Estos genes fueron analizados mediante técnicas de enriquecimiento funcional utilizando la paquetería gseapy en Python, en contraste con el conocimiento biológico previo a partir de las bases de datos de la Ontología Génica (GO) y la Enciclopedia de Genes y Genomas de Kioto (KEGG). Para el análisis de los resultados se extrajo los términos de mayor significancia estadística así también como relevancia biológica. Los análisis de enriquecimiento en los diferentes escenarios revelan un panorama complejo en el que la regulación del citoesqueleto, la homeostasis de metales, la respuesta al estrés y la señalización oncogénica que desempeñan roles críticos en la biología del glioblastoma multiforme. La recurrencia del EGFR en múltiples términos de GO sugiere que es un punto nodal en la regulación de diversas vías biológicas. Los inhibidores específicos del EGFR o de sus rutas de señalización asociadas podrían ser de gran interés terapéutico en el tratamiento de GBM. También se encontró que CALM1 regula el calcio y la transducción de señales hormonales. Estos genes no solo participan en vías de proliferación celular y oncogénesis, sino también en procesos de señalización hormonal, metabolismo, respuesta al estrés, e infección viral. Asimismo, se encontró el enriquecimiento de la ruta de la homeostasis de

minerales y el metabolismo del hierro. Estos resultados subrayan la importancia de desarrollar estrategias terapéuticas que puedan abordar estos procesos de manera integral para mejorar los resultados del tratamiento en pacientes con GBM.

Palabras clave: *Enriquecimiento de genes, Glioblastoma, análisis funcional, Ontología.*

2. ABSTRACT

Glioblastoma Multiforme (GBM) is an aggressive and common brain tumor in adults, characterized by its genomic heterogeneity and poor prognosis, with a median survival of approximately one year after diagnosis. GBM is classified into primary and secondary variants, differentiated by their origin and molecular characteristics. Machine Learning models have revolutionized the analysis of differential expression in these tumors; however, they require complementary analyses and validation with biological studies to better understand the biological functions, cellular processes, and metabolic pathways involved. This comprehensive approach would facilitate the identification of new therapeutic targets and improve the precision of treatment strategies in the era of Artificial Intelligence. In this study, gene enrichment analysis was performed to interpret the biological mechanisms underlying GBM heterogeneity. Using the differential expression analysis results from the study by Arteaga-Arteaga et al. (2023), a list of genes of interest that exhibit differential expression in GBM was identified. These genes were analyzed through functional enrichment techniques using the gseapy package in Python, in comparison with previous biological knowledge from the Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases. For the analysis of the results, the terms with the greatest statistical significance and biological relevance were extracted. Enrichment analyses across different scenarios reveal a complex landscape in which cytoskeletal regulation, metal homeostasis, stress response, and oncogenic signaling play critical roles in the biology of glioblastoma multiforme. The recurrence of EGFR in multiple GO terms suggests that it is a nodal point in the regulation of various biological pathways. Specific inhibitors of EGFR or its associated signaling pathways could be of great therapeutic interest in the treatment of GBM. CALM1 was also found to regulate calcium and hormonal signal transduction. These genes are involved not only in pathways related to cell proliferation and oncogenesis but also in processes such as hormonal signaling, metabolism, stress response, and viral infection. Furthermore, enrichment of mineral homeostasis and iron metabolism pathways was observed.

These results underscore the importance of developing therapeutic strategies that can comprehensively address these processes to improve treatment outcomes in patients with GBM.

Keywords: Gene enrichment, Glioblastoma, functional analysis, Ontology.

3. INTRODUCCIÓN

3.1. Justificación

Los tumores malignos que se originan en el cerebro, como los glioblastomas, son difíciles de tratar debido a la localización profunda en el cerebro, la capacidad de metastatizar rápidamente y el comportamiento agresivo, este hecho dificulta el pronóstico y a una supervivencia global reducida en los pacientes. En los últimos años, se han descubierto cada vez más marcadores moleculares relacionados con el pronóstico del glioma. Se puede entender mejor los mecanismos moleculares del glioma al descubrir estos marcadores, lo que luego puede ayudar en el diagnóstico y tratamiento clínico. Debido a que un solo índice no puede predecir con precisión el pronóstico del tumor, el análisis combinado de múltiples índices es cada vez más importante para mejorar la precisión de la predicción del pronóstico (Wang et al 2024).

El tratamiento estándar para el glioblastoma, que incluye temozolomida seguida de radioterapia, resulta en una corta supervivencia del paciente, con una media de solo 10-16 meses (Sarma y Kumar, 2023). Esta limitada eficacia terapéutica subraya la necesidad urgente de identificar genes asociados a la heterogeneidad del glioblastoma que puedan guiar estrategias de tratamiento más eficaces y personalizadas. Además, la Organización Mundial de la Salud (OMS) ha reclasificado los gliomas en 2016, incorporando mutaciones del isocitrato deshidrogenasa (IDH1/2) y el estado de codeleción 1p/19q, lo que resalta la importancia de las características moleculares en el pronóstico y tratamiento del GBM (Gray et al., 2022).

Sin embargo, más del 90% de los GBM primarios son de tipo salvaje para IDH1/2, lo que indica una clara necesidad de comprender mejor las vías de señalización molecular que conducen a la progresión del GBM para desarrollar estrategias terapéuticas integrales (Gray et al., 2022). La heterogeneidad molecular del GBM, con diferencias significativas en la expresión génica y las características celulares entre los tumores, sugiere que un enfoque basado en biomarcadores podría mejorar la estratificación de los pacientes y el pronóstico del tratamiento.

Por lo tanto, surge la necesidad de identificar genes de interés y comprender las vías de señalización que contribuyen a la progresión del GBM. Esto podría ser de aporte en la personalización de los tratamientos. También tiene el potencial de aumentar la eficacia de las terapias y prolongar la supervivencia de los pacientes.

3.2. Planteamiento del problema

El análisis de ARN de células individuales (single-cell RNA sequencing) permite la identificación de subpoblaciones de células dentro del tumor, revelando heterogeneidad intratumoral. Los modelos de aprendizaje automático (ML) aprenden a distinguir entre diferentes tipos de tejido, identificando la base sobre la comprensión detallada de la biología del tumor, su comportamiento y como actúa con el tejido cerebral sano. Los modelos ML se utilizan cada vez más para el análisis de expresión diferencial en la investigación genómica. Se han desarrollado diversos métodos basados en ML para detectar genes diferencialmente expresados (DEGs) en conjuntos de datos biológicos. Los modelos se han propuesto para evaluar patrones de interacción de la expresión diferencial, descubriendo patrones biológicos sutiles en fuentes de datos ómicos. Si bien los enfoques de ML muestran promesas en el análisis de expresión diferencial requieren de análisis complementarios con estudios biológicos para la comprensión de funciones biológicas, procesos celulares y vías metabólicas lo que facilita la identificación de nuevos objetivos terapéuticos y mejora la precisión de las estrategias de tratamiento (Tulika et al., 2019; Mahin et al 2022; Arteaga-Arteaga et al., 2023;). Además, la correcta identificación de la biología del tumor puede mejorar el diagnóstico y la planificación de intervenciones quirúrgicas (Arteaga-Arteaga et al., 2023).

Los tumores malignos que se originan en el cerebro son difíciles de tratar debido a su localización de origen, ya que están profundamente incrustados en el cerebro, a su capacidad de metastatizar rápidamente y a su comportamiento agresivo, lo que conlleva un mal pronóstico en los pacientes y una disminución de la supervivencia global. Los glioblastomas (GBM) y los epéndimos son tumores de grado IV que se originan en las células gliales y son los tumores más agresivos del sistema nervioso central (SNC) (Sarma y Kumar, 2023).

La terapia estándar en el glioblastoma sigue siendo la temozolomida seguida de radioterapia, que provoca una corta supervivencia y un mal pronóstico en los pacientes. La supervivencia media de los pacientes con glioblastoma es de sólo 10-16 meses tras el tratamiento (Sarma y Kumar, 2023).

El Glioblastoma según las directrices de la Organización Mundial de la Salud OMS se basa en necrosis y proliferación microvascular (Gray et al. 2022). Actualmente, el GBM se trata según el protocolo Stupp, que incluye la resección máxima segura seguida de radioterapia más quimioterapia concomitante y adyuvante con temozolomida. Las diferencias en los tiempos de supervivencia indican que un subconjunto de pacientes puede tener características moleculares que produzcan resultados más favorables. Esta heterogeneidad subyacente justifica el uso de biomarcadores moleculares para estratificar mejor a los pacientes para el tratamiento y determinar el pronóstico. La Organización Mundial de la Salud (OMS) ha reclasificado el GBM y el glioma de bajo grado en 2016 para incluir las mutaciones del isocitrato deshidrogenasa (IDH1/2) y el estado de codeleción 1p/19q (8). Por lo tanto, el estado de mutación IDH1/2 se ha utilizado de forma rutinaria en la clínica para predecir el pronóstico tumoral y guiar las estrategias de tratamiento de los pacientes con glioma. Sin embargo, más del 90% de los GBM primarios son IDH1/2 de tipo salvaje, por lo que es extremadamente necesario comprender claramente las vías de señalización molecular que conducen a la progresión del GBM para desarrollar estrategias terapéuticas integrales para el GBM (Gray et al. 2022).

3.3. Objetivos

3.3.1. Objetivo General

- Realizar un análisis de enriquecimiento de genes para interpretar los mecanismos biológicos asociados a la heterogeneidad en glioblastoma multiforme.

3.3.2. *Objetivos Específicos*

- Extraer y definir una lista de genes de interés que muestran expresión diferencial en glioblastoma, utilizando los resultados del estudio de Arteaga-Arteaga et al. (2023).
- Implementar pruebas estadísticas para evaluar el enriquecimiento de términos funcionales en la lista de genes de interés.
- Interpretar los términos funcionales enriquecidos y su relevancia biológica en el glioblastoma.

3.4. Alcance

Una de las principales fortalezas del estudio de Arteaga-Arteaga et al. (2023) es identificar con mayor precisión genes en cada uno de los tipos celulares que componen el GBM. Aunque los resultados de un solo estudio pueden no ser representativos de todas las variaciones posibles en los datos de GBM, sin embargo, la interpretación de los términos enriquecidos significativos puede ser un aporte adicional que complemente y corrobore los hallazgos basados en IA. La significancia estadística no siempre se traduce en relevancia biológica, y los resultados deben ser cuidadosamente contextualizados en el marco del GBM. Además, la complejidad de las vías metabólicas y los procesos biológicos puede dificultar la identificación de relaciones causales directas.

3.5. Revisión de la literatura

3.5.1. *Generalidades del Glioblastoma Multiforme (GBM)*

El GBM es un tumor cerebral agresivo de rápido crecimiento y presenta limitaciones en el diagnóstico y tratamiento debido a su heterogeneidad ya que presenta diversas características celulares y resistencia a las terapias convencionales como cirugía, radiación y química terapia. Representa aproximadamente el 80% de todos los gliomas cerebrales primarios y el 60% de todos los tumores cerebrales adultos (El Atat et al., 2022). La patogénesis del GBM involucra alteraciones complejas en la genética, epigenética y

transcriptoma que conducen a cambios significativos en las principales vías de señalización.

En las últimas décadas las tecnologías de secuenciación de última generación han permitido el análisis y clasificación de las huellas genéticas y moleculares del GBM. De acuerdo con el Atlas del Genoma del Cáncer, se define las siguientes subclases:

- Neuronal que representa el 16% del GBM y se caracteriza por la expresión de marcadores neuronales como NEFL, GABRA1, SYT1 y SLC1A5.
- Proneuronal con una alteración de PGFRA, una mutación puntual IDH1 y TP3 y una sobreexpresión en genes de desarrollo como NKX2-2, OLIG2. y una sobreexpresión en genes de desarrollo como NKX2-2, OLIG2.
- Mesenquimal que se distingue con alteraciones en la neurofibromatosis tipo 1 (NF1), mutación puntual del homólogo de fosfatasa y tensina (PTEN) y expresión de genes mesenquimales como MET, CD44. Subtipo clásico que muestra amplificación de EGFR, delección de CDKN2 A y mutaciones de p53.

El GBM se puede dividir en dos clases principales: GBM primario y GBM secundario, siendo la mayoría primarios. El GBM primario surge de *novo* sin un precursor clínico conocido, y la mayoría se presenta en adultos mayores (mayores de 50 años), mientras que el GBM secundario es el resultado de la progresión de un grado de malignidad menor preexistente y generalmente afecta a pacientes más jóvenes. El GBM primario y secundario son morfológicamente indistinguibles y responden de manera similar a la terapia convencional, pero tienen diferentes características moleculares y, por lo tanto, pueden responder de manera diferente a las terapias moleculares dirigidas.

3.5.2. Heterogeneidad Genómica y Subclases Moleculares en GBM

En un estudio de perfiles de expresión de microarrays GSE90604, GSE50601 y GSE134470 de glioblastoma revelaron genes diferencialmente expresados, como SPARC (proteína ácida y rica en cisteína secretada) y VIM (vimentina). Además, con el análisis de enriquecimiento lograron explicar cómo se detienen los puntos de control del ciclo celular y como se activan vías de señalización, lo que permitió la identificación de genes que podrían servir como biomarcadores predictivos de glioblastoma (Li et al., 2022).

Asimismo, Huawei et al. (2022) sugirieron que el GBM presentaba una sobreexpresión de múltiples vías relacionadas con la integridad del genoma y la infiltración de células inmunitarias. Además, identificaron genes con variantes de un solo nucleótido (SNV) o variantes en el número de copias (CNV) que podrían servir como marcadores potenciales para el pronóstico clínico.

El análisis de enriquecimiento funcional genes u ontología génica (GO) y las rutas metabólicas se utilizan para identificar y entender mejor las funciones biológicas, procesos celulares y funciones moleculares en las que están involucrados un conjunto específico de genes o proteínas (Wang et al., 2024). Darrak et al. (2023) utilizando enfoques canónicos y ML identificación de biomarcadores genómicos para el glioblastoma y sus interacciones, encontraron 42 genes enriquecidos en 7 vías como las proteínas ribosómicas citoplasmáticas, factores de traducción, cadena de transporte de electrones, ribosoma, enfermedad de Huntington, vías de inmunodeficiencia primaria y la vía de señalización del interferón tipo I que albergan tumores cuando están alteradas.

3.5.3. Generalidades y ventajas de ARN de Célula Única (scRNA-seq)

El ARN de Célula Única (scRNA-seq, por sus siglas en inglés) ha revolucionado la biología molecular al proporcionar una resolución sin precedentes en la comprensión de la expresión génica a nivel individual de células. A diferencia de las técnicas tradicionales de secuenciación de ARN (RNA-seq), que promedian la expresión génica en una población celular heterogénea, scRNA-seq permite el análisis detallado de la expresión génica en células individuales. Este método ofrece una visión más precisa y detallada de los mecanismos moleculares y la heterogeneidad celular dentro de los tejidos. El scRNA-seq permite el perfilado de miles de células individuales, revelando patrones de expresión génica diversos dentro de los tejidos.

El proceso de scRNA-seq comienza con la captura y el aislamiento de células individuales, lo que se puede lograr mediante diversas técnicas como la micromanipulación, la microfluídica, o el uso de gotas. Una vez aisladas, se extrae el ARN de cada célula y se convierte en ADN complementario (cDNA) a través de la transcripción inversa. Posteriormente, el cDNA se amplifica para generar cantidades suficientes de

material genético para la secuenciación. Finalmente, se secuencian el cDNA y los datos obtenidos se analizan para reconstruir los perfiles de expresión génica de cada célula individual.

Los primeros estudios en scRNA-seq identificaron la diversidad celular dentro de tejidos complejos. En enfermedades autoinmunes como el trastorno del espectro de neuromielitis óptica (NMOSD) ha revelado la regulación inmunitaria en cada tipo de célula, también evidenciaron el aumento en las células plasmáticas y las células T CD8+, así como una mayor expresión de marcadores inflamatorios (Fan et al., 2024).

Una de las mayores ventajas del scRNA-seq es su capacidad para resolver la heterogeneidad celular dentro de un tejido. En muchos casos, los tejidos están compuestos de múltiples subtipos celulares que pueden desempeñar funciones distintas o estar en diferentes estados fisiológicos o patológicos. La capacidad de analizar estas diferencias a nivel individual es crucial para entender procesos como la diferenciación celular, la respuesta inmune, y la progresión tumoral. En estudios de cáncer, el scRNA-seq ha permitido identificar subpoblaciones de células tumorales que pueden estar asociadas con resistencia a terapias o con un pronóstico desfavorable.

La técnica de scRNA-seq ha facilitado la identificación de nuevos tipos y estados celulares que no se habían caracterizado previamente. Esto es particularmente importante en tejidos complejos, como el cerebro, donde la diversidad celular es alta. La capacidad de identificar y caracterizar estas nuevas poblaciones celulares abre la puerta a nuevas áreas de investigación y a un mejor entendimiento de la biología del desarrollo y las enfermedades.

Además, permite el análisis de la expresión génica dinámica permite el estudio de cambios dinámicos en la expresión génica durante procesos biológicos como el desarrollo embrionario, la diferenciación celular, y la respuesta al estrés. Al poder analizar células individuales en diferentes puntos del tiempo, es posible construir trayectorias de diferenciación o de respuesta celular, lo que proporciona información sobre los mecanismos reguladores que controlan estos procesos.

3.5.4. Análisis de Expresión Diferencial (DE)

En la genómica funcional, las funciones de los genes se identifican utilizando tecnologías de alto rendimiento como los enfoques de microarrays y secuenciación de próxima generación (NGS). La bioinformática funcional decodifica cómo los genomas, proteomas y metabolomas dan lugar a diferentes fenotipos celulares y analiza las diferencias en cómo el mismo genoma funciona de manera diferente en diversos tipos de células y cómo los cambios en los genomas alteran las funciones celulares y moleculares a través de la expresión diferencial de transcripciones o genes (DEG), que a su vez regulan la expresión de proteínas y metabolitos en las células. En la bioinformática funcional se utilizan diversas herramientas computacionales para descifrar información biológica compleja en diversos conjuntos de datos y generar una comprensión biológica precisa e hipótesis sobre las funciones de los genes, la expresión de proteínas, las interacciones y las regulaciones tanto en la salud como en la enfermedad (Cremona et al., 2019).

El análisis de expresión diferencial es un aspecto crucial de los estudios de transcriptómica de células individuales y transcriptómica espacial, con el objetivo de identificar genes que se expresan diferencialmente en distintas condiciones o tipos de células (Wu et al., 2024). Esto permite la comprensión de la heterogeneidad celular y los procesos regulatorios.

3.5.5. Análisis de DE Machine Learning

El análisis de diferenciación de expresión (DE, por sus siglas en inglés) es un enfoque que permite identificar genes cuya expresión difiere significativamente entre diferentes condiciones biológicas o grupos de muestras. Tradicionalmente, el análisis DE se ha realizado utilizando técnicas estadísticas como el test t de Student, ANOVA, o modelos lineales generalizados. Sin embargo, con el auge de grandes volúmenes de datos y la complejidad inherente de los sistemas biológicos, el machine learning (ML) ha emergido como una herramienta poderosa para mejorar el análisis de DE, permitiendo un enfoque más robusto y predictivo (Cita).

El uso de ML en el análisis DE implica la aplicación de algoritmos de aprendizaje supervisado o no supervisado para clasificar y detectar genes diferencialmente expresados. En un enfoque supervisado, el modelo se entrena con un conjunto de datos

etiquetado donde las muestras se clasifican según su condición biológica (por ejemplo, tejido sano versus tejido tumoral). Los algoritmos de ML, como los árboles de decisión, las máquinas de soporte vectorial (SVM), y las redes neuronales, se utilizan para construir un modelo que pueda identificar patrones complejos en la expresión génica que distinguen estas condiciones (Arteaga-Arteaga et al., 2022).

Los modelos de aprendizaje automático ofrecen un enfoque integral para captar las complejidades de los datos de expresión génica superando a los métodos existentes en la estimación de cambios en la expresión génica e identificación de genes diferencialmente expresados en datos de scRNA-seq. Además, la integración de enfoques de bioinformática y aprendizaje automático ha permitido la identificación de genes y vías comunes entre enfermedades (Boyeau et al., 2023).

Una de las principales ventajas de utilizar ML en el análisis DE es su capacidad para capturar relaciones no lineales entre los genes y las condiciones biológicas. Esto es particularmente relevante en estudios transcriptómicos, donde las interacciones genéticas y la regulación de la expresión génica pueden ser altamente no lineales y complejas. Los algoritmos de ML pueden modelar estas relaciones de manera más efectiva que los métodos estadísticos tradicionales.

ML no solo identifica genes diferencialmente expresados, sino que también puede construir modelos predictivos que generalicen bien en nuevos conjuntos de datos. Esto es particularmente útil en el desarrollo de biomarcadores diagnósticos o pronósticos basados en perfiles de expresión génica. La capacidad de estos modelos para generalizar se puede evaluar utilizando técnicas como la validación cruzada o los conjuntos de datos independientes.

3.5.6. Resultados del Estudio de Arteaga-Arteaga et al. (2023).

Arteaga-Arteaga et al. 2023 utilizaron el aprendizaje automático en un conjunto de datos de GBM mediante secuenciación de ARN de célula única (scRNA-seq), con el objetivo de determinar genes que muestren diferencias significativas en su expresión entre células normales y neoplásicas de GBM. Identificaron marcadores para identificar características

específicas de las células cancerosas que podrían ser explotadas en la investigación y el desarrollo de nuevas terapias.

La fuente de los conjuntos de datos fue del Laboratorio de Investigación de Tumores Cerebrales de Gephart en Neurocirugía de Stanford. El conjunto de datos se dividió en un total de cuatro escenarios de clasificación. La división se realizó teniendo en cuenta los grupos a clasificar, que son Núcleo del Tumor (TC), Periferia del Tumor (TP) y Periferia Normal (NP). Se presentan conjuntos de datos con tres escenarios de clasificación binaria y uno de clasificación múltiple: (i)TPversusNP, (ii)TPversusTC, (iii)NPversusTPandTC(TPC) y (iv)NPversusTPversusTC.

Los modelos de aprendizaje automático (ML) lograron altas precisiones en la clasificación, con más del 97% de precisión en el escenario 1 (TP vs. TC), más del 96% en el escenario 2 (TP vs. NP), más del 99% en el escenario 3 (NP vs. TP y TC) y más del 97% en el escenario 4 (NP vs. TP vs. TC). Además Identificaron candidatos a biomarcadores relevantes para el estudio del GBM, incluyendo genes como ATP1A2, SPARCL1, FTL, EGFR, SPOCK1, ANXA1, APOD y TMSB4X, que podrían ser importantes para el desarrollo de tratamientos.

3.5.7. Análisis Funcional

El análisis funcional en biología implica descifrar las funciones de genes y proteínas utilizando tecnologías de alto rendimiento como los microarrays y la secuenciación de nueva generación, tal como se destaca en el contexto de la bioinformática funcional. Este análisis tiene como objetivo comprender cómo los genomas, proteomas y metabolomas contribuyen a los fenotipos celulares, revelando diferencias en la función génica a lo largo de diversos tipos de células y condiciones. Técnicas como los estudios de expresión génica, la edición del genoma y las herramientas computacionales desempeñan roles cruciales en la caracterización de genes, la predicción de funciones y el estudio de interacciones proteicas, tal como se discute en el contexto del análisis de la función génica. Además, la identificación de marcadores funcionales en la superficie celular para las células T reguladoras infiltrantes de tumores (TI-Tregs) mediante enfoques de biología de sistemas demuestra la importancia del análisis funcional en la identificación de objetivos

terapéuticos y biomarcadores para la inmunoterapia contra el cáncer. Adicionalmente, los avances en métodos biotecnológicos han mejorado la extracción y caracterización de metabolitos secundarios de plantas medicinales, mostrando la aplicación del análisis funcional en la validación de reclamos medicinales tradicionales (After et al. 2008; Li et al., 2022).

La bioinformática funcional es una parte de la biología computacional que utiliza la enorme cantidad de datos brutos derivados de la genómica, transcriptómica, proteómica, glicómica, lipidómica, metabolómica y otros experimentos "ómicos" a gran escala para decodificar las complejas funciones e interacciones de genes y proteínas en la salud y la enfermedad (Cremona et al., 2019).

3.5.8. Metodologías de Análisis Funcional

Una de las metodologías clave empleadas en estos estudios es el análisis de enriquecimiento de conjuntos de genes, que evalúa si ciertos conjuntos de genes definidos previamente están sobrerrepresentados en un grupo de células en comparación con otro. Este análisis se utiliza para identificar rutas metabólicas y procesos biológicos que están activos en diferentes estados celulares (Wang et al., 2024).

Las herramientas de enriquecimiento funcional ampliamente utilizadas se pueden clasificar en dos principales categorías: i) Análisis de Sobre-Representación (ASR), ii) Puntuación de Clases Funcionales (PFC) (Wijesooriya et al., 2021).

En el análisis de sobrerrepresentación los genes diferencialmente expresados (DEGs) que cumplen con un umbral de significancia y/o cambio de pliegue son consultados contra vías curadas (conjuntos de genes). Se realiza una prueba estadística para determinar si el número de DEGs que pertenecen a un conjunto de genes particular es mayor que el esperado por casualidad, comparado con una lista de genes de fondo. Estas herramientas pueden ser paquetes de software independientes o servicios web, y utilizan una o más pruebas estadísticas (por ejemplo, la prueba exacta de Fisher, la prueba hipergeométrica) (Wijesooriya et al., 2021).

La PFC implica asignar a cada gen detectado una puntuación de expresión diferencial y luego evaluar si las puntuaciones son más positivas o negativas de lo esperado por

casualidad para cada conjunto de genes. La popular herramienta de análisis de enriquecimiento de conjuntos de genes (GSEA) utiliza enfoques de permutación para establecer si un conjunto de genes está significativamente asociado con puntuaciones más altas o bajas, ya sea permutando las etiquetas de las muestras o permutando genes en el perfil de expresión diferencial (Wijesooriya et al., 2021).

3.5.9. Enriquecimiento de Genes

El análisis de enriquecimiento permite validar las técnicas de agrupamientos de genes, con un conocimiento biológico previo, extraído de las bases de datos biológicas disponibles en la web (Fernandez et al., 2014).

Este método bioinformático identifica vías biológicas o funciones que están significativamente enriquecidas en un conjunto dado de genes o regiones genómicas en comparación con lo que se esperaría por azar. Esta técnica implica comparar la lista de genes o regiones con bases de datos anotadas como Gene Ontology (GO) o Kyoto Encyclopedia of Genes and Genomes (KEGG) para determinar las vías más sobrerrepresentadas utilizando métodos estadísticos como la prueba exacta de Fisher (Chicco y Jurman, 2022). GO y KEGG son las herramientas más utilizadas en la anotación funcional de genes (Viera et al., 2010).

Los términos en GO se encuentran categorizados de la siguiente manera:

1. Funciones Moleculares. - realizan una descripción detallada de las actividades moleculares, sin detallar el contexto específico de la acción. Estas acciones a menudo son muy generales como la actividad catalítica, actividad del transportador o binding o específicas como actividad de la adenilatociclasa (Rodríguez, s/f).
2. Procesos biológicos. – Se refieren a un conjunto particular de Funciones Moleculares, que están altamente reguladas y siguen una secuencia temporal particular. Por ejemplo, la apoptosis, transducción de señales u otros ejemplos específicos son los procesos metabólicos de la pirimidina y el transporte de alfa-glucósidos (Rodríguez, s/f).

3. Componentes celulares. – Se detalla la componente de una célula que forma parte de una estructura mayor, como la estructura anatómica (núcleo o el retículo endoplasmático rugoso) o un conjunto de productos génicos (ribosoma, proteasoma, o un dímero proteico). Ejemplos de componentes celulares incluyen términos como parte citoplasmática de la membrana plasmática, mitocondria o ribosoma. A diferencia de las otras dos categorías de GO, los conceptos de componentes celulares se enfocan en la anatomía celular y no en los procesos (Rodríguez, s/f).

En general GO tiene la información de que genes que participan en las funciones moleculares, procesos biológicos, y componentes celulares. Un gen en particular puede participar en más de un término y categoría. Cada categoría está organizada como un grafo dirigido acíclico (árbol), donde cada termino tiene relaciones con uno o más términos del mismo dominio. Cada nodo del grafo tienes genes asociados que están involucrados en ese término. Los grafos se encuentran esquematizados jerárquicamente, y presentan conceptos biológicos más específicos a mayor profundidad, lo que implica una disminución en la cantidad de genes asociados a casa término. De este modo, un gen en un nodo también está presente en todos los nodos de sus ancestros como se ejemplifica en la Figura 1. (Rodríguez, s/f; López, et al.,2021).

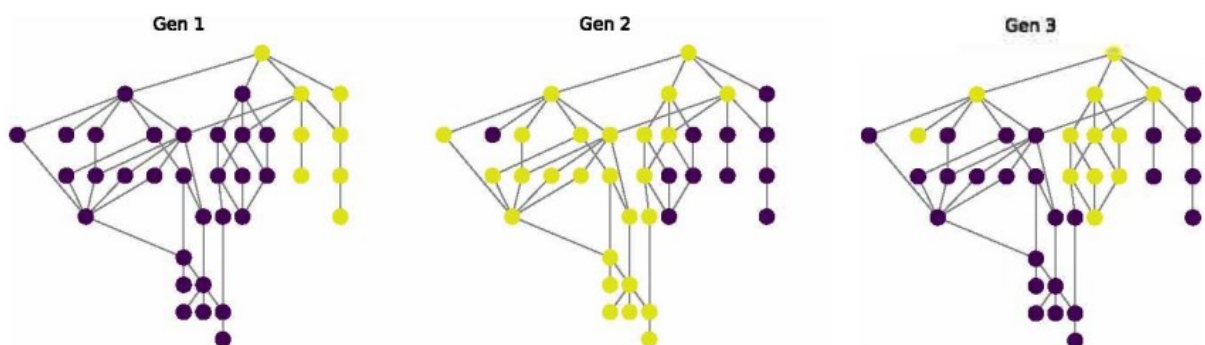


Figura. 1. Grafo Dirigido Acíclico, Fuente: (López, et al.,2021).

La ontología KEGG, compara los genomas en cuanto a su capacidad para codificar diferentes rutas metabólicas. A parte de identificar los genes para cada término o vía

metabólica, los resultados se visualizan en un diagrama visual que representa el conocimiento experimental génica y varias otras funciones que interactúan. Este diagrama contiene una red de interacciones y redes moleculares, de tal manera que pueden vincular genes del genoma con productos génicos de la vía (Rodríguez, s/f).

El análisis de enriquecimiento en la secuenciación de ARN de célula única (scRNA-seq) ha revolucionado la biología celular y molecular al permitir la desagregación de la heterogeneidad celular y la identificación de subpoblaciones celulares específicas. Este enfoque ha proporcionado una comprensión más profunda de las funciones celulares y sus cambios en diversas condiciones biológicas y patológicas (Jiang et al., 2024).

4. METODOLOGÍA

La metodología del análisis de enriquecimiento de genes se implementó para interpretar funcionalmente genes de Glioblastoma Multiforme, esta metodología se basó en la definición de una lista de genes de interés o candidatos que mostraron una expresión diferencial. Además, se seleccionó conocimiento biológico con grandes bases de datos como las ontologías, una vez determinada la información ontológica a utilizar, se aplicó técnicas estadísticas para evaluar si la relación observada en el experimento es significativa o simplemente un evento azaroso. Así, para cada término se obtiene un p-valor que indica si hay evidencia de que las proporciones son diferentes.

Este proceso se desarrolló en el lenguaje de programación Python utilizando la biblioteca gseapy, analizando genes funcionales a través de la API de Enrichr. Enrichr es una herramienta que proporciona acceso a múltiples bibliotecas de genes ayudando a identificar las vías de señalización, funciones biológicas y procesos moleculares relevantes en el conjunto de genes.

El esquema de la metodología se presentó en la Figura 2. y se detalla a continuación:

4.1. Definición de la lista de genes de interés

Los genes de interés se definieron del estudio de Arteaga-Arteaga et al. (2023), quienes identificaron genes relevantes de acuerdo con la expresión diferencial significativa a partir

de algoritmo ML para cada uno de los cuatro escenarios. Se considero una muestra de los 20 mejores genes. Estos resultados son disponibles en <https://github.com/BioAITeam/Machine-learning-for-glioblastoma-identification>.

4.1.1. Selección la base de datos de anotaciones funcionales

Se utilizo ontologías que asocian conjuntos de genes con funciones biológicas, localizaciones celulares y vías metabólicas, como la Ontología Génica (GO) y la Enciclopedia de Genes y Genomas de Kioto (KEGG).

4.1.2. Aplicación de una prueba estadístico de enriquecimiento

Se evaluó si los genes de interés estarán sobrerrepresentados en cada término funcional. Cada término biológico fue valorado mediante una medida estadística, el p-value, que indica la importancia de un término biológico con respecto al conjunto de genes analizado.

4.1.3. Interpretación de los términos enriquecidos significativos

Se interpreto y analizo los términos funcionales que presentaron valores estadísticamente significativos y así como relevancia biológica para las bases de datos GO y KEGG.

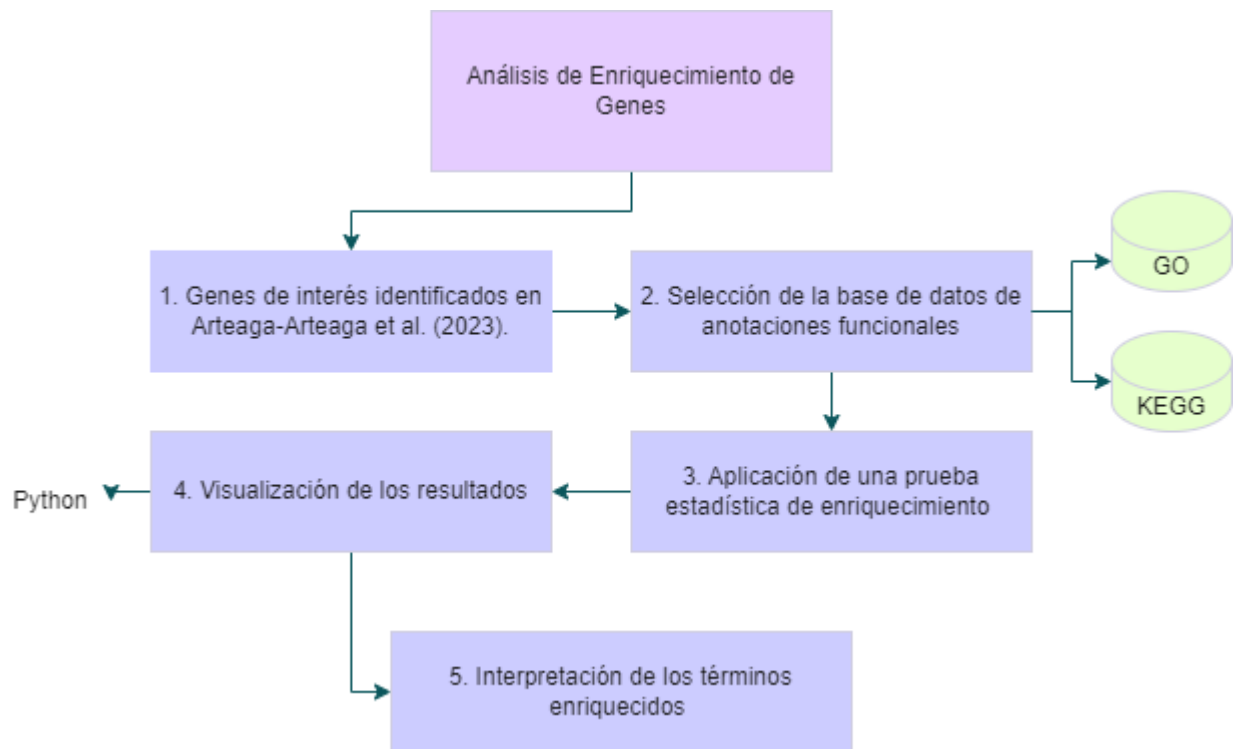


Figura 2. Pipeline de Análisis de Enriquecimiento de Genes

5. RESULTADOS Y DISCUSIÓN

Se realizó un análisis de enriquecimiento de genes, con una búsqueda de datos de KEGG y GO, se buscó identificar qué vías de señalización, procesos biológicos o funciones moleculares están significativamente enriquecidas en el conjunto de genes expresados con la siguiente clase Núcleo del Tumor (TC), Periferia del Tumor (TP) y Periferia Normal (NP). En los cuatro escenarios: TP vs TC; TP vs PN; NP vs TP y TC (TPC); NP vs TP vs TC.

Para este análisis se desarrolló e implementó el análisis de enriquecimiento de genes utilizando la biblioteca `gseapy` en Python de libre acceso y disponible en: <https://github.com/zqfang/GSEAPy>.

Para ello, se definió la lista de genes de interés. En este caso, los 20 genes altamente significativos en el análisis de expresión génica en los diferentes escenarios, estos análisis de obtuvieron de los mejores algoritmos ML de clasificación de genes (Anexo 1) obtenidos por Arteaga-Arteaga et al. (2023). El análisis se llevó a cabo utilizando las bases de datos de KEGG_2021_human que incluye vías de señalización y otros procesos

biológicos relevantes y GO_Biological_Process_2021 de la base de datos Gene Ontology (GO). Este proceso se contrastó con la anotación de todos los genes posibles. La implementación del proceso se encuentra disponible en: https://github.com/IsaBalboa/Enriquecimiento_Genes_GB.git.

Luego del análisis, los resultados de este proceso incluyeron términos de vías de señalización, p-valor, odds ratios y otros datos estadísticos (Anexo 2) que indicaron la relevancia de los genes proporcionados, como se detalla en la Tabla 1:

Tabla 1. Parámetros de salida de un análisis de enriquecimiento.

Nº	Parametros	Detalle
1	Gene_set	Nombre de la base de datos KEGG o GO.
2	Term	Nombre del término funcional o vías de señalización
3	Overlap	Genes de la lista de interés que coinciden con los genes anotados en el término específico
4	P-value	mide la significancia estadística de la sobre-representación de los genes en ese término.
5	Adjusted P-value	Mide la significancia estadística controlando los falsos positivos con el método de corrección Bonferroni.
6	Old P-value y Old Adjusted P-value	Valores de un análisis anterior para la comparación.
7	Odds Ratio	Razón de probabilidades, que mide la fuerza de la asociación entre los genes de interés y el término anotado.
8	Combined Score	Puntuación combinada que integra el valor P y el odds ratio para una medida global de enriquecimiento.
9	Genes	Lista de genes específicos que caen dentro de cada término o ruta.

Para el análisis de los resultados se ordenó por puntuación combinada (Combined Score) debido a que se priorizan términos que tienen tanto alta significancia estadística como relevancia biológica.

Escenario 1

En este escenario se analizó los 20 genes representativos identificados en la Periferia del Tumor (TP) y Periferia Normal (NP), como se indican en la Tabla 2. El algoritmo de enriquecimiento identificó 699 términos biológicos o rutas de señalización en los que se ha observado enriquecimiento significativo de genes. En GO se identificó 603 procesos biológicos y 96 vías metabólicas, redes de señalización celular, y funciones genómicas. El análisis de enriquecimiento de genes GO se indica en la Figura 3. Se muestra los procesos biológicos enriquecidos con respecto a la puntuación combinada (Combined Score) que indica que tan significativo son los términos biológicos, los valores más altos indican un enriquecimiento más fuerte con respecto a los genes analizados. Mientras, los Overlaps_Size indican el tamaño de cada punto y está relacionado con el número de genes que se superponen con los términos GEO. El punto más grande indica un mayor número de genes compartidos con el término respectivo. Por otra parte, los puntos están coloreados en base al p-valor, lo que indica significancia estadística con un nivel de significancia de 0.05, indicando que los términos GO son más relevantes.

Tabla 1. Listado de genes para el Escenario 1

Nº	Gen	Nombre completo
1	EGFR	Receptor del Factor de Crecimiento Epidérmico
2	CALM1	Calmodulina 1
3	APOD	Apolipoproteína D
4	SPOCK1	Proteoglicano de Oligodendrocito y Condroitín Sulfato 1
5	ANXA1	Anexina A1

6	PTGDS	Prostaglandina D Sintasa
7	EIF1	Factor de Iniciación de la Traducción Eucariota 1
8	VIM	Vimentina
9	MGLL	Monoacilglicerol Lipasa
10	ITM2C	Proteína Transmembrana Inducida por la β -amiloide 2C
11	PLLP	Proteína de la Línea del Oligodendrocito Lipidada
12	ITGB8	Integrina Subunidad β 8
13	HES6	Inhibidor Transcripcional Básico del Helix-loop-helix 6
14	RPS27L	Ribosomal Protein S27-like
15	GFAP	Proteína Fibrilar Ácida Glial
16	FTL	Ferritina, cadena ligera
17	TRIM2	Proteína Tripartita Motif Conteniendo 2
18	APOE	Apolipoproteína E
19	ANXA5	Anexina A5
20	NAV1	Navegador 1

Entre los términos más representativos se encuentran los siguientes:

El término de la regulación de la actividad de la proteína quinasa C (PKC), está involucrado el gen EGFR. El término PKC juega un papel clave en la transducción de señales celulares, afectando procesos como la proliferación celular, la diferenciación y la respuesta al estrés. En el Glioma, la disfuncionalidad en la señalización PKC puede contribuir al crecimiento tumoral y a la resistencia al tratamiento. El gen EGFR destaca la importancia de la señalización en la activación de las vías PKC, lo cual indica la progresión del glioma (Heckman et al., 2020).

La diferenciación de células T alfa-beta un proceso clave en la respuesta inmune adaptativa. El gen involucrado es ANXA1 conocido por su papel en la regulación de la respuesta inflamatoria y la activación de las células T, lo cual es importante en el contexto

del glioma debido a que puede evadir la vigilancia inmunológica (Aruga, 2016). Por tanto, el gen ANXA1 resalta la interacción entre el glioma y el sistema inmunológico. Este hallazgo sugiere que la modulación de la respuesta inmune puede ser un enfoque prometedor en el tratamiento de gliomas, particularmente mediante la activación o desregulación de células T en el microambiente tumoral.

Los genes APOE, CALM1, EGFR esta involucradas en la regulación de la actividad de la sintasa de óxido nítrico, una molécula señalizadora que influye en la vasodilatación, la proliferación celular y la apoptosis, importante en los procesos fisiológicos y patológicos de crecimiento tumoral (Ahmad et al., 1997). Estos procesos son cruciales en el crecimiento tumoral y la angiogénesis, sugiriendo que la modulación de estas vías podría ofrecer nuevos enfoques terapéuticos. En particular, el gen APOE, conocido por su papel en el metabolismo del colesterol, destaca la importancia de las vías metabólicas en la progresión del glioma. Estos hallazgos podrían orientar el desarrollo de terapias dirigidas que interfieran con el metabolismo lipídico y la señalización del óxido nítrico.

Los genes GFAP y VIM están involucrados en la organización de filamentos intermedios, que mantienen la integridad celular y la señalización. En los gliomas, los genes son marcadores de las células gliales y desempeñan un papel importante en la morfología celular y la invasión tumoral.

La homeostasis del hierro involucrada por el gen FTL, es fundamental para el metabolismo celular y la proliferación. La disfunción en este proceso puede llevar a estrés oxidativo, implicando a progresión del glioma (Palacios y Silva, 2020). El hierro es esencial para la proliferación celular, y su desregulación puede llevar a un estrés oxidativo significativo, contribuyendo a la progresión del glioma. Los resultados sugieren que la regulación del hierro y el manejo del estrés oxidativo podrían ser clave para limitar la invasión tumoral.

Por tanto, las mutaciones en la señalización de EGFR afectará la proliferación y supervivencia celular, además, en el microambiente tumoral, las respuestas inflamatorias y el metabolismo celular. El gen APOE involucrado en el metabolismo del colesterol y la señalización del óxido nítrico resalta las vías metabólicas y la señalización oncogénica en

la patogénesis del glioma. Por tanto, la modificación en estos procesos podría significar elementos fundamentales para la intervención terapéutica.

Uno de los hallazgos más significativos es la implicación del gen EGFR en la regulación de la actividad de la proteína quinasa C (PKC). Este receptor es conocido por su papel en la transducción de señales celulares, afectando procesos críticos como la proliferación celular, la diferenciación y la respuesta al estrés. En el contexto del glioma, la disfuncionalidad en la señalización de PKC, mediada por alteraciones en EGFR, puede contribuir al crecimiento tumoral y a la resistencia al tratamiento, lo que subraya la relevancia de EGFR como un objetivo terapéutico potencial en glioblastomas.

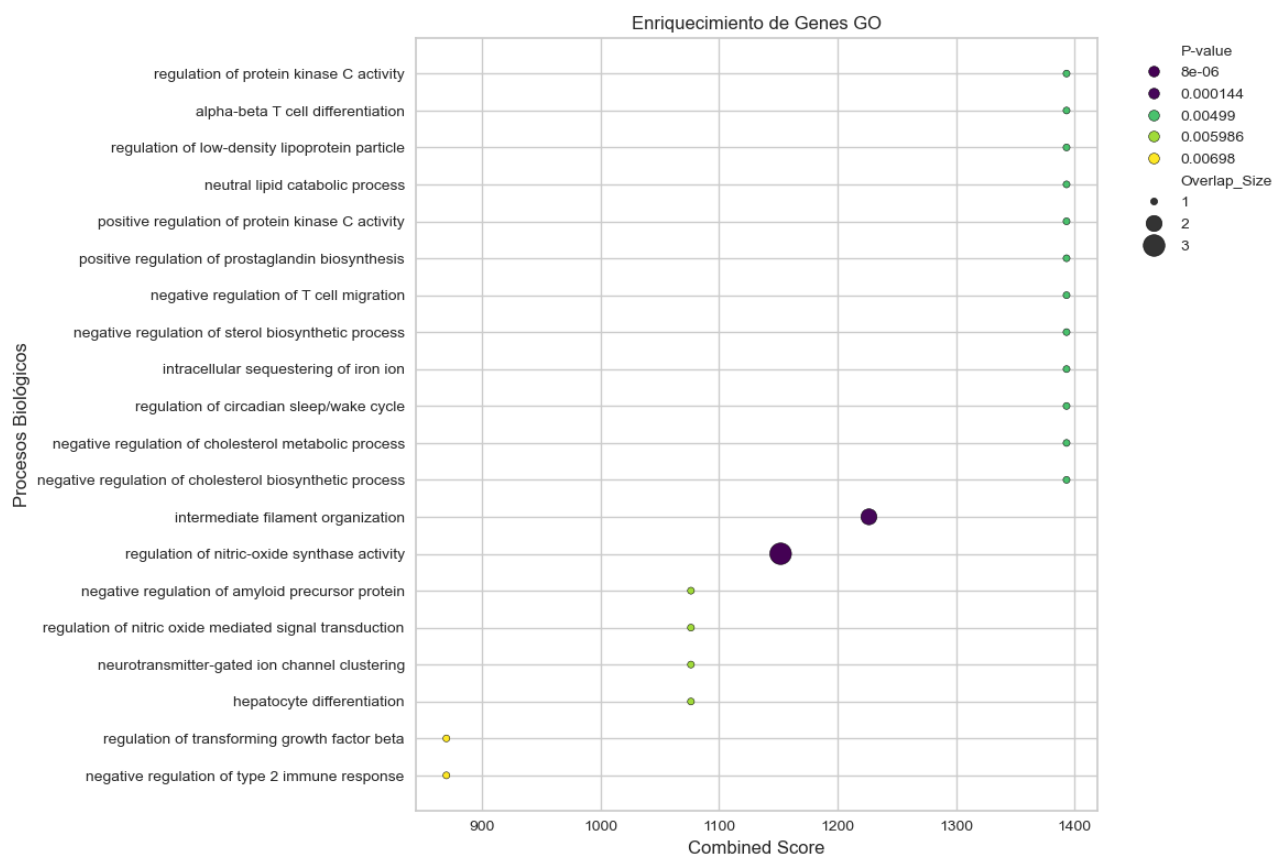


Figura 3. Enriquecimiento de genes GO, Escenario 1

En la base de datos KEGG, se identificó vías de señalización y procesos metabólicos clave en el cáncer, algunos términos tienen p-valor bajos, pero los valores ajustados relativamente altos, lo cual indica que los términos parecen ser significativos, aunque las

correcciones reducen la confianza en la significancia estadística como se visualiza en la Figura 4 y 5.

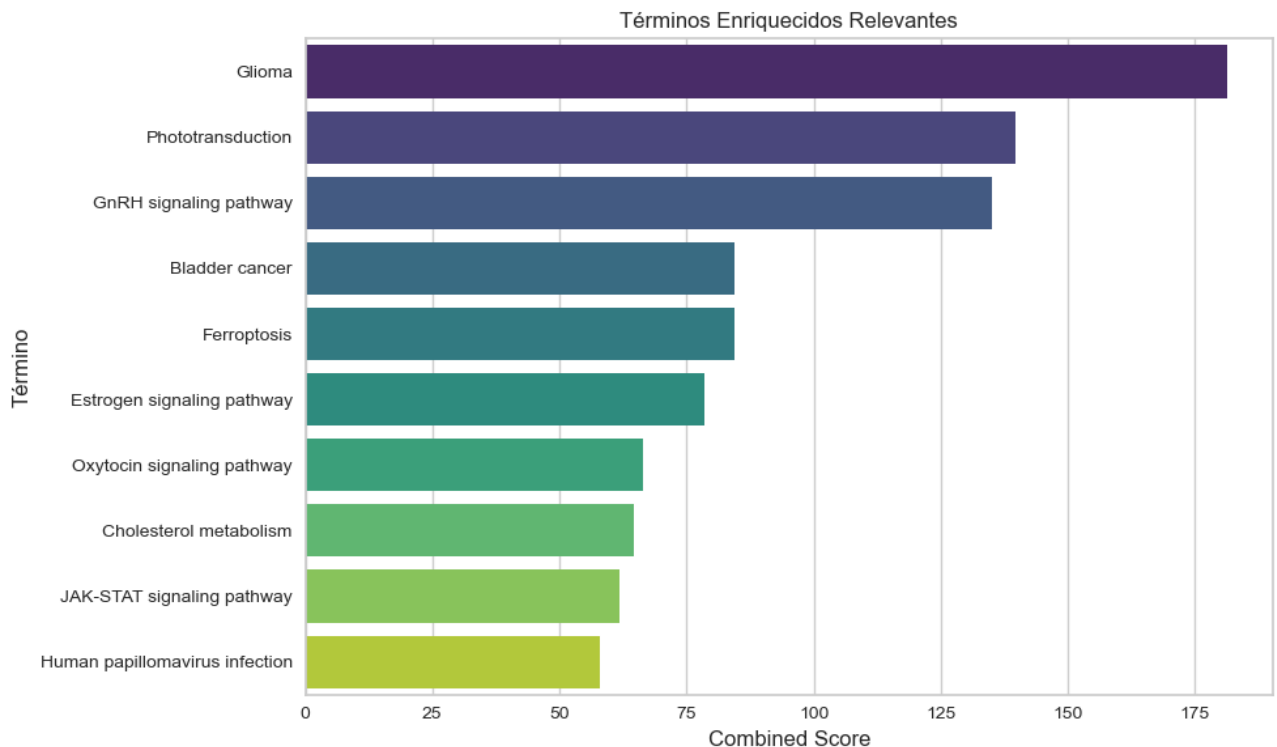


Figura 4. Análisis de enriquecimiento en la base de datos KEGG, Escenario 1.

En la Figura 4 se indica la puntuación combinada para cada termino, también se visualiza el `Overlap_size` que representa cuantos genes están presentes en cada término junto con el `p-value` que indica la significancia estadística del enriquecimiento, con colores más oscuros para valores significativos como el Glioma y el GnRH signaling pathway.

Glioma: Esta vía presenta un P-value bajo (0.002523) y una puntuación combinada alta (181.257422), lo que sugiere una asociación fuerte y significativa con los genes CALM1 y EGFR. Dado que Glioma es un tipo de cáncer cerebral, la presencia de EGFR es coherente con su papel conocido en la señalización y proliferación celular en tumores cerebrales.

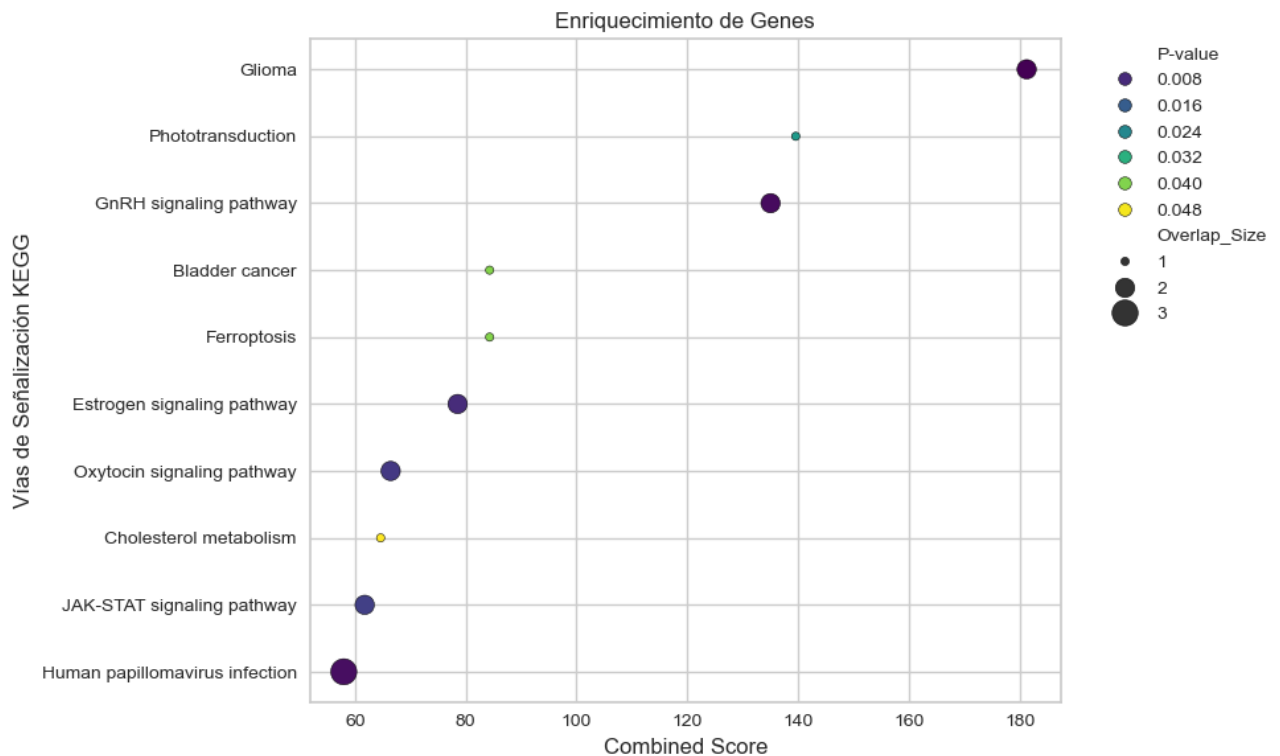


Figura 5. Términos KEGG enriquecidos, Escenario 1.

Fototransducción: Aunque tiene un solo gen (CALM1) involucrado y un valor P más alto (0.027643), su alta razón de probabilidades (38.894737) y puntuación combinada (139.568515) sugieren que CALM1 puede tener un papel significativo en la vía de fototransducción, aunque esto podría ser menos relevante para contextos de cáncer.

Vía de señalización de GnRH (Hormona Liberadora de Gonadotropina): Con un P-value de 0.003849 y una puntuación combinada de 135.023158, esta vía incluye también a CALM1 y EGFR. La señalización de GnRH está relacionada con la regulación hormonal, estudios sugieren que los receptores pueden estar presentes en células tumorales y podría inducir apoptosis o muerte celular, reduciendo así la capacidad de proliferación del tumor (Marelli et al., 2019).

Cáncer de vejiga y ferroptosis: Ambas vías tienen un valor P más alto (0.040230) y un valor P ajustado también alto (0.174456), lo que sugiere que su significancia es menos robusta tras la corrección por pruebas múltiples. Sin embargo, el hecho de que EGFR

aparezca en "Bladder cancer" y FTL en "Ferroptosis" puede indicar su relevancia potencial en diferentes tipos de cáncer y procesos de muerte celular regulada por hierro.

Vía de señalización de estrógenos y vía de señalización de oxitocina: Ambas vías incluyen CALM1 y EGFR y tienen valores P ajustados relativamente altos (0.174456). Estas vías están asociadas con señalización hormonal y podrían influir en procesos de crecimiento celular y regulación hormonal en el cáncer.

Cholesterol metabolism: Con el gen APOE involucrado, esta vía tiene una relevancia potencial en procesos de regulación lipídica y metabolismo de colesterol, lo cual es crítico en muchos procesos biológicos, incluyendo la progresión de ciertos tipos de cáncer.

Vía de señalización JAK-STAT: Esta vía, conocida por su papel en la señalización inmune y oncogénesis, muestra la implicación de EGFR y GFAP con un valor P ajustado de 0.174456. Esto sugiere su importancia en la regulación de respuestas inflamatorias y de proliferación celular en el contexto del cáncer.

Infección por el virus del papiloma humano: Involucra a tres genes (HES6, ITGB8, EGFR) y tiene un valor P bajo (0.004155), lo que indica una posible asociación significativa. La infección por VPH es un conocido factor de riesgo en ciertos tipos de cáncer, lo que podría conectar estos genes con mecanismos de oncogénesis viral.

Red de asociaciones entre genes y términos en KEGG

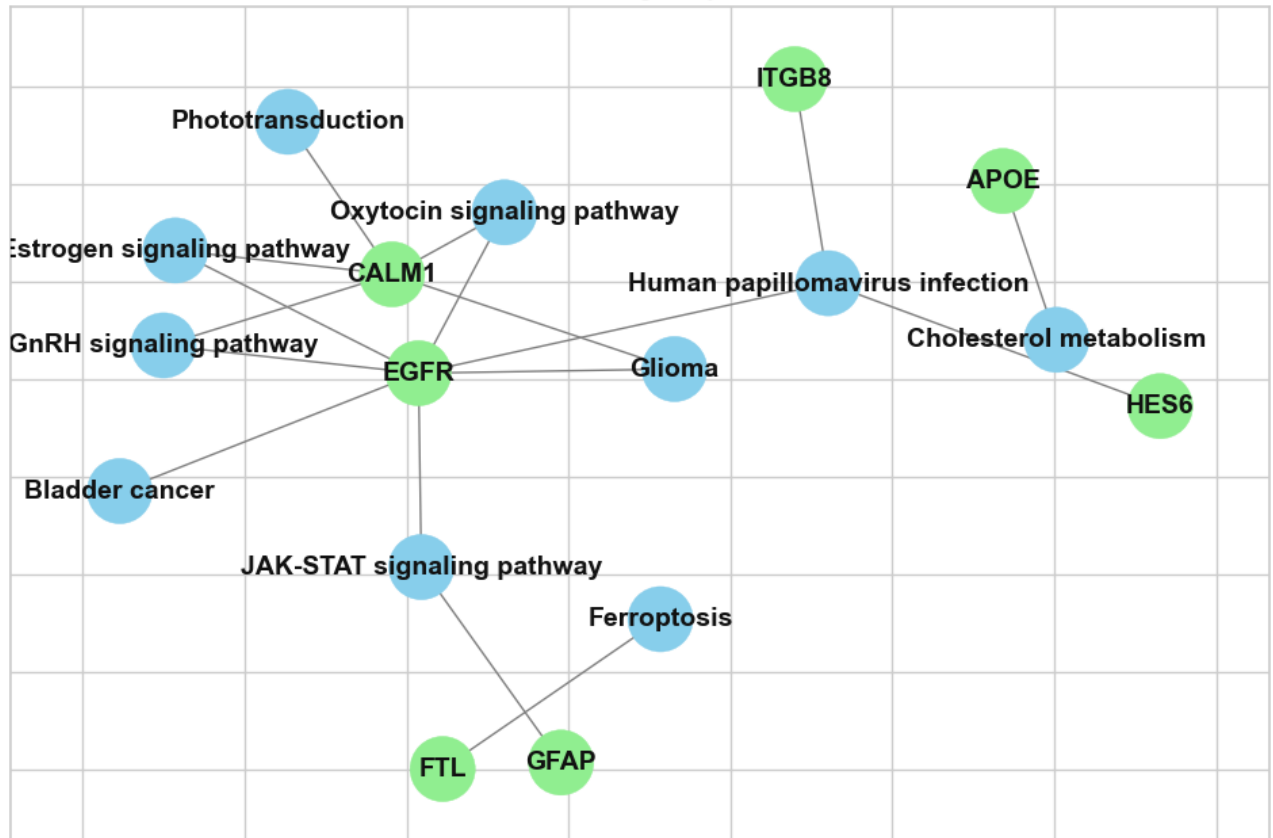


Figura 6. Rutas metabólicas KEGG, Escenario 1

El Glioma está asociada con los genes EGFR y CLAM1, como se observa en la Figura 6. El gen EGFR, un oncogen que juega un papel importante en la proliferación celular, supervivencia y diferenciación, particularmente en el glioblastoma multiforme, donde las mutaciones y los polimorfismos influyen significativamente en el comportamiento del tumor y en los resultados clínico del paciente. Los polimorfismos en el gen EGFR se han asociado con la susceptibilidad al glioma, con polimorfismo de un solo nucleótido (Matini et al., 2019). El gen CALM1 (Calmodulina 1), proteína reguladora de calcio en la transducción de señales y también en la proliferación celular y en la regulación de la respuesta al crecimiento celular. La presencia de estos genes sugiere su papel central en la señalización de crecimiento anormal y la oncogénesis (Liu et al., 2021).

Las vías de señalización hormonal y regulación: Vía de señalización de GnRH, Vía de señalización de estrógenos, y Vía de señalización de oxitocina. En estas vías siguen

asociados los genes CALM1 Y EGFR. La calmodulina al ser un modulador de la transducción de señales de calcio juegan un papel importante en la función de múltiples hormonas y factores de crecimiento, como los involucrados en las vías GnRH, estrógeno y oxitocina. Por otro lado, el gen EGFR influye en la función hormonal a través de la interacción con otras proteínas de señalización.

También se encontró que el gen FTL, es responsable del proceso de Ferroptosis, un tipo de muerte celular programada dependiente de la cantidad de hierro. La asociación. La vía del metabolismo y la regulación lipídica, están asociadas con el gen APOE (Apolipoproteínas E) debido al papel central en la homeostasis del colesterol y su implicación en condiciones metabólicas neurodegenerativas.

Por otro lado, la vía de señalización JAK-STAT, es fundamental en la transducción de señales para las citoquinas y factores de crecimiento regulando la inmunidad, inflamación y hematopoyesis. Los genes asociados son EGFR y GFAP. El gen EGFR, influye en esta vía a través de la activación de la activación de rutas de señalización que interactúan con JAK-SAT, por lo que puede tener implicaciones en la respuesta inflamatoria y en la oncogénesis. El gen GFAP es un marcador de las células gliales del sistema nervioso central, la expresión en relación con esta vía sugiere un papel en las respuestas inflamatorias del cerebro o en enfermedades neurológicas.

También se encontró infección viral y transformación celular, con los genes asociados HES6, ITGB8, EGFR. Estos genes asociados pueden ser regulados por las proteínas del HPV, que alteran las vías de señalización celular para promover la transformación maligna. El gen HES6, es un regulador transcripcional que puede interactuar con las vías de Notch y otros factores de crecimiento, relacionado con los mecanismos de oncogénesis por infección viral (Marelli et al., 2019).

Los análisis de enriquecimiento en la base de datos KEGG revelaron la implicación de vías de señalización hormonal (GnRH, estrógeno, oxitocina) y metabólicas en la patogénesis del glioma. La asociación de EGFR y CALM1 con estas vías subraya su papel en la regulación de la proliferación y diferenciación celular. La modulación de estas vías hormonales podría abrir nuevas oportunidades terapéuticas para intervenir en la progresión del glioma, particularmente en el contexto de resistencia a tratamientos convencionales.

Escenario 2

En el escenario 2 se trabajó con los 20 genes representativos identificados en el Núcleo del Tumor (TC) y Periferia del Tumor (TP) como se indica en la Tabla 3. El procedimiento de filtrado de los genes estadísticamente significativos y de importancia biológica se desarrolló de la misma manera que el escenario 1.

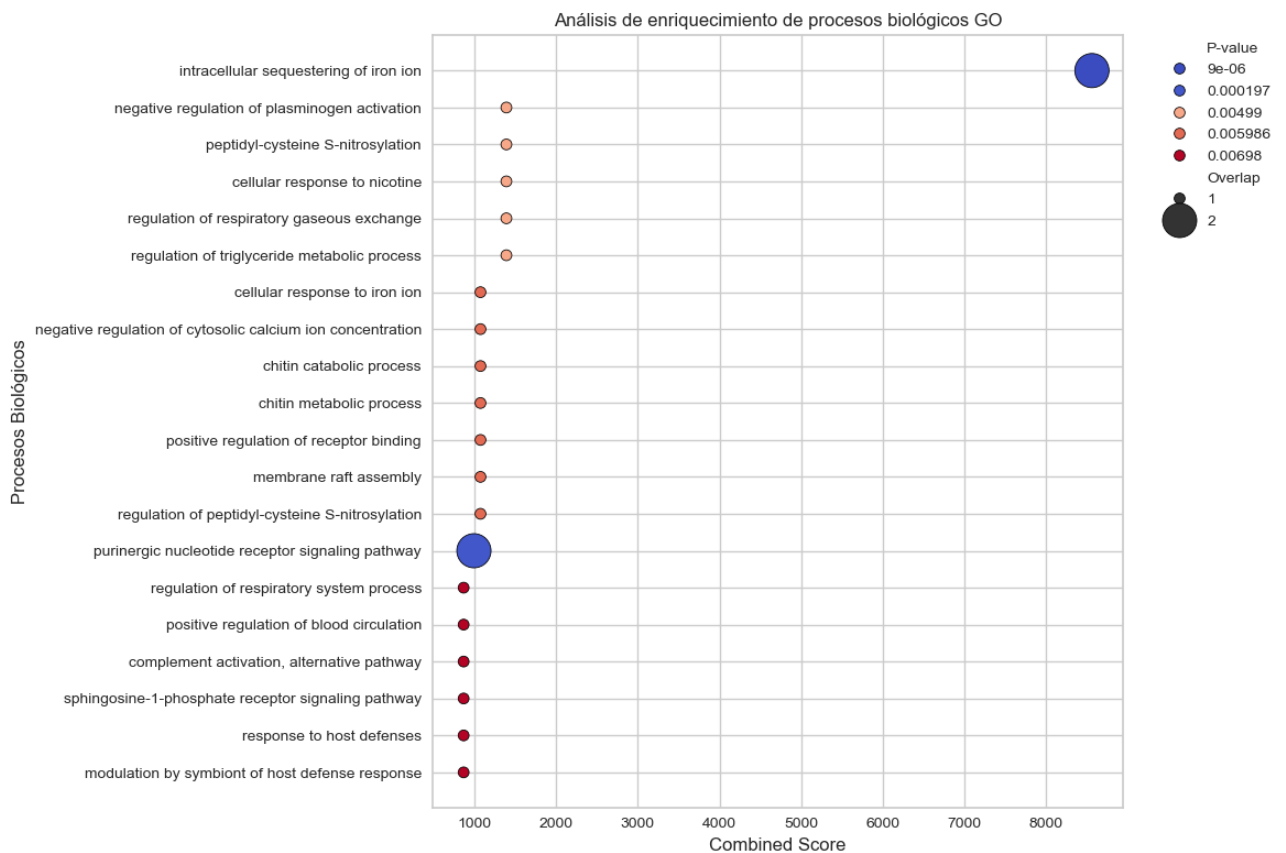


Figura 7. Enriquecimiento de genes GO, Escenario 2.

En la Figura 7, se muestra los procesos biológicos relacionados con un conjunto específico de 20 genes característicos. Cada punto representa los términos GO que están significativamente sobrerrepresentados en el conjunto de genes analizados. Se presentan los términos GO con un p-valor más bajo sugiriendo que los procesos biológicos podrían ser relevantes en el contexto de estudio.

Tabla 2. Listado de genes para el Escenario 2

Nº	Gen	Nombre completo
1	ATP1A2	ATPasa, Na ⁺ /K ⁺ transportadora, subunidad alfa 2
2	SPARCL1	Proteína similar a la ácido secretado y rico en cisteína 1
3	CPE	Carboxipeptidasa E
4	FTL	Ferritina ligera
5	FTH1	Ferritina pesada
6	S100A10	Proteína S100-A10
7	HES6	Homólogo de la familia hairy y enhancer of split 6
8	B2M	Microglobulina beta-2
9	PRODH	Prolina deshidrogenasa 1
10	SERPINE2	Serpina inhibidora de proteasas E2
11	RPS18	Proteína ribosomal S18
12	TMSB10	Timosina beta 10
13	SERPINA3	Serpina inhibidora de proteasas A3
14	GAPDH	Gliceraldehído-3-fosfato deshidrogenasa
15	GFAP	Proteína ácida fibrilar glial
16	RPLP1	Proteína ribosomal lateral P1
17	S1PR1	Receptor 1 de esfingosina-1-fosfato
18	TXN	Tiorredoxina
19	C3	Componente 3 del complemento
20	CHI3L1	Quitinasa 3 similar 1 (YKL-40)

El término más significativo fue el término de secuestro intracelular de iones de hierro, este proceso este asociado con el secuestro intracelular de iones de hierro, indicando un mecanismo celular de regulación para mantener la homeóstasis de hierro, importante para prevenir el estrés oxidativo y para el correcto funcionamiento de varias enzimas. Los genes que involucrados son los FTH1 y FTL, genes que codifican para subunidades de ferritina, una proteína clave en el almacenamiento de hierro dentro de las células. Este proceso es importante en la protección celular contra el daño de radicales libres. El término "Negative regulation of plasminogen activation" se refiere a la inhibición de la activación del plasminógeno, con implicaciones en procesos de coagulación y fibrinólisis y está relacionada con la respuesta inflamatoria. El gen que está involucrado es el

SERPINE2, indicando una respuesta de defensa al daño tisular o inflamación (Hoelzgen et al., 2024).

Otro termino significativo fu el término S-nitrosilación de peptidil-cisteína, que es una modificación post-traducciona que afecta la función de las proteínas. Los genes involucrados B2M pueden estar desempeñando un papel en la regulación de la señalización redox y en la modulación de funciones celulares bajo estrés oxidativo. La implicación en varios términos de GO sugiere una activación inmunológica o inflamatoria, posiblemente en respuesta a infección o daño tisular. Por tanto, los términos de GO asociados con el secuestro de iones de hierro, la regulación de la activación del plasminógeno y la S-nitrosilación sugieren una posible implicación en la homeostasis del hierro, respuestas a estrés oxidativo y regulación de la inflamación y la coagulación (Marelli et al., 2009).

En el análisis de enriquecimiento en la base de datos KEGG para el escenario 2, el conjunto de genes analizados está involucrados en una variedad de funciones celulares y mecanismos de respuesta, desde la homeostasis de minerales, el metabolismo del hierro hasta la respuesta a infecciones virales y bacterianas como se indica en la Figura 8. El p-valor es menor que el nivel de significancia lo que indica que los 10 primeros términos son significativos.

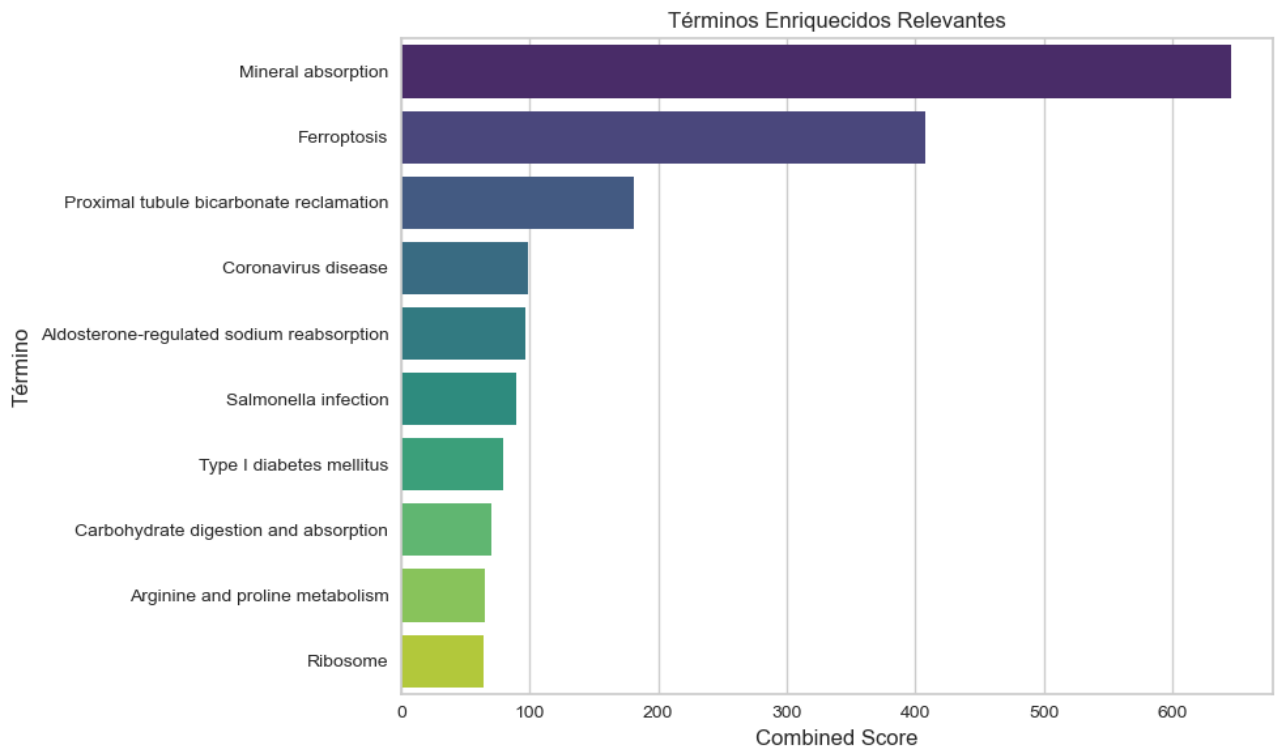


Figura 8. Análisis de enriquecimiento en la base de datos KEGG, Escenario 2

La ruta más significativa para este grupo fue la absorción de minerales con un valor P-value ajustado de 0.0016, los genes involucrados fueron FTH1, ATP1A2, FTL. Los genes FTH1 y FTL son fundamentales en la homeostasis del hierro y el almacenamiento de este mineral en las células. La implicación de ATP1A2, que codifica para una subunidad de la bomba de sodio-potasio ATPasa, sugiere un papel en la absorción de minerales que puede estar vinculado a la regulación del equilibrio de iones y metales en las células (Medeiros et al., 2014).

En la Ferroptosis los genes involucrados fueron FTH1, FTL. Ambos genes se relacionan con el metabolismo y almacenamiento en procesos de estrés oxidativo y respuesta inflamatoria. La relación de los genes C3, RPS18, RPLP1 con la respuesta inmune y la inflamación sugiere su posible papel en la respuesta celular al virus. C3 es un componente clave del sistema del complemento, implicado en la inmunidad innata, mientras que

RPS18 y RPLP1 son subunidades ribosomales, lo que podría implicar alteraciones en la síntesis de proteínas en la infección por coronavirus (Lazarowski et al., 2022).

La ruta de absorción de minerales es crucial para la homeostasis de metales en el organismo. Estas rutas en relación con los genes se visualizan en la Figura. 9. Los genes FTH1 y FTL codifican para las subunidades de la ferritina, una proteína esencial para el almacenamiento intracelular de hierro y la regulación de sus niveles en el organismo. La presencia de ATP1A2 en esta ruta sugiere una interrelación entre la homeostasis de hierro y la actividad de las bombas de sodio-potasio, lo cual podría influir en el transporte y la distribución de hierro a nivel celular. El ajuste de la absorción de minerales es esencial para el mantenimiento de la función celular, la prevención de estrés oxidativo y la regulación de la inmunidad innata.

La ferroptosis es un tipo de muerte celular regulada caracterizada por la acumulación de hierro y la peroxidación lipídica. La participación de FTH1 y FTL, genes que codifican las cadenas pesada y ligera de la ferritina, respectivamente, sugiere un papel protector en la mitigación del daño oxidativo al limitar la disponibilidad de hierro libre, el cual puede catalizar la formación de especies reactivas de oxígeno (ROS). La regulación de la ferroptosis es vital en el contexto de enfermedades neurodegenerativas, cáncer y daño isquémico, donde el balance de hierro y la respuesta antioxidante son críticos.

Red de asociaciones entre genes y términos en KEGG

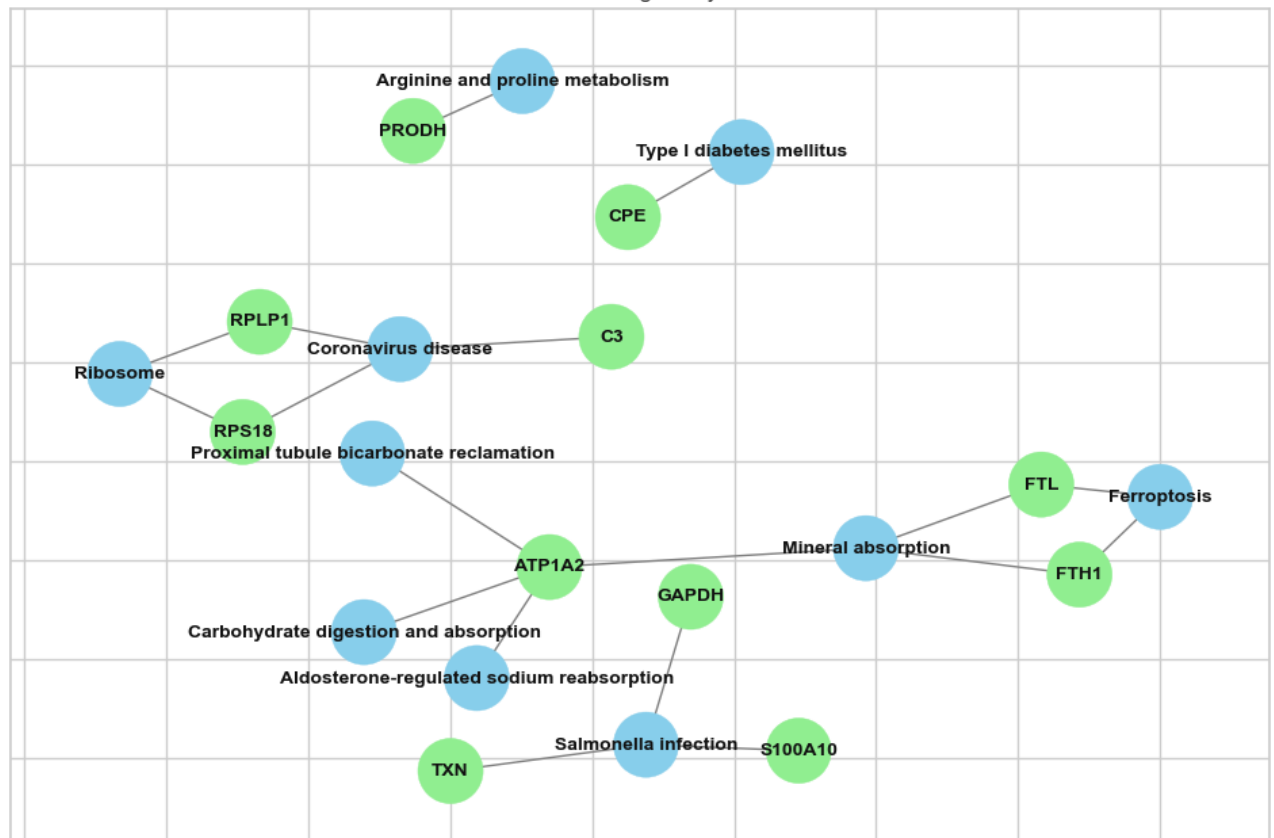


Figura 9. Rutas metabólicas KEGG, Escenario 2.

TP1A2 codifica una subunidad de la bomba de sodio-potasio-ATPasa, que es fundamental para el transporte iónico y la regulación del pH celular. En el contexto de la reclamación de bicarbonato en el túbulo proximal, ATP1A2 está involucrado en la mantención de un gradiente electroquímico que facilite la reabsorción de bicarbonato, esencial para la homeostasis ácido-base. Esta función es crítica para prevenir acidosis metabólica, una condición que puede tener repercusiones severas en la función celular y sistémica.

Escenario 3

Para el análisis GEO en el escenario 3, se analizó 20 genes representativos, como se enlista en la Tabla 4, después de la ejecución del algoritmo de enriquecimiento se filtró los 10 mejores términos en base a la puntuación combinada (Combined Score). Se encontró 304 procesos biológicos en GEO y 63 vías de señalización en la base de datos KEGG.

El análisis de enriquecimiento de los principales procesos biológicos se muestra en la Figura 10. Se visualiza que existen 7 términos con la misma puntuación combinada (1393.1238), así también con el mismo p-valor de 0.00490 estadísticamente significativa con respecto al nivel de confianza (0.05). Estos términos están relacionados con los diferentes genes.

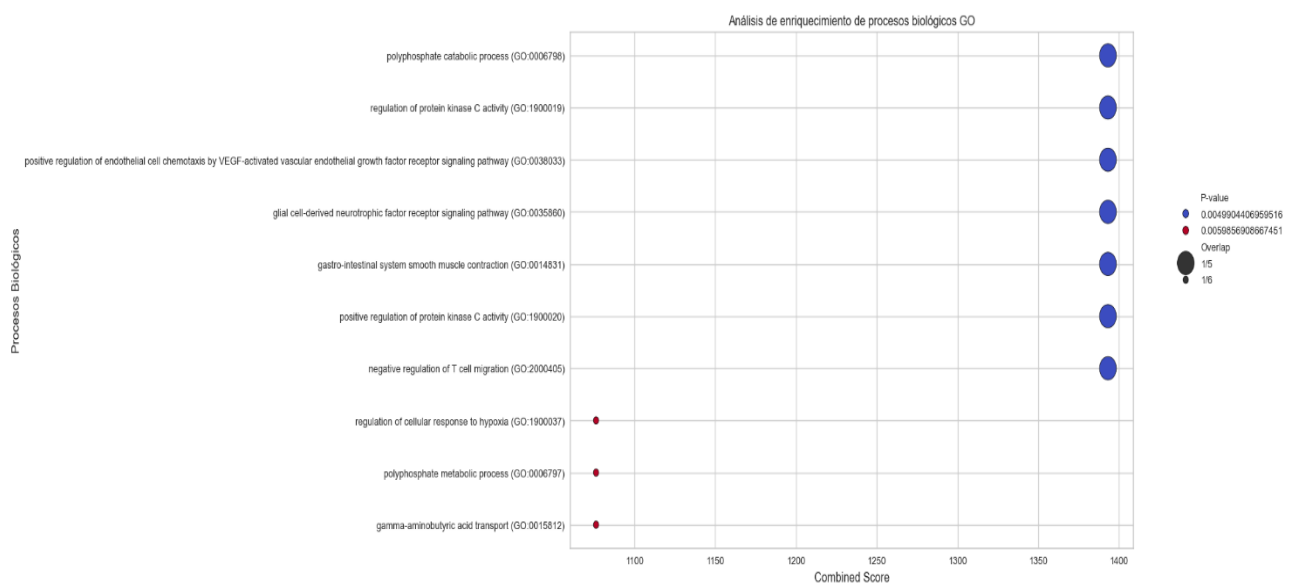


Figura 10. Enriquecimiento de genes GO, Escenario 3

Proceso catabólico de polifosfato: Este término está relacionado con la degradación de polifosfatos, compuestos involucrados en la regulación de diversas funciones celulares, incluyendo el almacenamiento de energía y la respuesta al estrés. La presencia de ATCAY, un gen implicado en este proceso sugiere que los mecanismos de regulación de polifosfatos podrían estar alterados en el glioblastoma, afectando potencialmente la homeostasis celular y la respuesta al estrés en las células tumorales (Siavashani et al., 2023).

Tabla 3. Listado de genes para el Escenario 3

Nº	Gen	Nombre completo
1	APOD	Apolipoproteína D

2	HTRA1	Proteasa de serina HTRA1, mitocondrial
3	SPOCK1	Proteoglicano de Oligodendrocito y Condroitín Sulfato 1
4	RPLP1	Proteína ribosomal lateral P1
5	SULF2	Sulfatasa 2
6	CDR1	Cadherina relacionada con la retinitis pigmentosa 1
7	GPR17	Receptor acoplado a proteínas G 17
8	TMSB10	Timosina beta 10
9	RPS18	Proteína ribosomal S18
10	POLR2J	Subunidad J de la ARN polimerasa II
11	SLC6A1	Transportador de solutos familia 6 miembro 1
12	ATCAY	Ataxina relacionada con cayado
13	NDUFB2	Subunidad B2 de la NADH deshidrogenasa (complejo I)
14	CHCHD2	Proteína de homología CHCHD2, mitocondrial
15	EGFR	Receptor del Factor de Crecimiento Epidérmico
16	CNTN1	Contactina 1
17	NKAIN4	Inhibidor de la ATPasa Na ⁺ /K ⁺ 4
18	RPL36	Proteína ribosomal L36
19	NDUFS5	Subunidad 5 de la NADH deshidrogenasa (complejo I)
20	HSPB1	Proteína de choque térmico beta-1 (HSP27)

Los términos relacionados con la regulación de la actividad de PKC (por ejemplo, GO:1904269, GO:0007204) resaltan la importancia de esta familia de enzimas en la señalización celular y la progresión del tumor. El gen EGFR está involucrado en esta vía y es un conocido impulsor del crecimiento tumoral y la resistencia a la terapia en GBM. Estos resultados indican el rol que juega este gen en la señalización mediada por PKC en la promoción de la proliferación celular y la supervivencia en glioblastoma (Heckman et al., 2020).

La regulación positiva de la quimiotaxis de células endoteliales: Este término, asociado al gen HSPB1, que sugiere la promoción de la migración de células endoteliales hacia el

tumor, un proceso crucial en la angiogénesis tumoral. La capacidad de los tumores de glioblastoma para atraer y formar nuevos vasos sanguíneos es fundamental para su crecimiento y diseminación, indicando que HSPB1 podría ser un mediador clave en estos procesos angiogénicos (Gimenez et al., 2015).

Transporte de ácido gamma-aminobutírico (GABA): El enriquecimiento en este término, con el gen SLC6A1, sugiere alteraciones en la neurotransmisión en el microambiente tumoral. Dado que el GABA es un importante neurotransmisor inhibitorio en el cerebro, su transporte y regulación anómala podrían influir en la comunicación célula-célula dentro del microambiente tumoral, afectando la interacción entre células tumorales y el entorno neural.

Respuesta a hipoxia: El término regulación de la respuesta celular a la hipoxia (GO:0071456), asociado con el gen CHCHD2, indica que la respuesta a la hipoxia, una condición común en microambientes tumorales podría estar regulada de manera anómala en el GBM. La hipoxia no solo induce cambios metabólicos que promueven la supervivencia celular, sino que también activa vías de señalización que pueden contribuir a la resistencia a terapias convencionales.

Estos resultados indican una adaptación metabólica y supervivencia tumoral. Los procesos relacionados con el catabolismo de polifosfato y la respuesta a hipoxia indican que las células tumorales en GBM están adaptándose a condiciones adversas mediante la modificación de su metabolismo y activando respuestas de supervivencia. Estos mecanismos permiten a las células tumorales sobrevivir a ambientes hostiles, como áreas con baja disponibilidad de oxígeno y nutrientes.

La regulación de la actividad de PKC y la quimiotaxis endotelial destacan la importancia de las vías de señalización en la proliferación y expansión del tumor. El gen EGFR, y HSPB1, relacionado con la migración celular, están implicados en estas vías, subrayando su rol en la promoción de la progresión tumoral. Los términos asociados con la regulación de la migración de células inmunes y la transmisión sináptica indican que las interacciones entre las células tumorales y su microambiente, incluidos los componentes inmunes y neurales, son importantes en la evolución del tumor. La alteración en el transporte de

GABA y la regulación negativa de la migración de células T podrían influir en la capacidad del tumor para evadir la respuesta inmunológica y modificar la comunicación celular (Marelli et al., 2009).

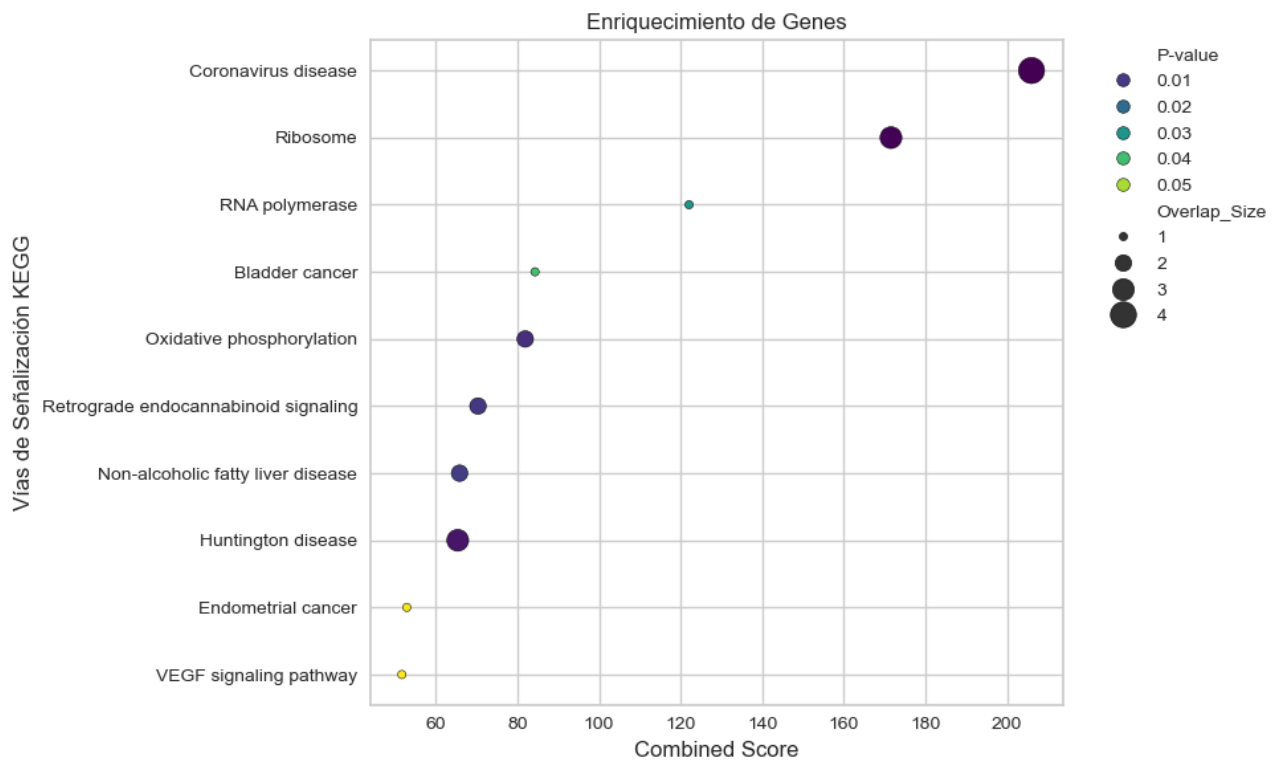


Figura 11. Términos KEGG enriquecidos, Escenario 3.

El término enfermedad por Coronavirus presento un alto puntaje combinado y un nivel de significancia menor, mostrando un enriquecimiento significativo con una superposición de 4 genes (RPS18, RPLP1, RPL36, EGFR), un valor p de 0.000074. Los genes ribosomales y el EGFR están relacionados con los mecanismos compartidos entre las respuestas del huésped al coronavirus y la biología del tumor. Debido a que EGFR está involucrado en la regulación de la proliferación y supervivencia celular en el GBM, y su relación con los procesos de infección podría indicar un papel en la modulación de la respuesta inmune y la inflamación en el contexto tumoral.

Ribosoma, este término también está significativamente enriquecido (valor p ajustado de 0.015738) y está asociado con genes ribosomales (RPS18, RPLP1, RPL36). Estas asociaciones se muestran en la Figura 11 y 12. La implicación de estos genes sugiere que la síntesis de proteínas y la maquinaria ribosomal están alteradas en el glioblastoma. Los ribosomas son cruciales para la traducción de proteínas, y su sobreexpresión o disfunción podría estar relacionada con el alto índice de proliferación celular y la adaptación metabólica en las células de GBM.

Fosforilación Oxidativa y Enfermedad de Huntington: Los genes NDUFS5 y NDUFB2 están relacionados tanto en la fosforilación oxidativa como en la enfermedad de Huntington, lo cual podría indicar un estrés metabólico elevado y disfunción mitocondrial en las células de GBM. La fosforilación oxidativa es crucial para la producción de ATP, y su desregulación podría promover un entorno pro-oxidativo, favoreciendo la proliferación tumoral y la resistencia a la apoptosis ((Marelli et al., 2009).

Por último, la señalización endocannabinoide retrógrada y enfermedad hepática grasa no alcohólica están implicadas en los mismos genes mitocondriales lo que sugiere una conexión entre la señalización celular y el metabolismo energético en glioblastoma. La señalización endocannabinoide puede regular procesos como la inflamación y el metabolismo energético, factores críticos en la progresión tumoral y la respuesta inmunitaria.

Existen cánceres asociados y señalización del EGFR en el cáncer de vejiga y el endometrio, ambos términos muestran enriquecimiento moderado con respecto al gen EGFR. Dado que el EGFR es un conocido oncogén en diversos tipos de cáncer, su activación en glioblastoma podría estar impulsando la proliferación celular, la angiogénesis y la resistencia a la apoptosis, mecanismos comúnmente alterados en otros cánceres sólidos (Liu et al., 2021).

La Vía de señalización de VEGF se encontró asociado al gen HSPB1, este término podría ser crucial en la angiogénesis, un proceso esencial para el suministro de nutrientes y oxígeno a las células tumorales. HSPB1 es una proteína de choque térmico que puede regular la respuesta al estrés y la supervivencia celular, y su implicación en la vía de señalización del VEGF destaca su potencial rol en la regulación del entorno angiogénico en el glioblastoma (Gimenez et al., 2015).

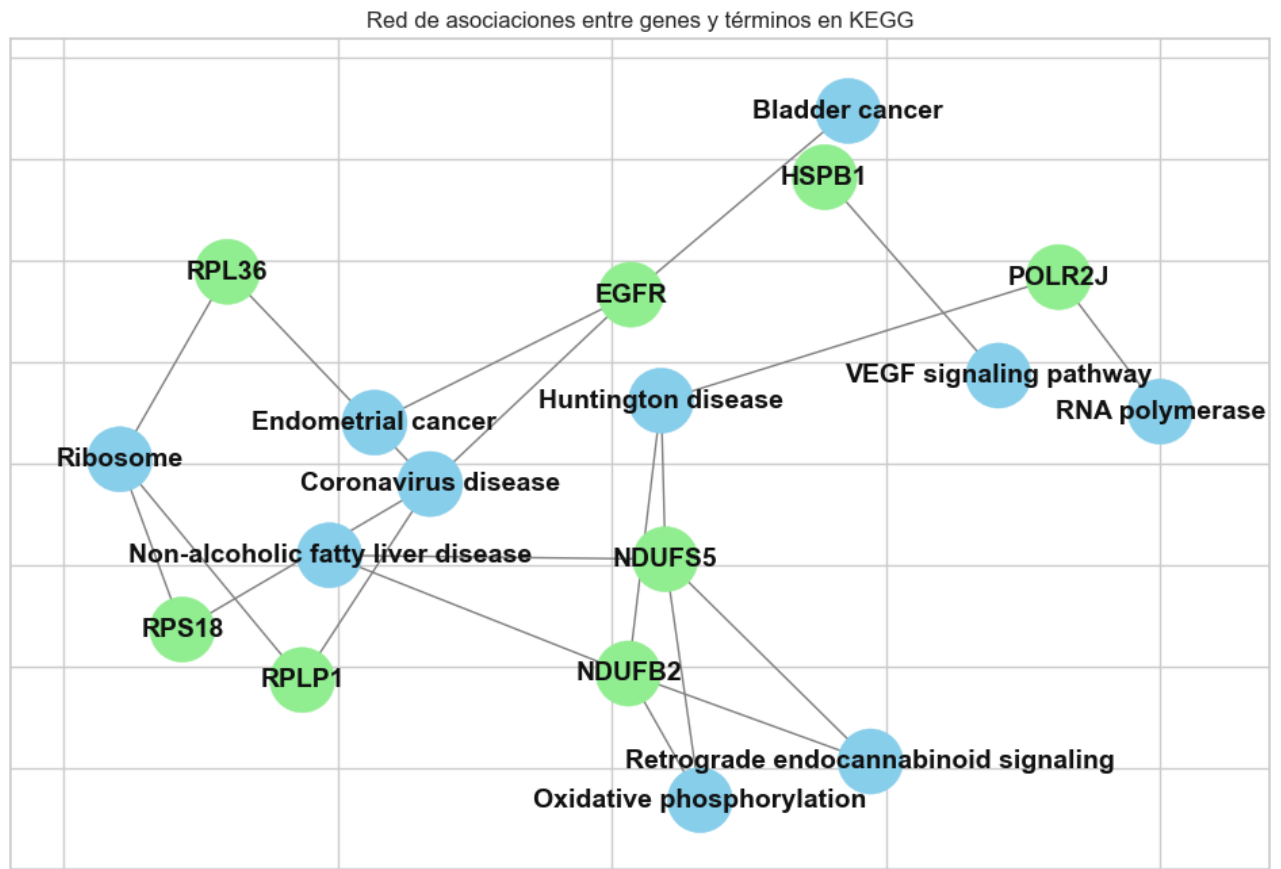


Figura 12. Rutas metabólicas KEGG, Escenario 3.

Escenario 4

En este escenario se obtuvieron 496 procesos biológicos de GO y 109 vías de señalización con la base de datos KEGG, para el conjunto de genes de Periferia Normal (NP) en comparación con Periferia del Tumor (TP) Núcleo del Tumor (TC), de los genes que se muestran en la Tabla 5. El proceso de filtrado para el análisis se siguió el mismo que de los escenarios anteriores.

En la Figura 13 se muestra los términos GO más representativos. El término secuestro de monómeros de actina, mostro un alto enriquecimiento con los genes TMSB4X y TMSB10, estos genes codifican para proteínas de timosina beta, que regulan la dinámica del citoesqueleto de la actina. En el Glioblastoma, la regulación de la actina es crucial

para la migración y invasividad celular, procesos que son importantes en la agresividad del tumor (Wirsching y Weller et al., 2017).

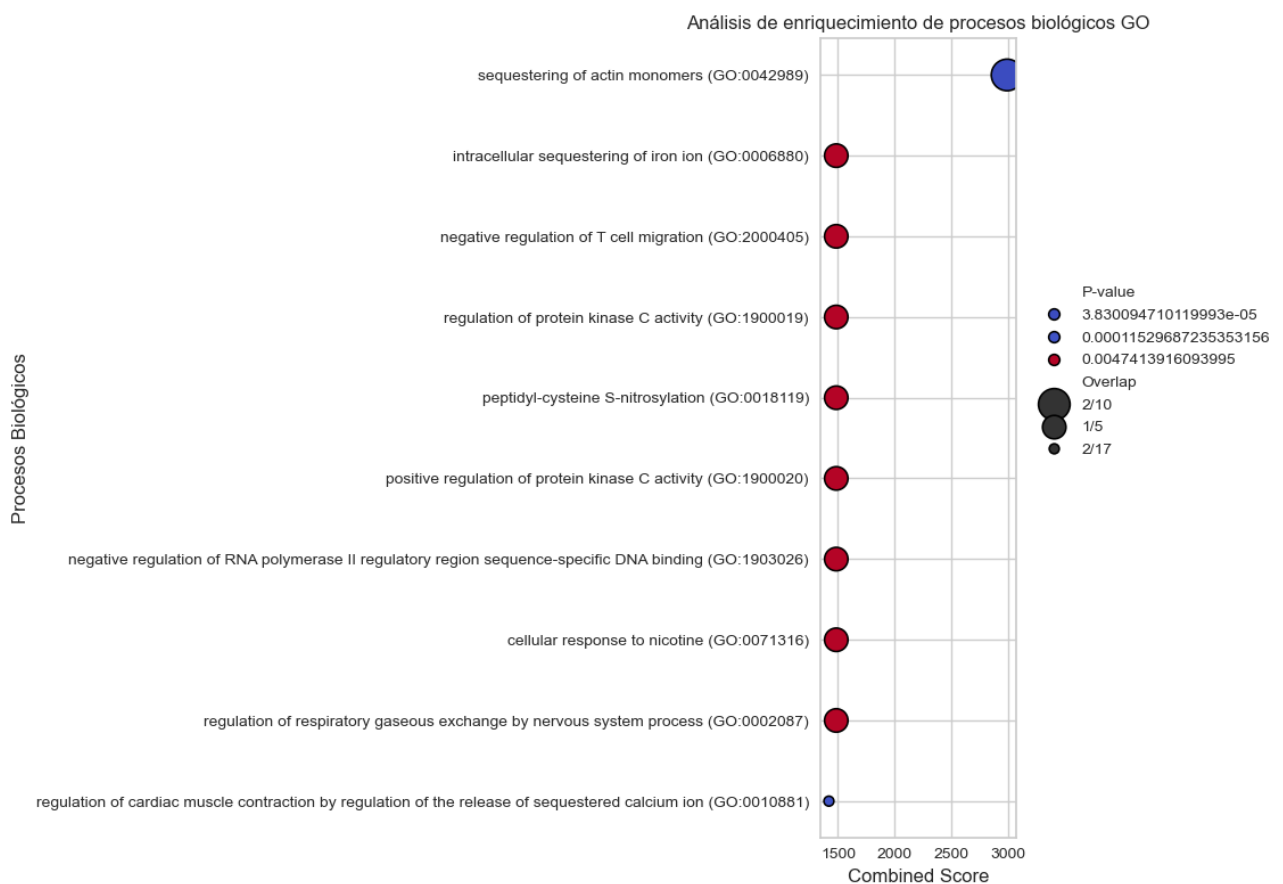


Figura 13. Enriquecimiento de genes GO, Escenario 4

También se encuentra el secuestro intracelular de iones de hierro, este término está asociada con el gen FTL, clave en la homeostasis de hierro. La regulación del hierro está relacionada con la prevención del estrés oxidativo al limita la disponibilidad del hierro libre que podría generar especies reactivas de oxígeno. Una alteración de hierro en el glioblastoma puede influir en la proliferación y supervivencia celular y resistencias a terapias.

Tabla 4. Listado de Genes para el Escenario 4

N°	Gen	Nombre completo
1	GFAP	Proteína ácida fibrilar glial
2	ATP1A2	ATPasa, Na ⁺ /K ⁺ transportadora, subunidad alfa 2
3	SPARCL1	Proteína similar a la ácido secretado y rico en cisteína 1
4	FTL	Ferritina ligera
5	S100A10	Proteína S100, miembro 10
6	CPE	Carboxipeptidasa E
7	TMSB4X	Timosina β4, isoforma X
8	CALM1	Calmodulina 1
9	APOD	Apolipoproteína D
10	CHI3L1	Proteína similar a la quimiosina 1
11	B2M	Beta-2-microglobulina
12	HES6	Factor de transcripción HES-6
13	TMSB10	Timosina β10
14	EGFR	Receptor del factor de crecimiento epidérmico
15	NTRK2	Receptor de la neurotrofina 2
16	GAPDH	Gliceraldehído-3-fosfato deshidrogenasa
17	RPS18	Proteína ribosómica S18
18	ALDOC	Aldolasa C
19	RPLP1	Proteína ribosómica L7/L12
20	SEC61G	Subunidad gamma del complejo de translocación SEC61

El término de la regulación negativa de la migración de las células T está asociado al gen APOD. Esa proteína regula los procesos inflamatorios y la respuesta inmune. La migración de las células T hacia el entorno tumoral permite la respuesta inmunitaria antitumoral. Sin embargo, la capacidad para regular negativamente este proceso

contribuye a la evasión inmune del GB promoviendo la supervivencia y la progresión tumoral (Wang et al., 2024).

En este escenario también se encontró el término de la regulación de la actividad de la proteína quinasa C (PKC) asociado con el gen EGFR. La activación de la PKC promueve la proliferación celular, la supervivencia y la migración, estos procesos son fundamentales en la progresión y agresividad del glioblastoma. La presencia de EGFR en múltiples términos de enriquecimiento refuerza su papel central en la biología del GBM.

También, el gen GAPDH este asociado con el término S-nitrosilación de peptidil-cisteína, es una enzima clave en el glucolisis, tiene funciones en la regulación de procesos celulares mediante modificaciones postraduccionales como la S-nitrosilación. Estas modificaciones están relacionadas con la señalización celular y la respuesta al estrés oxidativo, necesarios para la adaptación y supervivencia de las células tumorales.

Se evidencia que los procesos celulares y la respuesta al estrés, como la respuesta a la nicotina asociada con el gen B2M. La respuesta a la nicotina está asociada con la modulación de las vías de señalización que afectan el crecimiento tumoral y la inflamación (Li et al., 2022).

Los términos relacionados con el secuestro de actina y la regulación de la migración de células T sugieren que la dinámica del citoesqueleto y la evasión inmune son componentes clave en la progresión del glioblastoma. La comprensión de cómo estas vías son moduladas en el GBM podría abrir nuevas avenidas para terapias dirigidas que restrinjan la invasividad y mejoren la vigilancia inmunológica contra el tumor.

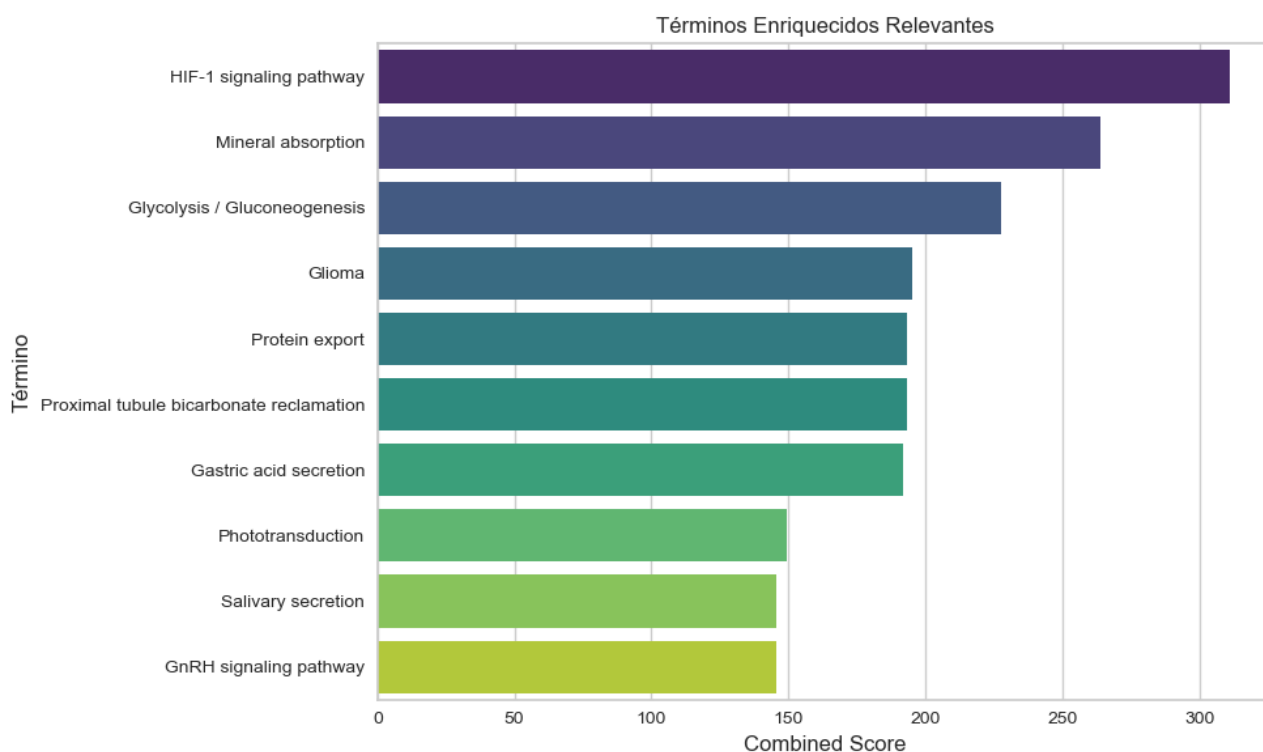


Figura 14. Análisis de enriquecimiento en la base de datos KEGG, Escenario 4

En la Figura 14 se presentan los 10 términos de vías de biológicas contrarrestados con los datos KEGG, estos están relacionados con la patogénesis y la progresión del tumor. La vía de señalización HIF-1 es estadísticamente significativo de acuerdo con la Figura 11, con un p-valor de 0.005 menor que el nivel de significancia. HIF-1 (Factor inducible por hipoxia 1) regula la expresión de genes que facilitan la adaptación a la hipoxia, promoviendo la angiogénesis, el metabolismo anaeróbico, y la supervivencia celular. ALDOC y GAPDH están involucrados en glicólisis, mientras que EGFR juega un papel en la señalización celular que puede contribuir a la proliferación tumoral. La activación de esta vía puede ayudar al GBM a adaptarse al microambiente hipóxico, favoreciendo su agresividad y resistencia a terapias (Wang et al., 2024), como se visualizan en la Figura 15 y 16.

Otra vía se relaciona con la regulación del transporte y absorción de minerales esenciales, como hierro y calcio. El gen FTL codifica la cadena ligera de la ferritina, una proteína clave en la homeostasis del hierro. El hierro es fundamental en varios procesos celulares,

incluido el metabolismo energético y la producción de radicales libres. ATP1A2 codifica una subunidad de la bomba de sodio/potasio ATPasa, crucial para mantener el equilibrio iónico en las células. Alteraciones en la absorción de minerales pueden influir en la viabilidad celular y en la proliferación tumoral (Hoelzgen et al., 2024).

Por otra parte, la glicólisis, es la principal vía metabólica para la producción de energía en las células tumorales, especialmente bajo condiciones hipóxicas. ALDOC y GAPDH son enzimas clave en este proceso. En el contexto del GBM, la preferencia por glicólisis incluso en presencia de oxígeno (conocido como el efecto Warburg) permite al tumor generar energía rápidamente y producir precursores biosintéticos necesarios para su crecimiento. Este enriquecimiento refuerza la idea de que el metabolismo glucolítico es un rasgo distintivo del GBM, contribuyendo a su agresividad (Wirsching et al., 2017).

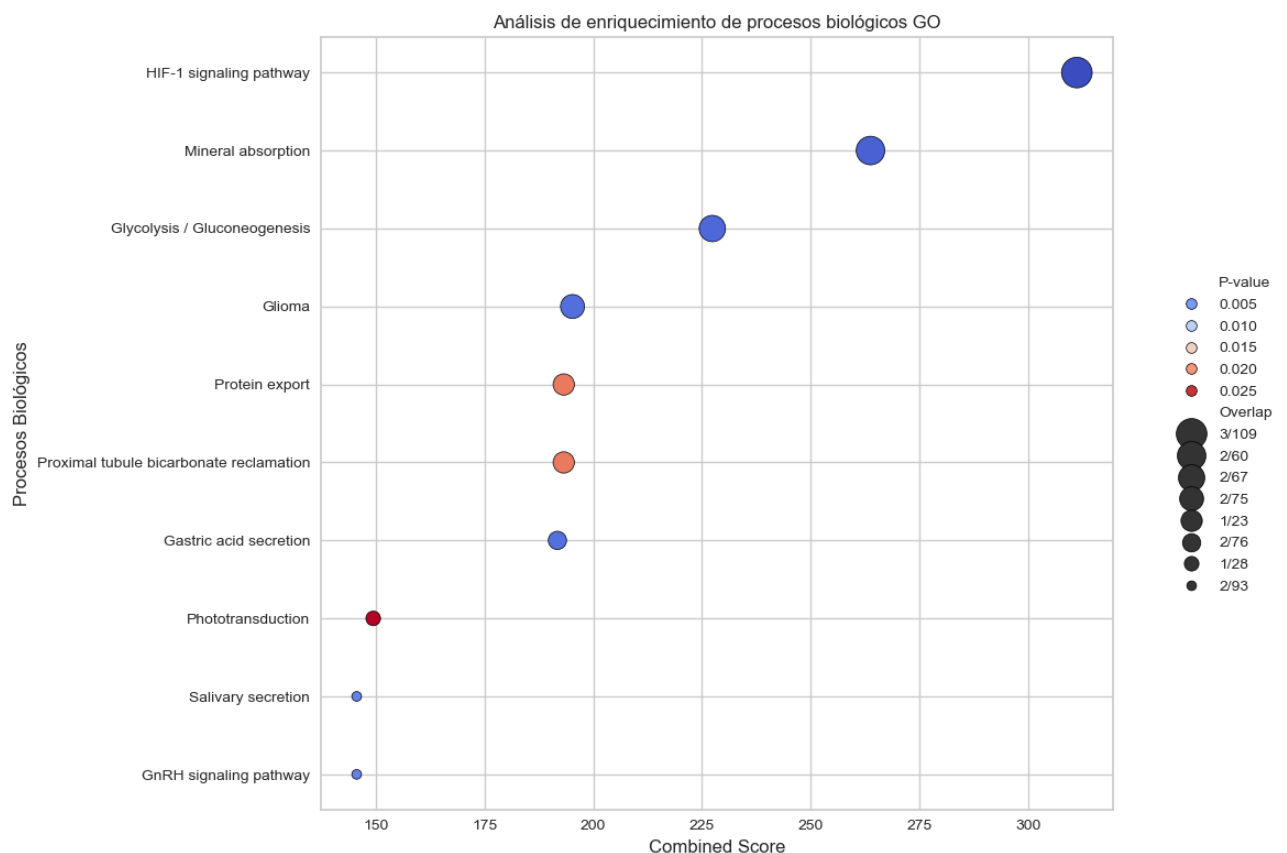


Figura 15. Términos KEGG enriquecidos, Escenario 4.

Este término directamente relacionado con el glioblastoma refuerza la relevancia de EGFR y CALM1 (calmodulina 1) en la patología del GBM. EGFR es un receptor tirosina quinasa frecuentemente sobreexpresado o mutado en GBM, promoviendo la proliferación, supervivencia, y migración celular. CALM1 está implicado en la señalización de calcio, que puede modular diversas vías oncogénicas, incluyendo aquellas mediadas por EGFR. La interacción entre EGFR y la señalización del calcio podría ser crítica en la regulación de procesos malignos en el GBM.

Aunque estos términos parecen menos directamente relacionados con el glioblastoma, los genes implicados (ATP1A2 y CALM1) desempeñan roles en el mantenimiento del equilibrio iónico y la señalización del calcio, procesos que también son importantes en la regulación de la función celular en el entorno tumoral. La secreción de ácido gástrico y la secreción salival están reguladas por la señalización de calcio, lo que sugiere que las alteraciones en estos procesos podrían tener implicaciones más amplias en la fisiología tumoral.

La vía de GnRH puede influir en la proliferación celular a través de mecanismos de señalización intracelular, la activación de esta vía podría afectar la expresión de genes relacionados con la proliferación y la supervivencia celular, en los que participan CALM1 y EGFR.

Red de asociaciones entre genes y términos en KEGG

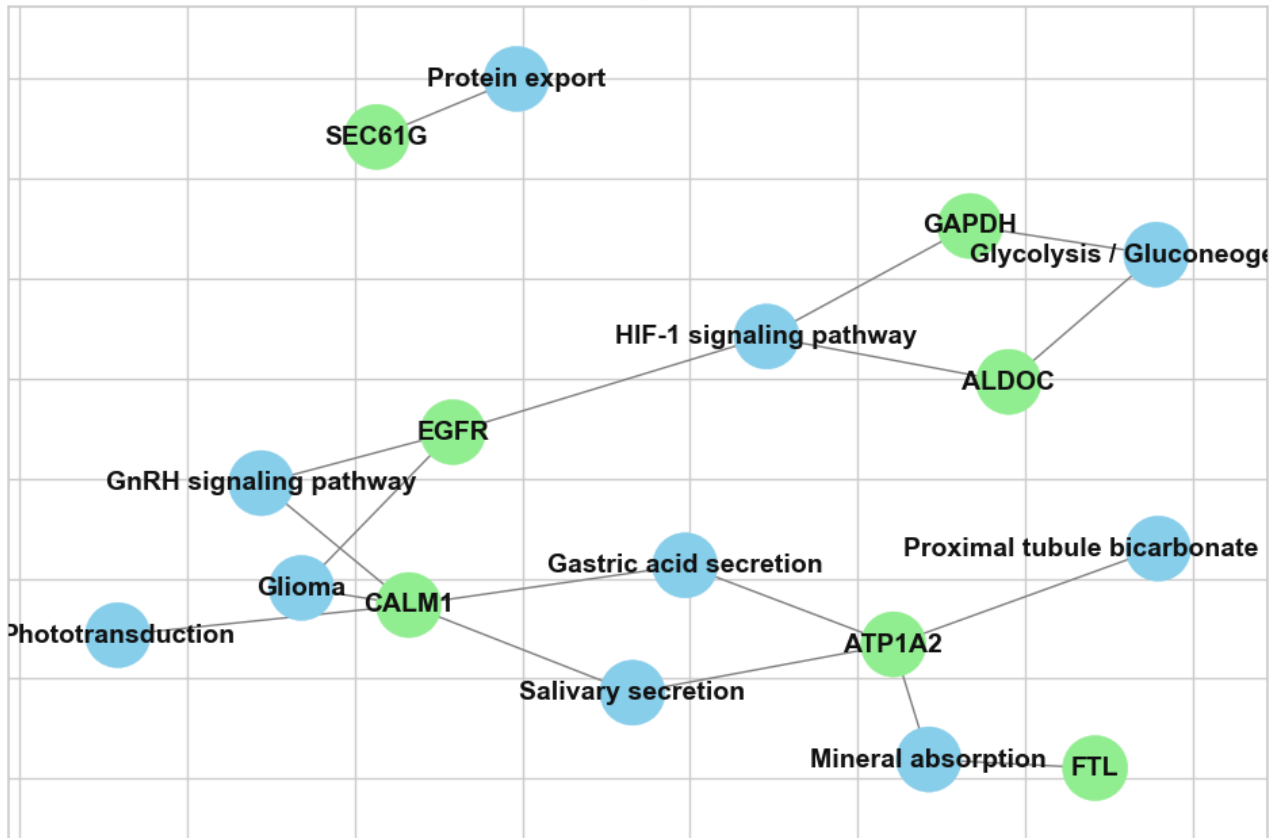


Figura 16. Rutas metabólicas KEGG, Escenario 4.

6. CONCLUSIONES

- A partir de los datos del estudio de Arteaga-Arteaga et al. (2023), se extrajo una lista de 20 genes que presentan expresión diferencial en glioblastoma multiforme en los diferentes escenarios con el mejor método de machine learning.
- Utilizando la paquetería gseapy en Python de libre acceso se realizó el enriquecimiento a la lista de genes definidas utilizando la base de datos KEGG y GO, obteniendo un enriquecimiento significativo en los términos biológicos de menor al nivel de confianza.
- El gen EGFR, está presentes en múltiples términos biológicos, mientras que CALM1 regula el calcio y la transducción de señales hormonales. Estos genes no solo participan en vías de proliferación celular y oncogénesis, sino también en procesos de señalización hormonal, metabolismo, respuesta al estrés, e infección viral. Además, se encontró el enriquecimiento de la ruta de la homeostasis de minerales y el metabolismo del hierro.
- La recurrencia del EGFR en múltiples términos de GO de los diferentes escenarios sugiere que es un punto nodal en la regulación de diversas vías biológicas críticas en el glioblastoma. Inhibidores específicos del EGFR o de sus rutas de señalización asociadas podrían ser de gran interés terapéutico en el tratamiento de GBM.

7. RECOMENDACIONES

- Se recomienda realizar estudios adicionales para investigar de manera más profunda las vías de señalización identificadas como enriquecidas en los análisis actuales. Experimentos funcionales, como la inhibición o activación de genes específicos en modelos celulares o animales, podrían proporcionar información valiosa sobre el papel exacto de estos genes en la progresión y heterogeneidad del glioblastoma.
- Los hallazgos de este análisis de enriquecimiento de genes sugieren que las terapias dirigidas a las vías de señalización celular anómalas y a los mecanismos de reparación del ADN podrían ser efectivas en el tratamiento del glioblastoma. Es esencial explorar la posibilidad de desarrollar inhibidores específicos o terapias combinadas que puedan contrarrestar estas vías y mejorar la respuesta al tratamiento.
- Dada la complejidad y heterogeneidad del glioblastoma, es fundamental integrar estos resultados en enfoques de medicina personalizada. Las pruebas genéticas y moleculares de los tumores individuales podrían utilizarse para identificar perfiles de expresión específicos, permitiendo una estratificación de pacientes y tratamientos más precisos basados en las características moleculares de sus tumores.

8. BIBLIOGRAFÍA

- Ahmad, Aljada., Paresh, Dandona. (1997). 5. Nitric Oxide Synthase. *Methods of Molecular Biology*, doi: 10.1385/0-89603-472-0:191.
- Arteaga-Arteaga, H. B., Candamil-Cortés, M. S., Breaux, B., Guillen-Rondon, P., Orozco-Arias, S., & Tabares-Soto, R. (2023). Machine learning applications on intratumoral heterogeneity in glioblastoma using single-cell RNA sequencing data. *Briefings in Functional Genomics*, 22(5), 428-441.
- Aruga, A. (2016). $\alpha\beta$ -T Cells. *Immunotherapy of Cancer: An Innovative Treatment Comes of Age*, 63-73.
- Alter, P. J., Conroy, M. A., Mancil, G. R., & Haydon, T. (2008). A comparison of functional behavior assessment methodologies with young children: Descriptive methods and functional analysis. *Journal of Behavioral Education*, 17, 200-219.
- Bhavnagari, H. M., & Shah, F. D. (2024). In silico analysis of hippo signaling pathway associated microRNAs in breast cancer. *Human Gene*, 39, 201269.
- Boyeau, P., Regier, J., Gayoso, A., Jordan, M. I., Lopez, R., & Yosef, N. (2023). An empirical Bayes method for differential expression analysis of single cells with deep generative models. *Proceedings of the National Academy of Sciences*, 120(21), e2209124120.
- Begagić, E., Pugonja, R., Bečulić, H., Čeliković, A., Tandir Lihčić, L., Kadić Vukas, S., ... & Pojskić, M. (2023). Molecular targeted therapies in glioblastoma multiforme: a systematic overview of global trends and findings. *Brain Sciences*, 13(11), 1602.
- Cremona, M. A., Xu, H., Makova, K. D., Reimherr, M., Chiaromonte, F., & Madrigal, P. (2019). Functional data analysis for computational biology. *Bioinformatics*, 35(17), 3211-3213.
- Darrak, M., Quddusi., Naim, Bajcinca. (2023). Identification of genomic biomarkers and their pathway crosstalks for deciphering mechanistic links in glioblastoma. *Iet Systems Biology*, doi: 10.1049/syb2.12066
- El Atat, O., Naser, R., Abdelkhalek, M., Habib, R. A., & El Sibai, M. (2022). Molecular targeted therapy: a new avenue in glioblastoma treatment. *Oncology Letters*, 25(2), 46.


- Fan, Q., Wang, Y., Cheng, J., Pan, B., Zang, X., Liu, R., & Deng, Y. (2024). Single-cell RNA-seq reveals T cell exhaustion and immune response landscape in osteosarcoma. *Frontiers in Immunology*, 15, 1362970.
- Fernández, A. L. (2014). BIGO: Mejora del análisis de enriquecimiento en grupos de genes. *MoleQla: revista de Ciencias de la Universidad Pablo de Olavide*, (14), 7-3.
- Gray, A., Cui, T., Bell, E. H., McElroy, J., Sebastian, E., Li, F., ... & Chakravarti, A. (2022). MicroRNA-575 acts as a novel oncogene via targeting multiple signaling pathways in glioblastoma. *Experimental and molecular pathology*, 128, 104813.
- Gimenez, M., Marie, S. K. N., Oba-Shinjo, S., Uno, M., Izumi, C., Oliveira, J. B., & Rosa, J. C. (2015). Quantitative proteomic analysis shows differentially expressed HSPB1 in glioblastoma as a discriminating short from long survival factor and NOVA1 as a differentiation factor between low-grade astrocytoma and oligodendroglioma. *BMC cancer*, 15, 1-13.
- hassan Matini, A., sadat Tayebi, M., Rezvani, Z., Vakili, Z., & Kashani, H. H. (2019). Association of EGFR gene mutations exons 18–21 with glioblastoma multiform cancer: A descriptive and cross-sectional study. *Gene Reports*, 17, 100526.
- Hoelzgen, F., Nguyen, T. T., Klukin, E., Boumaiza, M., Srivastava, A. K., Kim, E. Y., ... & Frank, G. A. (2024). Structural basis for the intracellular regulation of ferritin degradation. *Nature Communications*, 15(1), 3802.
- Heckman, C. A., Biswas, T., Dimick, D. M., & Cayer, M. L. (2020). Activated protein kinase C (PKC) is persistently trafficked with epidermal growth factor (EGF) receptor. *Biomolecules*, 10(9), 1288.
- Huawei, Jin., Tian, Tian., Guoping, Shen., Weian, Chen., Miao-Jing, Fan., Qun, He., Da-fei, Dai., Xuan, Zhang., Da, Jung, Liu. (2022). Integrative Genomic and Transcriptomic Analysis of Primary Malignant Gliomas Revealed Different Patterns Between Grades and Somatic Mutations Related to Glioblastoma Prognosis. *Frontiers in Molecular Biosciences*, 9 doi: 10.3389/fmolb.2022.873042
- Jiang, Y., Dai, S., Pang, R., Qin, L., Zhang, M., Liu, H., ... & Li, W. (2024). Single-cell RNA sequencing reveals cell type-specific immune regulation associated with

- human neuromyelitis optica spectrum disorder. *Frontiers in Immunology*, 15, 1322125.
- Lazarowski, A. J., Vitale, A. A., Auzmendi, J. A., & Pomilio, A. B. (2022). Hierro: desde la homeostasis a la muerte por ferroptosis. *Acta bioquímica clínica latinoamericana*, 56(4), 490-513.
- Li, Q., Aishwarya, S., Li, J. P., Pan, D. X., & Shi, J. P. (2022). Gene expression profiling of glioblastoma to recognize potential biomarker candidates. *Frontiers in genetics*, 13 doi: 10.3389/fgene.2022.832742
- Liu, T., Han, X., Zheng, S., Liu, Q., Tuerxun, A., Zhang, Q., ... & Lu, X. (2021). CALM1 promotes progression and dampens chemosensitivity to EGFR inhibitor in esophageal squamous cell carcinoma. *Cancer Cell International*, 21, 1-12.
- López, T., Di Persia, L. E., & Milone, D. H. (2021). Procesamiento de señales en grafos para medir distancias en la ontología de genes. In *XXIV Concurso de Trabajos Estudiantiles (EST 2021)-JAIIO 50 (Modalidad virtual)*.
- Mahin, K. F., Robiuddin, M., Islam, M., Ashraf, S., Yeasmin, F., & Shatabda, S. (2022). PanClassif: Improving pan cancer classification of single cell RNA-seq gene expression data using machine learning. *Genomics*, 114(2), 110264.
- Marelli, M. M., Moretti, R. M., Mai, S., Müller, O., Van Groeninghen, J. C., & Limonta, P. (2009). Novel insights into GnRH receptor activity: role in the control of human glioblastoma cell proliferation. *Oncology reports*, 21(5), 1277-1282.
- Medeiros, D. M. (2014). Role of the Menkes ATPase in the Absorption of Both Copper and Iron1. *The Journal of nutrition*, 144(1), 3-4.
- Palacios Paredes, L. F., & Silva, C. (2020). Gliomas de Alto Grado del Adulto, *Biología Molecular (Parte I)*. *Oncología (Guayaquil)*, 249-279.
- Rodriguez, J. C. (s/f). *Capítulo 1 Análisis funcional*. Github.io. Recuperado el 23 de julio de 2024, de https://jcrodriguez1989.github.io/tesis_doctoral/1-cap-af.html
- Siavashani, E. S., Ashrafí, M. R., Ghabeli, H., Heidari, M., & Garshasbi, M. (2023). Novel homozygous frameshift variant in the ATCAY gene in an Iranian patient with Cayman cerebellar ataxia; expanding the neuroimaging and clinical features: a case report. *BMC Medical Genomics*, 16(1), 226.

- The Cancer Genome Atlas Research Network. (2008). Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature*, 455(7216), 1061–1068. <https://doi.org/10.1038/nature07385>.
- Tulika, Kakati., Dhruva, K., Bhattacharyya., Jugal, Kalita. (2019). DEGnet: Identifying Differentially Expressed Genes Using Deep Neural Network from RNA-Seq Datasets. 130-138. doi: 10.1007/978-3-030-34872-4_15.
- Sharma, S., & Kumar, P. (2023). Dissecting the functional significance of HSP90AB1 and other heat shock proteins in countering glioblastomas and ependymomas using omics analysis and drug prediction using virtual screening. *Neuropeptides*, 102, 102383.
- Viera, I. C., Castellanos, M. M. S., Zardón, M. M. A., Ortiz, C. M., & Cáceres, D. J. L. H. (2010). Un acercamiento a la ontología de genes y sus aplicaciones. *Centro Nacional de Genética Médica MINSAP*, 1-25.
- Wang, Y., Wang, Y., Wang, S., Wang, C., Tang, Y., Zhang, C., ... & Lin, N. (2024). Comprehensive analysis of CYBB as a prognostic marker and therapeutic target in glioma: A bioinformatics approach. *Heliyon*, 10(8).
- Wirsching, H. G., & Weller, M. (2017). Glioblastoma. *Malignant Brain Tumors: State-of-the-Art Treatment*, 265-288.
- Wirsching, H. G., Galanis, E., & Weller, M. (2016). Glioblastoma. *Handbook of clinical neurology*, 134, 381–397. <https://doi.org/10.1016/B978-0-12-802997-8.00023-2>.
- Wijesooriya, K., Jadaan, S. A., Perera, K. L., Kaur, T., & Ziemann, M. (2021). Guidelines for reliable and reproducible functional enrichment analysis. *BioRxiv*, 2021-09.
- Wu, C. H., Zhou, X., & Chen, M. (2024). The curses of performing differential expression analysis using single-cell data. *bioRxiv*, 2024-05.

9. ANEXOS

ANEXO 1: Genes relevantes obtenidas del agrupamiento del modelo LRST por Arteaga-Arteaga et al. 2023 para el escenario 1.

jupyter Analisis de Enriquecimiento Escenario 1 (autosaved)  Logou

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3 (ipykernel)

```

In [14]: #dado que se usa model.coef_ se recomienda usar esta forma con el valor absoluto
#Porque Los positivos son para una clase, y Los negativos para La otra

feature_importance=pd.DataFrame({'feature':list(features2.columns),'feature_imp
feature_importance.sort_values('feature_importance',ascending=False)[:20]
    
```

Out[14]:

	feature	feature_importance
5392	EGFR	0.021252
2716	CALM1	0.018648
877	APOD	0.017311
19795	SPOCK1	0.016415
777	ANXA1	0.016401
16806	PTGDS	0.016139
5416	EIF1	0.015748
22155	VIM	0.015337
12064	MGLL	0.014881
8854	ITM2C	0.014774
16058	PLL	0.014773
8841	ITGB8	0.014581
7904	HES6	0.014277
17812	RPS27L	0.013682
7082	GFAP	0.013250
6799	FTL	0.013224
21374	TRIM2	0.012971
878	APOE	0.012918
787	ANXA5	0.012851
13994	NAV1	0.012836

ANEXO 2: Dataframe de la salida del análisis de enriquecimiento KEGG y GO.

jupyter Analisis de Enriquecimiento Escenario 1 (autosaved) Logou

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3 (ipykernel)

In [32]: enr.results

Out[32]:

	Gene_set	Term	Overlap	P-value	Adjusted P-value	Old P-value	Old Adjusted P-value	Odds Ratio
0	KEGG_2021_Human	Mineral absorption	3/60	0.000028	0.001608	0	0	61.681115
1	KEGG_2021_Human	Ferroptosis	2/41	0.000761	0.021689	0	0	56.811966
2	KEGG_2021_Human	Coronavirus disease	3/232	0.001518	0.026480	0	0	15.220396
3	KEGG_2021_Human	Salmonella infection	3/249	0.001858	0.026480	0	0	14.156385
4	KEGG_2021_Human	Ribosome	2/158	0.010733	0.103201	0	0	14.119658
5	KEGG_2021_Human	Necroptosis	2/159	0.010863	0.103201	0	0	14.029016
6	KEGG_2021_Human	Proximal tubule bicarbonate reclamation	1/23	0.022761	0.180482	0	0	47.746411
7	KEGG_2021_Human	Aldosterone-regulated sodium reabsorption	1/37	0.036374	0.180482	0	0	29.157895
8	KEGG_2021_Human	Type I diabetes mellitus	1/43	0.042152	0.180482	0	0	24.984962
9	KEGG_2021_Human	Neuroactive ligand-receptor interaction	2/341	0.044996	0.180482	0	0	6.437561
10	KEGG_2021_Human	Carbohydrate digestion and absorption	1/47	0.045987	0.180482	0	0	22.807780
11	KEGG_2021_Human	Arginine and proline metabolism	1/50	0.048853	0.180482	0	0	21.408163
12	KEGG_2021_Human	Endocrine and other factor-regulated calcium r...	1/53	0.051710	0.180482	0	0	20.170040
13	KEGG_2021_Human	Legionellosis	1/57	0.055508	0.180482	0	0	18.725564

	Gene_set	Term	Overlap	P-value	Adjusted P-value	Old P-value	Old Adjusted P-value
52	GO_Biological_Process_2021	regulation of protein kinase C activity (GO:19...	1/5	0.004990	0.050613	0	0 262.
41	GO_Biological_Process_2021	alpha-beta T cell differentiation (GO:0046632)	1/5	0.004990	0.050613	0	0 262.
51	GO_Biological_Process_2021	regulation of low-density lipoprotein particle...	1/5	0.004990	0.050613	0	0 262.
50	GO_Biological_Process_2021	neutral lipid catabolic process (GO:0046461)	1/5	0.004990	0.050613	0	0 262.
49	GO_Biological_Process_2021	positive regulation of protein kinase C activi...	1/5	0.004990	0.050613	0	0 262.
48	GO_Biological_Process_2021	positive regulation of prostaglandin biosynthe...	1/5	0.004990	0.050613	0	0 262.
47	GO_Biological_Process_2021	negative regulation of T cell migration (GO:20...	1/5	0.004990	0.050613	0	0 262.
46	GO_Biological_Process_2021	negative regulation of sterol biosynthetic pro...	1/5	0.004990	0.050613	0	0 262.
45	GO_Biological_Process_2021	intracellular sequestering of iron ion (GO:000...	1/5	0.004990	0.050613	0	0 262.