



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



Escola Tècnica Superior
d'Enginyeria Agronòmica i del Medi Natural

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Escuela Técnica Superior de Ingeniería Agronómica
y del Medio Natural

Caracterización funcional del Melanoma Metastásico
Cerebral (MMC) mediante el análisis masivo de datos de
expresión génica

Trabajo Fin de Grado

Grado en Biotecnología

AUTOR/A: Magán Torres, Cristina

Tutor/a: Forment Millet, José Javier

Cotutor/a externo: ANTON BERNAT, GONZALO

Director/a Experimental: PERPIÑA CLERIGUES, CARLA

CURSO ACADÉMICO: 2022/2023



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



Escola Tècnica Superior
d'Enginyeria Agronòmica
i del Medi Natural

Caracterización funcional del Melanoma Metastásico Cerebral (MMC) mediante el análisis masivo de datos de expresión génica

TRABAJO FINAL DE GRADO

Grado en Biotecnología

Curso académico 2022/2023

Universidad Politécnica de Valencia- Escuela Técnica Superior de Ingeniería
Agronómica y del Medio Natural ETSIAMN

Autor: Cristina Magán Torres

Tutor académico: Prof. D. José Javier Forment Millet

Tutores externos: Dña. Carla Perpiñá Clérigues

D. Gonzalo Antón Bernat

Centro de Investigación Príncipe Felipe



PRINCIPE FELIPE

CENTRO DE INVESTIGACION



Unidad de
Bioinformática y
Bioestadística

Valencia, Junio 2023

Caracterización funcional del Melanoma Metastásico Cerebral (MMC) mediante el análisis masivo de datos de expresión génica

RESUMEN

El melanoma es la tercera fuente más común de metástasis cerebrales, superada sólo por el cáncer de pulmón y de mama, y es el tumor sólido con mayor propensión a dirigirse al cerebro. La incidencia de metástasis cerebrales detectadas clínicamente en pacientes con melanoma avanzado es de aproximadamente el 40%, con un porcentaje mayor identificado en la autopsia. Estos pacientes, con melanoma metastásico cerebral (MMC), presentan la mayor morbilidad y mortalidad de los pacientes de melanoma.

Aunque en los últimos años se han realizado nuevos avances en la comprensión del melanoma metastásico cerebral que se pueden explotar con fines terapéuticos, sigue existiendo una gran necesidad médica no cubierta. Estos pacientes continúan teniendo pocas opciones clínicas, además, el conocimiento de los mecanismos moleculares y celulares responsables de la metástasis cerebral del melanoma sigue siendo limitado.

El objetivo de este trabajo es identificar biomarcadores de diagnóstico y pronóstico, que mejoren el tratamiento de los pacientes con MMC. Para ello, se llevó a cabo una revisión sistemática que permitió la selección de estudios centrados en MMC. A continuación, en cada uno de estos estudios se aplicó MetaFun, una herramienta web bioinformática para la caracterización molecular de enfermedades humanas. La estrategia bioinformática incluyó: un análisis exploratorio de los datos, un análisis de expresión diferencial entre los grupos de enfermos y controles, y un análisis de enriquecimiento funcional que permite detectar conjuntos de genes que se sobreexpresan o infraexpresan coordinadamente, compartiendo rutas y funciones biológicas.

Por último, el conjunto de estos resultados funcionales se ha metaanalizado obteniendo un perfil consenso sobre qué funciones están más o menos activadas en los pacientes con MMC. Analizando los resultados a nivel de términos GO, se detectó, a grandes rasgos, un enriquecimiento de las funciones relacionadas con la proliferación, replicación del ADN y división celular en los pacientes con melanoma. Frente a una relación de los procesos de queratinización, desarrollo epidermal y adhesión celular con los pacientes sanos.

Este mayor conocimiento de la enfermedad contribuye a un enfoque adaptado a la medicina personalizada de precisión, relacionando este trabajo con el Objetivo de Desarrollo Sostenible 3 (ODS 3) de Salud y Bienestar de la Agenda 2030 (Organización de las Naciones Unidas).

Palabras clave: Melanoma metastásico cerebral (MMC); metaanálisis; transcriptómica; medicina de precisión; biomarcadores.

Functional characterization of Melanoma Brain Metastases (MBM) by massive analysis of gene expression data

SUMMARY

Melanoma is the third most common cause of brain metastases, surpassed only by lung and breast cancer, and is the solid tumor with the highest propensity to target the brain. The incidence of clinically detected brain metastases in patients with advanced melanoma is approximately 40%, with an even higher percentage detected postmortem. These patients, with melanoma brain metastase (MBM), have the highest morbidity and mortality over all melanoma patients.

Although in recent years, new advances have been made in the understanding of brain metastatic melanoma that can be exploited for therapeutic purposes, there remains a large unmet medical need. These patients continue to have few clinical options; moreover, knowledge of the molecular and cellular mechanisms responsible for melanoma brain metastasis remains limited.

The aim of this work is to identify diagnostic and prognostic biomarkers that improve the treatment of patients with MBM. To this end, a systematic review was carried out to select studies focused on MBM. MetaFun, a bioinformatics web tool for the molecular characterization of human diseases, was then applied to each of these studies. The bioinformatics strategy included: an exploratory analysis of the data, a differential expression analysis between the disease and control groups, and a functional enrichment analysis to detect sets of genes that are coordinately over- or under-expressed, that share biological pathways and functions.

Finally, the set of these functional results has been meta-analyzed to obtain a consensus profile of which functions are more or less activated in patients with MBM. Analyzing the results at the level of GO terms, we detected, broadly speaking, an enrichment of functions related to proliferation, DNA replication and cell division in patients with melanoma. In contrast, the processes of keratinization, epidermal development and cell adhesion were related to healthy patients.

This further understanding of the disease would contribute to a tailored approach to personalized precision medicine, relating this work to Goal 3 on good health and well-being of the 2030 Agenda.

Keywords: Melanoma brain metastases (MBM); meta-analysis; transcriptomics; precision medicine; biomarkers.

Autor: Cristina Magán Torres

Tutor académico: Prof. D. José Javier Forment Millet

Tutores externos: Dña. Carla Perpiñá Clérigues y D. Gonzalo Antón Bernat

Valencia, Junio 2023

ÍNDICE

1. INTRODUCCIÓN	1
1.1. Enfermedad del cáncer de melanoma	1
1.1.1. Epidemiología y diferencias entre sexos	1
1.1.2. Melanoma Metastásico Cerebral (MMC).....	2
1.1.3. Alternativas terapéuticas	2
1.2. Datos ómicos y tecnologías de alto rendimiento	2
1.2.1. Transcriptómica.....	4
1.2.1.1. Tecnología microarrays	5
1.2.1.2. RNA-Sequencing	6
1.2.2. Análisis transcriptómicos	6
1.2.3. Caracterización funcional.....	7
1.2.4. Metaanálisis funcional	7
2. OBJETIVOS	9
3. MATERIAL Y MÉTODOS	10
3.1. Revisión sistemática y selección de estudios	10
3.2. Procesamiento y estandarización de los estudios	11
3.2.1. Evaluación de la señal y preprocesamiento	12
3.2.2. Anotación	13
3.2.3. Estandarización de grupos experimentales	13
3.3. MetaFun: herramienta web para la caracterización molecular de enfermedades humanas	14
3.3.1. Análisis individual de los estudios	14
3.3.1.1. Análisis exploratorio de los datos.....	14
3.3.1.2. Análisis de expresión diferencial	14
3.3.1.3. Análisis de enriquecimiento funcional	15
3.3.2. Metaanálisis funcional	15
4. RESULTADOS	16
4.1. Revisión sistemática y selección de estudios	16
4.2. Procesamiento y estandarización de los estudios	17
4.3. Análisis individual de los estudios con MetaFun	19
4.3.1. Procesamiento de los datos	19
4.3.2. Análisis de expresión diferencial.....	24
4.3.3. Análisis de enriquecimiento funcional	26
4.4. Metaanálisis con Meta Fun	26
5. DISCUSIÓN	33
6. CONCLUSIONES	36
7. REFERENCIAS	37
8. ANEXOS	41
Anexo I. Relación del trabajo con los objetivos de desarrollo sostenible de la agenda 2030.....	41
Anexo II. Tabla suplementaria.....	42

ÍNDICE DE FIGURAS

Figura 1. Dogma Central de la Biología Molecular. Sucesos que conforman el proceso de transmisión de la información genética desde la molécula de ADN hasta la codificación de proteínas.	3
Figura 2. Representación esquemática de la cascada ómica y de los diferentes aspectos estudiados en cada área de investigación.	4
Figura 3. Comparativa características de la herramienta de Microarray frente a RNAseq	5
Figura 4. Diagrama del flujo de información PRISMA 2020 para estudios identificados en bases de datos, registros y otros métodos a través de las diferentes fases de una nueva revisión sistemática	11
Figura 5. Diagrama de flujo de la revisión sistemática del estudio.....	17
Figura 6. Gráficos MA-plots para tres muestras aleatorias del estudio GSE46517 comparando dicha muestra con la de referencia.....	18
Figura 7. Diagramas de cajas del estudio GSE46517 antes de normalizar (izquierda) y tras aplicar la normalización del logaritmo en base 2 (derecha).....	18
Figura 8. Diagramas de cajas del estudio GSE60464 antes de normalizar (izquierda) y tras aplicar la normalización del logaritmo en base 2 (derecha).....	19
Figura 9. Diagrama de barras mostrando la distribución del tipo de muestras (caso control) en cada estudio.	19
Figura 10. Diagrama de cajas en el que están representados los valores normalizados de las muestras empleadas en el estudio GSE60464, coloreadas según caso y control.....	20
Figura 11. Análisis de componentes principales (PCA) de las muestras del estudio GSE60464.	20
Figura 12. Clustering de las muestras del estudio GSE60464. Cada grupo experimental aparece representado de un color.....	21
Figura 13. Diagrama de cajas en el que están representados los valores normalizados de las muestras empleadas en el estudio GSE46517 Comparación 1, coloreadas según caso y control.	21
Figura 14. Análisis de componentes principales (PCA) de las muestras del estudio GSE46517 Comparación 1.	22
Figura 15. Clustering de las muestras del estudio GSE46517 Comparación 1. Cada grupo experimental aparece representado de un color.	22

Figura 16. Diagrama de cajas en el que están representados los valores normalizados de las muestras empleadas en el estudio GSE46517 Comparación 2, coloreadas según caso y control.	23
Figura 17. Análisis de componentes principales (PCA) de las muestras del estudio GSE46517 Comparación 2 con la muestra GSM1131685.....	23
Figura 18. Análisis de componentes principales (PCA) de las muestras del estudio GSE46517 Comparación 2.	24
Figura 19. Clustering de las muestras del estudio GSE46517 Comparación 2. Cada grupo experimental aparece representado de un color.	24
Figura 20. Término GO:0003678 (DNA helicase activity). A: Gráfico de bosque; B: Funnel plot.	30
Figura 21. Término GO:0006260 (DNA replication). A: Gráfico de bosque; B: Funnel plot.	31
Figura 22. Término GO:000628 (DNA repair). A: Gráfico de bosque; B: Funnel plot.....	31
Figura 23. Término GO:0030216 (Keratinocyte differentiation). A: Gráfico de bosque; B: Funnel plot.	31
Figura 24. Término GO:0042611 (MHC protein complex). A: Gráfico de bosque; B: Funnel plot.	32
Figura 25. Término GO:0098632 (Cell-cell adhesion mediator activity). A: Gráfico de bosque; B: Funnel plot.	32

ÍNDICE DE TABLAS

Tabla 1. Criterios de inclusión y exclusión empleados en la identificación de estudios candidatos; n, tamaño muestral.	11
Tabla 2. Comparaciones realizadas en cada estudio, determinando los grupos experimentales según su diagnóstico.	13
Tabla 3. Estudios seleccionados tras la revisión sistemática.	16
Tabla 4. Resumen del número de genes diferencialmente expresados en cada estudio. Clasificados en UP Regulated y DOWN Regulated según su sobreexpresión. Tabla proporcionada por MetaFun.....	25
Tabla 5. Representación de los top 10 genes expresados diferencialmente en el estudio GSE60464 atendiendo al P value y logaritmo de la magnitud de cambio (logFC). Se indica también su identificador Entrez, nombre del gen y su P-valor ajustado.	25
Tabla 6. Representación de los top 10 genes expresados diferencialmente en el estudio GSE46517 Comparación 1 atendiendo al P value y logaritmo de la magnitud de cambio (logFC). Se indica también su identificador Entrez, nombre del gen y su P-valor ajustado.....	25
Tabla 7. Representación de los top 10 genes expresados diferencialmente en el estudio GSE46517 Comparación 2 atendiendo al P value y logaritmo de la magnitud de cambio (logFC). Se indica también su identificador Entrez, nombre del gen y su P-valor ajustado.....	26
Tabla 8. Resumen del número de funciones con diferencia de expresión significativa. Clasificados en UP Regulated y DOWN Regulated según su sobreexpresión. Tabla proporcionada por MetaFun.	26
Tabla 9. Resumen resultados de metaanálisis, con el número total de funciones identificadas y sus valores LOR (Logaritmos de Odds Ratios).	27
Tabla 10. Resumen de los términos GO significativos en el metaanálisis con LOR positivo, ordenados de mayor a menor.MF: molecular function, BP: biological process, CC: cellular component.	27
Tabla 11. Resumen de los términos GO significativos en el metaanálisis con LOR negativo, ordenados de mayor magnitud de cambio a menor. MF: molecular function, BP: biological process, CC: cellular component.	29

ÍNDICE DE ABREVIATURAS

ADN: ácido desoxirribonucleico

ARN: ácido ribonucleico

ARNm: ARN mensajero

ARNr: ARN ribosómico

ARNt: ARN transferente

GEO: *Gene Expression Omnibus*

GO: *Gene Ontology*

GSEA: *Gene Set Enrichment Analysis*

KEGG: *Kyoto Encyclopedia of Genes and Genomes*

LogFC: logaritmo de la magnitud de cambio

LOR: Logaritmo Odds Ratio

MMC: Melanoma Metastásico Cerebral

NCBI: *National Center for Biotechnology Information*

NGS: tecnología de alto rendimiento

ORA: Over-representation analysis

PCA: análisis de componentes principales

PRISMA: *Preferred Reporting Items for Systematic reviews and Meta-Analyses*

RNA-seq: *RNA Sequencing*

scRNA-seq : *Single-cell RNA sequencing*

SD: desviación estándar

SNV: *Single Nucleotide Variant*

WBRT: radioterapia total del cerebro

1. INTRODUCCIÓN

1.1. Enfermedad del cáncer de melanoma

El cáncer comprende un grupo de enfermedades, caracterizadas por la proliferación autónoma de células neoplásicas que presentan varias alteraciones, incluyendo mutaciones e inestabilidad genética (Monzón et al., 2011). Esta enfermedad es una de las principales causas de muerte en todo el mundo. Según datos de la Agencia Internacional de Investigación sobre el Cáncer, en el año 2020 se diagnosticaron aproximadamente 19,3 millones de nuevos casos en el mundo, y se prevé que dicha cifra alcance en las dos próximas décadas los 27,0 millones (Sung et al., 2021).

El número de casos y muertes por cáncer aumenta rápidamente a medida que la población crece, envejece y adopta estilos de vida que incrementan el riesgo de padecer esta enfermedad (Torre et al., 2016). Por lo que nos enfrentamos a un enorme problema de salud pública, obligando a adoptar nuevas medidas preventivas.

Existen muchos tipos de cáncer, cada uno con características propias. Sin embargo, todos los tipos comparten una característica fundamental, la proliferación celular descontrolada. El melanoma es un tipo de cáncer derivado de la transformación maligna de los melanocitos, células pigmentadas de tipo dendrítico localizadas en varios lugares anatómicos (Chang et al., 1998). Estas células productoras de pigmento derivan embriológicamente de células madre pluripotentes de la cresta neural, las cuales durante el desarrollo embrionario migran predominantemente a la epidermis diferenciándose en ella, aunque también migran a otros sitios como los ojos, meninges, esófago y mucosas (Ali et al., 2013).

Respecto a la clasificación del melanoma se puede atender a distintos parámetros, por una parte, en función de la localización primaria del melanoma se encuentran: cutáneo, el más frecuente derivado de los melanocitos de la epidermis; mucoso, a partir de los melanocitos que residen en las membranas mucosas; y úveal, de los melanocitos que residen en el estroma ocular y de sitio primario desconocido (Ali et al., 2013). Por otra parte, atendiendo a las características fenotípicas e histológicas, se obtiene la clasificación más empleada en la práctica clínica, que incluyen los melanomas lentigo maligno, de extensión superficial, nodular y lentiginoso acral (Chang et al., 1998; Scatena et al., 2021).

Hoy en día, el melanoma se considera una enfermedad multifactorial que surge de la interacción entre la susceptibilidad genética y la exposición ambiental. Debido a su efecto genotóxico, la exposición a los rayos UV es considerado el factor de riesgo ambiental más importante, potencialmente relacionado con el desarrollo de melanoma maligno. Por parte del huésped, los factores de riesgo más destacados son el número de nevos melanocíticos, la historia familiar y la susceptibilidad genética (Rastrelli et al., 2014). Un mayor conocimiento de estos factores de riesgo mejorará el diagnóstico y prevención de la enfermedad.

1.1.1. Epidemiología y diferencias entre sexos

El melanoma cutáneo es, con diferencia, el subtipo más frecuente, ya que representa más del 90% de los casos (Rastrelli et al., 2014). En contraste con el continuo aumento de melanomas

cutáneos, la incidencia de melanomas oculares y mucosos han permanecido estables (Chang et al., 1998). El melanoma cutáneo es el decimonoveno cáncer más común en todo el mundo, con unas tasas de incidencia de 3.0-3.8 por 100.000 personas (Sung et al., 2021). Sin embargo, la incidencia varía mucho de un país a otro. Estas diferencias se atribuyen a variaciones en el fenotipo racial de la piel, así como, a diferencias en la exposición al sol (Ali et al., 2013).

A diferencia de otros tumores sólidos afecta sobre todo a personas jóvenes y de mediana edad. La incidencia aumenta linealmente a partir de los 25 años hasta la edad de 50, y después disminuye, principalmente en el sexo femenino (Leonardi et al., 2018). Analizando los datos de incidencia en relación con el sexo, las mujeres son más frecuentes en los grupos de edad más jóvenes, mientras que el sexo masculino prevalece a partir de los 55 años (Rastrelli et al., 2014).

1.1.2. Melanoma Metastásico Cerebral (MMC)

El melanoma cutáneo es una de las formas más agresivas de cáncer de piel y una de las principales causas de mortalidad debido a su poder metastásico (Leonardi et al., 2018), que se define como la capacidad de las células cancerígenas de invadir tejidos próximos y diseminarse hacia órganos distantes, reproduciendo el tumor (Massagué, 2009).

En concreto, el Melanoma Metastásico Cerebral (MMC) objeto de nuestro estudio, puede ser un paso mortal en la progresión del cáncer. El melanoma es la tercera fuente más común de metástasis cerebrales, sólo superada por el cáncer de pulmón y de mama, y es el tumor sólido con mayor propensión a dirigirse al cerebro. La incidencia de metástasis cerebrales detectadas clínicamente en pacientes con melanoma avanzado es de aproximadamente el 40%, con un porcentaje mayor identificado en la autopsia. Estos pacientes presentan la mayor morbilidad y mortalidad de los pacientes de melanoma (Karz et al., 2022).

1.1.3. Alternativas terapéuticas

Las opciones de tratamiento para los pacientes con metástasis cerebrales son limitadas. Dependen en gran medida del número y el tamaño de las lesiones, así como de la extensión extracraneal de la metástasis (Bafaloukos & Gogas, 2004). Las recomendaciones generales basadas en informes de series clínicas incluyen la cirugía del tumor, la radioterapia total del cerebro (WBRT), la radiocirugía y la quimioterapia o un enfoque combinado que mejore la recuperación y aumente la esperanza de vida (Bafaloukos & Gogas, 2004).

Aunque en los últimos años la inclusión de pacientes con metástasis cerebrales en ensayos clínicos, el estudio del papel de los astrocitos y las adaptaciones metabólicas al microambiente cerebral han permitido detectar nuevas vulnerabilidades que podrían explotarse con fines terapéuticos, sigue existiendo una gran necesidad médica no cubierta. Estos pacientes cuentan con pocas opciones de tratamiento, además, el conocimiento de los mecanismos moleculares y celulares responsables de la metástasis cerebral del melanoma continúa siendo limitado (Karz et al., 2022).

1.2. Datos ómicos y tecnologías de alto rendimiento

La información genética se almacena en el ácido desoxirribonucleico (ADN), una molécula constituida por dos cadenas antiparalelas de nucleótidos sucesivos que consisten en una

molécula de azúcar-fosfato con una base nitrogenada unida (Adenina (A), Timina (T), Guanina (G) o Citosina (C)) (Watson y Crick, 1953). En las células se localiza en el núcleo y en las mitocondrias, y tiene la capacidad de generar copias de sí misma con el objetivo de conservar la información genética, proceso conocido como replicación. Además, mediante la transcripción, el ADN proporciona el molde para la producción del ácido ribonucleico (ARN) (Salazar Montes et al., 2013), el cual se diferencia del ADN por su azúcar, que es una ribosa en lugar de una desoxirribosa, por estar formado por una sola cadena y por sustituir la base nitrogenada Timina (T) por Uracilo (U). Dentro del RNA existen varios tipos, si la fracción del ADN utilizada corresponde con un gen, el ARN obtenido se denomina ARN mensajero (ARNm), el cual se utiliza de nuevo como molde para codificar proteínas por medio del proceso de traducción. Este proceso consiste en la síntesis de una cadena polipeptídica, dentro de la cual el orden de aminoácidos refleja la sucesión de bases en la molécula de ARNm. Estas proteínas son las encargadas de realizar la gran mayoría de funciones celulares.

Todas las células, desde las bacterias hasta los seres humanos, expresan su información genética de este modo, un principio tan fundamental que se denomina el dogma central de la biología molecular, postulado por F.Crick y más tarde modificado por H.Temin (Lozano Teruel, 2005) (Figura 1).

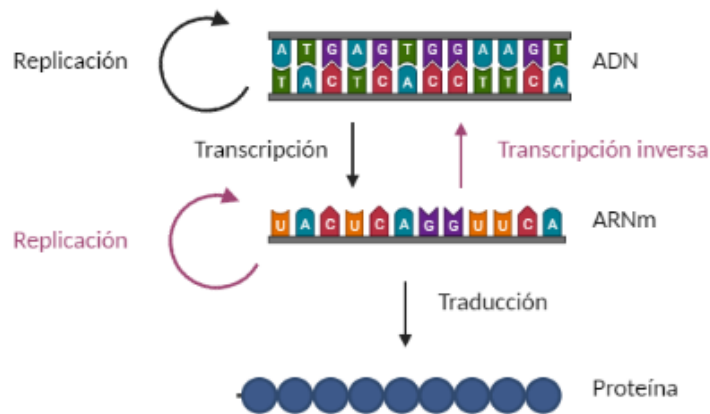


Figura 1. Dogma Central de la Biología Molecular. Sucesos que conforman el proceso de transmisión de la información genética desde la molécula de ADN hasta la codificación de proteínas. (Elaboración propia con BioRender.com adaptada de Lozano Teruel (2005)).

Durante finales del siglo XX, se desarrollaron tecnologías basadas en estos procesos de la biología molecular, estas tecnologías se conocen como ciencias ómicas y constituyen un conjunto de metodologías destinadas a la obtención masiva de datos del tipo de molécula de interés. La integración e interacción de los distintos componentes de la cascada ómica han dado lugar al concepto de biología de sistemas que permite obtener una perspectiva holística del individuo (Figura 2). Estas ciencias presentan un enorme potencial en el campo de la salud, en concreto en la biomedicina, contribuyendo a establecer las bases para la aplicación de la conocida como Medicina Personalizada de Precisión (Tebani et al., 2016).

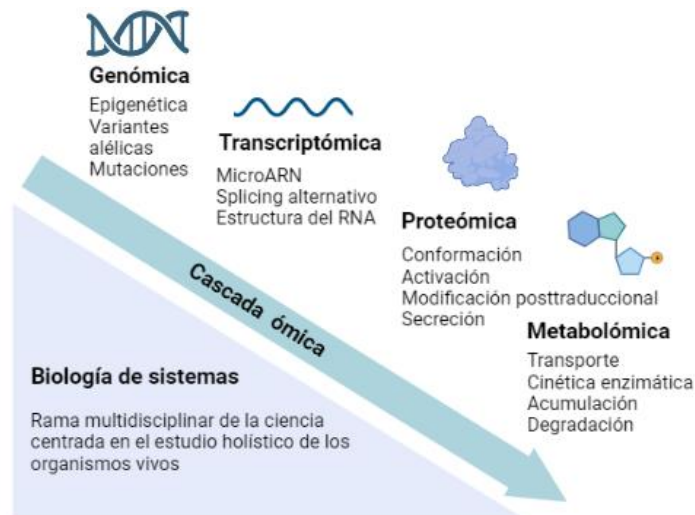


Figura 2. Representación esquemática de la cascada ómica y de los diferentes aspectos estudiados en cada área de investigación. (Elaboración propia con Biorender.com basada en Bedia (2018)).

Existen tantas ciencias ómicas como elementos biológicos o moleculares susceptibles de ser estudiados por estas tecnologías ómicas. Las ciencias ómicas establecidas incluyen la genómica (detección de genes mediante el estudio del genoma completo), la epigenómica (estudio de los cambios epigenéticos), la transcriptómica (estudio del ARN de un individuo), la proteómica (estudios de las proteínas) y la metabolómica (estudio de los metabolitos producidos). La lista se ha ido ampliando pero todas ellas se caracterizan por ser tecnologías de alto rendimiento, generando un gran volumen de datos (también denominado *big data*) en un solo experimento realizado a partir de una única muestra (Bedia, 2018).

Debido a la gran cantidad de información biológica generada, se necesitan bases de datos donde poder almacenarlos, clasificarlos y ordenarlos. Por ello, se han generado diferentes repositorios como Gene Expression Omnibus (GEO) (Barrett et al., 2012) y ArrayExpress (Athar et al., 2019). Además, la gestión de esta información requiere del apoyo fundamental de la bioinformática y la bioestadística, enfrentado problemas como la construcción del genoma, selección entre múltiples bases de datos a gran escala y la predicción del efecto de variantes (Tebani et al., 2016).

1.2.1. Transcriptómica

La transcriptómica es la disciplina que estudia el total de los transcritos de ARN de una célula, tanto los ARN codificantes como los no codificantes, es decir, el transcriptoma (Manzoni et al., 2018). El análisis de la medida y de los cambios de expresión de los genes de un organismo (en distintas condiciones, tejidos o tiempos) aporta información sobre la regulación de los mismos y sobre la dinámica del metabolismo celular y tisular (Lowe et al., 2017).

Además, ha sido fundamental para comprender las enfermedades humanas, destacando su uso en diferentes tipos de cáncer, ayudando a entender los procesos de carcinogénesis y de desarrollo y diferenciación celular (Lowe et al., 2017). Los avances tecnológicos de la transcriptómica, junto con la bioinformática, han supuesto un gran impulso en la investigación, con importantes consecuencias clínicas en cuanto al diagnóstico, el pronóstico y el tratamiento

de los pacientes (Cos, 2010). Por ello, la transcriptómica ha sido una disciplina muy extendida, desarrollando técnicas fundamentales para su análisis destacando los microarrays y la secuenciación del ARN (RNA-Seq) (Manzoni et al., 2018). La Figura 3 compara las principales diferencias entre las dos tecnologías que se detallarán en los siguientes apartados.

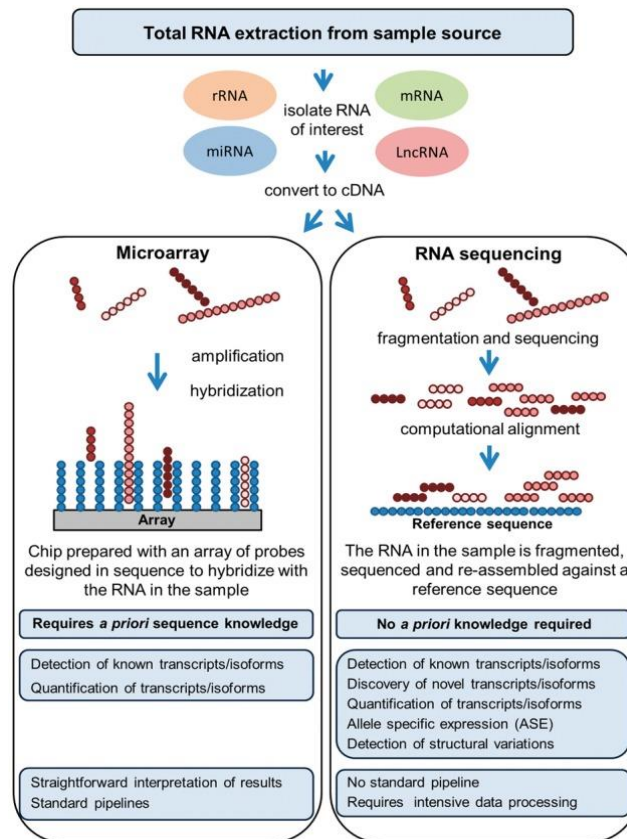


Figura 3. Comparativa características de la herramienta de Microarray frente a RNAseq (Manzoni et al., 2018).

1.2.1.1. Tecnología microarrays

La tecnología de microarrays permite estudiar la expresión de un conjunto definido de transcritos mediante la hibridación de sondas complementarias conocidas en un soporte de vidrio (chip). Según el apareamiento de las bases complementarias entre la muestra y las secuencias de genes en el chip se emite una cantidad de luz que se puede medir, relacionando la emisión con la expresión de cada área conocida del chip, es decir, con la expresión de cada gen (Lowe et al., 2017). Existen diversos tipos de microarrays según las sondas y la biomolécula utilizada, como los arrays de oligonucleótidos punteados o arrays de alta densidad de Affymetrix. Debido a los avances en el diseño y fabricación de arrays se ha mejorado la especificidad de las sondas permitiendo analizar más genes en un único array. También, como consecuencia de la mejora en la detección por fluorescencia, se ha aumentado la precisión de las mediciones de transcritos de baja abundancia (Heller, 2002).

Los microarrays son más limitados en comparación con las estrategias de secuenciación de alto rendimiento (NGS), ya que se basan en el conocimiento *a priori* de la secuencia y en los single nucleotide variant (SNV), mientras que la NGS permite la detección de cambios novedosos. La

tecnología RNA-seq es más exhaustiva, proporcionando conocimientos más profundos y detallados sobre la caracterización y cuantificación de los transcriptomas (Manzoni et al., 2018). Estas comparativas se muestran en la Figura 3.

1.2.1.2. RNA-Sequencing

La tecnología RNA-Seq se basa en el estudio de la secuenciación de ADNc de transcritos obteniendo la abundancia a partir del número de recuentos de cada transcrito, utilizando la secuenciación de alto rendimiento para capturar todas las secuencias. Esta técnica permite evaluar la composición celular completa del ARN, incluyendo el ARNm, el ARNr (ARN ribosómico) y el ARNt (ARN transferente) de manera cuantitativa y con un alto rendimiento (Lowe et al., 2017).

Esta técnica ha avanzado rápidamente, convirtiéndose en la técnica dominante, aunque aún requiere de un complejo procesamiento que incluye el enriquecimiento de los transcritos, la fragmentación, la amplificación y la elección del tipo de secuenciación (single o paired-end, específica de hebra o no específica) (Conesa et al., 2016). Sin embargo, también cuenta con limitaciones debido a la obtención de los resultados como el promedio de todas las células analizadas, rebajando la robustez de los resultados, para ello se han desarrollado nuevas técnicas como Single-cell RNA sequencing (scRNA-seq) (Kaisa et al., 2020).

Existen varias plataformas de secuenciación NGS como Roche, illumina, PacBio, IonTorrent... Todas estas técnicas tienen diferentes químicas de secuenciación y rendimientos que varían en consecuencia en el análisis final de los experimentos (Kaisa et al., 2020). El rendimiento actual de esta técnica permite cuantificar con precisión transcriptomas humanos completos (Lowe et al., 2017).

1.2.2. Análisis transcriptómicos

El rápido avances de las tecnologías de NGS ha traído a la par el desarrollo de numerosos métodos computacionales para el análisis de datos ómicos, variando en función del objetivo del estudio y del tipo de datos a analizar (Nekrutenko & Taylor, 2012).

En los análisis transcriptómicos se requiere del pre-procesamiento de los datos antes del análisis y la interpretación de los resultados. Este incluye un proceso de normalización que elimina la variación sistemática, la transformación de los datos y su filtrado (Allison et al., 2006), obteniendo una matriz de expresión que será el punto de partida para la aplicación de uno o varios de los siguientes métodos de análisis (Tarca et al., 2008):

- Estudio de comparación de clases o expresión diferencial: se realizan para comparar los perfiles de expresión génica de dos o más grupos de estudio, identificando genes diferencialmente expresados en las distintas condiciones experimentales (Phipson et al., 2016).
- Descubrimiento de clases o análisis cluster: consiste en analizar un conjunto dado de perfiles de expresión génica con el objetivo de descubrir subgrupos que compartan características comunes.

- Predicción de clases: el objetivo es clasificar las muestras en función de su perfil de expresión en grupos predefinidos mediante modelos matemáticos.

1.2.3. Caracterización funcional

A partir del listado de genes resultante del análisis de expresión diferencial, es interesante conocer la función o funciones que realizan en la célula. El objetivo de la caracterización funcional es la asociación de cada gen, transcrito, proteína o producto con, por ejemplo, un proceso biológico, una función molecular específica o una ruta de señalización o metabólica. Algunas de las bases de datos que contienen esta información son:

- The Gene Ontology (GO): emplea un vocabulario dinámico y controlado, organiza los términos jerárquicamente en tres categorías: procesos biológicos, componentes celulares y funciones moleculares (Ashburner et al., 2000).
- Kyoto Encyclopedia of Genes and Genomes (KEGG): proporciona información sobre las rutas de señalización, organizándolas en forma de red, asociadas a genes metabolitos (Kanehisa et al., 2021).

Existen distintos métodos para llevar a cabo la caracterización funcional, combinando la información disponible en las bases de datos y los resultados obtenidos en el análisis de expresión diferencial:

- Over-representation analysis (ORA): en este análisis se caracterizan las funciones de un grupo específico de genes y se comparan con las funciones de otro grupo de genes de referencia, buscando determinar si una serie de funciones biológicas están sobrerrepresentadas o enriquecidas en el grupo de genes específicos (Maleki et al., 2020).
- Gene set enrichment analysis (GSEA): este análisis indica cuál es la función biológica, proceso molecular o componente celular detrás de un conjunto de genes con un patrón similar de expresión. Permite enfocarse en dichos grupos de genes y determinar si tienden a sobreexpresarse o subexpresarse coordinadamente (Korotkevich et al., 2016).

1.2.4. Metaanálisis funcional

El metaanálisis es un conjunto de técnicas que buscan agregar y contrastar los resultados de varios estudios relacionados con el mismo tema y analizar las causas de variación entre ellos (Gurevitch et al., 2018). Los resultados del metaanálisis pueden ser cuantitativos o cualitativos, permitiendo revelar los sesgos, puntos fuertes y puntos débiles de los estudios existentes. Estos resultados se pueden utilizar para formular recomendaciones de tratamiento o para orientar el futuro el diseño de ensayos clínicos (Russo, 2007).

El metaanálisis incluye un proceso de revisión sistemática como primer paso, constituyendo una búsqueda exhaustiva de estudios referentes al tema de interés (Gurevitch et al., 2018). Esta revisión se basa en el uso de directrices metodológicas formales para el cribado de los estudios,

que están bien establecidos en muchos campos, por ejemplo, la guía de PRISMA (The Preferred Reporting Items for Systematic reviews and Meta-Analyses) (Liberati et al., 2009).

Tras la selección de los estudios, se determina la magnitud del efecto de la variable de interés y su variabilidad, ya que en función de esta se selecciona el tipo de metaanálisis, siguiendo un modelo de efectos fijos o un modelo de efectos aleatorios (Langan et al., 2019).

La mayoría de los métodos de metaanálisis se han centrado a nivel genómico, aunque dirigir esta metodología a nivel de función biológica permite realizar una mejor interpretación de los resultados, solventando problemas como un reducido tamaño muestral o la heterogeneidad en el diseño de los estudios (García et al., 2016).

2. OBJETIVOS

Aunque en los últimos años se han realizado nuevos avances en la comprensión del melanoma metastásico cerebral que se pueden emplear con fines terapéuticos, sigue existiendo una gran necesidad médica no cubierta. Estos pacientes continúan teniendo pocas opciones clínicas, por ello, el objetivo de este trabajo es *identificar nuevos biomarcadores de diagnóstico y pronóstico, que mejoren el tratamiento de los pacientes con MMC*. Relacionando este trabajo con el ODS 3 de salud y bienestar de la agenda 2030, que busca garantizar una vida sana y promover el bienestar para todas las personas de todas las edades. Este mayor conocimiento de la enfermedad contribuiría a un enfoque adaptado a la medicina personalizada de precisión.

El objetivo general se cumplirá siguiendo los siguientes objetivos específicos:

1. Revisión sistemática y selección de estudios de melanoma metastásico cerebral con datos de expresión de microarrays en los repositorios de Gene Expression Omnibus (GEO).
2. Análisis individual de cada estudio seleccionado mediante MetaFun, una herramienta web bioinformática para la caracterización molecular de enfermedades humanas, incluyendo un análisis de expresión diferencial y un análisis de enriquecimiento funcional.
3. Metaanálisis funcional con la herramienta MetaFun que permitirá la identificación de biomarcadores de diagnóstico y pronóstico de la enfermedad, en el conjunto de los estudios evaluados.

3. MATERIAL Y MÉTODOS

El análisis bioinformático descrito en esta sección se realizó en el lenguaje de programación R (versión 4.2.2) (2022-10-31). El código para cada estudio está disponible en github (https://github.com/cristinamagant/TFG_meta-analysis_MMC.git). Los softwares y paquetes empleados se recogen en el anexo II (Tabla A2).

3.1. Revisión sistemática y selección de estudios

La revisión sistemática es el procedimiento que busca resumir y recopilar la información científica existente respecto de un tema particular (Ferreira González et al., 2011). Este proceso se ha implementado en el trabajo siguiendo las directrices estipuladas en la declaración PRISMA, resultado de una colaboración internacional que tiene como objetivo estandarizar la metodología de las revisiones sistemáticas (Liberati et al., 2009).

El primer paso a realizar corresponde con la definición de la pregunta clínica de interés. Una vez establecida se determinan los criterios de inclusión y exclusión necesarios para la selección de los estudios. Seguidamente, se realiza una búsqueda exhaustiva de literatura en las distintas bases de datos, filtrando de acuerdo con los criterios establecidos (fase de identificación). Tras su selección, se evalúan de manera individual, descartando aquellos que presenten baja calidad o riesgo de sesgo (fases de cribado y elegibilidad). Por último, se descargan los datos de los estudios incluidos para su posterior análisis bioinformático (fase de inclusión) y realización del metaanálisis, en este caso, e interpretación de los resultados.

En detalle, se han recopilado estudios que permitan la caracterización funcional del melanoma metastásico cerebral. La base de datos pública del NCBI (National Center for Biotechnology Information), llamada GEO (Gene Expression Omnibus), se utilizó con la finalidad de buscar estudios de expresión génica que cumplan los requisitos de inclusión mostrados en la Tabla 1.

Para la búsqueda se emplearon palabras clave como "*brain metastatic melanoma*" siendo este un MeSH Term (Medical Subject Headings), determinado tras una búsqueda general de papers relacionados con el tema y fijándose en las palabras claves, siendo "*brain metastatic melanoma*" el término médico que engloba más estudios de interés acorde a los objetivos. En cuanto al tipo de tecnología estudio se incluyeron en un primer momento tanto RNA-seq como microarray, pero tras una exhaustiva búsqueda y revisión de los estudios de RNA-seq se descartaron por el uso de tecnología single-cell, aplicando finalmente el filtro de tipo de estudio como "*expression profiling by array*", trabajando solo con estudios de microarray. También se aplicó el filtro de related series con "*gse*", para centrarnos en la búsqueda de estudios y no otros elementos de las base de datos como muestras individuales, o plataformas tecnológicas. Se seleccionaron exclusivamente estudios en humanos "*Homo sapiens*", se incluyeron los basados en tejidos humanos, excluyendo *cell lines* u organoides, y aquellos estudios donde el diseño experimental se basaba en tratamientos. Respecto al tamaño muestral (n), para el estudio estadístico se requerían un mínimo de 5 muestras (Tabla 1).

Tabla 1. Criterios de inclusión y exclusión empleados en la identificación de estudios candidatos; n, tamaño muestral.

Criterios de inclusión	Criterios de exclusión
"Melanoma Brain Metastases (MBM)"	Otras enfermedades
$n \geq 5$	$n < 5$
<i>Homo sapiens</i>	Otros organismos
Células humanas	Cell lines
Tecnología de Microarrays y RNA-seq	Otras tecnologías

Finalmente, se evaluaron todos los estudios de acuerdo con los criterios de inclusión en la Tabla 1, con el fin de seleccionar aquellos que cumplieran todos los requisitos. El flujo de este proceso se recoge en un diagrama PRISMA (Figura 4).

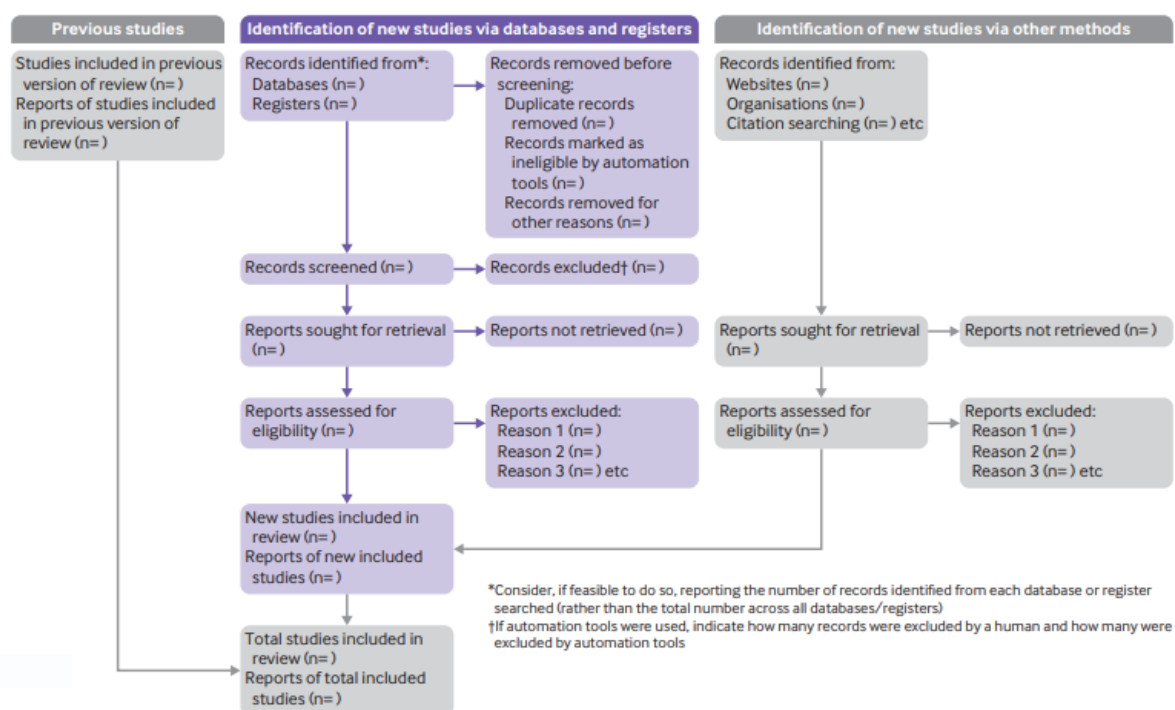


Figura 4. Diagrama del flujo de información PRISMA 2020 para estudios identificados en bases de datos, registros y otros métodos a través de las diferentes fases de una nueva revisión sistemática (Liberati et al., 2009).

3.2. Procesamiento y estandarización de los estudios

Para llevar a cabo el metaanálisis, previamente se procesaron y analizaron los estudios individualmente mediante lenguaje de programación R.

A continuación se obtuvieron los datos de los estudios seleccionados. En este caso los dos estudios se encuentran en la base de datos GEO (Barrett et al., 2013). Para ello se utilizó la función `getGEO()` del paquete `GEOquery` (Sean & Meltzer, 2007). Este paquete sirve como interfaz entre GEO y BioConductor (Huber et al., 2015), un entorno de software de código abierto basado en R diseñado específicamente para el análisis de datos genómicos. De esta manera, se establece una conexión fluida y eficiente entre los datos almacenados en GEO y las herramientas de análisis proporcionadas por BioConductor. Además, como cada estudio pertenece a una plataforma de microarray diferente (GSE46517: Affymetrix,

GSE60464:illumina), se utilizaron los paquetes Affy (Gautier et al., 2004) y lumi (Du et al., 2008) para su lectura y procesado.

3.2.1. Evaluación de la señal y preprocesamiento

Se realizó una evaluación de la señal sobre las muestras de los estudios seleccionados, con el objetivo de determinar la distribución de la expresión de las muestras, valorar la calidad de los datos del estudio y comprobar que no existen patrones anómalos que puedan afectar a los análisis posteriores.

Para este análisis se estudió la calidad de las muestras mediante varios métodos:

- MA-plots: el gráfico obtenido permite visualizar la variabilidad intra y entre estudios comparando todas las muestras con la de referencia. Representa en el eje x la distribución de la relación de intensidad de una muestra en escala logarítmica ('M') trazada en función de la intensidad promedio ('A'), eje y. Su finalidad es determinar si alguna muestra debe ser descartada y si se necesita homogeneizar los valores para las distintas muestras. La mayoría de los puntos debería tener un valor de 0 ajustándose al eje X, si no es el caso se debería aplicar un método de normalización adecuado a los datos ajustando las intensidades desequilibradas (Vafae et al., 2019).
- Diagramas de cajas o *boxplots*: es un método análogo de visualización de la distribución de los niveles de expresión de las muestras en cajas de Tukey. Son gráficos que muestran la distribución de una variable usando cuartiles, de manera visual se puede inferir datos como su dispersión, ubicación y simetría (Vega, n.d.).

En función de los resultados, se procede a completar el proceso de normalización de los datos. Este paso es esencial en el procesamiento de los datos ya que permite su comparabilidad, tras ajustar la expresión de las muestras en la misma escala (Quackenbush, 2002).

Existen varios métodos de normalización en función del tipo de tecnología y ómica. En este trabajo se completó la normalización de los datos seleccionados con la transformación logarítmica en base 2 para homogeneizar las escalas de los distintos grupos de datos. Contando este método con la ventaja de producir un espectro continuo de valores y tratar los genes regulados al alza y a la baja de forma similar (Quackenbush, 2002).

Tras completar el proceso de normalización, se repitieron los análisis de calidad de los estudios (MA-plot y diagramas de cajas) para evaluar el impacto de los cambios realizados. Todos estos gráficos fueron obtenidos mediante RStudio y se guardaron en formato png para poder ser analizados posteriormente mediante las funciones base de R `png()` y `dev.off()`.

3.2.2. Anotación

Dado que los estudios seleccionados han sido realizados en diferentes plataformas tecnológicas, se estandarizó la nomenclatura a un identificador único común para su interpretación y análisis conjunto. En este caso, el análisis posterior en MetaFun requiere el código ENTREZID, que sirve como identificador de los genes y secuencias en la base de datos del NCBI (Maglott et al., 2011). Para ello, se necesita conocer la correspondencia entre los identificadores de los genes en las

distintas bases de datos (ENTREZID) y los identificadores de las sondas de cada plataforma de microarray (PROBEID). Esta información la proporcionan los paquetes de anotación de Bioconductor. Al trabajar con *Homo Sapiens* la base de datos de anotación empleada fue hgu133a.db (Carlson M, 2016) para el estudio GSE46517 (plataforma Affymetrix) y illuminaHumanv4.db (Dunning et al., 2015) para el estudio GSE604646 (plataforma Illumina).

En esta sección, se debe tener en cuenta el problema de las correspondencias múltiples, que se da cuando distintas sondas corresponden a un mismo gen y distintos genes corresponden a la misma sonda, esta redundancia es una fuente de ruidos y errores en análisis posteriores (Zhang et al., 2009). Existen distintas maneras para solventarlo como la selección de la primera aparición y una sola sonda por identificador, eliminando los repetidos. Otro método para evitar eliminar información del estudio es la obtención de la mediana entre los distintos valores de expresión para cada gen repetido, obteniendo un valor representativo de la expresión del gen en las muestras. En este trabajo se empleó el método de la mediana.

3.2.3. Estandarización de grupos experimentales

Para facilitar la integración de los datos en el metaanálisis posterior, se renombraron los grupos clínicos en función de su diagnóstico (caso y control) se realizaron 3 comparativas (Tabla 2).

- Para el estudio GSE60464, se marcaron las muestras “non-cerebrotopic” como controles y las “cerebrotopic” como casos. En este estudio ambas muestras son de pacientes con estadio muy avanzado de la enfermedad y se diferencia en la progresión de la metástasis al cerebro o no.
- Para el estudio GSE46517 se realizaron 2 comparativas. En la primera se identificó “metastatic melanoma” como caso, frente a “primary” como control, comparando pacientes enfermos pero con grados muy distintos de la enfermedad. En la segunda comparativa se mantuvo “metastatic melanoma” como caso, y se tomaron los “nevus” y “normal skin” como control. Se enfrentan muestras de pacientes enfermos frente a pacientes sanos.

Tabla 2. Comparaciones realizadas en cada estudio, determinando los grupos experimentales según su diagnóstico.

Estudio	Caso	Control
GSE46064	Cerebrotopic	Non-cerebrotopic
GSE46517 Comparación 1	Metastatic Melanoma	Primary Melanoma
GSE46517 Comparación 2	Metastatic Melanoma	Nevus y normal skin

El bajo número de estudios encontrados y la heterogeneidad de sus diseños experimentales, obliga a establecer comparaciones a primera vista dispares. Sin embargo, aportan resultados interesantes detectando funciones comunes entre los estadios iniciales (control, nevus, primary) frente a los niveles más graves de enfermedad.

3.3. MetaFun: herramienta web para la caracterización molecular de enfermedades humanas

Para realizar la caracterización funcional se empleó la herramienta de MetaFun “*cuyo objetivo es proporcionar un recurso bioinformático a la comunidad científica que permita una mejor comprensión de los resultados obtenidos en la integración de diversos estudios, con técnicas de metaanálisis*” (Malmierca, 2019).

La herramienta web MetaFun actúa combinando la información de la actividad de todos los genes de los estudios seleccionados. Para ello se aplican técnicas de metaanálisis que proporcionan indicadores de consenso sobre la actividad clínica y biológica de las enfermedades humanas. Además genera de una forma sencilla, una imagen común de los estudios de la enfermedad seleccionada.

En los apartados posteriores se explican los análisis desarrollados por la herramienta que incluyen: el análisis individual de los estudios (análisis exploratorio de los datos, análisis de expresión diferencial y análisis de enriquecimiento funcional) y la realización del metaanálisis funcional.

3.3.1. Análisis individual de los estudios

La primera parte del procesamiento de los datos por parte de MetaFun incluye el procesamiento de los datos. La información que se muestra a continuación ha sido extraída mayoritariamente de las instrucciones que la propia herramienta web ofrece en GitLab (*Summary · Wiki · UBB-CIPF / MetaFunWeb · GitLab, 2022*).

3.3.1.1. Análisis exploratorio de los datos

Esta sección de la herramienta bioinformática permite saber si contamos con una cantidad y calidad de datos suficientes para el posterior metaanálisis. La pestaña de análisis exploratorio se compone de tres tipos de gráficos:

- Diagramas de cajas o *boxplot*: representan la variabilidad de los datos utilizando cuartiles, dividiendo las cajas en primer cuartil (parte inferior) y tercer cuartil (parte superior) y la línea que atraviesa la caja es la mediana de los datos, también pueden mostrar la presencia de puntos extremos (valor atípicos).
- PCA Plot: es un análisis de componentes principales, permite simplificar la complejidad de espacios muestrales con muchas dimensiones mientras que conserva su información. Se obtiene un gráfico con dos dimensiones, seleccionando los componentes principales que muestran la mayor varianza entre datos.
- Cluster Plot: es una técnica de agrupación, clasificando (clustering) los datos en grupos homogéneos, es similar al PCA.

3.3.1.2. Análisis de expresión diferencial

La herramienta lleva a cabo un análisis de expresión diferencial que consiste en identificar genes expresados diferencialmente entre condiciones experimentales, este análisis se ejecuta para

identificar los genes que posteriormente se van a caracterizar funcionalmente (Phipson et al., 2016).

3.3.1.3. Análisis de enriquecimiento funcional

Tras el análisis de expresión diferencial, los genes se evalúan con la opción de caracterización funcional seleccionada por el usuario al principio del análisis, en este caso GSEA (Montaner & Dopazo, 2010). El objetivo de estos análisis es indicar cuál es la función biológica, proceso molecular o componente celular detrás del conjunto de genes con un patrón similar de expresión (Korotkevich et al., 2016). Se asocia cada gen de la matriz de expresión con su término GO, ordenándolos.

3.3.2. Metaanálisis funcional

Una vez completado el análisis bioinformático individual para cada estudio, se realiza el metaanálisis funcional a partir de los resultados obtenidos en el análisis de enriquecimiento funcional. MetaFun permite además seleccionar el método de metaanálisis siguiendo un modelo de efectos fijos si los estudios son homogéneos o un modelo de efectos aleatorios para estudios con elevada heterogeneidad.

El metaanálisis consta de dos partes, la primera acerca de la configuración y exploración de matrices de entrada y el consiguiente análisis de heterogeneidad y determinación de la medida combinada del efecto. Se utilizaron los logaritmos de odds ratios (LOR) como valor de medida del efecto y la desviación estándar (SD) de los LOR para cada una de las funciones evaluadas, procedentes de los resultados del GSEA de los estudios individuales. De esta forma se evaluaron las funciones con la anotación GO con sus 3 ontologías (procesos biológicos, componentes celulares y funciones moleculares).

La herramienta MetaFun representa los resultados con dos tipos de gráficos:

- Gráfico de bosque: estos representan en el eje vertical los estudios seleccionados para el metaanálisis y en el eje horizontal el valor de la variable de interés, que en el caso de MetaFun es el valor del Log Odds Ratio (LOR) de la función seleccionada y su intervalo de confianza. La línea roja debajo del eje X, representa el LOR resumido de la función entre todos los estudios.
- Gráfico de embudo o Funnel Plot: se representa el error estándar (variabilidad) en el eje Y frente al tamaño del efecto, el valor de LOR en este caso en el eje X. Por tanto, los estudios que aparecen alejados horizontalmente del centro son estudios menos precisos. En el eje vertical, cuanto más cerca del 0 se encuentre su desviación estándar será menor.

4. RESULTADOS

A continuación, se muestran los principales resultados obtenidos en el estudio.

4.1. Revisión sistemática y selección de estudios

En el primer paso de la revisión sistemática, llevada a cabo en marzo de 2023 en la base de datos GEO, se identificaron 23 estudios candidatos para su inclusión en el metaanálisis. Tras una revisión exhaustiva para su cribado, se excluyeron 17 por no cumplir los siguientes requisitos: no tratarse de la enfermedad estudiada (melanoma metastásico cerebral), trabajar con cell lines en lugar de muestras de tejido humanas, utilizar otras tecnologías (scRNA-seq, snRNA-seq).

Los 7 estudios restantes fueron evaluados para su elegibilidad. De ellos se descartaron 4 por los siguientes motivos: no alcanzar el mínimo tamaño muestral; incluir tratamiento de los pacientes; uso de muestras congeladas. Por ello, finalmente fueron seleccionados 2 estudios para el metaanálisis (Tabla 3). En la Figura 5 se muestra el flujo del proceso llevado a cabo mediante el diagrama PRISMA.

Tabla 3. Estudios seleccionados tras la revisión sistemática.

Identificador del estudio	Plataforma	Artículo
GSE60464	Illumina HumanHT-12 WG-DASL V4.0 R2 expression beadchip	(Jilaveanu et al., 2015) https://www.ncbi.nlm.nih.gov/pubmed/25316811
GSE46517	Affymetrix Human Genome U133A Array	(Kabbarah et al., 2010) https://www.ncbi.nlm.nih.gov/pubmed/20520718

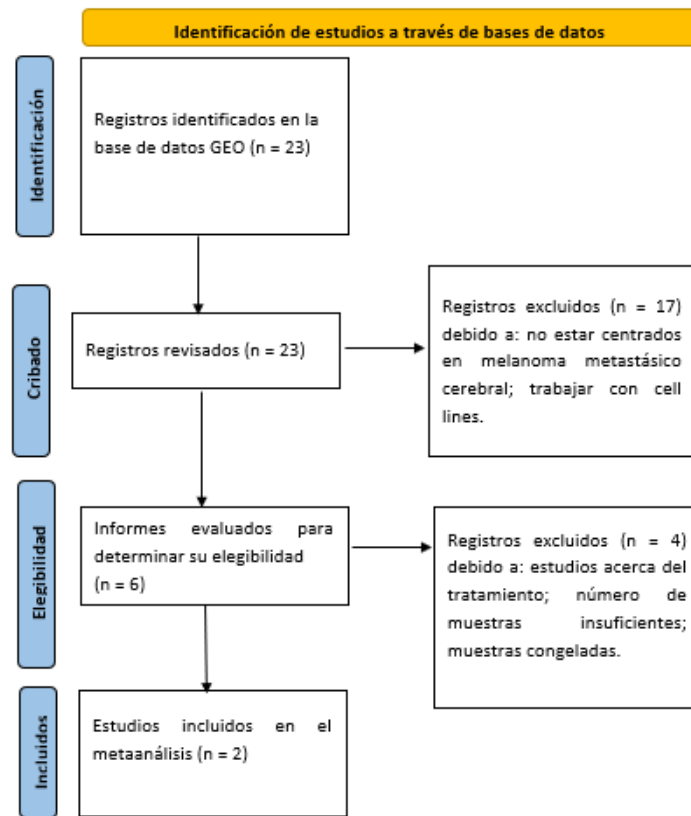


Figura 5. Diagrama de flujo de la revisión sistemática del estudio.

4.2. Procesamiento y estandarización de los estudios

Los datos de los estudios seleccionados fueron descargados y analizados en primer lugar mediante MA-plots y diagramas de cajas.

Para el estudio GSE46517, se observaron primero los gráficos MA-plots (Figura 6). En general, se observa una similar variabilidad de las muestras, sin la presencia de valores extraños, ya que la mayoría de puntos tienen un valor cercano a 0 ajustándose al eje X. El diagrama de cajas muestra una distribución bastante heterogénea y un gran número de valores atípicos (Figura 7, izquierda). Por ello, se aplicó la normalización de logaritmo en base 2 y se ajustaron los valores para que todos fueran positivos, obteniendo una distribución más homogénea (Figura 7, derecha) que permitió los análisis posteriores.

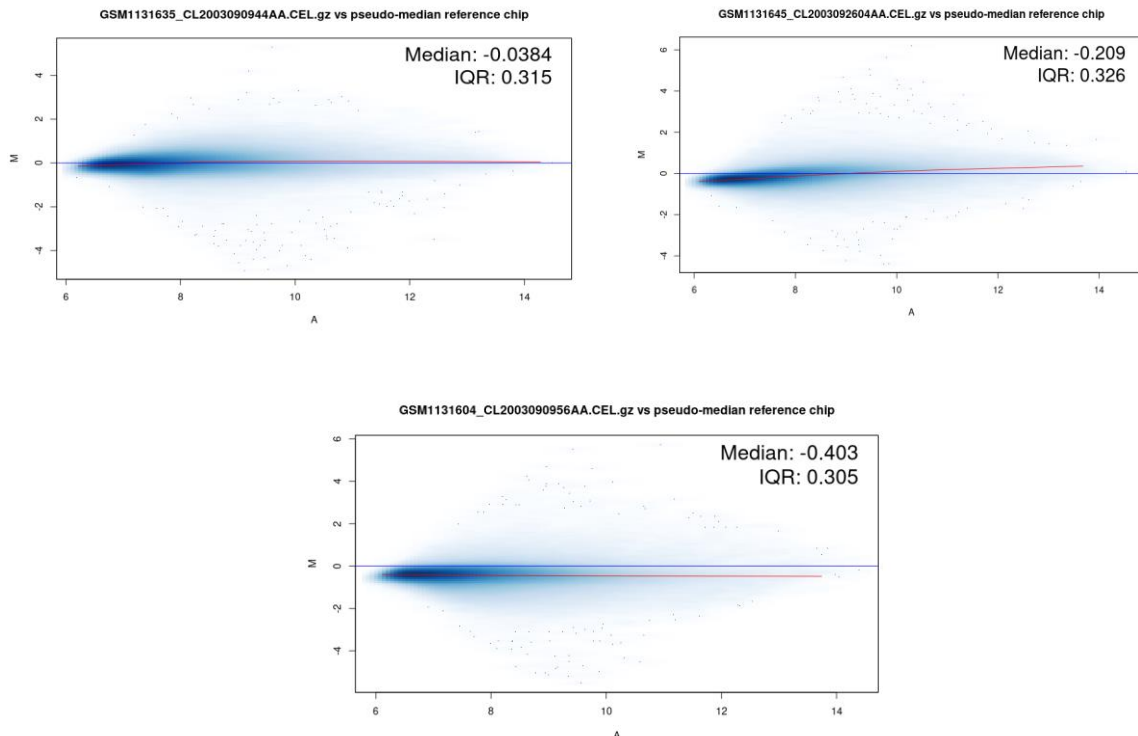


Figura 6. Gráficos MA-plots para tres muestras aleatorias del estudio GSE46517 comparando dicha muestra con la de referencia.

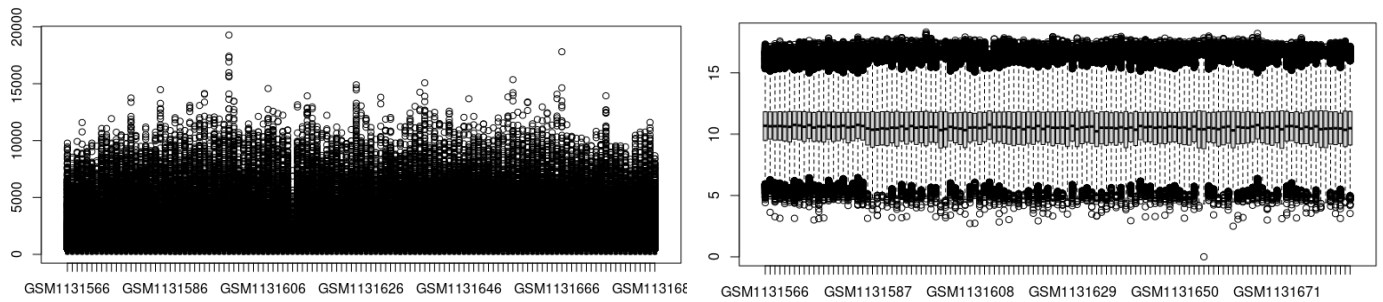


Figura 7. Diagramas de cajas del estudio GSE46517 antes de normalizar (izquierda) y tras aplicar la normalización del logaritmo en base 2 (derecha).

Para el estudio GSE60464 se emplearon los mismos gráficos, mostrando de nuevo una falta en el proceso de normalización que se completa con la aplicación del método de normalización del logaritmo en base 2 (Figura 8).

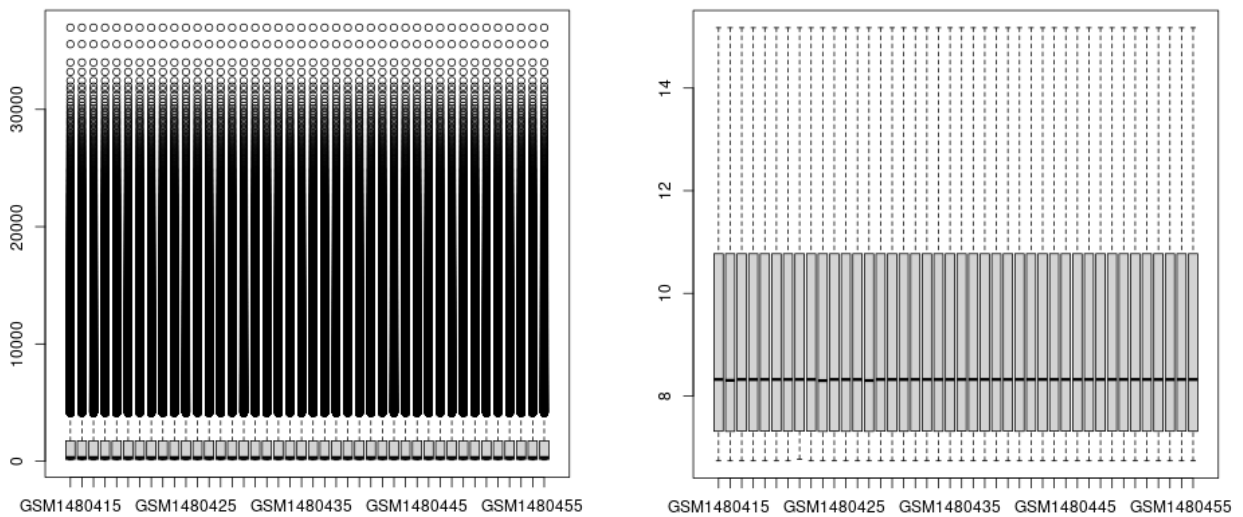


Figura 8. Diagramas de cajas del estudio GSE60464 antes de normalizar (izquierda) y tras aplicar la normalización del logaritmo en base 2 (derecha).

4.3. Análisis individual de los estudios con MetaFun

4.3.1. Procesamiento de los datos

Tras el preprocesado de los datos y la adaptación al formato requerido mediante el lenguaje de programación R, se subieron a la herramienta MetaFun. En el primer apartado de resumen del análisis, muestra gráficos con la distribución de tipo de muestras (caso, control) en cada estudio (Figura 9). Siendo en el estudio GSE46517 el número de muestras caso mucho mayor que el de controles para ambas comparativas, mientras que en el estudio GSE60464 el número de muestras de cada tipo es casi equitativo.

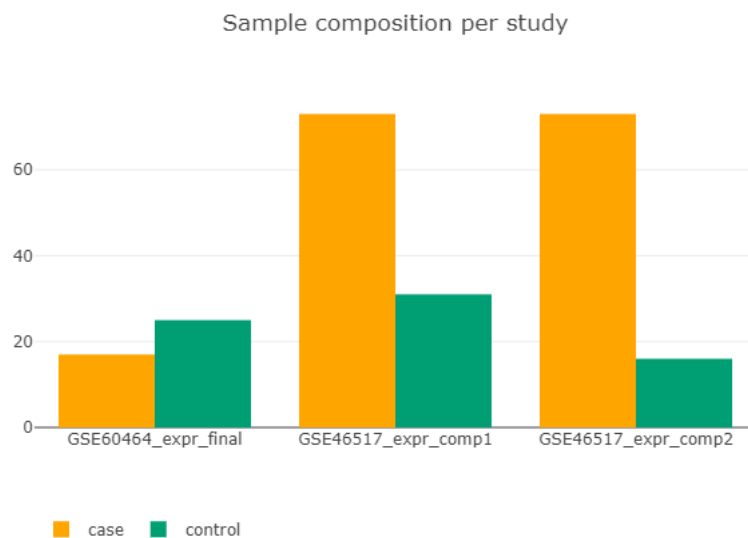


Figura 9. Diagrama de barras mostrando la distribución del tipo de muestras (caso control) en cada estudio.

La herramienta MetaFun realiza un nuevo análisis exploratorio de los datos ya normalizados y se obtuvieron diagramas de cajas, PCA y cluster plot de cada estudio:

GSE60464: Cerebrotopico (case) vs. Non-cerebrotopico (control)

Para el estudio GSE60464 se observa el diagrama de cajas que muestra una distribución homogénea de los datos con casi los mismos valores máximos y mínimos (representados por los bigotes de las cajas) al igual que la mediana representada por la línea horizontal que cruza la caja (Figura 10).

El análisis exploratorio del PCA muestra un alto solapamiento entre las muestras caso y control (Figura 11). En parte es lo esperado ya que en este caso estamos comparando grupos muy similares, con la única diferencia el lugar de progresión de la metástasis (Figura 11). Lo mismo sucede con la agrupación de clustering, que no se observan grupos claros separados del resto (Figura 12).

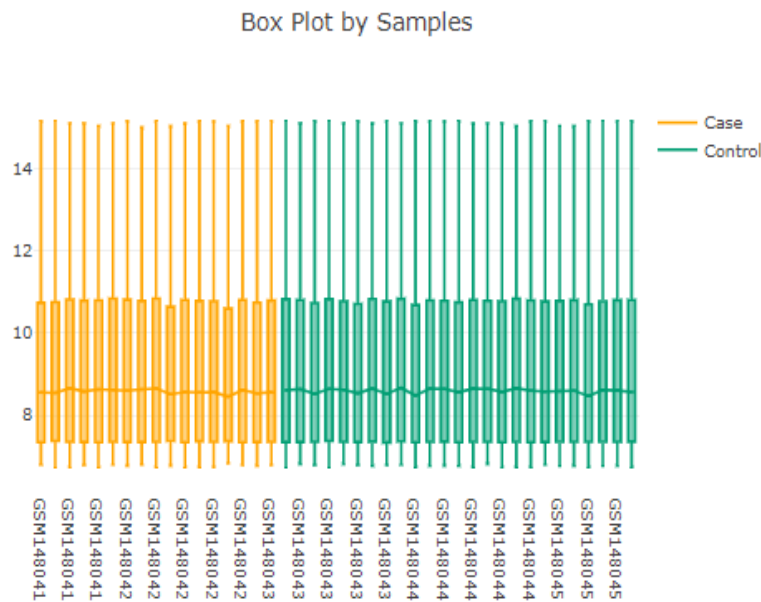


Figura 10. Diagrama de cajas en el que están representados los valores normalizados de las muestras empleadas en el estudio GSE60464, coloreadas según caso y control.

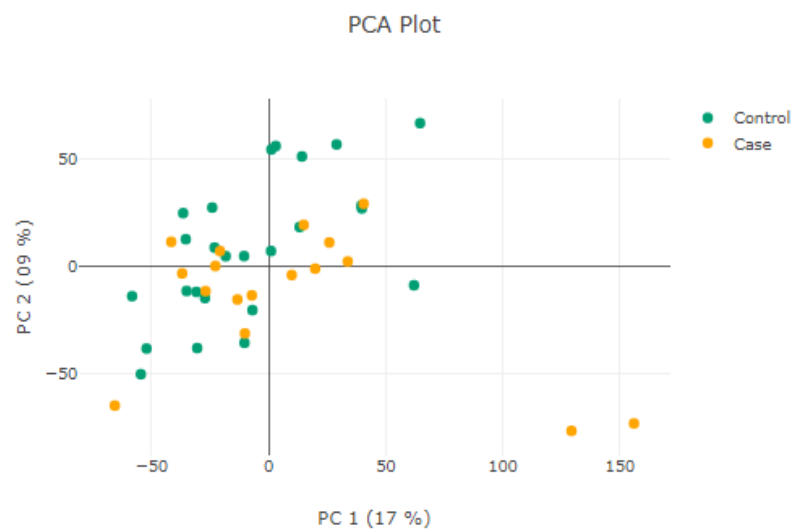


Figura 11. Análisis de componentes principales (PCA) de las muestras del estudio GSE60464.

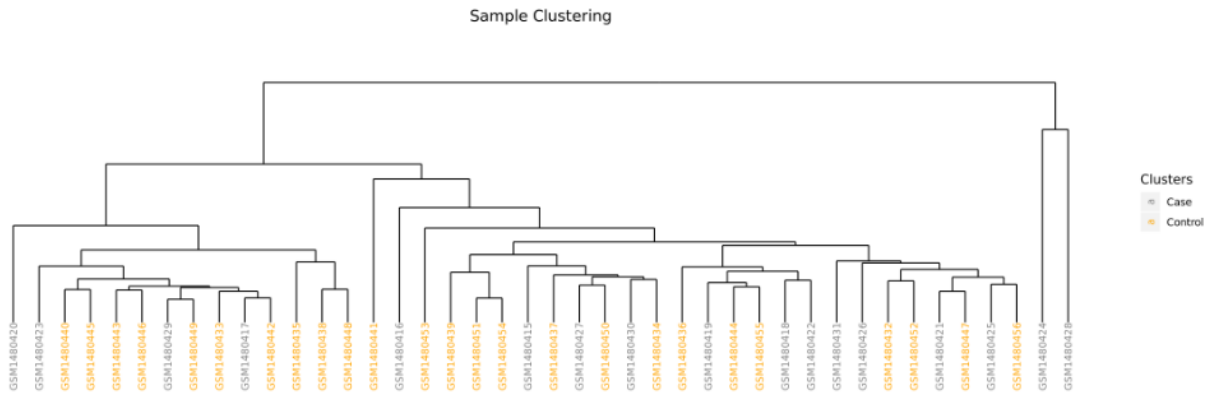


Figura 12. Clustering de las muestras del estudio GSE60464. Cada grupo experimental aparece representado de un color.

GSE46517 Comparación 1: Metastatic Melanoma (case) vs. Primary (control)

Para este estudio se realizan dos comparaciones, en esta primera la comparación se hace entre individuos enfermos con un grado más o menos avanzado de la enfermedad. El diagrama de cajas muestra una distribución homogénea de los valores de expresión, junto con la mediana que es muy similar en todos. Los valores máximos y mínimos sí que varían más en este caso pero nada relevante (Figura 13).

El PCA muestra una buena separación de las muestras según el diagnóstico pero no es del todo perfecta, lo que cuadra con la comparación que se ha hecho, son muestras más similares que las del estudio GSE60464 pero menos diferentes que las de la comparación 2 de este mismo estudio (Figura 14). Al igual que en el PCA, la agrupación de las muestras en el clustering es bastante buena pero no perfectamente homogénea (Figura 15).

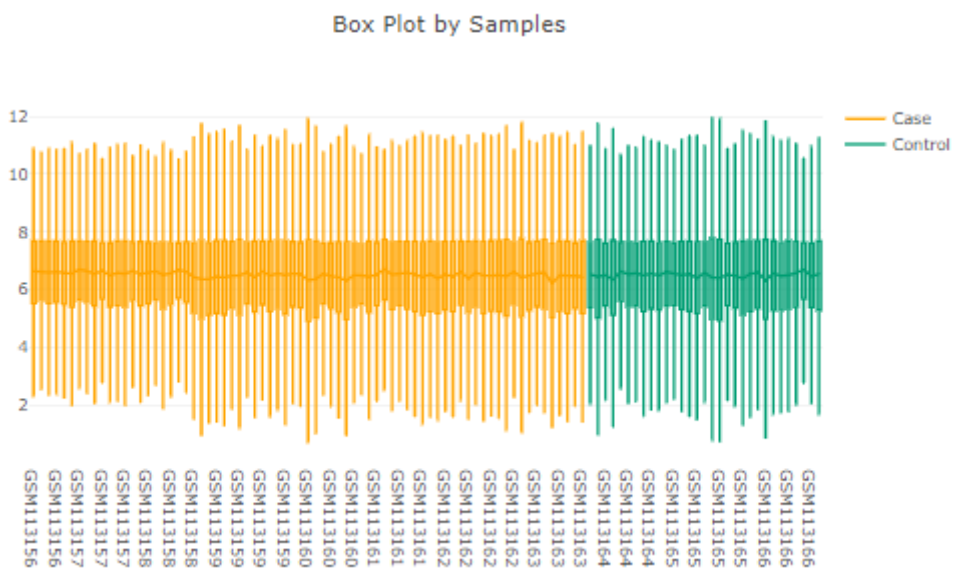


Figura 13. Diagrama de cajas en el que están representados los valores normalizados de las muestras empleadas en el estudio GSE46517 Comparación 1, coloreadas según caso y control.

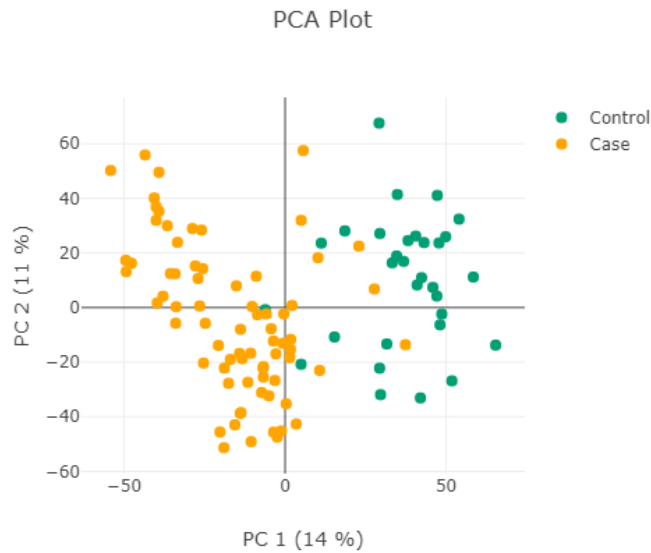


Figura 14. Análisis de componentes principales (PCA) de las muestras del estudio GSE46517 Comparación 1.

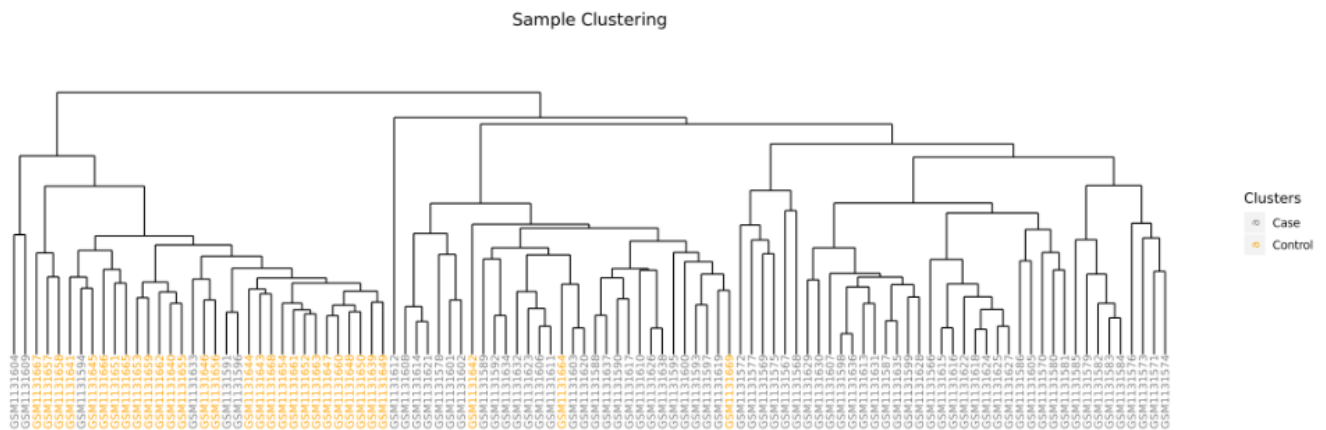


Figura 15. Clustering de las muestras del estudio GSE46517 Comparación 1. Cada grupo experimental aparece representado de un color.

GSE46517 Comparación 2: Metastatic Melanoma (case) vs. nevus + normal skin (control)

La segunda comparación del estudio GSE46517 sí que se ajusta al estudio de comparación caso y control. El diagrama de cajas muestra una distribución homogénea de valores, con una mediana similar en todas las muestras y valores máximos y mínimos más variables pero dentro de un rango normal (Figura 16). En este gráfico no aparecen todas las muestras del estudio ya que tras un primer análisis exploratorio con todos los datos se decidió eliminar la muestra GSM1131685, ya que presentaba una distribución diferente al resto.

En un primer momento al observar el PCA había una agrupación casi perfecta de las muestras a excepción de nuevo de la GSM1131685, marcada como control que quedaba agrupada como caso (Figura 17). Se buscó de nuevo en GEO la identificación de la muestra y apareció como “normal epithelial melanocytes”, siendo la única muestra con ese tipo de caracterización en el

estudio. Por ello, como se ha introducido anteriormente se decidió eliminarla del estudio. Se realizó de nuevo el metaanálisis definitivo sin esa muestra (Figura 18). El clustering muestra una agrupación casi perfecta de las muestras en función de su diagnóstico (Figura 19).

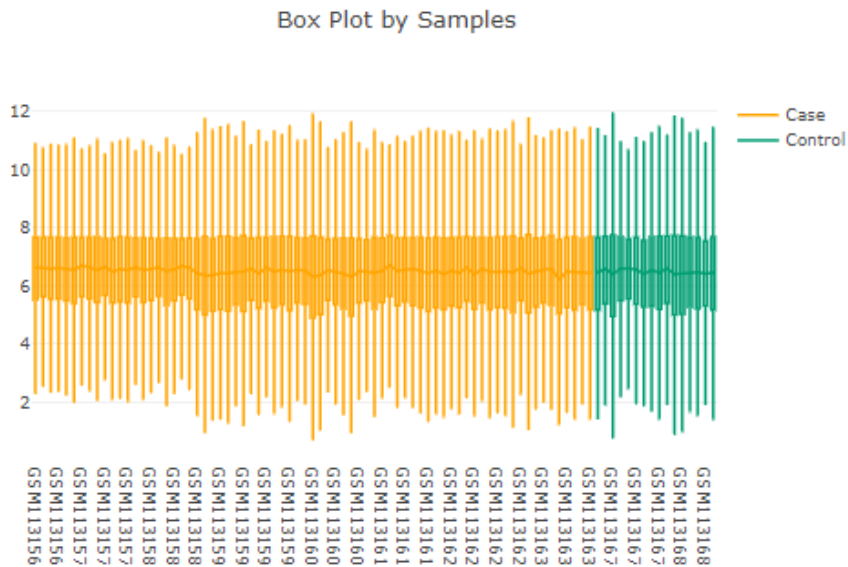


Figura 16. Diagrama de cajas en el que están representados los valores normalizados de las muestras empleadas en el estudio GSE46517 Comparación 2, coloreadas según caso y control.

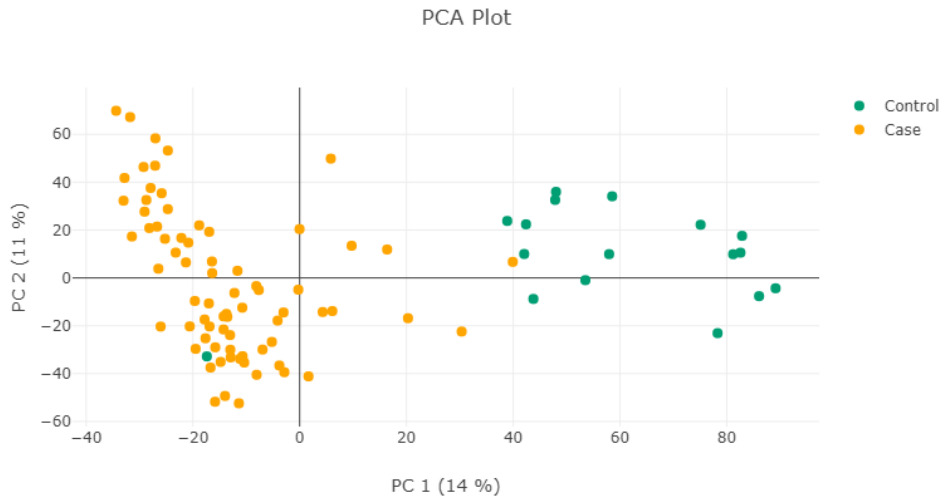


Figura 17. Análisis de componentes principales (PCA) de las muestras del estudio GSE46517 Comparación 2 con la muestra GSM1131685.

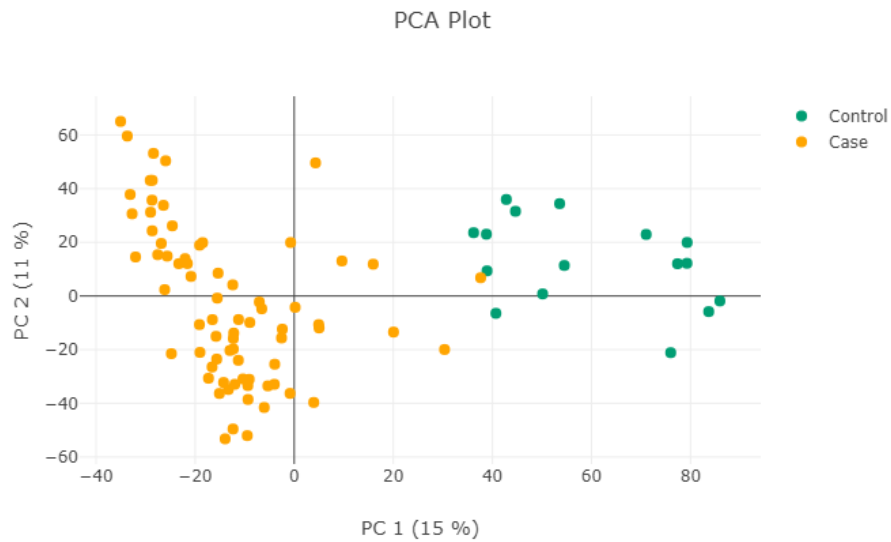


Figura 18. Análisis de componentes principales (PCA) de las muestras del estudio GSE46517 Comparación 2.

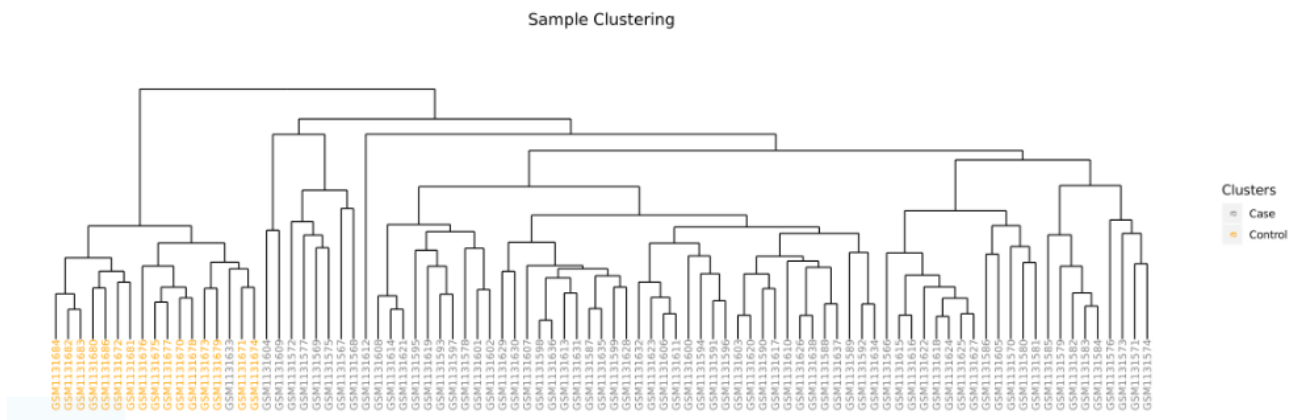


Figura 19. Clustering de las muestras del estudio GSE46517 Comparación 2. Cada grupo experimental aparece representado de un color.

4.3.2. Análisis de expresión diferencial

MetaFun proporciona una tabla resumen con el número total de genes diferencialmente expresados en cada uno de los estudios, distinguiendo entre genes sobreexpresados (UP Regulated) y genes subexpresados (DOWN Regulated) (Tabla 4). En los estudios evaluados hay una clara diferencia de expresión entre controles y casos.

Tabla 4. Resumen del número de genes diferencialmente expresados en cada estudio. Clasificados en UP Regulated y DOWN Regulated según su sobreexpresión. Tabla proporcionada por MetaFun.

Differential Expression Summary				
Study Name	Total Genes	Total Significant	UP Regulated	DOWN Regulated
GSE60464_expr_final	19680	1375	881	494
GSE46517_expr_comp1	13039	6273	3964	2309
GSE46517_expr_comp2	13039	5986	3687	2299

Además de esta tabla resumen, MetaFun proporciona para cada uno de los estudios una tabla con el identificador y nombre de los genes de los 20 mejores resultados obtenidos. Estas tablas ajustadas se muestran a continuación (Tabla 5 a 7).

Tabla 5. Representación de los top 10 genes expresados diferencialmente en el estudio GSE60464 atendiendo al P value y logaritmo de la magnitud de cambio (logFC). Se indica también su identificador Entrez, nombre del gen y su P-valor ajustado.

Entrez ID	Gene Name	logFC	Statistic	P value	Adjusted P value
590	butyrylcholinesterase	1.988	4.580	4.0797e-5	3.1895e-1
2770	G protein subunit alpha i1	1.187	4.736	2.4737e-5	3.1895e-1
83938	leucine rich melanocyte differentiation associated	1.068	3.557	9.4200e-4	6.8583e-1
10404	carboxypeptidase Q	0.883	3.482	1.1731e-3	6.8583e-1
90557	coiled-coil domain containing 74A	0.713	3.468	1.2208e-3	6.8583e-1
79814	agmatinase	-1.421	-4.504	5.1922e-5	3.1895e-1
5330	phospholipase C beta 2	-1.328	-4.175	1.4620e-4	5.7543e-1
255231	mucolipin 2	-1.241	-3.932	3.0832e-4	6.8583e-1
5153	phosphodiesterase 1B	-1.089	-3.568	9.1276e-4	6.8583e-1
474170	leucine rich repeat containing 37 member A2	-0.945	-4.434	6.4827e-5	3.1895e-1

Tabla 6. Representación de los top 10 genes expresados diferencialmente en el estudio GSE46517 Comparación 1 atendiendo al P value y logaritmo de la magnitud de cambio (logFC). Se indica también su identificador Entrez, nombre del gen y su P-valor ajustado.

Entrez ID	Gene Name	logFC	Statistic	P value	Adjusted P value
8364	H4 clustered histone 3	1.809	10.523	3.4806e-18	4.3638e-16
6637	small nuclear ribonucleoprotein polypeptide G	1.083	8.961	1.1663e-14	1.0862e-12
55151	transmembrane protein 38B	1.346	8.944	1.2762e-14	1.1802e-12
6782	heat shock protein family A (Hsp70) member 13	1.055	8.792	2.8036e-14	2.4049e-12
23165	nucleoporin 205	0.736	8.536	1.0465e-13	8.27029e-12
3872	keratin 17	-3.905	-17.325	8.9502e-33	5.9351e-29
6278	S100 calcium binding protein A7	-8.238	-17.321	9.1036e-33	5.9351e-29
5268	serpin family B member 5	-4.696	-17.142	2.0590e-32	6.7118e-29

3854	keratin 6B	-4.393	-17.151	1.9751e-32	6.7118e-29
27076	LY6/PLAUR domain containing 3	-4.529	-16.676	1.7641e-31	4.6006e-28

Tabla 7. Representación de los top 10 genes expresados diferencialmente en el estudio GSE46517 Comparación 2 atendiendo al P value y logaritmo de la magnitud de cambio (logFC). Se indica también su identificador Entrez, nombre del gen y su P-valor ajustado.

Entrez ID	Gene Name	logFC	Statistic	P value	Adjusted P value
6696	secreted phosphoprotein 1	5.652	11.803	4.3261e-20	5.9377e-18
9787	DLG associated protein 5	1.839	10.040	1.9707e-16	1.4853e-14
701	BUB1 mitotic checkpoint serine/threonine kinase B	1.836	9.966	2.8142e-16	2.0968e-14
23532	preferentially expressed antigen in melanoma	4.209	9.694	1.0474e-15	7.1882e-14
1164	CDC28 protein kinase regulatory subunit 2	2.703	9.507	2.5827e-15	1.6190e-13
728113	annexin A8 like 1	-4.107	-19.967	4.0020e-35	5.2183e-31
63928	calcineurin like EF-hand protein 2	-3.259	-19.658	1.2788e-34	8.3373e-31
10850	C-C motif chemokine ligand 27	-4.059	-19.155	8.6443e-34	3.7571e-30
1825	desmocollin 3	-3.529	-17.700	2.6202e-31	6.4860e-28
7123	C-type lectin domain family 3 member B	-2.719	-17.665	3.0136e-31	6.4860e-28

4.3.3. Análisis de enriquecimiento funcional

Del mismo modo que en el apartado anterior, MetaFun nos proporciona una tabla resumen con el número total de funciones con mayor sobrerrepresentación en cada uno de los estudios (Tabla 8). También aparece el número total de funciones significativas distinguiendo las funciones que están sobreexpresadas (UP Regulated) y subexpresadas (DOWN Regulated). Estos resultados se utilizaron en el metaanálisis.

Tabla 8. Resumen del número de funciones con diferencia de expresión significativa. Clasificados en UP Regulated y DOWN Regulated según su sobreexpresión. Tabla proporcionada por MetaFun.

Functional Profiling Summary				
Study Name	Total Functions	Total Significant	UP Regulated	DOWN Regulated
GSE60464_expr_final	8363	202	91	111
GSE46517_expr_comp1	7783	1234	562	672
GSE46517_expr_comp2	7783	958	449	509

4.4. Metaanálisis con MetaFun

En este apartado MetaFun proporciona una tabla con el resultado final de las funciones identificadas tras realizar el metaanálisis (Tabla 9). Además para cada una de las funciones

aporta los gráficos de bosque y embudo permitiendo observar la aportación de cada estudio en cada función y la robustez del análisis.

Tabla 9. Resumen resultados de metaanálisis, con el número total de funciones identificadas y sus valores LOR (Logaritmos de Odds Ratios).

Meta-Analysis Summary			
Ontology	significant Positive (LOR)	significant Negative (LOR)	Total significant Functions
BP	22	9	31
MF	15	17	33
CC	15	18	35

En general los términos GO identificados concuerdan con las características propias de la enfermedad (Tabla 10 y 11). Las funciones sobreexpresadas en los casos, es decir, pacientes enfermos con melanoma, están relacionados con una mayor replicación de ADN (GO:0006260) y división celular (GO:0051301), sobreexpresión de los componentes que participan en el ciclo celular como helicasas (GO:0003678). También aparece un incremento en los mecanismos de reparación del ADN (GO:0006281) (Tabla 10).

Por otro lado en las muestras marcadas como controles, que incluyen pacientes con estadios iniciales de la enfermedad o pacientes sanos muestran se relacionan con la diferenciación de los queratinocitos (GO:0030216) que sería el proceso de desarrollo de la epidermis. Además, destaca la actividad del Complejo Mayor de Histocompatibilidad (MHC), lo que concuerda con pacientes sanos en los que el sistema inmune funciona de manera correcta (GO:0042611). Las moléculas de adhesión célula a célula y célula con la matriz son mayores en pacientes sanos que en enfermos (GO:0098632) (Tabla 11).

Tabla 10. Resumen de los términos GO significativos en el metaanálisis con LOR positivo, ordenados de mayor a menor. MF: molecular function, BP: biological process, CC: cellular component.

GO term	name	Ontology	LOR
GO:0043138	3'-5' DNA helicase activity	MF	1.335
GO:0003688	DNA replication origin binding	MF	1.204
GO:0035173	histone kinase activity	MF	1.176
GO:0017116	single-stranded DNA-dependent ATP-dependent DNA helicase activity	MF	1.14
GO:0043142	single-stranded DNA-dependent ATPase activity	MF	1.14
GO:0004003	ATP-dependent DNA helicase activity	MF	1.003
GO:0000400	four-way junction DNA binding	MF	1.001
GO:0000777	condensed chromosome kinetochore	CC	0.91
GO:0000779	condensed chromosome, centromeric region	CC	0.89
GO:0033045	regulation of sister chromatid segregation	BP	0.878
GO:0003678	DNA helicase activity	MF	0.857
GO:0051983	regulation of chromosome segregation	BP	0.851
GO:0045132	meiotic chromosome segregation	BP	0.84

GO:0008094	DNA-dependent ATPase activity	MF	0.832
GO:0000070	mitotic sister chromatid segregation	BP	0.787
GO:0000724	double-strand break repair via homologous recombination	BP	0.785
GO:0000725	recombinational repair	BP	0.785
GO:0017056	structural constituent of nuclear pore	MF	0.774
GO:0006261	DNA-dependent DNA replication	BP	0.762
GO:0000819	sister chromatid segregation	BP	0.746
GO:0000793	condensed chromosome	CC	0.735
GO:0000776	kinetochore	CC	0.734
GO:0005657	replication fork	CC	0.716
GO:0098813	nuclear chromosome segregation	BP	0.685
GO:0005876	spindle microtubule	CC	0.649
GO:0007059	chromosome segregation	BP	0.638
GO:0006302	double-strand break repair	BP	0.63
GO:0000775	chromosome, centromeric region	CC	0.618
GO:0004386	helicase activity	MF	0.603
GO:0000794	condensed nuclear chromosome	CC	0.594
GO:0098687	chromosomal region	CC	0.571
GO:0000723	telomere maintenance	BP	0.562
GO:0032200	telomere organization	BP	0.543
GO:0140097	catalytic activity, acting on DNA	MF	0.534
GO:0006310	DNA recombination	BP	0.527
GO:0005643	nuclear pore	CC	0.525
GO:0000781	chromosome, telomeric region	CC	0.515
GO:0071103	DNA conformation change	BP	0.506
GO:0006260	DNA replication	BP	0.505
GO:0000784	nuclear chromosome, telomeric region	CC	0.482
GO:0140014	mitotic nuclear division	BP	0.471
GO:0000280	nuclear division	BP	0.465
GO:0006281	DNA repair	BP	0.452
GO:0003697	single-stranded DNA binding	MF	0.425
GO:0048285	organelle fission	BP	0.423
GO:0051301	cell division	BP	0.357
GO:0044772	mitotic cell cycle phase transition	BP	0.353
GO:0005819	spindle	CC	0.35
GO:0016887	ATPase activity	MF	0.312
GO:0044454	nuclear chromosome part	CC	0.302
GO:0003729	mRNA binding	MF	0.246
GO:0005813	centrosome	CC	0.243

Tabla 11. Resumen de los términos GO significativos en el metaanálisis con LOR negativo, ordenados de mayor magnitud de cambio a menor. MF: molecular function, BP: biological process, CC: cellular component.

GO term	name	Ontology	LOR
GO:0030280	structural constituent of epidermis	MF	-1.882
GO:0010839	negative regulation of keratinocyte proliferation	BP	-1.605
GO:0070268	cornification	BP	-1.581
GO:0031424	keratinization	BP	-1.524
GO:0030216	keratinocyte differentiation	BP	-1.174
GO:0045095	keratin filament	CC	-1.116
GO:0098641	cadherin binding involved in cell-cell adhesion	MF	-1.101
GO:0043588	skin development	BP	-0.975
GO:0009913	epidermal cell differentiation	BP	-0.962
GO:0042613	MHC class II protein complex	CC	-0.955
GO:0008544	epidermis development	BP	-0.924
GO:0042611	MHC protein complex	CC	-0.88
GO:0098632	cell-cell adhesion mediator activity	MF	-0.728
GO:0030020	extracellular matrix structural constituent conferring tensile strength	MF	-0.675
GO:0044420	extracellular matrix component	CC	-0.674
GO:0045111	intermediate filament cytoskeleton	CC	-0.673
GO:0005201	extracellular matrix structural constituent	MF	-0.635
GO:0005604	basement membrane	CC	-0.634
GO:0033293	monocarboxylic acid binding	MF	-0.554
GO:0019199	transmembrane receptor protein kinase activity	MF	-0.54
GO:0005200	structural constituent of cytoskeleton	MF	-0.509
GO:0005913	cell-cell adherens junction	CC	-0.499
GO:0019838	growth factor binding	MF	-0.459
GO:0062023	collagen-containing extracellular matrix	CC	-0.45
GO:0034330	cell junction organization	BP	-0.436
GO:0008236	serine-type peptidase activity	MF	-0.43
GO:0031012	extracellular matrix	CC	-0.425
GO:0017171	serine hydrolase activity	MF	-0.418
GO:0005911	cell-cell junction	CC	-0.385
GO:0031589	cell-substrate adhesion	BP	-0.369
GO:0009897	external side of plasma membrane	CC	-0.337
GO:0019955	cytokine binding	MF	-0.332
GO:0008201	heparin binding	MF	-0.328
GO:1901681	sulfur compound binding	MF	-0.325
GO:0005912	adherens junction	CC	-0.316
GO:0050839	cell adhesion molecule binding	MF	-0.303
GO:0045296	cadherin binding	MF	-0.297
GO:0098552	side of membrane	CC	-0.284

GO:0030055	cell-substrate junction	CC	-0.27
GO:0005924	cell-substrate adherens junction	CC	-0.259
GO:0003779	actin binding	MF	-0.259
GO:0005925	focal adhesion	CC	-0.258
GO:0030139	endocytic vesicle	CC	-0.233
GO:0010008	endosome membrane	CC	-0.227

En las figuras de los gráficos de bosque (Figuras A de la 20 a la 25) se observa que la mayoría de los resultados están asociados al estudio GSE46517 con sus dos comparaciones, observando en todos ellos una mayor aportación de la comparativa 1, representado por el cuadrado negro. Otras veces aparece el estudio GSE60464 frente al GSE46517 con su primera comparativa (Figura 24). Pero nunca aparecen juntos el estudio GSE60464 frente a la comparación 2 del GSE46517.

Por otra parte, los gráficos de embudo o funnel plots muestran que todos los estudios se encuentran dentro de la región de confianza delimitada por la zona blanca, lo que confirma la robustez del metaanálisis.

A continuación, se muestran los gráficos de bosque y funnel plots para algunas de las funciones sobreexpresadas en las muestras caso:

GO:0003678 (DNA helicase activity)

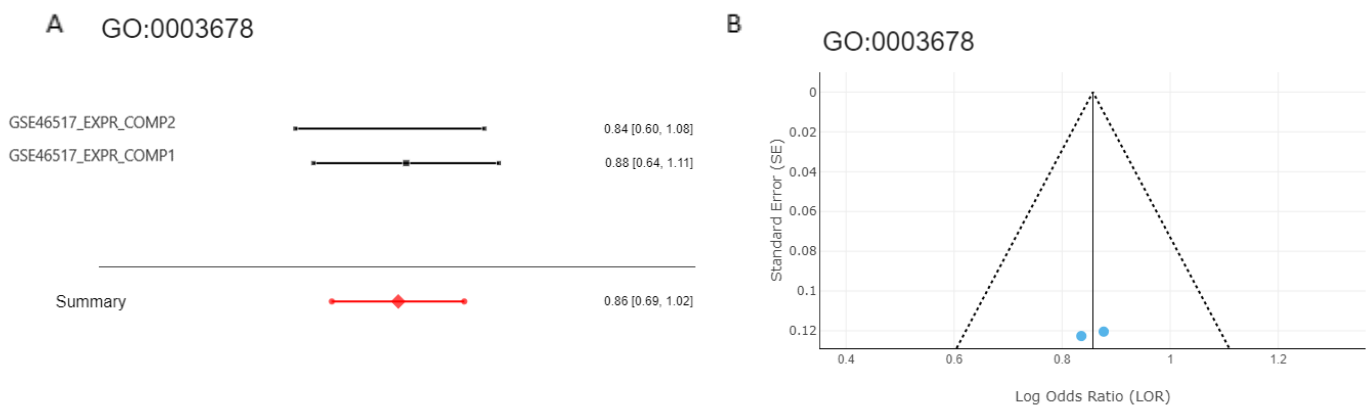


Figura 20. Término GO:0003678 (DNA helicase activity). A: Gráfico de bosque; B: Funnel plot.

GO:0006260 (DNA replication)

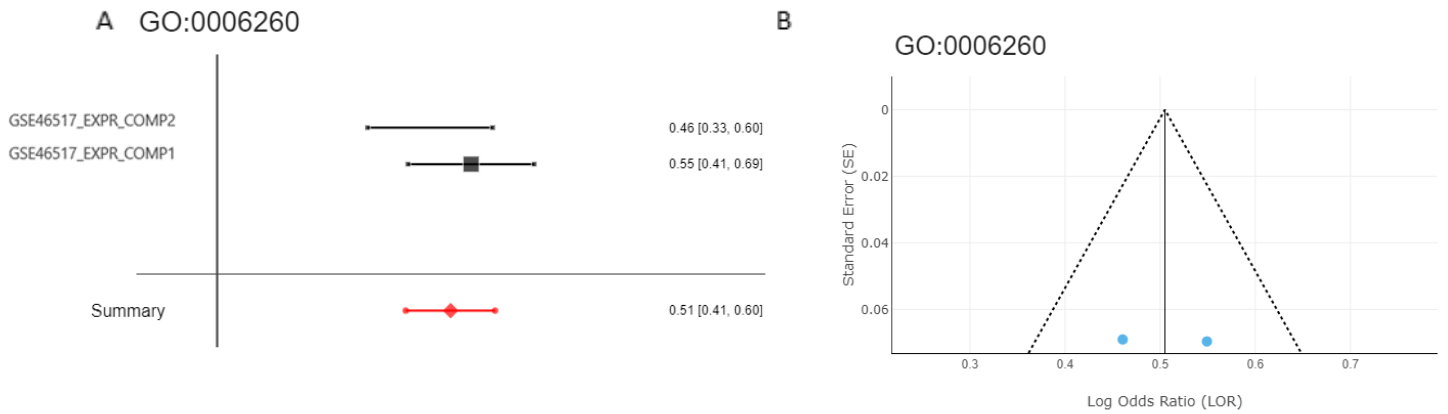


Figura 21. Término GO:0006260 (DNA replication). A: Gráfico de bosque; B: Funnel plot.

GO:0006281 (DNA repair)

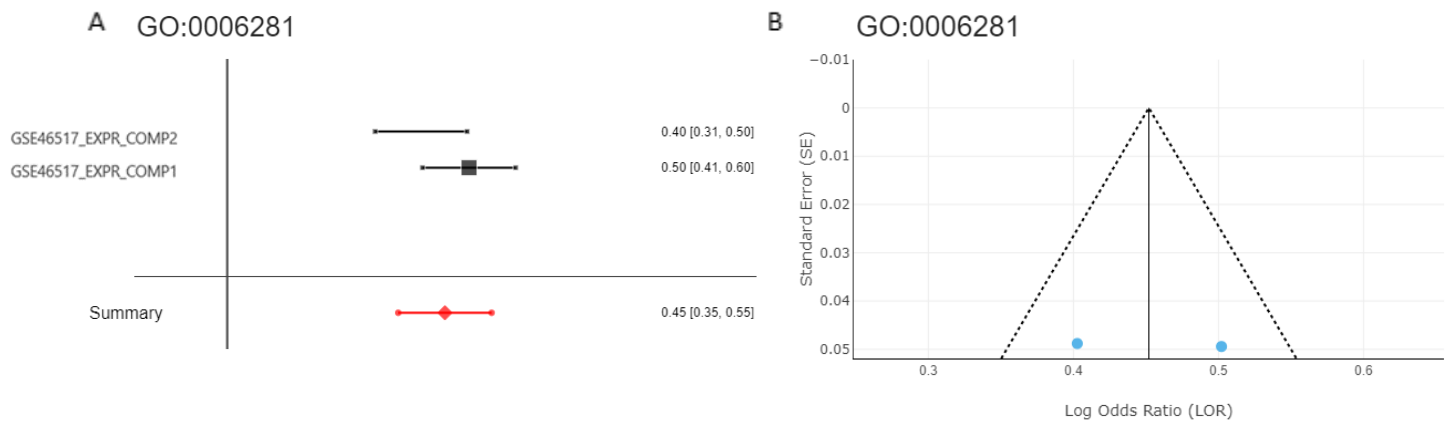


Figura 22. Término GO:000628 (DNA repair). A: Gráfico de bosque; B: Funnel plot.

Por otro lado tenemos funciones sobreexpresadas en los controles, que presentan un valor LOR negativo. Se representan los siguientes gráficos de bosque y funnel plots:

GO:0030216 (Keratinocyte differentiation)

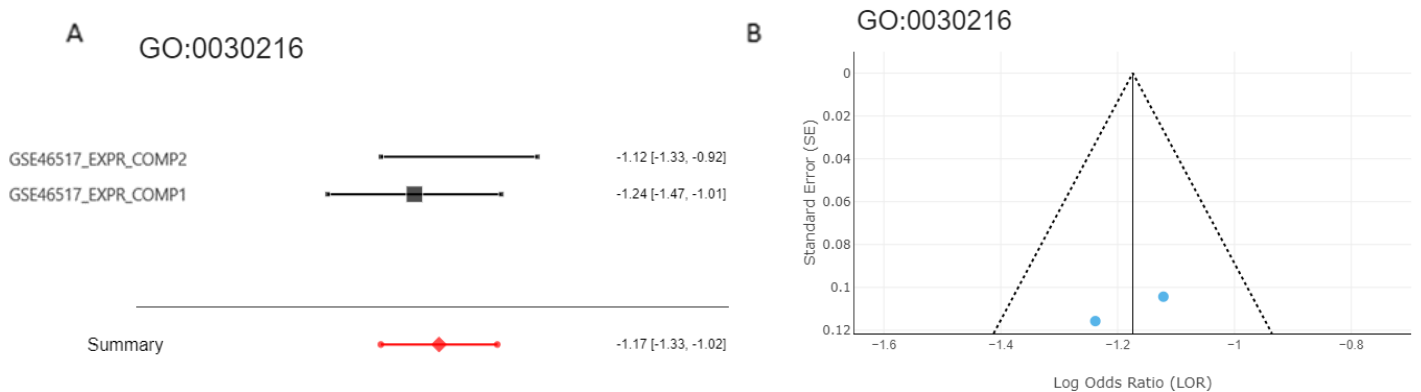


Figura 23. Término GO:0030216 (Keratinocyte differentiation). A: Gráfico de bosque; B: Funnel plot.

GO:0042611 (MHC protein complex)

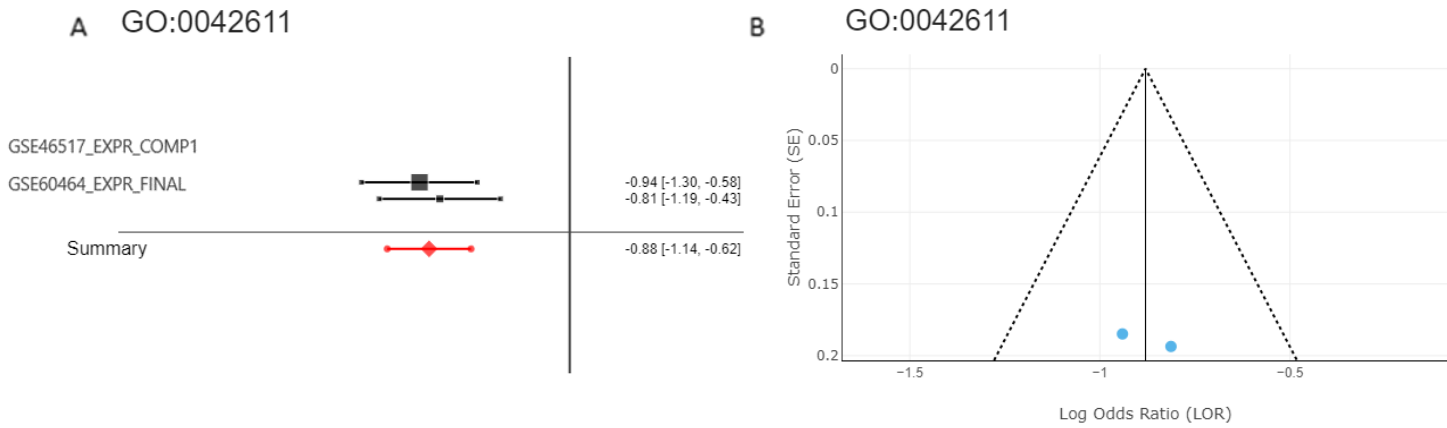


Figura 24. Término GO:0042611 (MHC protein complex). A: Gráfico de bosque; B: Funnel plot.

GO:0098632 (Cell-cell adhesion mediator activity)

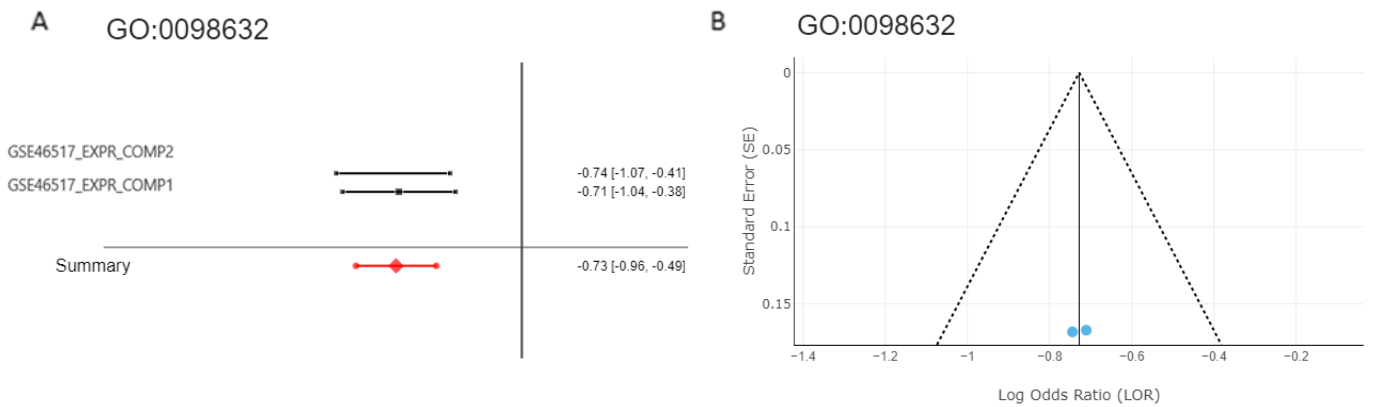


Figura 25. Término GO:0098632 (Cell-cell adhesion mediator activity). A: Gráfico de bosque; B: Funnel plot.

5. DISCUSIÓN

El Melanoma Metastásico Cerebral (MMC) es un tipo de cáncer de melanoma con tendencia a metastatizar al cerebro, siendo un paso mortal en la progresión del cáncer (Leonardi et al., 2018). Estos pacientes presentan la mayor morbilidad y mortalidad de los pacientes de melanoma. Sin embargo, las opciones de tratamiento son aún muy limitadas debido a falta de un mayor conocimiento de los mecanismos moleculares y celulares responsables.

Los avances tecnológicos de la transcriptómica, junto con la bioinformática, han supuesto un gran impulso en la investigación de las enfermedades humanas, con importantes consecuencias clínicas en cuanto al diagnóstico, el pronóstico y el tratamiento de los pacientes (Lowe et al., 2017). Estas ciencias generan una gran cantidad de información conocida como *big data*, necesitando para su gestión bases de datos como Gene Expression Omnibus (GEO). Destacar los principios FAIR (Findable, Accesible, Interoperable, Reusable) (Wilkinson et al., 2016) y el proceso de investigación de acceso público (Open Science) en la ciencia, que pretenden crear estándares para que los datos se depositen en los repositorios de forma homogénea.

Con estos datos se pueden realizar distintos abordajes destacando en este trabajo el metaanálisis funcional. El metaanálisis es un conjunto de técnicas que buscan agregar y contrastar los resultados de varios estudios relacionados con el mismo tema y analizar las causas de variación entre ellos (Gurevitch et al., 2018). En este caso, se empleó la herramienta web MetaFun para realizar un metaanálisis a nivel de función que ayude a determinar biomarcadores de diagnóstico y progresión del melanoma metastásico cerebral.

Tras la revisión sistemática se seleccionaron 2 estudios para su análisis (GSE46517 y GSE60464). Esta búsqueda presentó limitaciones debido al bajo número de estudios relacionados, la disponibilidad de los datos y las distintas plataformas empleadas. A partir de los estudios seleccionados se realizaron 3 comparaciones de tipo caso control. Analizando los resultados del metaanálisis funcional a nivel de términos GO, se ha detectado, a grandes rasgos, un enriquecimiento de las funciones relacionadas con la proliferación, replicación del ADN y división celular en los pacientes con melanoma. Frente a una relación de los procesos de queratinización, desarrollo epidermal y adhesión celular con los pacientes sanos.

Concretamente, el proceso de replicación del ADN se ve sobreexpresado en pacientes enfermos como muestran los términos GO:0006261, GO:0006260 junto con una mayor actuación de los procesos de reparación (GO:0006281). También aparece un aumento de la división celular con el término GO:0051301. Los términos GO:0017116, GO:0004003, GO:0003678, GO:0004386 hacen referencia a la actividad helicasa que se ve aumentada en los pacientes enfermos. Estos procesos muestran una correcta segregación de las muestras ya que son todos procesos característicos y determinantes en el origen y progresión del cáncer en general.

Estas funciones pueden ser puntos de estudio para un abordaje terapéutico, mejorando las terapias actuales. Por ejemplo, Datta & Brosh (2018) se centran en estudiar las helicasas como dianas farmacológicas para el tratamiento del cáncer. Investigando la inhibición de la función de la helicasa inducida por pequeñas moléculas mediante cribados de compuestos utilizando enfoques bioquímicos y celulares. Otro abordaje presenta *“La replicación del ADN, clave para*

evitar que el cáncer se dispare” (Ecancer, 2014). En este estudio Arturo Calzada, se centra en estudiar los mecanismos básicos de división celular, y también los mecanismos de replicación del ADN como puntos de control clave que eviten la escalada de las mutaciones producidas. Respecto al genoma del melanoma, está muy reordenado y es muy heterogéneo. Estudios de caracterización genómica de melanomas metastásicos y líneas celulares de melanoma han descubierto muchas regiones de aberraciones cromosómicas recurrentes con pocos genes reconocibles o validados relevantes para el cáncer (Kabbarah et al., 2010). Existe aún gran potencial de eventos genéticos por descubrir que ayuden a determinar la patogénesis del melanoma, constituyendo una posible línea de investigación.

Otro fallo en los mecanismos de la célula se da en la reparación del ADN. La identificación de los procesos de reparación del ADN es crucial para el hallazgo de tratamientos más eficaces y específicos contra el cáncer (*Los mecanismos de reparación del ADN, una esperanza contra el cáncer*, 2022). Estudios en el ciclo celular y su regulación, como el de Peralta-Zaragoza et al. (1997), han permitido conocer cómo conservar la fidelidad e integración de los sistemas de replicación y reparación del genoma.

Por otro lado, los términos GO:0030280, GO:0043588, GO:0009913 hacen referencia al desarrollo de la piel y la diferenciación celular de la epidermis que se relaciona con los pacientes sanos. También tiene importancia la diferenciación de los queratinocitos y la queratinización mostrada por los términos GO:0030216 y GO:0031424 respectivamente. Por último, las uniones célula a célula y la organización de la unión celular (GO:0098632, GO:0034330 respectivamente). Además, se relaciona la actuación del complejo MHC con los pacientes controles (GO:0042611).

Esto corresponde con los resultados esperados, la piel está sana y tiene un desarrollo y proliferación normal en los pacientes controles. El término GO:0030216 hace referencia al proceso de diferenciación de los queratinocitos que forman las capas de la epidermis envejeciendo según alcanzan la superficie hasta ser células muertas. Los nevus epidérmicos histológicamente presentan papilomatosis e hiperqueratosis, engrosamiento de la granulosa y aumento de la melanina en la lámina basal (Garcias-Ladaria et al., 2018). Estos nevus suelen mantenerse estables pero en ocasiones pueden transformarse en melanoma dando lugar al cáncer, por ello el Instituto Nacional del Cáncer (NCI) recomienda un control y estudio rutinario de los nevus del paciente ayudando a la detección precoz del cáncer.

La adhesión celular entre superficies de células y a matrices extracelulares representa un mecanismo fundamental en la organización de los tejidos e influye en el comportamiento biológico y la arquitectura de los tumores (Cerón et al., 2014). Las Moléculas de Adhesión Celular (CAM), incluyen selectinas, súper familia de las inmunoglobulinas, cadherinas e integrinas, teniendo cada una de ellas un papel distinto en la progresión del tumor (Brooks et al., 2010). En la etapa de migración se busca la pérdida de adhesión celular para conseguir abandonar el tejido. El hecho de que el melanoma tenga gran capacidad metastásica llama la atención y es debido entre otros a que las células de melanoma comparten muchos antígenos con las células endoteliales vasculares. Entre estas moléculas se encuentran las moléculas de adhesión celular del melanoma (MCAM), que permite a las células de melanoma sobrevivir a las fuerzas físicas dentro de la circulación, adherirse a la pared vascular dentro de órganos distantes y extravasarse

al parénquima (Braeuer et al., 2014). Se identificó por primera vez como un antígeno de melanoma humano altamente expresado en melanomas primarios y metastásicos avanzados, pero raramente expresado en melanocitos epidérmicos normales (Lehmann et al., 1987). Por tanto, la adhesión celular es una función que varía según el estadio de la enfermedad, en nuestro caso esta sobreexpresión en los controles frente a los casos se puede deber a trabajar con pacientes metastásicos, en los cuales suelen disminuir para dar lugar a la migración.

El término GO:0042611 está relacionado con el Complejo Mayor de Histocompatibilidad (MHC) que está conformado por un conjunto de genes cuyos productos son expresados en la superficie de las células del sistema inmune. Este complejo tiene un papel fundamental en la detección de las células tumorales mediante la acción de reconocimiento por parte de los linfocitos T CD8+ de los antígenos asociados a tumor (AAT). En melanoma por ejemplo, existen antígenos inmunodominantes derivados de proteínas involucradas en la síntesis de melanina y expresados tanto en el tumor como en los melanocitos normales. Esto produce fallos en el reconocimiento del antígeno tumoral por parte del sistema inmune, de ahí la sobreexpresión de este término en pacientes control. Recalcando su importancia en numerosos estudios como Salazar (2000) definiendo al sistema inmune como *“herramienta estratégica en la batalla contra el cáncer”*.

En definitiva, las funciones identificadas en este estudio pueden proporcionar nuevas preguntas de investigación para ser abordadas y contestadas en otros trabajos posteriores sobre esta enfermedad. Este mayor conocimiento del melanoma metastásico cerebral contribuiría a un enfoque adaptado a la medicina personalizada y de precisión.

6. CONCLUSIONES

1. La metodología de metaanálisis empleada en este estudio es un método robusto y eficaz, que ha permitido evaluar e integrar datos de distintos estudios de transcriptómica tras una exhaustiva revisión sistemática.
2. La revisión sistemática realizada pone de manifiesto la falta de investigación en el ámbito. La escasez de estudios, disponibilidad de datos y uso de plataformas obliga a mezclar comparativas diferentes.
3. La herramienta web Metafun ha demostrado su sencillez y correcto funcionamiento para caracterizar enfermedades humanas. Sin embargo, al ser una herramienta web en desarrollo, cuenta con problemas que deben ser resueltos.
4. Analizando los resultados del metaanálisis funcional a nivel de términos GO, se ha identificado, a grandes rasgos, un enriquecimiento de las funciones relacionadas con la proliferación y división celular en los pacientes con melanoma. Frente a una relación de los procesos de queratinización, desarrollo epidermal y adhesión celular con los pacientes sanos.
5. Este estudio ha permitido caracterizar los mecanismos del cáncer, frente a funciones que protegen a los individuos sanos. Identificando nuevas preguntas de investigación para ser abordadas y contestadas en otros trabajos sobre esta enfermedad.

7. REFERENCIAS

- Ali, Z., Yousaf, N., & Larkin, J. (2013). Melanoma epidemiology, biology and prognosis. *European Journal of Cancer Supplements*, 11(2), 81–91. <https://doi.org/10.1016/j.ejcsup.2013.07.012>
- Allison, D. B., Cui, X., Page, G. P., & Sabripour, M. (2006). Microarray data analysis: From disarray to consolidation and consensus. *Nature Reviews Genetics*, 7(1), 55–65. <https://doi.org/10.1038/nrg1749>
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., & Sherlock, G. (2000). Gene Ontology: Tool for the unification of biology. *Nature Genetics*, 25(1), 25–29. <https://doi.org/10.1038/75556>
- Athar, A., Füllgrabe, A., George, N., Iqbal, H., Huerta, L., Ali, A., Snow, C., Fonseca, N. A., Petryszak, R., Papatheodorou, I., Sarkans, U., & Brazma, A. (2019). ArrayExpress update – from bulk to single-cell expression data. *Nucleic Acids Research*, 47(D1), D711–D715. <https://doi.org/10.1093/nar/gky964>
- Bafaloukos, D., & Gogas, H. (2004). The treatment of brain metastases in melanoma patients. *Cancer Treatment Reviews*, 30(6), 515–520. <https://doi.org/10.1016/j.ctrv.2004.05.001>
- Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., Marshall, K. A., Phillippy, K. H., Sherman, P. M., Holko, M., Yefanov, A., Lee, H., Zhang, N., Robertson, C. L., Serova, N., Davis, S., & Soboleva, A. (2012). NCBI GEO: Archive for functional genomics data sets—update. *Nucleic Acids Research*, 41(D1), D991–D995. <https://doi.org/10.1093/nar/gks1193>
- Bedia, C. (2018). Experimental Approaches in Omic Sciences. In *Comprehensive Analytical Chemistry* (Vol. 82, pp. 13–36). Elsevier. <https://doi.org/10.1016/bs.coac.2018.07.002>
- Brauer, R. R., Watson, I. R., Wu, C.-J., Mobley, A. K., Kamiya, T., Shoshan, E., & Bar-Eli, M. (2014). Why is melanoma so metastatic? *Pigment Cell & Melanoma Research*, 27(1), 19–36. <https://doi.org/10.1111/pcmr.12172>
- Brooks, S. A., Lomax-Browne, H. J., Carter, T. M., Kinch, C. E., & Hall, D. M. S. (2010). Molecular interactions in cancer cell metastasis. *Acta Histochemica*, 112(1), 3–25. <https://doi.org/10.1016/j.acthis.2008.11.022>
- Carlson, M (2016). hgu133a.db: Affymetrix Human Genome U133 Set annotation data (chip hgu133a). R package version 3.2.3.
- Cerón, L. F. Z., Bernal, S. I. F., Navia, C. A., Mv, A., & Torres, J. S. S. (2014). *Adhesión celular: El ensamblaje de la vía al cáncer*. 6.
- Chang, A. E., Karnell, L. H., & Menck, H. R. (1998). The National Cancer Data Base report on cutaneous and noncutaneous melanoma: A summary of 84,836 cases from the past decade. *Cancer*, 83(8), 1664–1678. [https://doi.org/10.1002/\(SICI\)1097-0142\(19981015\)83:8<1664::AID-CNCR23>3.0.CO;2-G](https://doi.org/10.1002/(SICI)1097-0142(19981015)83:8<1664::AID-CNCR23>3.0.CO;2-G)
- Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szcześniak, M. W., Gaffney, D. J., Elo, L. L., Zhang, X., & Mortazavi, A. (2016). A survey of best practices for RNA-seq data analysis. *Genome Biology*, 17(1), 13. <https://doi.org/10.1186/s13059-016-0881-8>
- Cos, M. (2010). Transcriptómica (mARN y miR). Nuevos métodos de diagnóstico molecular Vol. 9 (4), 160-164. [https://doi.org/10.1016/S1578-1550\(10\)70035-8](https://doi.org/10.1016/S1578-1550(10)70035-8)
- Datta, A., & Brosh, R. M. (2018). New Insights Into DNA Helicases as Druggable Targets for Cancer Therapy. *Frontiers in Molecular Biosciences*, 5. <https://www.frontiersin.org/articles/10.3389/fmolb.2018.00059>
- Du, P., Kibbe, W. A., & Lin, S. M. (2008). lumi: A pipeline for processing Illumina microarray.

- Bioinformatics* (Oxford, England), 24(13), 1547–1548. <https://doi.org/10.1093/bioinformatics/btn224>
- Dunning, M, Lynch A, Eldridge M (2015). illuminaHumanv4.db: Illumina HumanHT12v4 annotation data (chip illuminaHumanv4). R package version 1.26.0.
- Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D. M., Forman, D., & Bray, F. (2015). Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012: Globocan 2012. *International Journal of Cancer*, 136(5), E359–E386. <https://doi.org/10.1002/ijc.29210>
- Ferreira González, I., Urrútia, G., & Alonso-Coello, P. (2011). Revisiones sistemáticas y metaanálisis: Bases conceptuales e interpretación. *Revista Española de Cardiología*, 64(8), 688–696. <https://doi.org/10.1016/j.recesp.2011.03.029>
- García, F. G., García, G. G., & Biològiques, U. de V. F. de C. (2016). *Métodos de análisis de enriquecimiento funcional de estudios genómicos*. Universitat de València, Facultat de Ciències Biològiques. <https://books.google.es/books?id=SXS4QAACAAJ>
- Garcias-Ladaria, J., Cuadrado Rosón, M., & Pascual-López, M. (2018). Nevus epidérmicos y síndromes relacionados. Parte 1: Nevus queratinocíticos. *Actas Dermo-Sifiligráficas*, 109(8), 677–686. <https://doi.org/10.1016/j.ad.2018.05.005>
- Gautier, L., Cope, L., Bolstad, B. M., & Irizarry, R. A. (2004). affy—Analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* (Oxford, England), 20(3), 307–315. <https://doi.org/10.1093/bioinformatics/btg405>
- Gurevitch, J., Koricheva, J., Nakagawa, S., & Stewart, G. (2018). Meta-analysis and the science of research synthesis. *Nature*, 555(7695), 175–182. <https://doi.org/10.1038/nature25753>
- Heller, M. J. (2002). DNA Microarray Technology: Devices, Systems, and Applications. *Annual Review of Biomedical Engineering*, 4(1), 129–153. <https://doi.org/10.1146/annurev.bioeng.4.020702.153438>
- Jilaveanu, L. B., Parisi, F., Barr, M. L., Zito, C. R., Cruz-Munoz, W., Kerbel, R. S., Rimm, D. L., Bosenberg, M. W., Halaban, R., Kluger, Y., & Kluger, H. M. (2015). PLEKHA5 as a Biomarker and Potential Mediator of Melanoma Brain Metastasis. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, 21(9), 2138–2147. <https://doi.org/10.1158/1078-0432.CCR-14-0861>
- Kabbarah, O., Nogueira, C., Feng, B., Nazarian, R. M., Bosenberg, M., Wu, M., Scott, K. L., Kwong, L. N., Xiao, Y., Cordon-Cardo, C., Granter, S. R., Ramaswamy, S., Golub, T., Duncan, L. M., Wagner, S. N., Brennan, C., & Chin, L. (2010). Integrative genome comparison of primary and metastatic melanomas. *PLoS One*, 5(5), e10770. <https://doi.org/10.1371/journal.pone.0010770>
- Kaisa, K., Kumar, H., Panigrahi, M., Dutt, T., & Bhushan, B. (2020). RNA Sequencing: A Potent Transcription Profiling Tool. *International Journal of Current Microbiology and Applied Sciences*, 9(10), 891–905. <https://doi.org/10.20546/ijcmas.2020.910.107>
- Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M., & Tanabe, M. (2021). KEGG: Integrating viruses and cellular organisms. *Nucleic Acids Research*, 49(D1), D545–D551. <https://doi.org/10.1093/nar/gkaa970>
- Karz, A., Dimitrova, M., Kleffman, K., Alvarez-Breckenridge, C., Atkins, M. B., Boire, A., Bosenberg, M., Brastianos, P., Cahill, D. P., Chen, Q., Ferguson, S., Forsyth, P., Glitza Oliva, I. C., Goldberg, S. B., Holmen, S. L., Knisely, J. P. S., Merlino, G., Nguyen, D. X., Pacold, M. E., ... Hernando, E. (2022). Melanoma central nervous system metastases: An update to approaches, challenges, and opportunities. *Pigment Cell & Melanoma Research*, 35(6), 554–572. <https://doi.org/10.1111/pcmr.13059>
- Korotkevich, G., Sukhov, V., Budin, N., Shpak, B., Artyomov, M. N., & Sergushichev, A. (2016). *Fast gene set enrichment analysis* [Preprint]. *Bioinformatics*. <https://doi.org/10.1101/060012>
- Langan, D., Higgins, J. P. T., Jackson, D., Bowden, J., Veroniki, A. A., Kontopantelis, E., Viechtbauer, W., & Simmonds, M. (2019). A comparison of heterogeneity variance

- estimators in simulated random-effects meta-analyses. *Research Synthesis Methods*, 10(1), 83–98. <https://doi.org/10.1002/jrsm.1316>
- Lehmann, J. M., Holzmann, B., Breitbart, E. W., Schmiegelow, P., Riethmüller, G., & Johnson, J. P. (1987). Discrimination between benign and malignant cells of melanocytic lineage by two novel antigens, a glycoprotein with a molecular weight of 113,000 and a protein with a molecular weight of 76,000. *Cancer Research*, 47(3), 841–845.
- Leonardi, G., Falzone, L., Salemi, R., Zanghì, A., Spandidos, D., McCubrey, J., Candido, S., & Libra, M. (2018). Cutaneous melanoma: From pathogenesis to therapy (Review). *International Journal of Oncology*. <https://doi.org/10.3892/ijo.2018.4287>
- Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P. A., Clarke, M., Devereaux, P. J., Kleijnen, J., & Moher, D. (2009). The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventions: Explanation and Elaboration. *PLoS Medicine*, 6(7), e1000100. <https://doi.org/10.1371/journal.pmed.1000100>
- Los mecanismos de reparación del ADN, una esperanza contra el cáncer*. (2022, enero 28). www.nationalgeographic.com.es.
- Lowe, R., Shirley, N., Bleackley, M., Dolan, S., & Shafee, T. (2017). Transcriptomics technologies. *PLoS Computational Biology*, 13(5), e1005457. <https://doi.org/10.1371/journal.pcbi.1005457>
- Lozano Teruel, J. A. (2005). *Bioquímica y biología molecular para ciencias de la salud* (3a ed.). Interamericana-McGraw-Hill.
- Lunares comunes, nevos displásicos y el riesgo de melanoma—NCI* (nciglobal,ncienterprise). (2018, May 9). [CgvArticle]. <https://www.cancer.gov/espanol/tipos/piel/hoja-informativa-lunares>
- Maglott, D., Ostell, J., Pruitt, K. D., & Tatusova, T. (2011). Entrez Gene: Gene-centered information at NCBI. *Nucleic Acids Research*, 39(suppl_1), D52–D57. <https://doi.org/10.1093/nar/gkq1237>
- Maleki, F., Ovens, K., Hogan, D. J., & Kusalik, A. J. (2020). Gene Set Analysis: Challenges, Opportunities, and Future Research. *Frontiers in Genetics*, 11, 654. <https://doi.org/10.3389/fgene.2020.00654>
- Malmierca, P. (2019) Metafun: herramienta web para la integración y caracterización funcional de estudios ómicos con técnicas de metaanálisis. Trabajo Final de Máster. Máster Universitario en Bioinformática, Universitat de València. pp 60.
- Manzoni, C., Kia, D. A., Vandrovcova, J., Hardy, J., Wood, N. W., Lewis, P. A., & Ferrari, R. (2018). Genome, transcriptome and proteome: The rise of omics data and their integration in biomedical sciences. *Briefings in Bioinformatics*, 19(2), 286–302. <https://doi.org/10.1093/bib/bbw114>
- Massagué, J. (2009). *Evolución y metástasis del cáncer*.
- Montaner, D., & Dopazo, J. (2010). Multidimensional Gene Set Analysis of Genomic Data. *PLoS ONE*, 5(4), e10348. <https://doi.org/10.1371/journal.pone.0010348>
- Monzón, O. G., Mora Padilla, E., Torres Tobar, L., Gutiérrez, L. D., & Rubi, C. (2011). Bases moleculares del cáncer. *Revista Repertorio de Medicina y Cirugía*, 20(4), 210–216. <https://doi.org/10.31260/RepertMedCir.v20.n4.2011.768>
- Nekrutenko, A., & Taylor, J. (2012). Next-generation sequencing data interpretation: Enhancing reproducibility and accessibility. *Nature Reviews Genetics*, 13(9), 667–672. <https://doi.org/10.1038/nrg3305>
- Peralta-Zaragoza, O., Bahena-Román, M., Díaz-Benítez, C. E., & Madrid-Marina, V. (1997). Regulación del ciclo celular y desarrollo de cáncer: Perspectivas terapéuticas. *Salud Pública de México*, 39, 451–462. <https://doi.org/10.1590/S0036-36341997000500008>
- Phipson, B., Lee, S., Majewski, I. J., Alexander, W. S., & Smyth, G. K. (2016). Robust hyperparameter estimation protects against hypervariable genes and improves power to detect differential expression. *The Annals of Applied Statistics*, 10(2).

- <https://doi.org/10.1214/16-AOAS920>
- Quackenbush, J. (2002). Microarray data normalization and transformation. *Nature Genetics*, 32(S4), 496–501. <https://doi.org/10.1038/ng1032>
- Rastrelli, M., Tropea, S., Rossi, C. R., & Alaibac, M. (2014). Melanoma: Epidemiology, Risk Factors, Pathogenesis, Diagnosis and Classification. *In Vivo*.
- Russo, M. W. (2007). How to Review a Meta-analysis. *Gastroenterology & Hepatology*, 3(8), 637–642.
- Salazar Montes, A., Sandoval Rodríguez, A., & Armendáriz Borunda, J. (Eds.). (2013). *Biología molecular: Fundamentos y aplicaciones en las ciencias de la salud*. McGraw-Hill Interamericana.
- Salazar O., F. (2000). El sistema inmune, herramienta estratégica en la batalla contra el cáncer. *Revista Chilena de Pediatría*, 71(4), 296–306. <https://doi.org/10.4067/S0370-41062000000400003>
- Scatena, C., Murtas, D., & Tomei, S. (2021). Cutaneous Melanoma Classification: The Importance of High-Throughput Genomic Technologies. *Frontiers in Oncology*, 11, 635488. <https://doi.org/10.3389/fonc.2021.635488>
- Summary · Wiki · UBB-CIPF / MetaFunWeb · GitLab. (2022, September 15). GitLab. <https://gitlab.com/ubb-cipf/metafunweb/-/wikis/Summary>
- Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*, 71(3), 209–249. <https://doi.org/10.3322/caac.21660>
- Tarca, A. L., Romero, R., & Draghici, S. (2008). *Analysis of microarray experiments of gene expression profiling*.
- Tebani, A., Afonso, C., Marret, S., & Bekri, S. (2016). Omics-Based Strategies in Precision Medicine: Toward a Paradigm Shift in Inborn Errors of Metabolism Investigations. *International Journal of Molecular Sciences*, 17(9), 1555. <https://doi.org/10.3390/ijms17091555>
- Torre, L. A., Siegel, R. L., Ward, E. M., & Jemal, A. (2016). Global Cancer Incidence and Mortality Rates and Trends—An Update. *Cancer Epidemiology, Biomarkers & Prevention*, 25(1), 16–27. <https://doi.org/10.1158/1055-9965.EPI-15-0578>
- Vafae, F., Dashti, H., & Alinejad-Rokny, H. (2019). Transcriptomic Data Normalization. In *Encyclopedia of Bioinformatics and Computational Biology* (pp. 364–371). Elsevier. <https://doi.org/10.1016/B978-0-12-809633-8.20209-4>
- Vega, J. B. M. (n.d.). *R para principiantes*. Retrieved May 15, 2023, from <https://bookdown.org/jboscomendoza/r-principiantes4/diagramas-de-caja.html>
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1), Article 1. <https://doi.org/10.1038/sdata.2016.18>
- Zhang, Y., Szustakowski, J., & Schinke, M. (2009). Bioinformatics Analysis of Microarray Data. In K. DiPetrillo (Ed.), *Cardiovascular Genomics* (Vol. 573, pp. 259–284). Humana Press. https://doi.org/10.1007/978-1-60761-247-6_15
- Watson, J., Crick, F. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* 171, 737–738 (1953). <https://doi.org/10.1038/171737a0>

8. ANEXOS

Anexo I. Relación del trabajo con los objetivos de desarrollo sostenible de la agenda 2030

Los objetivos de desarrollo sostenible (ODS) fueron establecidos en 2015 por la Asamblea General de las Naciones Unidas (AG-ONU) como un llamamiento universal para poner fin a la pobreza, proteger el planeta y garantizar que para el 2030 todas las personas disfruten de paz y prosperidad. Constituyen 17 objetivos globales interconectados. Siguiendo estos objetivos, la Universidad Politécnica de Valencia (UPV) se une al desarrollo de los ODS a través de sus alumnos, buscando relacionar sus trabajos de final de grado con las ODS.

En el presente trabajo sobre la caracterización funcional del Melanoma Metastásico Cerebral (MMC) mediante el análisis masivo de datos de expresión génica se ve una clara relación con el ODS 3 de salud y bienestar. El ODS 3 busca luchar contra varias de las principales causas de mortalidad y enfermedad, mediante enfoques multisectoriales, con perspectiva de género abordando desigualdades y asegurando una buena salud para las personas. Concordando con el objetivo de este trabajo que es identificar biomarcadores de diagnóstico y pronóstico, que mejoren el tratamiento y el bienestar de los pacientes con MMC.

Tabla A1. Relación del trabajo con los objetivos de desarrollo sostenibles.

Objetivos de Desarrollo Sostenibles	Alto	Medio	Bajo	No Procede
ODS 1. Fin de la pobreza.				X
ODS 2. Hambre cero.				X
ODS 3. Salud y bienestar.	X			
ODS 4. Educación de calidad.				X
ODS 5. Igualdad de género.				X
ODS 6. Agua limpia y saneamiento.				X
ODS 7. Energía asequible y no contaminante.				X
ODS 8. Trabajo decente y crecimiento económico.				X
ODS 9. Industria, innovación e infraestructuras.				X
ODS 10. Reducción de las desigualdades.				X
ODS 11. Ciudades y comunidades sostenibles.				X
ODS 12. Producción y consumo responsables.				X
ODS 13. Acción por el clima.				X
ODS 14. Vida submarina.				X
ODS 15. Vida de ecosistemas terrestres.				X
ODS 16. Paz, justicia e instituciones sólidas.				X
ODS 17. Alianzas para lograr objetivos.				X

Anexo II. Tabla suplementaria

Tabla A2. Softwares empleados y sus versiones.

Software y paquetes	Versión
R	4.2.2
affy	1.76.0
AnnotationDbi	1.60.2
Biobase	2.58.0
geneplotter	1.76.0
GEOquery	2.66.0
hgu133a.db	3.13.0
illuminaHumanv4.db	1.26.0
illuminaio	0.40.0
limma	3.54.2
lumi	2.50.0
medianreps.R	Función paquete MetanIP