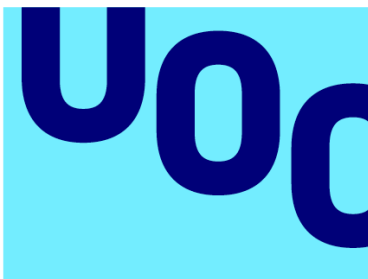


# Estudio de los cambios moleculares del melanoma a lo largo de su evolución



Universitat  
Oberta  
de Catalunya



UNIVERSITAT DE  
BARCELONA

**Nombre Estudiante:**

**Estefanía Gómez Guil**

MU Bioinf. y Bioest.

Desarrollo de Software y  
Aplicaciones

**Nombre Tutor/a de TF**

Teresa Torres Moral

**Profesor/a responsable de la  
asignatura**

Antoni Pérez Navarro

**06/2023**



Esta obra está sujeta a una licencia de  
Reconocimiento-NoComercial-

SinObraDerivada [3.0 España de Creative  
Commons](https://creativecommons.org/licenses/by-nc-nd/3.0/es/)

**FICHA DEL TRABAJO FINAL**

<b>Título del trabajo:</b>	<i>Estudio de los cambios moleculares del melanoma a lo largo de su evolución</i>
<b>Nombre del autor:</b>	<i>Estefanía Gómez Guil</i>
<b>Nombre del consultor/a:</b>	<i>Teresa Torres Moral</i>
<b>Nombre del PRA:</b>	<i>Antoni Pérez Navarro</i>
<b>Fecha de entrega (mm/aaaa):</b>	<i>06/2023</i>
<b>Titulación o programa:</b>	Bioestadística y Bioinformática
<b>Área del Trabajo Final:</b>	<i>Desarrollo de Software y Aplicaciones</i>
<b>Idioma del trabajo:</b>	<i>Castellano</i>
<b>Palabras clave</b>	<i>Melanoma, metástasis, tumor primario</i>

**Resumen del Trabajo**

El melanoma, un tipo agresivo de cáncer de piel, representa un desafío en la actualidad a nivel médico y científico. Este trabajo final de máster se realiza con el objetivo de comprender mejor esta enfermedad y qué cambios moleculares y de expresión génica se dan en su evolución.

Identificar estos posibles cambios en la expresión génica permitirá revelar alteraciones en los procesos moleculares y celulares que contribuyen al desarrollo del melanoma y permitirá obtener pistas sobre posibles dianas terapéuticas y biomarcadores.

Para ello, se ha realizado un estudio comparativo en el lenguaje de programación R de una muestra de datos de pacientes que padecen melanoma y se encuentran en estadios diferentes de la enfermedad, pudiendo haber recaído.

La metodología utilizada ha sido una combinación de la exploración y selección manual de los datos de expresión génica y de su posterior procesamiento en RStudio mediante un pipeline de análisis especializado para este tipo de datos.

Los resultados obtenidos de las diferentes comparativas realizadas revelan hallazgos significativos y consistentes con los estudios realizados en la actualidad. Se encuentran genes diferencialmente expresados entre los estadios finales del melanoma 1 y 4, y en el melanoma ulcerado y el no ulcerado, que es un indicador claro de evolución de la enfermedad.

De este trabajo se puede concluir que existen diferencias en la expresión génica a lo largo de la evolución del melanoma, que estas diferencias se reflejan también en las características que están asociadas a la evolución como la ulceración, y, de manera más general, la necesidad de realizar enfoques multidisciplinarios en la investigación del melanoma, integrando datos genéticos, clínicos, ambientales y personales de cada paciente para una comprensión mayor de la enfermedad.

## **Abstract**

Melanoma, an aggressive type of skin cancer, poses a challenge in the medical and scientific fields today. This final master's project is carried out with the aim of gaining a better understanding of this disease and the molecular and gene expression changes that occur during its progression.

Identifying these potential changes in gene expression will help reveal alterations in the molecular and cellular processes that contribute to the development of melanoma. It will also provide clues about possible therapeutic targets and biomarkers.

To achieve this, a comparative study has been conducted using the R programming language on a sample of data from patients who suffer from melanoma and are in different stages of the disease, including potential relapses.

The methodology used involved a combination of manual exploration and selection of gene expression data, followed by their processing in RStudio using a specialized analysis pipeline for this type of data.

The results obtained from the different comparisons reveal significant and consistent findings with current studies. Differentially expressed genes are found between the final stages of melanoma 1 and 4, as well as between ulcerated and non-ulcerated melanoma, which is a clear indicator of disease progression.

From this work, it can be concluded that there are differences in gene expression throughout the evolution of melanoma. These differences are also reflected in characteristics associated with progression, such as ulceration. Furthermore, it emphasizes the need for multidisciplinary approaches in melanoma research, integrating genetic, clinical, environmental, and personal data from each patient to achieve a better understanding of the disease.

# Índice

1. Introducción.....	1
1.1. Contexto y justificación del Trabajo .....	1
1.2. Objetivos del Trabajo .....	2
1.3. Impacto en sostenibilidad, ético-social y de diversidad .....	3
1.4. Enfoque y método seguido .....	4
1.5. Planificación del Trabajo.....	4
1.7. Breve resumen de productos obtenidos .....	8
1.8. Breve descripción de los otros capítulos de la memoria.....	8
2. Estado del arte .....	9
2.1 Introducción.....	9
2.2 Estudios sobre la evolución de la biología molecular.....	9
2.3 Terapias y tratamientos.....	13
3. Materiales y métodos .....	16
3.1 Conjunto de datos .....	16
3.2 Entorno.....	19
4. Resultados .....	26
5. Conclusiones y trabajos futuros .....	33
6. Glosario .....	36
7. Bibliografía .....	38
8. Anexos .....	40

# Lista de figuras

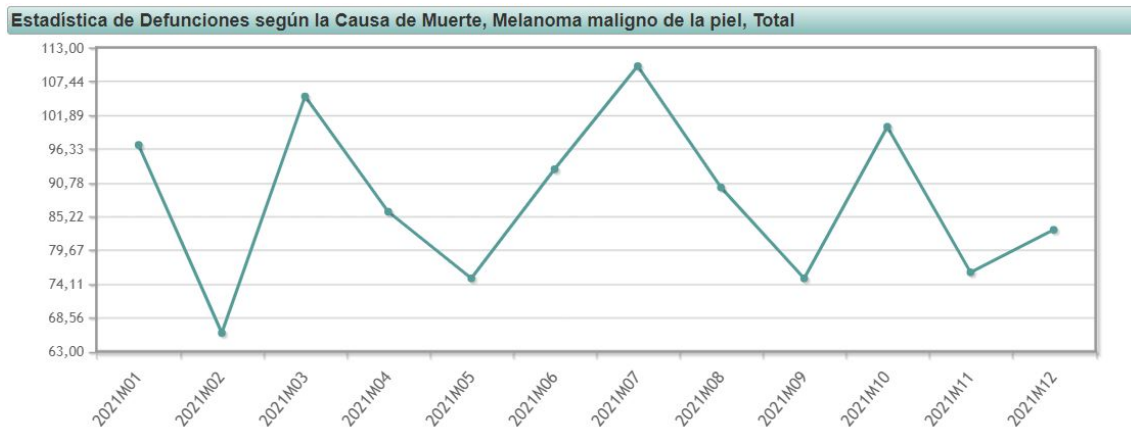
Figura 1 - Tabla INE. Defunciones melanoma maligno de piel, 2021 .....	1
Figura 2 - Diagrama de Gantt.....	6
Figura 3. Noticia vacuna Moderna/Merck. Fuente: La Vanguardia.....	15
Figura 4. Librerías utilizadas .....	22
Figura 5. Importación de los datos .....	22
Figura 6. Control de calidad de los datos.....	23
Figura 7. Parte del preprocesado: Normalización .....	23
Figura 8. Análisis de componentes principales.....	24
Figura 9. Parte del análisis de expresión diferencial: Modelización y Regresión .....	24
Figura 10. Obtención y muestra de los resultados.....	25
Figura 11. Análisis de enriquecimiento génico.....	25
Figura 12. Resultado Estadios .....	29
Figura 13. Resultado Fototipo .....	28
Figura 14. Resultado color de Ojos.....	28
Figura 15. Resultado color Pelo .....	28
Figura 16. Resultado Sexos .....	27
Figura 17. Resultado Recaída.....	30
Figura 18. Resultado de Subtipos .....	30
Figura 19. Matriz de contingencia .....	31
Figura 20. Resultado Chi Cuadrado.....	31
Figura 21. Resultado Ulceración .....	32

# 1. Introducción

## 1.1. Contexto y justificación del Trabajo

El cáncer de piel es un tema cada vez más relevante para la sociedad actual, ya que se ha incrementado su incidencia en los últimos años. La falta de conocimientos, malos hábitos como la sobreexposición a los rayos UV, ausencia de prevención y detección temprana son factores que pueden contribuir en el desarrollo del melanoma, haciendo que pueda causar metástasis. El desarrollo del melanoma es multifactorial, se debe a la interacción del ambiente y la susceptibilidad genética de cada persona.

Hace un siglo, el melanoma tenía una incidencia de 1 cada 500 personas, actualmente esta probabilidad se ha reducido y es de 1 cada 50 personas(1). En datos a nivel estatal, en 2021 causó la muerte de 1056 personas según el INE(2).



**Figura 1 - Tabla INE. Defunciones melanoma maligno de piel, 2021**

En la última década el uso de microarrays de ADN para monitorizar perfiles de tumores de ARN ha definido el funcionamiento del cáncer, ayudando a identificar nuevos fármacos y tratamientos, para saber cómo funciona la resistencia a estos.

El descubrimiento de microRNAs (miRNA) y su funcionamiento aporta luz a la forma de entender los cambios dados en las expresiones proteínicas en el



proceso de la metástasis. Datos recientes sugieren que los miRNAs están funcionalmente involucrados en numerosos procesos fisiológicos, la patogénesis de nevus y de la metástasis en el melanoma maligno.(3)

Este trabajo final de máster que pertenece al ámbito de desarrollo de software y aplicaciones para la salud surge de la necesidad de analizar y comprender el funcionamiento de los tumores primarios. En la actualidad, el cáncer en continúa siendo una de las principales causas de mortalidad en la población.

El resultado que se busca obtener con este estudio es el conocimiento de cómo puede evolucionar un tumor primario hasta llegar a la metástasis, contribuyendo la mejora de la toma de decisiones y futuros diagnósticos.

## 1.2. Objetivos del Trabajo

El objetivo principal que tiene este trabajo es ver la evolución de los tumores primarios hasta pasar a metástasis con datos de expresión génica, mediante la identificación de sus cambios y alteraciones epigenéticas que se dan en la evolución del melanoma.

Como objetivos específicos tenemos el tratamiento y homogeneización de los datos, recabar información para la realización del estudio del arte, planteamiento de estrategias y analizar los resultados obtenidos.

Estos objetivos iniciales pueden variar durante el proceso de análisis debido a que puedan surgir nuevas preguntas o descubrimientos a lo largo de este trabajo.

### 1.3. Impacto en sostenibilidad, ético-social y de diversidad

El presente trabajo final de máster, en el cual se busca ampliar el conocimiento del que se dispone de la evolución del melanoma, se realiza desde una investigación sostenible, ética y socialmente responsable.

En el apartado de la sostenibilidad, cabe destacar que este trabajo utiliza datos ya existentes, reduciendo así la necesidad de la obtención de nuevas muestras y a su vez la generación de residuos y emisiones asociadas a esta recopilación. Por otro lado, toda la documentación utilizada, el tratamiento y análisis de los datos y la generación de esta memoria se realiza de forma completamente digital, minimizando así el uso de recursos físicos.

En cuanto al compromiso ético y social, se toman todas las medidas recomendadas por la legislación vigente necesarias para garantizar la privacidad y la protección de los datos utilizados. En el tratamiento de los datos no se hace ningún tipo de diferenciación en función de orígenes étnicos o raciales, el proyecto abordará la diversidad genética y étnica en los datos y en los análisis, lo que permitirá una mejor comprensión del melanoma en diferentes poblaciones y una atención médica más equitativa y personalizada.

En conclusión, este estudio no solo pretende avanzar en el conocimiento científico, sino también ser un ejemplo de investigación sostenible, ética y socialmente responsable, contribuyendo así al desarrollo de una investigación más responsable y comprometida con el bienestar de la sociedad.

#### 1.4. Enfoque y método seguido

En este trabajo se podrían seguir varios tipos de enfoque de los que se destacan:

1. Enfoque epidemiológico: centrado en analizar la incidencia, prevalencia y mortalidad del melanoma.
2. Enfoque clínico: centrado en analizar factores clínicos asociados a la evolución del melanoma, como pueden ser edad, género, tipo...
3. Enfoque molecular: centrado en analizar factores moleculares asociados con la evolución del melanoma como mutaciones genéticas o alteraciones en la expresión génica.

Debido al tipo de datos obtenidos este análisis se va a realizar mediante el enfoque molecular.

El método por seguir se compone de los siguientes pasos:

1. Exploración de los datos.
2. Preparación de los datos.
3. Análisis y representación gráfica de los datos.
4. Identificación y búsqueda de diferencias entre los diferentes grupos.
5. Validación y verificación de resultados

Al igual que en los objetivos este apartado puede ir variando a lo largo de la realización de este estudio.

#### 1.5. Planificación del Trabajo

##### 1.5.1. Tareas

Para conseguir llegar a realizar los objetivos planteados previamente, se proponen las siguientes tareas:

1. Búsqueda bibliográfica
2. Exploración de datos
3. Tratamiento y homogeneización de los datos
4. Análisis estadístico de los datos
5. Análisis de los resultados
6. Redacción de la memoria
7. Elaboración de la presentación

### 1.5.2. Calendario e Hitos

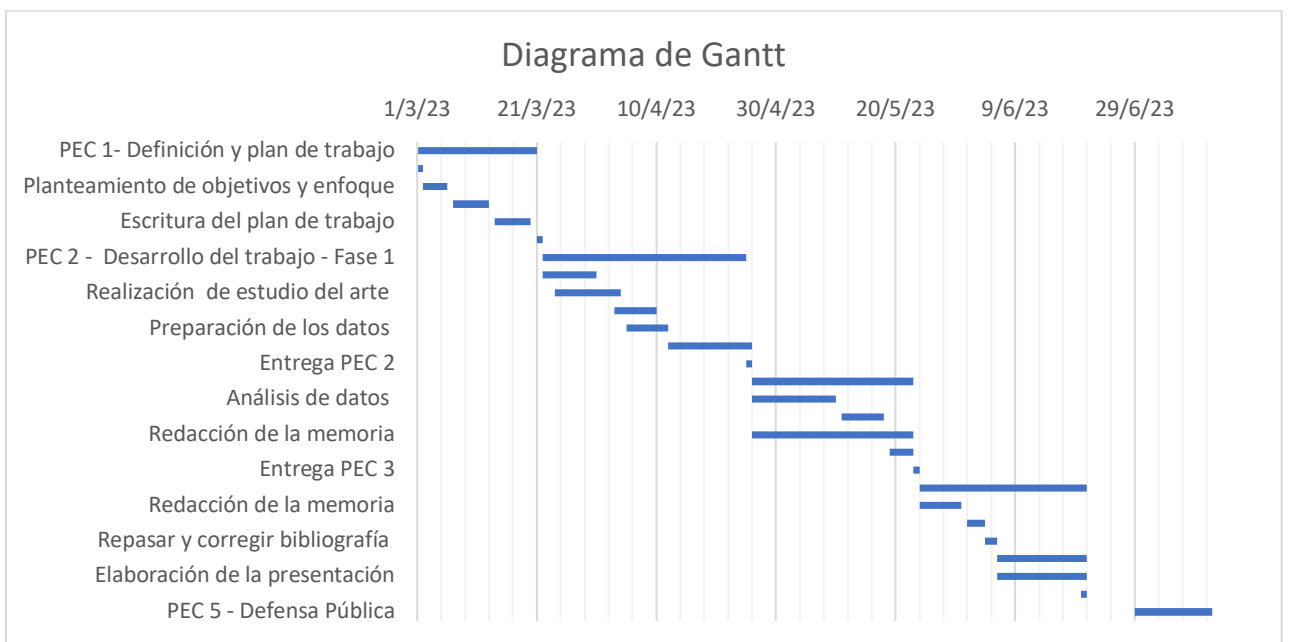
El calendario inicial planteado se basa en las fechas límite de cada prueba de evaluación continuada propuesta dividiendo cada una de las tareas dentro de los períodos de tiempo marcados.

Dentro de este calendario se encuentran marcados los hitos a realizar, las tareas propuestas y la duración prevista de cada una. Como en el resto de los apartados, estas fechas y objetivos son variables en el tiempo según vaya avanzando el trabajo o surjan nuevos hitos a realizar.

**Tabla 1 - Calendario de Tareas**

TAREAS	Inicio	Fin	Duración prevista
<b>PEC 1- Definición y plan de trabajo</b>	1/3/23	21/3/23	20
Lectura de los requisitos de entrega	1/3/23	1/3/23	1
Planteamiento de objetivos y enfoque	2/3/23	6/3/23	4
Búsqueda de información y documentación	7/3/23	13/3/23	6
Escritura del plan de trabajo	14/3/23	20/3/23	6
Entrega PEC 1	21/3/23	21/3/23	1
<b>PEC 2 - Desarrollo del trabajo - Fase 1</b>	22/3/23	25/4/23	34
Documentación, búsqueda de información y artículos	22/3/23	31/3/23	9
Realización de estudio del arte	24/3/23	4/4/23	11
Exploración de los datos	3/4/23	10/4/23	7

Preparación de los datos	5/4/23	12/4/23	7
Primeros análisis de los datos	12/4/23	25/4/23	14
Entrega PEC 2	25/4/23	25/4/23	1
<b>PEC 3 - Desarrollo del trabajo Fase 2</b>	26/4/23	23/5/23	27
Análisis de datos	26/4/23	10/5/23	14
Validación y verificación de resultados	11/5/23	18/5/23	7
Redacción de la memoria	26/4/23	23/5/23	27
Conclusiones y resultados	19/5/23	22/5/23	4
Entrega PEC 3	23/5/23	23/5/23	1
<b>PEC 4 - Cierre de la memoria y de la presentación</b>	24/5/23	20/6/23	28
Redacción de la memoria	24/5/23	31/5/23	7
Corrección	1/6/23	3/6/23	3
Repasar y corregir bibliografía	4/6/23	5/6/23	2
Cierre de la memoria	6/6/23	20/6/23	15
Elaboración de la presentación	6/6/23	20/6/23	15
Entrega PEC 4	20/6/23	20/6/23	1
<b>PEC 5 - Defensa Pública</b>	29/6/23	12/7/23	13



**Figura 2 - Diagrama de Gantt**

### 1.5.3. Análisis de riesgos

Durante la realización de este trabajo se pueden encontrar diversos riesgos, los que se pueden esperar que surjan son principalmente relacionados con el tiempo disponible y la organización. Estos contratiempos se pueden ocasionar por compaginar una jornada laboral completa con sus posibles inconvenientes asociados como estrés laboral, falta de tiempo y energía y cambios de horarios inesperados con la ejecución del trabajo.

Por otro lado, la gran cantidad de datos a observar, analizar y procesar, a pesar de que ya tienen previsto un tiempo extenso, puede ser variable y generar el riesgo de no cumplir con alguno de los hitos a tiempo.

También, durante el proceso de análisis de los resultados la no obtención de coherencia en ellos podría alterar la ejecución prevista del calendario. Para un buen funcionamiento y consecución de tareas e hitos, será esencial la capacidad de adaptación y flexibilidad en horarios y fechas propuestas previamente.

### 1.6. Resultados esperados

#### A. Plan de trabajo

De la organización pensada y presentada previamente, se espera principalmente que esta sea útil y holgada en cuanto al tiempo dedicado para cada tarea. Sería ideal poder cumplir a tiempo con todo de forma que los resultados puedan ser repasados y validados juntamente con la tutora del proyecto.

#### B. Memoria

El objetivo y el resultado de la memoria esperado es plasmar paso a paso la organización que se ha realizado para el proceso, cómo se han tratado y analizado los datos y mostrar clara y concisamente que resultados se han obtenido. También, se intenta transmitir el contexto en el que se desarrolla el trabajo y su importancia y utilidad posterior.

### C. Presentación virtual

La presentación virtual se espera que sea amena y entendible, haciendo un repaso de toda la memoria generada y dando énfasis a los resultados finales de la comparación realizada.

#### 1.7. Breve resumen de productos obtenidos

El producto final de este trabajo final de máster se basa en el análisis y la comparación de datos previamente recopilados, relacionados con el melanoma. Cabe destacar que este trabajo no tiene como objetivo la obtención de un producto a partir de él como por ejemplo un software o un método de análisis.

Los resultados de este TFM podrían servir como base para otros trabajos complementarios y futuras investigaciones en el campo del melanoma y su desarrollo.

#### 1.8. Breve descripción de los otros capítulos de la memoria

En el segundo punto de este trabajo llamado Estudio del Arte se realizará una recopilación de datos sobre el melanoma en la actualidad y se aportará el contexto necesario para entender los resultados que se obtendrán a posteriori.

El tercer apartado del trabajo contiene la explicación del material y los métodos utilizados para el análisis. Se explica de dónde provienen los datos, cómo están realizados y qué contienen. También, se explica el pipeline o guion de análisis utilizado.

El cuarto y quinto apartados aportarán los resultados obtenidos de las comparativas realizadas y las conclusiones que se pueden obtener de la realización del trabajo.

## 2. Estado del arte

### 2.1 Introducción

El melanoma es un tipo de cáncer cutáneo agresivo y potencialmente mortal que se origina en los melanocitos, las células encargadas de la producción de melanina. A pesar de que el melanoma solo representa el 1% de los cánceres de piel, es el responsable de causar el 75% de las muertes relacionadas con esta enfermedad(4). Además, la incidencia del melanoma está aumentando rápidamente en todo el mundo, lo que lo convierte en un problema de salud cada vez más alarmante a nivel mundial.

La evolución molecular del melanoma es un campo de investigación que se encuentra en evolución constante. Su investigación ha permitido la identificación de nuevas dianas terapéuticas, estrategias de tratamiento, el desarrollo de nuevas terapias dirigidas y de inmunoterapia. Se espera que la investigación futura siga identificando nuevas mutaciones y vías de señalización implicadas en la progresión del melanoma, lo que puede llevar al desarrollo de nuevas terapias y estrategias para tratar la enfermedad.

Una de las características más distintivas del melanoma es su alta tasa de mutación. La exposición a la radiación ultravioleta (UV) del sol es la principal causa de estas mutaciones. Los melanomas pueden surgir de forma espontánea debido a mutaciones aleatorias en el ADN. Estas mutaciones pueden activar o desactivar ciertos genes que regulan el crecimiento y la muerte celular programada, lo que puede llevar a la formación de tumores.

### 2.2 Estudios sobre la evolución de la biología molecular

En los últimos años, se ha avanzado en la comprensión de la biología molecular y celular del melanoma. De manera que se han identificado varios genes y vías de señalización implicadas en su desarrollo y progresión. Uno de los hallazgos más significativos ha sido la identificación de mutaciones en los genes BRAF y



NRAS en muchos pacientes con melanoma. Las mutaciones de estos genes activan la vía de la señalización MAPK, que promueven el crecimiento y la supervivencia celular, la cual, está implicada en la proliferación y metástasis del melanoma(5).

La mayoría de los melanomas contienen mutaciones en los genes BRAF y NRAS, que forman parte de la vía de señalización RAS/RAF/MAPK. Estos genes codifican proteínas que regulan el crecimiento y la supervivencia celular. Las mutaciones en BRAF y NRAS pueden activar la vía de la señalización RAS/RAF/MAPK de manera constante, lo que resulta en un crecimiento celular descontrolado y la formación de tumores. Por esta razón, la vía RAS/RAF/MAPK se ha convertido en un objetivo importante para el tratamiento del melanoma. Además, de la vía RAS/RAF/MAPK, también se han identificado otras vías de señalización importantes en el melanoma, como la vía PI3K/AKT/mTOR y la vía Wnt/ $\beta$ -catenina.(6) Estas vías también regulan procesos celulares como la supervivencia, la diferenciación y la apoptosis. La activación de estas vías puede contribuir al desarrollo del melanoma y la resistencia a los tratamientos.

Los pares de ligando-receptor son una forma de comunicación celular que desempeña un papel crucial en la génesis, crecimiento, metástasis y pronóstico del tumor. El estudio realizado por Lin Z. et al. (7), examinó 131 pares de ligando-receptor correspondientes al pronóstico del melanoma mediante el análisis de datos del TCGA( The Cáncer Genome Atlas) del melanoma y los 2293 pares de ligando-receptor recuperados de la base de datos connectomeDB2020. Los resultados de este estudio demostraron que los pares de ligando-receptores son vitales para la heterogeneidad molecular del melanoma y se caracterizaron 3 subtipos diferentes. Se encontró que el subtipo C1 tiene el peor pronóstico y el C3 un mejor pronóstico que los otros. Esto podría estar relacionado con la activación e inhibición diferidas en el ciclo celular y las vías relacionadas con el sistema inmunológico. En este estudio también se ha demostrado que los pares de ligando-receptor están significativamente asociados con el pronóstico y la terapia del melanoma. El modelo de puntuación basado en pares de ligando-receptor mostró potencial para la evaluación del pronóstico del melanoma y

predicción del resultado a la terapia inmunológica, lo que es crucial para la supervivencia de los pacientes.

Otro tema relevante en la investigación del melanoma es la secuenciación de la terapia dirigida y la inmunoterapia, que son dos modalidades de tratamiento aprobadas para pacientes con melanoma avanzado. Estudios recientes han abordado este tema y han demostrado que la secuenciación óptica depende de la presencia o ausencia de las mutaciones en los genes BRAF del tumor del paciente. En pacientes con mutaciones BRAF, la terapia dirigida tiene que ser la primera línea de tratamiento, seguida de la inmunoterapia. Por otro lado, en pacientes sin estas mutaciones, la inmunoterapia en cambio tiene que ser la primera opción. El enfoque personalizado del tratamiento puede mejorar la eficacia de este y por lo tanto, la supervivencia de los pacientes con un melanoma en estado avanzado(8). La identificación de estas mutaciones ha llevado al desarrollo de terapias dirigidas que tienen como objetivo específico estas vías de señalización. Los inhibidores de BRAF y MEK, que bloquean la vía MAPK, se han utilizado con éxito en el tratamiento del melanoma metastásico con mutaciones en BRAF(9). También se están desarrollando nuevas terapias dirigidas que tienen como objetivo otros genes y vías de señalización implicados en la progresión del melanoma.

Se han descubierto nuevos microARN relacionados directamente con la expresión de genes involucrados en el melanoma. Gracias al uso de diversas herramientas informáticas, se han identificado genes y vías que podrían ser importantes. Los miARNs impactan en la progresión, iniciación y respuesta al tratamiento del melanoma al afectar a una amplia gama de procesos biológicos(6). Entre los microARN más destacados se encuentran el miR-181 a/b, miR-633, miR-128-3p, miR-137 y miR-534-5p, los cuales podrían ser objetivo de potenciales terapias contra el melanoma y útiles en su diagnóstico y tratamiento.

Por otro lado, la identificación de comunidades celulares y de interacciones en el microambiente se lleva a cabo utilizando técnicas de análisis de red, como la llamada RCN (Redes de co-ocurrencia de células). Las RCN se construyen a

partir de la coocurrencia de células dentro de pequeñas vecindades y se ha demostrado que son útiles para identificar subpoblaciones celulares y su interacción con por ejemplo la exposición a los rayos UVB(10). En general, estas técnicas están permitiendo una mejor comprensión de la complejidad del microambiente tumoral del melanoma y la interacción entre las células tumorales y las no tumorales.

Los avances recientes en la secuenciación del ADN han permitido una mejor comprensión de la complejidad genética del melanoma. Se han identificado varias mutaciones y aberraciones cromosómicas que pueden contribuir al desarrollo y progresión. Por ejemplo, se ha demostrado que las mutaciones en los genes CDKN2A, TP53 y PTEN están asociadas con un mayor riesgo de melanoma. Además, se han identificado mutaciones en otros genes como TP53, PTEN, CDKN2A y TERT, que también están implicados en la progresión del melanoma. Estas mutaciones pueden actuar de forma sinérgica para promover el crecimiento y supervivencia celular, se cree que son responsables de la heterogeneidad tumoral observada en el melanoma(1).

El proceso de metástasis del cáncer requiere que las células tumorales superen una serie de barreras para reproducirse en otros órganos del cuerpo, lo que requiere de propiedades celulares que vienen reguladas por mecanismos epigenéticos y de transcripción, como la formación de protuberancias en la membrana nuclear y la liberación de presión intranuclear en las células causada por la interacción con la proteína LAP1(11). Uno de los factores que intervienen es el CDR1as, un tipo de ARN circular que actúa como regulador del miR-7, un micro-ARN que regula diferentes procesos celulares. Los diferentes niveles de CDR1as se asocian con diferentes estados del melanoma, sea estado de metástasis o estados donde se está tratando terapéuticamente(12).

### 2.3 Terapias y tratamientos

En las últimas décadas se han realizado importantes avances en el diagnóstico y tratamiento del melanoma. Diversos estudios han demostrado que la detección temprana es esencial para la supervivencia del paciente a largo plazo. A causa de ello, se han desarrollado diversas técnicas de diagnóstico, como la dermatoscopia y la microscopía confocal, para mejorar la detección temprana del melanoma. También, se ha observado un aumento en el uso de las técnicas de imagenología avanzadas y análisis computacionales para el estudio del microambiente tumoral del melanoma. La microscopía de imagen de alta resolución combinada con el análisis cuantitativo de imagen ha permitido una evaluación detallada del comportamiento celular y la interacción en el microambiente tumoral. Además, el uso de métodos de aprendizaje automático y análisis de datos ha permitido una mejor comprensión de la heterogeneidad del melanoma y como las células tumorales interactúan con su microambiente.

Las técnicas de imagenología avanzada más utilizadas son la microscopía de fluoresceína multiplexada y la CyCIF (citometría de imagen basada en ciclos)(13). Este enfoque, combina la microscopía de fluoresceína con el etiquetado secuencial de múltiples anticuerpos para permitir la evaluación simultánea de múltiples biomarcadores en una sola muestra. Además, se han utilizado algoritmos de análisis de imagen para cuantificar la expresión de biomarcadores específicos en diferentes células del melanoma y de su microambiente.

Además de la terapia dirigida, la inmunoterapia se ha convertido en una herramienta importante en el tratamiento del melanoma. La inmunoterapia aprovecha el sistema inmunológico del cuerpo para combatir el cáncer y ha demostrado una tasa de respuesta significativa en pacientes con melanoma avanzado.

La utilización de miARNs como fármacos terapéuticos en el tratamiento de enfermedades cancerosas es una estrategia emergente, pero presenta algunos desafíos en términos de respuesta inmunitaria del huésped. Se ha demostrado que el ARN bicatenario tratado como patógeno puede activar vías de señalización que llevan a la liberación de citocinas proinflamatorias. Sin embargo, las terapias combinadas de miARNs con otras estrategias terapéuticas ya han demostrado su eficacia. Además, la presencia de miARNs en fluidos corporales sugiere su posible utilidad en la detección temprana de diversos tipos de cáncer. Es crucial que se siga investigando para mejorar la eficacia y seguridad de las terapias de miARNs, ya que, su potencial en el futuro del tratamiento oncológico es prometedor. (14)

En la actualidad, en estas últimas semanas, Moderna ha publicado un estudio clínico que se encuentra en fase 2b sobre un tratamiento para el melanoma. Los resultados de este estudio muestran que el tratamiento con mRNA-4157/V940, una vacuna personalizada contra el cáncer de ARN mensajero(mRNA), en combinación con KEYTRUDA (un tratamiento anti-PD-1 de Merck), redujo significativamente el riesgo de recurrencia del melanoma o muerte en pacientes con melanoma de estadio III/IV después de una resección completa. La combinación de los dos tratamientos redujo el riesgo de recurrencia o muerte en un 44% en comparación con el de KEYTRUDA aplicado individualmente. Los resultados demuestran también que la terapia de vacunación personalizada con mRNA-4157/ V940 puede ser beneficiosa para pacientes con melanoma. Se espera que este estudio se amplie a otros tipos de cáncer, no únicamente para el melanoma. Estas empresas planean entrar en un estudio de fase 3 este año(15).

## Una vacuna de Moderna contra el cáncer mejora la eficacia de la inmunoterapia



- Dos tratamientos que estimulan la inmunidad de maneras complementarias demuestran su eficacia cuando se administran juntos a pacientes con melanoma



Investigación en un laboratorio de Moderna en EE.UU. (Moderna / Bearwalk)

**Figura 3. Noticia vacuna Moderna/Merck. Fuente: La Vanguardia**

El estudio anterior tiene una gran importancia teniendo en cuenta el factor de la resistencia a la terapia en casos de melanoma, que es un desafío importante para los médicos y pacientes. Una de las causas de la resistencia es la heterogeneidad tumoral, lo que significa que el melanoma puede tener diferentes subpoblaciones celulares con diferentes perfiles genéticos y moleculares. Esto puede hacer que algunas células sean resistentes al tratamiento mientras que otras no lo sean.

En resumen, en las últimas décadas ha habido un gran avance en la comprensión de la biología del melanoma, lo que ha llevado al desarrollo de nuevos tratamientos. A pesar de estos avances, el desarrollo de resistencia a los tratamientos continúa siendo un problema importante principalmente en casos de recurrencia del melanoma. Para solucionar esto, se están investigando nuevas estrategias, combinando diferentes terapias, terapias más antiguas con terapias que surgen en la actualidad, con el objetivo de encontrar nuevas opciones terapéuticas y mejorar los resultados en pacientes con melanoma.

## 3. Materiales y métodos

### 3.1 Conjunto de datos

El conjunto de datos utilizado se basa en un conjunto de datos provenientes del estudio “Implementation of personalized medicine in cutaneous melanoma based on genomic susceptibility and tumor signatures: Identification of new targets for melanoma treatment” financiado por la fundación La Marató de TV3 y realizado por S.Puig en el año 2013. Son datos clínicos y de arrays de expresión (SurePrint G3 Human Gene Expression v3 8x60K, Agilent) de 41 pacientes de melanoma donde se contiene información sobre tumores primarios en diferentes estadios. Concretamente tenemos 32 melanomas primarios y 9 metástasis cutáneas, sin embargo, trabajaremos con datos de 38 pacientes de los que se dispone de información clínica completa.

Estos datos vienen acompañados de un archivo en el cual se pueden relacionar cada una de las muestras con los datos clínicos de cada uno de los pacientes. A continuación, se comentan las columnas y los datos más relevantes:

- TP\_ARRAY: el número de identificación del array.
- CODI\_Anaxomics: código asignado por la empresa de análisis a la muestra.
- NHC: número de historia clínica del cliente.
- Patient\_gender: género del paciente.
- Updated Genetic Results: resultados genéticos actualizados de cada muestra.
- MDK Expression Profile: perfil de expresión MDK de cada muestra.
- TP\_driver y M1\_driver: genes diferencialmente expresados en tumor primario y metástasis.
- Pipeline used for the analysis: método usado para analizar las muestras.
- Características fenotípicas del paciente: como el color de ojos, del pelo, o la edad en la que se diagnostica el melanoma.

- Datos del tumor primario: el conteo de nevus y su tamaño en la escala de Breslow, si está ulcerado o no, si es mitótico, etc.
- Los estadios iniciales y finales del tumor.
- Si hay recaída del paciente, la fecha y el tiempo que ha tardado en reaparecer.
- Supervivencia del paciente.
- Tratamientos utilizados, inicio y final de la terapia y número de terapias utilizadas.

Los datos sobre los cuales se realiza el análisis vienen datos en formato RAW (en crudo) y contienen un archivo .pdf donde se puede observar el microarray y el resultado obtenido de la reacción a la fluorescencia y unos archivos .txt dónde se procesan estos resultados. Los archivos .txt, que se importan a R para analizar, se componen de las siguientes columnas:

- TYPE: tipo de dato.
- FeatureNum: número de identificación del microarray de la muestra.
- Row: fila.
- Col: columna.
- SubTypeMask: código o máscara que representa el subtipo de sonda utilizada.
- ControlType: tipo de control utilizado en el experimento.
- ProbeName: nombre de la sonda utilizada.
- SystematicName: nombre sistemático, se puede observar el nombre del gen.
- PostionX y PositionY: coordenadas espaciales de la ubicación de la sonda.
- gProcessedSignal: señal procesada, nivel de expresión génica.
- gProcessedSignalError: error de la señal procesada.
- gMedianSignal: mediana de la señal procesada obtenida.
- gBGMedianSignal: mediana de la señal de fondo obtenida.
- gBGPixSDev: desviación estándar de los píxeles de fondo.
- gIsSaturated: indicador booleano de la saturación de la señal.



- gIsFeatNonUnifOL: indicador booleano que indica si el gen se encuentra por encima del fondo no uniforme.
- gIsFeatPopnOL: indicador booleano que indica si el gen pertenece a una población específica.
- gIsBGPpopnOL: indicador booleano de si el fondo pertenece a una población específica.
- IsManualFlag: indicador booleano que indica si hay una marca o “bandera” asociada al gen.
- gBGSubSignal: señal del gen tras la extracción del fondo.
- gIsPosAndSignif: indicador booleano de si la señal es positiva y significativa.
- gIsWellAboveBG: indicador booleano de si la señal está significativamente por encima del fondo.
- SpotExtentX: extensión del punto o sonda en la dirección X.
- gBGMeanSignal: media de la señal basal del gen.

También, teniendo en cuenta y a partir de los datos en formato RAW que contienen la información de expresión génica en crudo, se cuenta con la recodificación de los mismos, donde los datos han sido ordenados y transformados para poder trabajar mejor con ellos y obtener comparaciones y resultados más precisos.

Así pues, para la realización del estudio se trabajará con los datos recodificados y utilizando el cuadro de información en Excel donde se puede identificar qué características corresponden a cada variable.

## 3.2 Entorno

### 3.2.1 Entorno de trabajo

Para la realización de este estudio se ha trabajado con el sistema operativo MacOS y con el IDE RStudio.

### 3.2.2 Lenguaje de programación

El lenguaje de programación utilizado en este proyecto ha sido R, y su utilización viene dada por diversas razones. Primeramente, es un lenguaje de programación de código abierto, lo que lo hace gratuito y accesible. Por otro lado, es un software pensado para el análisis estadístico de datos, permitiendo su visualización de forma óptima. Por último, tiene integración con el sistema de visualización de datos RMarkdown, lo que facilita la generación de informes posteriores al análisis de los datos.

### 3.2.3 Librerías utilizadas

Las librerías utilizadas para el desarrollo del proyecto han sido:

- limma
- readR
- stringR
- Bioconductor
- GEOquery
- ggplot2
- HsAligentDesign026652.db
- xfun
- rlang
- G0.db
- EnhancedVolcano

### 3.3 Métodos

Antes de realizar las comparativas principales de este estudio, se han querido analizar las posibles variables confusoras y fenotípicas que pudiesen estar influyendo en el análisis:

- Sexo: Se buscan diferencias en la expresión génica entre sexos.
- Datos fenotípicos: Según el color del pelo, ojos y fototipo del paciente. Se estudia en función a la clasificación fototípica, clasificación utilizada para categorizar la respuesta de la piel humana a la exposición solar y su capacidad para broncearse o quemarse.

A continuación, se ha realizado un estudio comparando las variables ya asociadas a la evolución del melanoma:

- Estadios del melanoma: Comparativa entre los diferentes estadios con el objetivo de encontrar la diferencia y cuanto de notoria es en los estadios más avanzados.
- Recaída: Es importante examinar como varía la actividad genética en caso de que el paciente experimente una recaída, en comparación a cuando no lo hay.

Además, se han estudiado variables que pudiesen estar también asociadas a la evolución del melanoma:

- Datos del tumor primario: Según el tamaño del melanoma, el conteo de nevus, o si está ulcerado o no.

Una vez realizadas las diferentes comparativas y tras la búsqueda de los genes diferencialmente expresados, se buscará analizar los resultados y obtener patrones y relaciones que se den entre estas características de los pacientes.

### 3.4 Separación de los datos en grupos

Antes de comenzar a comparar los diferentes grupos de interés, se necesita realizar la separación y categorización de los datos. Sin embargo, y como se ha detallado previamente en la explicación de los datos, no se dispone de una variable específica que indique a qué grupo pertenece cada muestra. Ante esta situación, se ha tenido que realizar la separación en carpetas de forma manual para después poder importar los datos en el orden deseado.

Esta separación manual ha requerido de un análisis más exhaustivo de los datos, teniendo que examinar y analizar cuidadosamente las características de cada uno. Aunque esto haya requerido de un esfuerzo adicional, se ha podido asegurar que cada muestra esté separada correctamente en su grupo asignado.

### 3.5 Pipeline

Tras la examinación inicial de los datos y considerar diversas maneras de poder analizarlos, el pipeline seguido para el análisis de los datos se ha dividido en los siguientes bloques (se adjuntan ejemplos de uno de las comparativas realizadas):

#### 1. Configuración inicial:

En este primer bloque se cargan los paquetes necesarios para el análisis y se establecen las configuraciones iniciales en el entorno de ejecución.

```

library(xfun)
library(rlang)
library(limma)
library(readr)
library(stringr)
library(HsAgilentDesign026652.db)
library(GEOquery)
library(ggplot2)
library(EnhancedVolcano)
library(GO.db)

```

Figura 4. Librerías utilizadas

## 2. Importación de datos:

Se definen los parámetros y rutas de directorios para la importación de los datos. Se asignan los nombres de los archivos a las rutas especificadas, se crea el vector Grupo con las etiquetas de cada componente y se leen los microarrays mediante la función “read.maimages”.

```

## 1.Importación de datos
```{r}
params <- list(
  estadio1="Est1",
  estadio4="Est4",
  directorio= "/Users/estefaniagomezguil/Desktop/Estadios Recod"
)
ruta_est1<- file.path(params$directorio, "1")
ruta_est4<- file.path(params$directorio, "4")

files_est1 <- list.files(path = ruta_est1, pattern = ".txt", full.names = TRUE)
files_est4 <- list.files(path = ruta_est4, pattern = ".txt", full.names = TRUE)
Grupo <- c(rep(params$estadio1, length(files_est1)), rep(params$estadio4, length(files_est4)))
files <- c(files_est1, files_est4)
print(files)
targets = read.maimages(file=files, source="agilent", green.only=TRUE)
targets$targets$class <- Grupo
files2<- str_remove_all(string=files, pattern= "US85003608_253949442215")
files2<- str_remove_all(string=files2, patter= "_S01_GE1_107_Sep09")
```

```

Figura 5. Importación de los datos

### 3. Control de Calidad

En este bloque se dedica a la visualización de los datos con el objetivo de realizar un control de calidad. Se generan gráficos MA y MDS.

```
## 2. Control de calidad
```{r}
limma::plotMA3by2(targets, status = targets$genes$ControlType, prefix="MA", device = "pdf")
plotMDS(targets, top =100, gene.selection="pairwise", plot = TRUE, labels = targets$targets$class)
plotMDS(targets, top =100, gene.selection="pairwise", plot = TRUE, labels = files2)
```
```

Figura 6. Control de calidad de los datos

### 4. Preprocesado

Mediante los métodos “normexp” y “quantile” se realiza la corrección de fondo y la normalización de los arrays. A continuación, se realiza la anotación de los genes con identificadores Entrez.

```
### 3.2 Normalización
```{r}
MA.p<-normalizeBetweenArrays(BK, method = 'quantile')
transposed <- t(MA.p$E)
df<-data.frame(value = transposed, group = as.character(Grupo))
plotDensities(MA.p, legend = FALSE, type="1")
```
```

Figura 7. Parte del preprocesado: Normalización

### 5. Filtrado

En este apartado se utilizan diferentes filtros para seleccionar genes de interés. Se crean condiciones lógicas basadas en el tipo de control, la existencia de símbolos y Entrez ID. Por último, se crea una matriz de datos llamada “yfilt” y se genera un gráfico de caja para visualizar el resultado.

## 6. Análisis de componentes principales

Mediante la función “percomp” se realiza un análisis de componentes principales. Seguidamente, se muestra un resumen del resultado obtenido y se visualiza en un gráfico de dispersión.

```
```{r}
pcomp=prcomp(yfilt$E, scale. = TRUE)
eso=summary(pcomp)
eso
```

```{r}
com=as.data.frame(pcomp$x)
ggplot(com, aes(PC1, PC2)) + geom_point()+theme_bw(base_size=32)+xlab(paste("PC1:", eso$importance[2,1])) +
ylab(paste("PC2:", eso$importance[2,2]))+theme(legend.position="top")
```
```

Figura 8. Análisis de componentes principales

## 7. Análisis de expresión diferencial

En este bloque se realiza el análisis de expresión diferencial utilizando el paquete “limma”. Para ello, se construye la matriz de diseño y se ajusta un modelo lineal mediante la función “lmfit”. Por último, se realiza un contraste con la función “makeContrasts” y se realiza la estimación empírica mediante la función “eBayes”.

### ### 4.1 Modelización i Regresión

```
```{r}
design <- model.matrix(~0+Grupo)
colnames(design) = c(params$estadio1, params$estadio4)
design
fit = lmFit(yfilt, design)
```

```{r}
parametre1 = params$estadio1
parametre2 = params$estadio4

contraste = makeContrasts(Est1-Est4, levels = design)
fit2 = contrasts.fit(fit, contraste)
fit2 = eBayes(fit2)
```
```

Figura 9. Parte del análisis de expresión diferencial: Modelización y Regresión

## 8. Resultados

Se muestran las primeras filas de resultados utilizando la función “topTable” y se muestra un gráfico mediante la función “EnhancedVolcano”. Para acabar, se genera un diagrama de Venn donde se puede visualizar los genes significativos.

```
### Resultados

```{r}
topTable(fit2, adjust.method = "bonferroni", n=30)
tT=topTable(fit2, adjust.method = "bonferroni", n=3000)

results = decideTests(fit2, p.value = 0.1)
summary(decideTests(fit2), p.value = 0.1)
```

```{r}
EnhancedVolcano(tT, lab = tT$Symbol, x = "logFC", y = "adj.P.Val", pCutoff = 0.05, FCcutoff=2, labSize = 6.0, ylim=0.95)
```

```{r}
vennDiagram(results, include=c("both"), circle.col = "red", counts.col = "blue", names = c("CHD9-CTL"))
```
```

Figura 10. Obtención y muestra de los resultados

## 9. Análisis de enriquecimiento génico

En este último bloque se realiza el análisis de enriquecimiento génico utilizando las bases de datos GEO y KEGG.

```
## 5. Gene Enrichment Analysis

### Gene Ontology

```{r}
g = goana(fit2, species="Hs", geneid = fit2$genes$EntrezID, FDR = 0.05)
topGO(g, n=20)
```

### KEGG

```{r}
k = kegg(fit2, species="Hs", geneid = fit2$genes$EntrezID, FDR = 0.05)
topKEGG(k, number = 20)
```
```

Figura 11. Análisis de enriquecimiento génico



## 4. Resultados

### 4.1 Análisis exploratorio de los datos

Inicialmente y antes de realizar la comparación comentada previamente en el apartado de métodos, mediante grupos y características de los pacientes, se ha realizado una comparación general con el objetivo de hacer una exploración inicial de los datos.

En esta primera comparativa, se han importado todos los datos de los que se dispone de forma conjunta, se han separado en dos grupos:

- Los estadios 1 y 2.
- Los estadios 3 y 4.

Se han realizado dos pruebas estadísticas, la prueba T-Test y la prueba de Wilcoxon-Bonferroni.

El resultado del T-Test indica que hay diferencias estadísticamente significativas en los resultados obtenidos al comparar los estadios 1 y 2 con los estadios 3 y 4. Se ha comparado el valor de la señal procesada, el p-valor obtenido es muy pequeño ( $6.532e-06$ ) y sugiere que la diferencia entre las medias de los dos grupos es improbable que sea debido al azar. Por lo tanto, se puede concluir que hay diferencias significativas en los resultados obtenidos entre los dos grupos de estadios de melanoma.

Por otro lado, y en la misma línea de resultados, del resultado del test de Wilcoxon se obtiene un p-valor de  $2.2e-16$ , lo cual indica que es extremadamente pequeño. Un p-valor tan pequeño nuevamente sugiere que hay una diferencia estadísticamente significativa entre los dos grupos en términos de la ubicación (media) de los datos.

## 4.2. Análisis de posibles variables confusoras

### 4.2.1. Comparación por sexos

En el análisis de expresión génica diferencial por sexos, no se han observado diferencias significativas. Esto sugiere que, dentro del alcance de este estudio, la expresión génica de los pacientes con melanoma no se ve afectada por el género, por lo que, no es necesario realizar ajustes teniendo en cuenta esta variable.

```
##           H - M
## Down           0
## NotSig 36037
## Up             0
```

**Figura 12. Resultado Sexos**

A pesar de que la incidencia del melanoma pueda ser mayor en un sexo que en el otro no se han observado diferencias a nivel de expresión génica. Esta diferencia entre la prevalencia y los resultados obtenidos se puede deber a diversos factores como la regulación hormonal, respuestas inmunitarias o interacciones genéticas y el factor ambiental.

### 4.2.2. Comparación fenotípica

Para el análisis de expresión génica comparando rasgos fenotípicos se han realizado tres tipos de comparativas:

- Color de ojos: Se han separado los datos en color de ojos marrón, el más oscuro, contra azul y verde, los más claros.
- Color de pelo: Se han separado los datos en color de pelo oscuro, marrón y negro, contra rubio y pelirrojo, los más claros.
- Fototipos: Considerando la clasificación existente por fototipos donde el tipo 1 son personas con piel muy clara y pálida, el 2 corresponde a personas con piel clara, el 3 a personas con tono de piel medio, y el 4 con

piel morena. Se han separado los dos grupos conteniendo el fototipo 1 y 2, contra el 3 y 4.

Tras la realización de estas comparativas no se han observado diferencias significativas en la expresión génica.

```
##          Fot12 - Fot34
## Down                0
## NotSig            36037
## Up                 0
```

**Figura 13. Resultado Fototipo**

```
##          Brown - Rest
## Down                0
## NotSig            36037
## Up                 0
```

**Figura 14. Resultado color de Ojos**

```
##          Brown - Rest
## Down                0
## NotSig            36037
## Up                 0
```

**Figura 15. Resultado color Pelo**

Podrían existir diversas explicaciones a esta falta de diferencias:

- Complejidad de los fenotipos: Este tipo de rasgos, como el color de ojos o pelo, son rasgos multifactoriales, influenciados por una combinación en el tiempo de factores genéticos ambientales.
- Variabilidad genética y heterogeneidad: La expresión génica puede variar entre individuos debido a diferencias genéticas y a influencia de otros factores no genéticos.

- Tamaño de la muestra: Si el tamaño de la muestra es pequeño es posible que no se hayan detectado diferencias, especialmente, si estas diferencias son pequeñas entre los grupos.

En resumen, esta falta de diferencias significativas en la expresión génica sugiere que la contribución de los rasgos fenotípicos puede ser limitada o estar influenciada por otros factores genéticos o ambientales.

### 4.3. Comparación de la evolución

#### 4.3.1. Comparación por estadios

En este análisis se realiza la comparativa entre dos grupos formados por el estadio final del melanoma, agrupados por los estadios 1 y 4, ya que estos son los más extremos entre ellos. En este caso, se obtiene como resultado un gen significativamente disminuido; el gen UNCX.

| ##        | Est1 - Est4 |
|-----------|-------------|
| ## Down   | 1           |
| ## NotSig | 36036       |
| ## Up     | 0           |

**Figura 16. Resultado Estadios**

Este gen también conocido como Unc-homeobox desempeña un papel crucial en la regulación del desarrollo embrionario y la diferenciación celular, pero también es objeto de la investigación del cáncer. La reducción de la expresión de este gen puede contribuir a la desregulación de vías moleculares claves involucradas en la proliferación celular, la invasión tumoral y la resistencia a la terapia.

En la actualidad encontramos el estudio de Zhou et al. (16) en el que se investiga y se proponen diferentes genes como candidatos involucrados en el proceso de la metástasis, entre ellos encontramos el gen UNCX. Aun así, esta relación se

encuentra en etapas de investigación temprana y se requieren estudios adicionales para su validación y comprensión.

#### 4.3.2 Recaída

En este análisis se han separado los datos en función de si el paciente ha sufrido recaída del melanoma o no. El resultado obtenido indica que no hay diferencias detectables en los niveles de expresión génica. Igual que en los casos anteriores, la recaída puede estar asociada a eventos moleculares y cambios epigenéticos que no implican cambios en la expresión génica.

```
##          Rec0 - Rec1
## Down                0
## NotSig            36037
## Up                  0
```

**Figura 17. Resultado Recaída**

El resultado obtenido de este análisis en este contexto descartaría los cambios de expresión génica como un causante de la recaída.

#### 4.4. Comparación de variables asociadas a la evolución

##### 4.4.1. Comparación de subtipos

El siguiente análisis que se ha realizado ha sido mediante la separación de los datos según el subgrupo al que pertenecen. En este caso, debido a la distribución de las muestras se ha separado entre los subtipos de melanoma superficial y melanoma nodular.

```
##          Sup - Nod
## Down                0
## NotSig            36037
## Up                  0
```

**Figura 18. Resultado de Subtipos**

Tras la realización del pipeline de análisis no se obtienen diferencias significativas en la expresión génica. Esto puede ser debido a que estos subtipos de melanoma pueden compartir características genéticas y moleculares. Además, la heterogeneidad del melanoma en sí y su variabilidad entre pacientes puede influir en ello.

#### 4.4.2. Ulceración

La última de las comparativas realizadas separa los grupos en función de si el paciente ha sufrido ulceración o no. Primeramente, se realiza un breve análisis previo en el que se realiza una matriz de contingencia y se realiza la prueba de Chi cuadrado para observar la relación entre la ulceración y los estadios del melanoma:

#### Análisis previo: Chi cuadrado

Los datos se distribuyen de esta manera:

| Estadio   | Ulceración presente | Ulceración ausente |
|-----------|---------------------|--------------------|
| Estadio 1 | 0                   | 10                 |
| Estadio 2 | 3                   | 4                  |
| Estadio 3 | 9                   | 5                  |
| Estadio 4 | 10                  | 12                 |

**Figura 19. Matriz de contingencia**

```
##
## Pearson's Chi-squared test
##
## data:  tabla
## X-squared = 10.234, df = 3, p-value = 0.01668
```

**Figura 20. Resultado Chi Cuadrado**

De la realización de esta prueba se puede observar que el p-valor es de 0.01668, lo que indica que hay evidencias para rechazar la hipótesis nula, y que indica que existe una asociación significativa entre la presencia de ulceración y el estadio del melanoma.

| ##        | ulcAus - ulcPres |
|-----------|------------------|
| ## Down   | 605              |
| ## NotSig | 34278            |
| ## Up     | 1154             |

**Figura 21. Resultado Ulceración**

En este caso, sí que se obtienen diferencias significativas en la expresión génicas. Los resultados muestran que 605 genes se encuentran disminuidos en pacientes con ulceración y 1154 se encuentran aumentados en estos mismos pacientes.

Los genes que muestran mayores cambios en su expresión son:

- **ANGPTL5:** Este gen que se encuentra disminuido está involucrado en la regulación del metabolismo de lípidos y asociado a enfermedades metabólicas y a la respuesta inflamatoria.
- **LIFR:** Este gen también se encuentra disminuido en pacientes con ulceración y desempeña un papel clave en la regulación del crecimiento, la diferenciación celular y la respuesta inmunológica.
- **SLC22A3:** En este caso, el gen se encuentra aumentado en pacientes con ulceración. Este gen codifica el transportador de solutos orgánicos y su expresión se asocia a la absorción y eliminación de diversos compuestos, principalmente, fármacos y toxinas.

En resumen, la comparativa de estos dos grupos revela diferencias significativas en la expresión génica. Este hecho aporta sospechas sobre cómo afecta la presencia de ulceración en los mecanismos moleculares.

## 5. Conclusiones y trabajos futuros

Basado en los resultados obtenidos de las comparaciones realizadas entre los grupos seleccionados y cómo se ve afectada la expresión génica, se pueden concluir diversos aspectos.

En relación a las variables confusoras, la falta de diferencias significativas en la expresión génica relacionada con características fenotípicas de los pacientes contrasta con la idea inicial y algunos estudios recientes que sugieren posibles asociaciones entre estas características y los perfiles de expresión génica en el melanoma (17). El estudio de Mercieca et al. encontró que los melanomas en pacientes con fototipos I y II tenían más probabilidades de ser invasivos en comparación con pacientes con fototipos de piel III y IV ( $P = 0.00027$ ) (17).

En relación a la evolución del melanoma, el descubrimiento del gen UNCX disminuido en pacientes con estadio final 4 de melanoma es consistente con las investigaciones previas que destacan la relación entre la progresión del melanoma en estados avanzados y la alteración genética, como el realizado por Daniele et al. donde los resultados mostraron que la activación del gen UNCX está relacionada de manera significativa con un mayor nivel de metilación del ADN(18). Este gen que pertenece a la familia de genes homeobox, desempeña un papel importante en la regulación de la expresión génica durante el desarrollo de diferentes tejidos y órganos. En particular, se ha encontrado que está involucrado en la formación y diferenciación de los tejidos del sistema nervioso, incluyendo el cerebro y la médula espinal.

En relación a las variables asociadas a la evolución, en relación con los subtipos de melanoma analizados, nodular y superficial, no se han obtenido diferencias significativas en la expresión génica de estos pacientes. Se reconocen y observan diferencias clínicas y de comportamiento entre estos tipos de melanoma, estudios recientes resaltan que sus características moleculares tienen ciertas diferencias y patrones de mutación genética diferentes, como mutaciones distintas en el gen BRAF, o una mutación mayor de los genes PTEN



y TP53 en el melanoma nodular (19). Esto contrasta con el resultado obtenido en este estudio. Existen varias razones por las cuales no se han obtenido diferencias significativas en la expresión génica, una posible explicación puede ser el tamaño de la muestra realizada, que tal vez no sea el adecuado a la hora de detectar estas diferencias debido a la falta de poder estadístico. Otra posible razón puede ser la heterogeneidad intratumoral del melanoma, lo que puede contribuir a la variabilidad de los resultados y dificultar la detección de diferencias significativas.

Por último, de la comparación entre pacientes los cuales su melanoma ha sufrido ulceración o no, comparativa interesante ya que la ulceración es un factor pronóstico importante para los pacientes con melanoma y, curiosamente, también es un marcador predictivo para la respuesta a la terapia (20), la identificación de genes diferencialmente expresados, como el ANGPTL5, el LIFR o el SLC33A3, proporcionan información relevante sobre los procesos moleculares asociados a la aparición de úlcera. Estos resultados también están en consonancia con investigaciones recientes donde se sugiere un papel clave en la respuesta inflamatoria, la angiogénesis y la regulación del crecimiento celular en la progresión de úlceras en el melanoma (21).

En cuanto a las fortalezas de este estudio, cabe destacar el método seguido para todas las comparativas. Este método, que ha sido riguroso y replicable para todos los casos, genera consistencia en los datos y en los resultados obtenidos. La adopción de un enfoque metodológico sólido es crucial para garantizar la validez y la fiabilidad de los hallazgos.

Por otro lado, también es importante reconocer las limitaciones y debilidades que han podido influir en este trabajo. Una de las debilidades identificadas tiene que ver con la experiencia previa en este tipo de análisis, aunque se han realizado comparativas exhaustivas, es posible que existan otras vías de análisis que no se han podido o sabido explorar en profundidad. También, la falta de tiempo debido a las responsabilidades diarias de trabajo ha podido ser una limitación.

En resumen, estos resultados concuerdan en su mayoría con estudios realizados previamente, y refuerzan la necesidad de enfoques multidisciplinarios en la investigación del melanoma, integrando datos genéticos, clínicos, ambientales y personales para una comprensión mayor de la enfermedad. A partir de ello, se pueden proponer varias direcciones para trabajos futuros que podrían enfocarse en:

- Validación y replicación: repetición de análisis de índole similar con otros datos aportados e incluso con una muestra de datos mayor, con el objetivo de fortalecer la evidencia y confirmar la relevancia de los resultados.
- Investigación de mecanismos subyacentes: a través de la identificación de cambios de la expresión de algunos genes, sería interesante analizar los mecanismos subyacentes implicados en ellos.
- Evaluación de biomarcadores: igual que en el caso previo, a través de los genes identificados, se podrían llevar a cabo estudios clínicos para poder localizar herramientas de diagnóstico, de pronóstico e incluso de tratamiento.
- Estudios de terapia dirigida: exploración de fármacos existentes y su relación con los resultados obtenidos de este estudio.
- Investigación de factores de riesgo: a partir del estudio de los grupos y de sus características, se podrían realizar estudios adicionales para analizar estos factores que han aportado resultados.

## 6. Glosario

- miRNA: *microRNA*, moléculas de ARN pequeñas y no codificantes que desempeñan un papel crucial en la regulación de la expresión génica. Estas moléculas se unen a secuencias específicas de ARN mensajero, lo que puede resultar en una disminución de la proteína codificada por este ARN mensajero.

- BRAF: gen que codifica la proteína homónima y que está involucrada en la proliferación y la supervivencia celular. Las mutaciones de este gen son comunes en diversos tipos de cáncer.

- NRAS: gen que codifica la proteína homónima y que actúa como transmisora de señales para el crecimiento y la división celular en respuesta a estímulos externos. También, la mutación de este gen es común en diferentes tipos de cáncer.

- MAPK: *mitogen-activated protein kinases*, proteínas quinasas activadas por mitógenos. Desempeñan un papel clave en la transmisión de señales intracelulares y regulan una amplia variedad de procesos celulares. Las mutaciones que se dan en los genes BRAF y NRAS, componentes clave de esta vía de señalización, son frecuentes en el cáncer.

- Wnt/ $\beta$ -catenina: vía de señalización celular, es una vía de transducción que desempeña un papel crucial en el desarrollo embrionario, la homeóstasis de los tejidos y la regulación de la expresión génica.

- TCGA: *The Cancer Genome Atlas*, atlas genómico del cáncer. Proyecto de investigación colaborativo que tiene como objetivo caracterizar molecularmente los diferentes tipos de cáncer.

- CDKN2A: gen supresor de tumores. Este gen produce la proteína p16INK4a que inhibe la actividad de las ciclinas dependientes de quinasas (CDKs). La

alteración de este gen puede alterar la producción de la proteína y así aumentar el riesgo de que se de crecimiento celular y la aparición del cáncer.

- TP53: gen supresor de tumores. Este gen produce la proteína p53 que tiene como función principal la prevención de la proliferación celular. Cuando se encuentran daños en el ADN, esta proteína se activa y desencadena respuestas celulares protectoras.

- PTEN: *phosphatase and tensin homolog*. Proteína cuya principal función es inhibir la vía de señalización PI3K/AKT/mTOR, involucrada en el crecimiento celular y la supervivencia.

- TERT: *telomerase reverse transcriptase*. Estructura que juega un papel crucial en el mantenimiento y extensión de los telómeros, estructuras protectoras del cromosoma.

- CDR1AS: *cerebellar degeneration-related protein 1 antisense RNA*. ARN no codificante que se encuentra en el cerebelo y que se asocia con diversas enfermedades neurológicas y trastornos del sistema nervioso central. Desempeña un papel importante en la regulación génica y la función neuronal.

- CyCIF: *cyclic immunofluorescence*. Técnica de imagen multiplexada que combina la inmunofluorescencia y la microscopía de alta resolución para el análisis de muestras biológicas.

## 7. Bibliografía

1. Dzwierzynski WW. Melanoma Risk Factors and Prevention. *Clin Plast Surg.* octubre de 2021;48(4):543-50.
2. INEbase. CONSUL [Internet]. INE. [citado 20 de marzo de 2023]. Disponible en: <https://www.ine.es/consul/serie.do?d=true&s=ECM453205>
3. Puig Sarda S. Implementation of personalized medicine in cutaneous melanoma based on genomic susceptibility and tumor signatures: Identification of new targets for melanoma treatment. 2013.
4. Potrony M, Badenas C, Aguilera P, Puig-Butille JA, Carrera C, Malveyh J, Puig S. Update in genetic susceptibility in melanoma - PubMed [Internet]. [citado 23 de abril de 2023]. Disponible en: <https://pubmed.ncbi.nlm.nih.gov/26488006/>
5. Podlipnik S, Potrony M, Puig S. Genetic markers for characterization and prediction of prognosis of melanoma subtypes: a 2021 update. *Ital J Dermatol Venereol* [Internet]. julio de 2021 [citado 10 de abril de 2023];156(3). Disponible en: <https://www.minervamedica.it/index2.php?show=R23Y2021N03A0322>
6. Abd-Allah GM, Ismail A, El-Mahdy HA, Elsakka EGE, El-Husseiny AA, Abdelmaksoud NM, et al. miRNAs as potential game-changers in melanoma: A comprehensive review. *Pathol - Res Pract.* 1 de abril de 2023;244:154424.
7. Lin Z, Lin X, Sun Y, Lei S, Cai G, Li Z. Melanoma molecular subtyping and scoring model construction based on ligand-receptor pairs. *Front Genet.* 26 de enero de 2023;14:1098202.
8. Chandra S, Choi JS, Sosman JA. Melanoma: Does Sequencing Really Matter? *J Clin Oncol.* 10 de enero de 2023;41(2):167-9.
9. Adams R, Coumbe JEM, Coumbe BGT, Thomas J, Willsmore Z, Dimitrievska M, et al. BRAF inhibitors and their immunological effects in malignant melanoma. *Expert Rev Clin Immunol.* 3 de abril de 2022;18(4):347-62.
10. Jin SG, Padron F, Pfeifer GP. UVA Radiation, DNA Damage, and Melanoma. *ACS Omega.* 20 de septiembre de 2022;7(37):32936-48.
11. Jung-Garcia Y, Maiques O, Monger J, Rodriguez-Hernandez I, Fanshawe B, Domart MC, et al. LAP1 supports nuclear adaptability during constrained melanoma cell migration and invasion. *Nat Cell Biol.* enero de 2023;25(1):108-19.
12. Hanniford D, Ulloa-Morales A, Karz A, Berzoti-Coelho MG, Moubarak RS, Sánchez-Sendra B, et al. Epigenetic Silencing of CDR1as Drives IGF2BP3-Mediated Melanoma Invasion and Metastasis. *Cancer Cell.* 13 de enero de 2020;37(1):55-70.e15.
13. Nirmal AJ, Maliga Z, Vallius T, Quattrochi B, Chen AA, Jacobson CA, et al. The Spatial Landscape of Progression and Immunoediting in Primary Melanoma at Single-Cell Resolution. *Cancer Discov.* 2 de junio de 2022;12(6):1518-41.
14. Poniewierska-Baran A, Zadroga Ł, Danilyan E, Małkowska P, Niedźwiedzka-Rystwej P, Pawlik A. MicroRNA as a Diagnostic Tool, Therapeutic Target and Potential Biomarker in Cutaneous Malignant Melanoma Detection—Narrative Review. *Int J Mol Sci.* enero de 2023;24(6):5386.
15. Moderna and Merck. Moderna and Merck Announce mRNA 4157V940 an Investigational Personalized mRNA Cancer Vaccine [Internet]. 2023. Disponible en: <https://www.merck.com/news/moderna-and-merck-announce-mrna-4157-v940-an-investigational-personalized-mrna-cancer-vaccine-in-combination-with-keytruda-pembrolizumab-met-primary-efficacy-endpoint-in-phase-2b-keynote-94/>

16. Zhou Y, Zang Y, Yang Y, Xiang J, Chen Z. Candidate genes involved in metastasis of colon cancer identified by integrated analysis. *Cancer Med.* 18 de marzo de 2019;8(5):2338-47.
17. Mercieca L, Aquilina S, Calleja N, Boffa MJ. Cutaneous Melanoma More Likely to Be Invasive in Fairer Skin Phototypes: A Retrospective Observational Study. *Skinmed.* 2021;19(4):280-3.
18. Daniele G, Simonetti G, Fusilli C, Iacobucci I, Lonoce A, Palazzo A, et al. Epigenetically induced ectopic expression of UNCX impairs the proliferation and differentiation of myeloid cells. *Haematologica.* julio de 2017;102(7):1204-14.
19. Changes in the Presentation of Nodular and Superficial Spreading Melanomas Over 35 Years - PMC [Internet]. [citado 17 de junio de 2023]. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3624077/>
20. Bønnelykke-Behrndtz ML, Steiniche T. Ulcerated Melanoma: Aspects and Prognostic Impact. En: Ward WH, Farma JM, editores. *Cutaneous Melanoma: Etiology and Therapy* [Internet]. Brisbane (AU): Codon Publications; 2017 [citado 17 de junio de 2023]. Disponible en: <http://www.ncbi.nlm.nih.gov/books/NBK481861/>
21. Barricklow Z, DiVincenzo MJ, Angell CD, Carson WE. Ulcerated Cutaneous Melanoma: A Review of the Clinical, Histologic, and Molecular Features Associated with a Clinically Aggressive Histologic Phenotype. *Clin Cosmet Investig Dermatol.* 30 de agosto de 2022;15:1743-57.

## 8. Anexos

Se adjunta como anexo uno de los análisis completos realizados en RStudio, en este caso el que compara los estadios 1 y 4 de melanoma. El resto se pueden encontrar en GitHub, en el enlace:

[https://github.com/egomezguil/eguil\\_fm](https://github.com/egomezguil/eguil_fm)

### Análisis estadios con datos recodificados

Estefanía Gómez Guil

2023-06-02

#### Table of Contents

|                                   |    |
|-----------------------------------|----|
| 1.Importación de datos.....       | 40 |
| 2.Quality Control.....            | 41 |
| 3. Preprocesado.....              | 43 |
| 3.1 Background.....               | 43 |
| 3.2 Normalización.....            | 43 |
| 3.3 Anotación.....                | 45 |
| 3.3 Filtrado.....                 | 46 |
| 4. Expresión Diferencial.....     | 47 |
| 4.1 Modelización i Regresión..... | 47 |
| Resultados.....                   | 48 |
| 5. Gene Enrichment Analysis.....  | 52 |
| Gene Ontology.....                | 52 |
| KEGG.....                         | 53 |

#### 1.Importación de datos

```
params <- list(
  estadio1="Est1",
  estadio4="Est4",
  directorio= "/Users/estefaniagomezguil/Desktop/Estadios Recod"
)
ruta_est1<- file.path(params$directorio, "1")
ruta_est4<- file.path(params$directorio, "4")

files_est1 <- list.files(path = ruta_est1, pattern = ".txt", full.names = TRUE)
files_est4 <- list.files(path = ruta_est4, pattern = ".txt", full.names = TRUE)
Grupo <- c(rep(params$estadio1, length(files_est1)), rep(params$estadio4, length(files_est4)))
```

```

files <- c(files_est1, files_est4)
print(files)

## [1] "/Users/estefaniagomezguil/Desktop/Estadios Recod/1/2701.txt"
## [2] "/Users/estefaniagomezguil/Desktop/Estadios Recod/1/2858.txt"
## [3] "/Users/estefaniagomezguil/Desktop/Estadios Recod/1/2894.txt"
## [4] "/Users/estefaniagomezguil/Desktop/Estadios Recod/1/2907.txt"
## [5] "/Users/estefaniagomezguil/Desktop/Estadios Recod/1/2912.txt"
## [6] "/Users/estefaniagomezguil/Desktop/Estadios Recod/4/2708.txt"
## [7] "/Users/estefaniagomezguil/Desktop/Estadios Recod/4/2730.txt"
## [8] "/Users/estefaniagomezguil/Desktop/Estadios Recod/4/2736.txt"
## [9] "/Users/estefaniagomezguil/Desktop/Estadios Recod/4/2801.txt"
## [10] "/Users/estefaniagomezguil/Desktop/Estadios Recod/4/2823.txt"
## [11] "/Users/estefaniagomezguil/Desktop/Estadios Recod/4/2824.txt"
## [12] "/Users/estefaniagomezguil/Desktop/Estadios Recod/4/2891.txt"
## [13] "/Users/estefaniagomezguil/Desktop/Estadios Recod/4/2893.txt"
## [14] "/Users/estefaniagomezguil/Desktop/Estadios Recod/4/2911.txt"
## [15] "/Users/estefaniagomezguil/Desktop/Estadios Recod/4/2934.txt"
## [16] "/Users/estefaniagomezguil/Desktop/Estadios Recod/4/2948.txt"

targets = read.maimages(file=files, source="agilent", green.only=TRUE)

## Read /Users/estefaniagomezguil/Desktop/Estadios Recod/1/2701.txt
## Read /Users/estefaniagomezguil/Desktop/Estadios Recod/1/2858.txt
## Read /Users/estefaniagomezguil/Desktop/Estadios Recod/1/2894.txt
## Read /Users/estefaniagomezguil/Desktop/Estadios Recod/1/2907.txt
## Read /Users/estefaniagomezguil/Desktop/Estadios Recod/1/2912.txt
## Read /Users/estefaniagomezguil/Desktop/Estadios Recod/4/2708.txt
## Read /Users/estefaniagomezguil/Desktop/Estadios Recod/4/2730.txt
## Read /Users/estefaniagomezguil/Desktop/Estadios Recod/4/2736.txt
## Read /Users/estefaniagomezguil/Desktop/Estadios Recod/4/2801.txt
## Read /Users/estefaniagomezguil/Desktop/Estadios Recod/4/2823.txt
## Read /Users/estefaniagomezguil/Desktop/Estadios Recod/4/2824.txt
## Read /Users/estefaniagomezguil/Desktop/Estadios Recod/4/2891.txt
## Read /Users/estefaniagomezguil/Desktop/Estadios Recod/4/2893.txt
## Read /Users/estefaniagomezguil/Desktop/Estadios Recod/4/2911.txt
## Read /Users/estefaniagomezguil/Desktop/Estadios Recod/4/2934.txt
## Read /Users/estefaniagomezguil/Desktop/Estadios Recod/4/2948.txt

targets$targets$class <- Grupo
files2<- str_remove_all(string=files, pattern= "US85003608_25394944221
5")
files2<- str_remove_all(string=files2, patter= "_S01_GE1_107_Sep09")

```

## 2. Quality Control

```

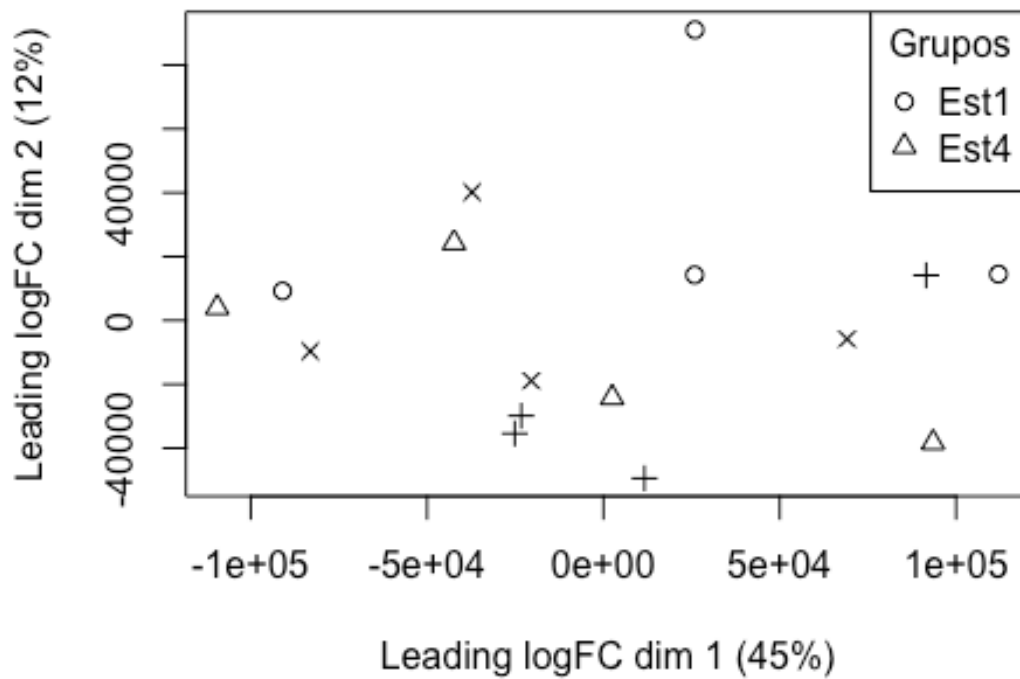
limma::plotMA3by2(targets, status = targets$genes$ControlType, prefix
= "MA", device = "pdf")
# Obtener Los grupos únicos y sus correspondientes colores
grupos_unicos <- unique(Grupo)
colores <- 1:4
Grupo2<-colores
plotMDS(targets, top = 100, gene.selection = "pairwise", pch=colores,
type="p", plot = TRUE)

# Agregar La Leyenda

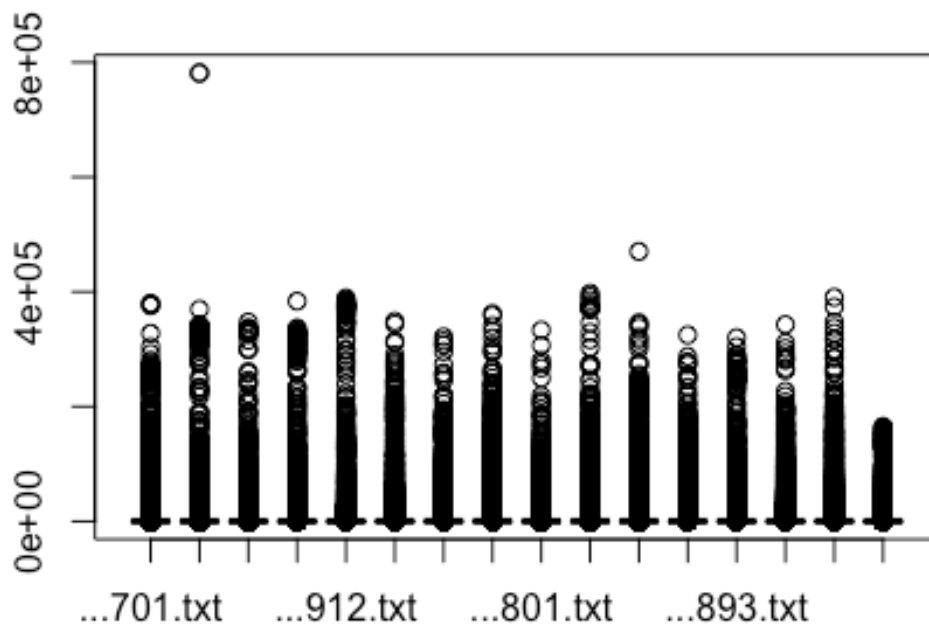
```



```
legend("topright", legend = grupos_unicos, pch = colores, title = "Grupos")
legend("topright", legend = grupos_unicos, pch = colores, title = "Grupos")
```



```
boxplot(targets$E, col = "royalblue", names = str_trunc(targets$target  
s$FileName, width = 10, side = "left"))
```



### 3. Preprocesado

#### 3.1 Background

```
BK <- backgroundCorrect(targets, method = "normexp", printer=targets$printer, verbose=TRUE)
```

```
## Array 1 corrected
## Array 2 corrected
## Array 3 corrected
## Array 4 corrected
## Array 5 corrected
## Array 6 corrected
## Array 7 corrected
## Array 8 corrected
## Array 9 corrected
## Array 10 corrected
## Array 11 corrected
## Array 12 corrected
## Array 13 corrected
## Array 14 corrected
## Array 15 corrected
## Array 16 corrected
```

#### 3.2 Normalización

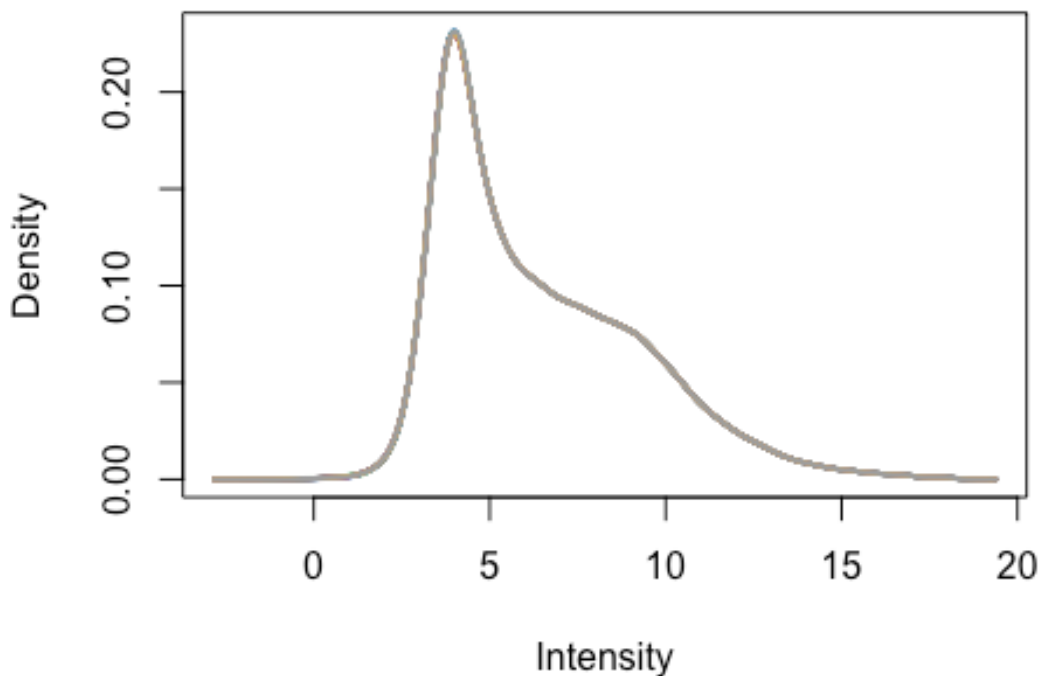
```
MA.p <- normalizeBetweenArrays(BK, method = 'quantile')
transposed <- t(MA.p$E)
```



```
## Warning: In density.default(E[, a], n = npoint, na.rm = TRUE, ...)
:
## extra argument 'type' will be disregarded

## Warning: In density.default(E[, a], n = npoint, na.rm = TRUE, ...)
:
## extra argument 'type' will be disregarded

## Warning: In density.default(E[, a], n = npoint, na.rm = TRUE, ...)
:
## extra argument 'type' will be disregarded
```



### 3.3 Anotación

```
MA.p$genes$EntrezID<- mapIds(HsAgilentDesign026652.db, MA.p$genes$ProbeName, keytype = "PROBEID", column = "ENTREZID")

## 'select()' returned 1:1 mapping between keys and columns

MA.p$genes$Symbol <- mapIds(HsAgilentDesign026652.db, MA.p$genes$ProbeName, keytype = "PROBEID", column = "SYMBOL")

## 'select()' returned 1:1 mapping between keys and columns

map=getGEO("GPL20844")
```

### 3.3 Filtrado

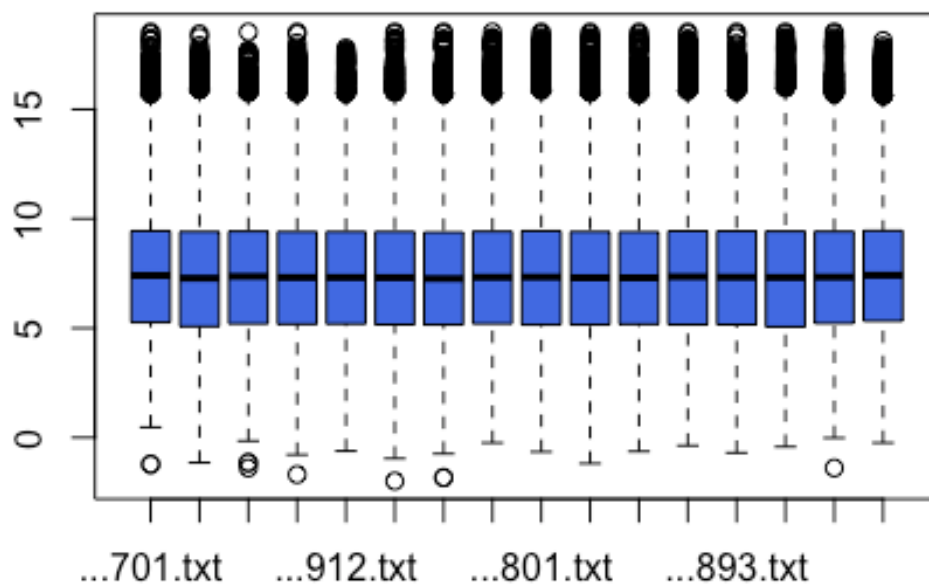
```
control_1 =MA.p$genes$ControlType == 1
control_b= MA.p$genes$ControlType == -1

NoSym = is.na(MA.p$genes$Symbol)
NoEnID= is.na(MA.p$genes$EntrezID)
table(NoEnID)

## NoEnID
## FALSE TRUE
## 36037 26939

yfilt = MA.p[!control_1& !control_b & !NoSym & !NoEnID,]

boxplot(yfilt$E, col = "royalblue", names = str_trunc(targets$targets$
FileName, width = 10, side = "left"))
```

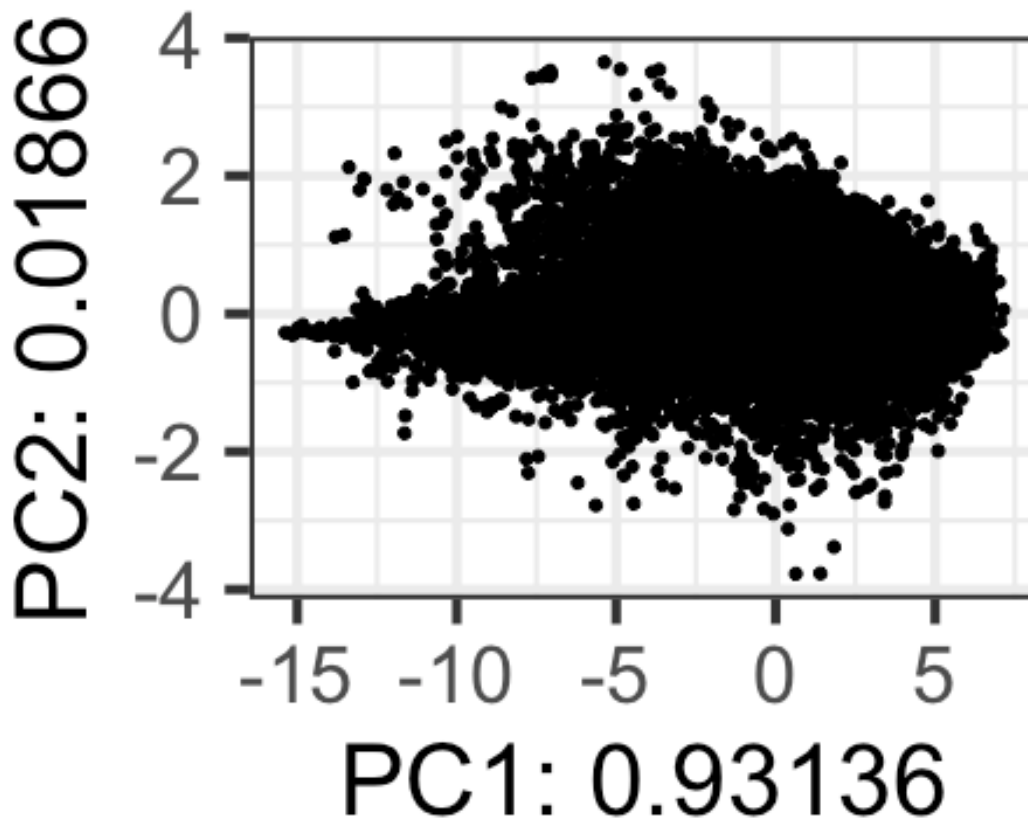


```
pcomp=prcomp(yfilt$E, scale. = TRUE)
eso=summary(pcomp)
eso

## Importance of components:
##              PC1      PC2      PC3      PC4      PC5      P
C6      PC7
## Standard deviation      3.8603 0.54638 0.37118 0.32929 0.29281 0.283
41 0.25835
## Proportion of Variance 0.9314 0.01866 0.00861 0.00678 0.00536 0.005
02 0.00417
```

```
## Cumulative Proportion 0.9314 0.95002 0.95863 0.96541 0.97076 0.975
79 0.97996
##          PC8      PC9      PC10      PC11      PC12      P
C13  PC14
## Standard deviation 0.25461 0.22436 0.20594 0.19374 0.17947 0.16
337 0.1598
## Proportion of Variance 0.00405 0.00315 0.00265 0.00235 0.00201 0.00
167 0.0016
## Cumulative Proportion 0.98401 0.98715 0.98981 0.99215 0.99416 0.99
583 0.9974
##          PC15      PC16
## Standard deviation 0.14591 0.14095
## Proportion of Variance 0.00133 0.00124
## Cumulative Proportion 0.99876 1.00000
```

```
com=as.data.frame(pcomp$x)
ggplot(com, aes(PC1, PC2)) + geom_point()+theme_bw(base_size=32)+xlab(
paste("PC1:",eso$importance[2,1])) + ylab(paste("PC2:",eso$importance[
2,2]))+theme(legend.position="top")
```



#### 4. Expresión Diferencial

##### 4.1 Modelización i Regresión

```
design <- model.matrix(~0+Grupo)
colnames(design) = c(params$estadio1, params$estadio4)
design
```

```

##      Est1 Est4
## 1      1    0
## 2      1    0
## 3      1    0
## 4      1    0
## 5      1    0
## 6      0    1
## 7      0    1
## 8      0    1
## 9      0    1
## 10     0    1
## 11     0    1
## 12     0    1
## 13     0    1
## 14     0    1
## 15     0    1
## 16     0    1
## attr("assign")
## [1] 1 1
## attr("contrasts")
## attr("contrasts")$Grupo
## [1] "contr.treatment"

fit = lmFit(yfilt, design)

parametre1 = params$estadio1
parametre2 = params$estadio4

contraste = makeContrasts(Est1-Est4, levels = design)
fit2 = contrasts.fit(fit, contraste)
fit2 = eBayes(fit2)

```

## Resultados

```
topTable(fit2, adjust.method = "bonferroni", n=30)
```

| ## | Row      | Col | ControlType | ProbeName       | SystematicName | EntrezID |
|----|----------|-----|-------------|-----------------|----------------|----------|
| ## | 14027    | 86  | 87          | 0 A_33_P3411388 | NM_001080461   | 340260   |
|    | UNCX     |     |             |                 |                |          |
| ## | 46463    | 284 | 51          | 0 A_33_P3358457 | NM_005270      | 2736     |
|    | GLI2     |     |             |                 |                |          |
| ## | 58569    | 358 | 21          | 0 A_23_P123704  | NR_024032      | 58483 L  |
|    | INC00474 |     |             |                 |                |          |
| ## | 45947    | 281 | 27          | 0 A_33_P3292064 | BC062656       | 254240   |
|    | BPIFC    |     |             |                 |                |          |
| ## | 48397    | 296 | 17          | 0 A_33_P3381478 | NM_001105659   | 127255   |
|    | LRRIQ3   |     |             |                 |                |          |
| ## | 35434    | 217 | 10          | 0 A_33_P3348352 | NR_015407      | 339535 L |
|    | INC01139 |     |             |                 |                |          |
| ## | 57904    | 354 | 12          | 0 A_23_P112296  | NM_000787      | 1621     |
|    | DBH      |     |             |                 |                |          |
| ## | 36631    | 224 | 59          | 0 A_33_P3296427 | NR_015407      | 339535 L |
|    | INC01139 |     |             |                 |                |          |
| ## | 46435    | 284 | 23          | 0 A_23_P1102    | NM_001100      | 58       |

|          |            |           |           |               |                 |           |   |
|----------|------------|-----------|-----------|---------------|-----------------|-----------|---|
| ACTA1    |            |           |           |               |                 |           |   |
| ## 6403  | 40         | 7         | 0         | A_23_P120428  | NM_014477       | 27296     |   |
| TP53TG5  |            |           |           |               |                 |           |   |
| ## 13838 | 85         | 62        | 0         | A_33_P3388183 | AK126237        | 4057      |   |
| LTF      |            |           |           |               |                 |           |   |
| ## 13906 | 85         | 130       | 0         | A_33_P3382867 | NR_027420       | 389834    | L |
| OC389834 |            |           |           |               |                 |           |   |
| ## 42053 | 257        | 69        | 0         | A_33_P3356990 | NM_130464       | 23117     |   |
| NPIP3    |            |           |           |               |                 |           |   |
| ## 7120  | 44         | 68        | 0         | A_33_P3546363 | BQ950045        | 400128    |   |
| TUSC8    |            |           |           |               |                 |           |   |
| ## 50931 | 311        | 91        | 0         | A_23_P1102    | NM_001100       | 58        |   |
| ACTA1    |            |           |           |               |                 |           |   |
| ## 51722 | 316        | 62        | 0         | A_23_P1102    | NM_001100       | 58        |   |
| ACTA1    |            |           |           |               |                 |           |   |
| ## 49376 | 302        | 12        | 0         | A_24_P366652  | NM_005353       | 3681      |   |
| ITGAD    |            |           |           |               |                 |           |   |
| ## 47072 | 288        | 4         | 0         | A_33_P3397545 | NM_001033017    | 613209    |   |
| DEFB135  |            |           |           |               |                 |           |   |
| ## 13611 | 83         | 163       | 0         | A_23_P1102    | NM_001100       | 58        |   |
| ACTA1    |            |           |           |               |                 |           |   |
| ## 12032 | 74         | 60        | 0         | A_23_P1102    | NM_001100       | 58        |   |
| ACTA1    |            |           |           |               |                 |           |   |
| ## 31774 | 194        | 122       | 0         | A_23_P1102    | NM_001100       | 58        |   |
| ACTA1    |            |           |           |               |                 |           |   |
| ## 20025 | 123        | 17        | 0         | A_23_P1102    | NM_001100       | 58        |   |
| ACTA1    |            |           |           |               |                 |           |   |
| ## 36707 | 224        | 135       | 0         | A_23_P1102    | NM_001100       | 58        |   |
| ACTA1    |            |           |           |               |                 |           |   |
| ## 43645 | 267        | 21        | 0         | A_23_P66593   | NM_002055       | 2670      |   |
| GFAP     |            |           |           |               |                 |           |   |
| ## 62794 | 383        | 146       | 0         | A_23_P35782   | NM_024771       | 79829     |   |
| NAA40    |            |           |           |               |                 |           |   |
| ## 11591 | 71         | 111       | 0         | A_33_P3267185 | NM_001001323    | 490       |   |
| ATP2B1   |            |           |           |               |                 |           |   |
| ## 21404 | 131        | 84        | 0         | A_23_P1102    | NM_001100       | 58        |   |
| ACTA1    |            |           |           |               |                 |           |   |
| ## 26345 | 161        | 105       | 0         | A_23_P1102    | NM_001100       | 58        |   |
| ACTA1    |            |           |           |               |                 |           |   |
| ## 60401 | 369        | 49        | 0         | A_33_P3275226 | ENST00000414750 | 101927156 | L |
| INC01934 |            |           |           |               |                 |           |   |
| ## 60921 | 372        | 77        | 0         | A_33_P3217495 | NM_003041       | 6524      |   |
| SLC5A2   |            |           |           |               |                 |           |   |
| ##       | logFC      | AveExpr   | t         | P.Value       | adj.P.Val       |           |   |
| B        |            |           |           |               |                 |           |   |
| ## 14027 | -2.2720234 | 13.209101 | -8.443014 | 1.413578e-07  | 0.005094109     | 4.486     |   |
| 8391     |            |           |           |               |                 |           |   |
| ## 46463 | -1.2069692 | 5.267545  | -6.484043 | 4.909597e-06  | 0.176927144     | 2.553     |   |
| 0295     |            |           |           |               |                 |           |   |
| ## 58569 | -1.2854009 | 3.643822  | -6.073209 | 1.103130e-05  | 0.397534883     | 2.063     |   |
| 5939     |            |           |           |               |                 |           |   |
| ## 45947 | -1.3244308 | 3.307873  | -5.629304 | 2.712331e-05  | 0.977442620     | 1.500     |   |
| 0412     |            |           |           |               |                 |           |   |
| ## 48397 | -1.8589308 | 3.580196  | -5.609381 | 2.825746e-05  | 1.000000000     | 1.473     |   |



```

9049
## 35434 -1.3502628 3.902014 -5.465142 3.806826e-05 1.000000000 1.282
5423
## 57904 -0.9751315 4.096276 -5.329178 5.053085e-05 1.000000000 1.098
7463
## 36631 -0.9799991 3.770903 -5.269513 5.725650e-05 1.000000000 1.017
0574
## 46435 3.2923483 7.156777 5.243505 6.046939e-05 1.000000000 0.981
2534
## 6403 -1.0694685 4.305319 -5.236029 6.142664e-05 1.000000000 0.970
9407
## 13838 0.9245064 4.514476 5.222460 6.320404e-05 1.000000000 0.952
1970
## 13906 -1.5415469 3.930114 -5.149474 7.371099e-05 1.000000000 0.850
8280
## 42053 -1.2769467 4.106202 -5.137577 7.558622e-05 1.000000000 0.834
2178
## 7120 -1.7213817 12.313771 -5.108727 8.033933e-05 1.000000000 0.793
8369
## 50931 3.1631896 7.386266 4.966034 1.087638e-04 1.000000000 0.592
0462
## 51722 3.1561055 7.382460 4.944645 1.138366e-04 1.000000000 0.561
5061
## 49376 -1.0980037 5.163349 -4.926253 1.183916e-04 1.000000000 0.535
1863
## 47072 -1.2091881 3.490359 -4.918647 1.203293e-04 1.000000000 0.524
2857
## 13611 3.0010779 7.160533 4.912428 1.219379e-04 1.000000000 0.515
3645
## 12032 3.0831708 7.337670 4.908267 1.230263e-04 1.000000000 0.509
3934
## 31774 2.9856592 7.348353 4.894459 1.267100e-04 1.000000000 0.489
5555
## 20025 3.0869178 7.125093 4.847597 1.400726e-04 1.000000000 0.422
0021
## 36707 3.1956223 7.221084 4.813673 1.506362e-04 1.000000000 0.372
8815
## 43645 -1.0144588 3.425805 -4.800023 1.551129e-04 1.000000000 0.353
0649
## 62794 -0.8808876 6.220432 -4.743723 1.750590e-04 1.000000000 0.271
0219
## 11591 0.7490237 5.726711 4.704939 1.903034e-04 1.000000000 0.214
2186
## 21404 2.9290395 7.639089 4.675732 2.026709e-04 1.000000000 0.171
2908
## 26345 3.0460645 7.585673 4.652421 2.131267e-04 1.000000000 0.136
9374
## 60401 -1.2492521 3.980746 -4.644361 2.168685e-04 1.000000000 0.125
0400
## 60921 -0.8253286 6.267019 -4.631065 2.231878e-04 1.000000000 0.105
3922

```

```
tT=topTable(fit2, adjust.method = "bonferroni", n=3000)
```

```

results = decideTests(fit2, p.value = 0.1)
summary(decideTests(fit2), p.value = 0.1)

##           Est1 - Est4
## Down                1
## NotSig             36036
## Up                  0

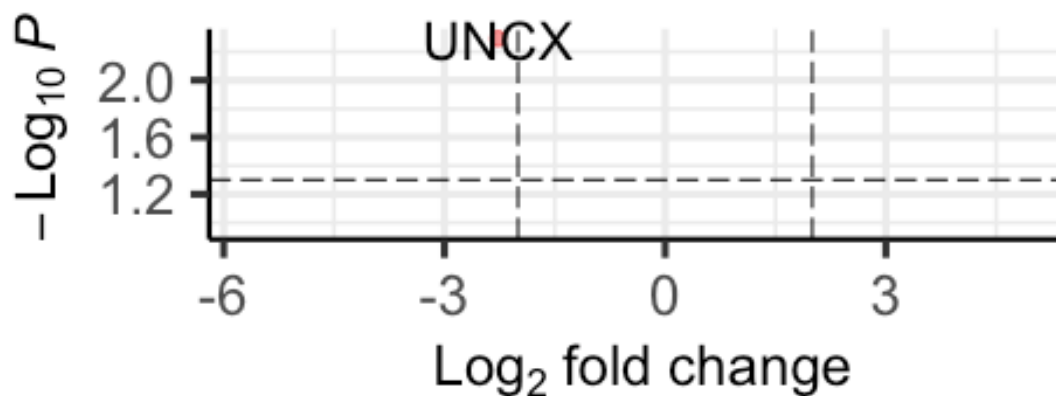
EnhancedVolcano(tT, lab = tT$Symbol, x = "logFC", y = "adj.P.Val", pCutoff = 0.05, FCcutoff=2, labSize = 6.0, ylim=0.95)

```

## Volcano plot

*EnhancedVolcano*

● NS ●  $\text{Log}_2$  FC ● p – value and  $\text{log}_2$  FC

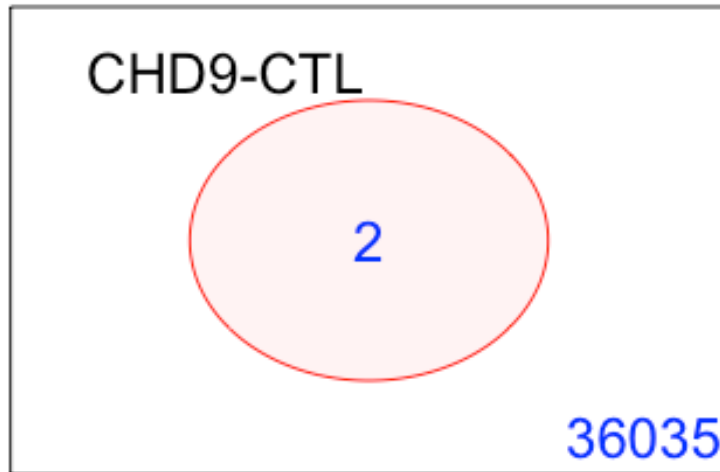


total = 3000 variables

```

vennDiagram(results, include=c("both"), circle.col = "red", counts.col = "blue", names = c("CHD9-CTL"))

```



## 5. Gene Enrichment Analysis

### Gene Ontology

```
g = goana(fit2, species="Hs", geneid = fit2$genes$EntrezID, FDR = 0.05)
topGO(g, n=20)

##
## Term
## GO:0035726 common myeloid progenitor cell
## GO:0021889 proliferation
## GO:0021889 olfactory bulb interneuron differentiation
## GO:0001502 cartilage condensation
## GO:0021516 dorsal spinal cord development
## GO:0098743 cell aggregation
## GO:0021772 olfactory bulb development
## GO:0021988 olfactory lobe development
## GO:0021510 spinal cord development
## GO:0051216 cartilage development
```

```

## GO:0048705 skeletal system
morphogenesis
## GO:0021537 telencephalo
n development
## GO:0061448 connective tissu
e development
## GO:0030900 forebrai
n development
## GO:0007389 pattern specific
ation process
## GO:0001501 skeletal syste
m development
## GO:0007420 brai
n development
## GO:0060322 hea
d development
## GO:0009887 animal organ
morphogenesis
## GO:0007417 central nervous syste
m development
## GO:0000981 DNA-binding transcription factor activity, RNA polymeras
e II-specific
##
## Ont N Up Down P.Up P.Down
## GO:0035726 BP 8 0 1 1 0.0003843198
## GO:0021889 BP 11 0 1 1 0.0005284397
## GO:0001502 BP 20 0 1 1 0.0009607994
## GO:0021516 BP 20 0 1 1 0.0009607994
## GO:0098743 BP 24 0 1 1 0.0011529593
## GO:0021772 BP 33 0 1 1 0.0015853190
## GO:0021988 BP 35 0 1 1 0.0016813989
## GO:0021510 BP 102 0 1 1 0.0049000769
## GO:0051216 BP 199 0 1 1 0.0095599539
## GO:0048705 BP 226 0 1 1 0.0108570331
## GO:0021537 BP 266 0 1 1 0.0127786318
## GO:0061448 BP 270 0 1 1 0.0129707917
## GO:0030900 BP 393 0 1 1 0.0188797079
## GO:0007389 BP 464 0 1 1 0.0222905457
## GO:0001501 BP 531 0 1 1 0.0255092237
## GO:0007420 BP 770 0 1 1 0.0369907763
## GO:0060322 BP 818 0 1 1 0.0392966949
## GO:0009887 BP 1021 0 1 1 0.0490488086
## GO:0007417 BP 1056 0 1 1 0.0507302075
## GO:0000981 MF 1136 0 1 1 0.0545734051

```

## KEGG

```

k = kegg(k, species="Hs", geneid = fit2$genes$EntrezID, FDR = 0.05
)
topKEGG(k, number = 20)

```

```

## Pathway
N Up Down
## hsa00780 Biotin metabolism
3 0 0
## hsa00290 Valine, leucine and isoleucine biosynthesis

```

|    |          |      |  |  |
|----|----------|------|--|--|
| 4  | 0        | 0    |  |  |
| ## | hsa03267 |      | Virion - Adenovirus  |  |
| 4  | 0        | 0    |  |  |
| ## | hsa03264 |      | Virion - Flavivirus  |  |
| 4  | 0        | 0    |  |  |
| ## | hsa03265 |      | Virion - Lyssavirus  |  |
| 4  | 0        | 0    |  |  |
| ## | hsa00232 |      | Caffeine metabolism  |  |
| 5  | 0        | 0    |  |  |
| ## | hsa00524 |      | Neomycin, kanamycin and gentamicin biosynthesis            |  |
| 5  | 0        | 0    |  |  |
| ## | hsa03260 |      | Virion - Human immunodeficiency virus                      |  |
| 5  | 0        | 0    |  |  |
| ## | hsa00470 |      | D-Amino acid metabolism                                    |  |
| 6  | 0        | 0    |  |  |
| ## | hsa00400 |      | Phenylalanine, tyrosine and tryptophan biosynthesis        |  |
| 6  | 0        | 0    |  |  |
| ## | hsa00440 |      | Phosphonate and phosphinate metabolism                     |  |
| 6  | 0        | 0    |  |  |
| ## | hsa00750 |      | Vitamin B6 metabolism                                      |  |
| 6  | 0        | 0    |  |  |
| ## | hsa00740 |      | Riboflavin metabolism                                      |  |
| 8  | 0        | 0    |  |  |
| ## | hsa04122 |      | Sulfur relay system  |  |
| 8  | 0        | 0    |  |  |
| ## | hsa03266 |      | Virion - Herpesvirus                                       |  |
| 9  | 0        | 0    |  |  |
| ## | hsa00920 |      | Sulfur metabolism  |  |
| 10 | 0        | 0    |  |  |
| ## | hsa00130 |      | Ubiquinone and other terpenoid-quinone biosynthesis        |  |
| 11 | 0        | 0    |  |  |
| ## | hsa03450 |      | Non-homologous end-joining                                 |  |
| 13 | 0        | 0    |  |  |
| ## | hsa00533 |      | Glycosaminoglycan biosynthesis - keratan sulfate           |  |
| 14 | 0        | 0    |  |  |
| ## | hsa00603 |      | Glycosphingolipid biosynthesis - globo and isoglobo series |  |
| 14 | 0        | 0    |  |  |
| ## |          | P.Up | P.Down   |  |
| ## | hsa00780 | 1    | 1  |  |
| ## | hsa00290 | 1    | 1  |  |
| ## | hsa03267 | 1    | 1  |  |
| ## | hsa03264 | 1    | 1  |  |
| ## | hsa03265 | 1    | 1  |  |
| ## | hsa00232 | 1    | 1  |  |
| ## | hsa00524 | 1    | 1  |  |
| ## | hsa03260 | 1    | 1  |  |
| ## | hsa00470 | 1    | 1  |  |
| ## | hsa00400 | 1    | 1  |  |
| ## | hsa00440 | 1    | 1  |  |
| ## | hsa00750 | 1    | 1  |  |
| ## | hsa00740 | 1    | 1  |  |
| ## | hsa04122 | 1    | 1  |  |
| ## | hsa03266 | 1    | 1  |  |
| ## | hsa00920 | 1    | 1  |  |

|             |   |   |
|-------------|---|---|
| ## hsa00130 | 1 | 1 |
| ## hsa03450 | 1 | 1 |
| ## hsa00533 | 1 | 1 |
| ## hsa00603 | 1 | 1 |