

Impacto de la mutación del gen supresor tumoral *PTEN* en el tratamiento en cáncer.

Laura Gallardo Alfaro

Máster Universitario en Bioinformática y Bioestadística
Genoma y enfermedades oncológicas

Nombre Consultor/a: Dra. Laia Rosich Moya

4 de junio de 2019



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

Título del trabajo:	Impacto de la mutación del gen supresor tumoral <i>PTEN</i> en el tratamiento en cáncer
Nombre del autor:	<i>Laura Gallardo Alfaro</i>
Nombre del consultor/a:	<i>Dra. Laia Rosich Moya</i>
Nombre del PRA:	<i>Dr. David Merino Arranz</i>
Fecha de entrega (mm/aaaa):	06/2019
Titulación:	<i>Máster Universitario en Bioinformática y Bioestadística</i>
Área del Trabajo Final:	<i>Genoma y enfermedades oncológicas</i>
Idioma del trabajo:	<i>Español</i>
Palabras clave	PTEN, cáncer, diana terapéutica.
<p>Resumen del Trabajo (máximo 250 palabras): <i>Con la finalidad, contexto de aplicación, metodología, resultados i conclusiones del trabajo.</i></p>	
<p>Introducción: El cáncer es la segunda causa de muerte en el mundo. La aparición de mutaciones en ciertos genes es clave para el desarrollo tumoral, siendo una de las mutaciones más frecuentes la alteración del gen supresor tumoral <i>PTEN</i>.</p> <p>Objetivo: Identificar una vía metabólica/señalización implicada en la alteración de <i>PTEN</i> y que sea común en diversos cánceres, y los genes que la componen, susceptibles de ser una nueva diana terapéutica.</p> <p>Metodología: Se han utilizado bases de datos de microarrays (GEO data). Se obtienen 5 tipos de cánceres con <i>PTEN knock-out</i> (cáncer de colon, mama, epidermoide y pulmón no microcítico en <i>Homo sapiens</i> y cáncer de próstata en <i>Mus musculus</i>). Se analiza la diferente expresión entre la muestra <i>PTEN knock-out</i> y control y se extraen las vías de señalización estadísticamente significativas. Se analizan los genes que las componen.</p> <p>Resultados: La vía de señalización del cáncer de pulmón de células pequeñas (KEGGID: 05222), la vía de señalización de p53 (KEGGID: 04115) y las vías de señalización implicadas en el cáncer (KEGGID: 05200) son las vías de señalización comunes en el cáncer de colon, epidermoide y mama y estadísticamente significativas al comparar el cáncer con delección del supresor tumoral <i>PTEN</i> con su control. Los genes implicados en estas vías y en común en los diferentes cánceres son <i>TRAF4</i>, <i>TP53</i> y <i>CDKN1A</i>.</p> <p>Conclusiones: <i>TRAF4</i>, <i>TP53</i> y <i>CDKN1A</i> son susceptibles de ser nuevas dianas terapéuticas en aquellos cánceres con delección del gen <i>PTEN</i>.</p>	

Abstract (in English, 250 words or less):

Introduction: Cancer is the second leading cause of death globally. It develops mutations in order to proliferate and one of the most frequent one is the *PTEN* mutation.

Objective: To identify a metabolic/signalling pathway common in various cancers that plays a role in the alteration of *PTEN*. To identify genes included in these pathways susceptible to be new therapeutic targets.

Methodology: Microarray databases (GEO data) have been used. Five types of cancers were obtained with *PTEN* deletion (colon, breast, epidermoid and non-small cell lung cancer in *Homo sapiens* and prostate cancer in *Mus musculus*). Different expression between *PTEN* knock-out and control cancer samples was analysed, and statistically significant signalling pathways were extracted. Genes included in these pathways were analysed.

Results: The small cell lung cancer (KEGGID: 05222), the p53 signalling (KEGGID: 04115) and pathways involved in cancer (KEGGID: 05200) were statistically significant pathways in colon, epidermoid and breast cancer, when comparing *PTEN* knock-out and control. Common genes in these pathways were *TRAF4*, *TP53* and *CDKN1A*.

Conclusions: *TRAF4*, *TP53* and *CDKN1A* are susceptible to be therapeutic targets in those cancers with *PTEN* deletion.

Índice

1. INTRODUCCIÓN	1
1.1. INTRODUCCIÓN	1
1.1.1. <i>Definición de cáncer</i>	1
1.1.2. <i>Diferencias entre células cancerosas y normales</i>	2
1.1.3. <i>Origen de las células tumorales</i>	2
1.1.4. <i>Epidemiología</i>	3
1.1.5. <i>Definición y función de PTEN</i>	4
1.1.6. <i>Mutaciones en PTEN y su papel en el desarrollo tumoral</i>	5
1.1.7. <i>Mutaciones en PTEN y resistencia a tratamientos</i>	6
1.1.8. <i>Cáncer de próstata y PTEN</i>	7
1.1.9. <i>Cáncer de colon y PTEN</i>	8
1.1.10. <i>Carcinoma epidermoide y PTEN</i>	8
1.1.11. <i>Cáncer de pulmón no microcítico y PTEN</i>	8
1.1.12. <i>Cáncer de mama y PTEN</i>	9
1.2. CONTEXTO Y JUSTIFICACIÓN DEL TRABAJO	10
1.2.1. <i>Descripción general</i>	10
1.2.2. <i>Justificación del TFG</i>	10
1.3. OBJETIVOS DEL TRABAJO	11
1.3.1. <i>Objetivos generales</i>	11
1.3.2. <i>Objetivos específicos</i>	11
1.4. ENFOQUE Y MÉTODO SEGUIDO	12
1.5. PLANIFICACIÓN DEL TRABAJO	13
1.5.1. <i>Tareas</i>	13
1.5.2. <i>Calendario</i>	14
1.5.3. <i>Hitos</i>	14
1.5.4. <i>Análisis de riesgos</i>	14
1.5.5. <i>Acciones de mitigación</i>	15
1.6. BREVE SUMARIO DE PRODUCTOS OBTENIDOS.....	16
1.7. BREVE DESCRIPCIÓN DE LOS OTROS CAPÍTULOS DE LA MEMORIA.....	17
2. METODOLOGÍA	18
2.1. LECTURA DE LOS DATOS	19
2.2. PREPROCESADO: EXPLORACIÓN, CONTROL DE CALIDAD Y NORMALIZACIÓN.....	21
2.2.1. <i>Exploración de los datos</i>	21
2.2.2. <i>Control de calidad</i>	25
2.2.3. <i>Normalización</i>	27
2.2.4. <i>Exploración de los datos normalizados</i>	28
2.2.5. <i>Filtraje no específico</i>	29
2.3. SELECCIÓN DE GENES DIFERENCIALMENTE EXPRESADOS	29
2.4. ANOTACIÓN DE RESULTADOS.....	29
2.5. GENES SOBREENSPRESADOS Y DOWNREGULADOS O INHIBIDOS.	31
2.6. VISUALIZACIÓN DE LOS PERFILES DE EXPRESIÓN.	33
2.7. ANÁLISIS DE SIGNIFICACIÓN BIOLÓGICA.	36
2.7.1. <i>Análisis de enriquecimiento</i>	36
2.7.2. <i>Ontología Génica (GO)</i>	36

2.7.3. Vías de señalización o pathways (KEGG).....	38
2.8. GENES IMPLICADOS EN LAS VÍAS METABÓLICAS SELECCIONADAS	41
3. RESULTADOS Y DISCUSIÓN	42
3.1. VÍAS DE SEÑALIZACIÓN IMPLICADAS EN LA MUTACIÓN DE <i>PTEN</i>	42
3.1.1 Vía de señalización del cáncer de pulmón de células pequeñas.	42
3.1.2 Vía de señalización de p53.	43
3.1.3 Vías de señalización implicadas en el cáncer.	44
3.2. GENES EN COMÚN EN LAS DIFERENTES VÍAS DE SEÑALIZACIÓN SELECCIONADAS.	45
3.3. POSIBLES DIANAS TERAPÉUTICAS.	46
3.3.1. Factor 4 asociado a receptor de TNF (<i>TRAF4</i>).	46
3.3.2. Proteína tumoral 53 (<i>TP53</i>).	48
3.3.3. Inhibidor de la quinasa dependiente de ciclina (<i>CDKN1A</i>).....	49
3.4. IMPLICACIONES.....	51
4. LIMITACIONES	52
5. CONCLUSIONES	53
5.1. CONCLUSIONES.....	53
5.2. REFLEXIÓN CRÍTICA.....	53
5.3. ANÁLISIS CRÍTICO	54
5.4. LÍNEAS FUTURAS DE TRABAJO	54
6. GLOSARIO.....	55
7. BIBLIOGRAFÍA	56
ANEXO: PIPELINE.....	62

Lista de figuras

Figura 1. División celular en células normales y en células cancerosas.....	1
Figura 2. Localización molecular de <i>PTEN</i>	4
Figura 3. Vía de señalización de PTEN/PI3K/Akt.....	5
Figura 4. Fosforilación de PTEN en la tirosina 240 y posterior reparación del ADN.....	7
Tabla 1. Tareas.....	13
Figura 5. Diagrama de Gannt.....	14
Figura 6. Gráficos de densidad.....	22
Figura 7. Boxplots o diagramas de cajas.....	23
Figura 8. Dendogramas.....	24
Figura 9. Gráficos de degradación de cada array.....	25
Figura 10. Gráficos de expresiones relativas.....	26
Figura 11. Gráficos del error no estandarizado y normalizado o NUSE.....	27
Figura 12. Diagrama de cajas tras la normalización de los datos.....	28
Figura 13. Genes diferencialmente expresados cáncer de colon.....	30
Figura 14. Genes diferencialmente expresados cáncer de epidermoide.....	30
Figura 15. Genes diferencialmente expresados cáncer de mama.....	30
Figura 16. Genes diferencialmente expresados cáncer de próstata.....	30
Figura 17. Genes diferencialmente expresados cáncer de pulmón no microcítico.....	30
Figura 18. HeatMap para genes diferencialmente expresados cáncer de colon.....	33
Figura 19. HeatMap para genes diferencialmente expresados cáncer de epidermoide.....	34
Figura 20. HeatMap para genes diferencialmente expresados cáncer de mama.....	34
Figura 21. HeatMap para genes diferencialmente expresados cáncer de próstata.....	35
Figura 22. HeatMap para genes diferencialmente expresados cáncer de pulmón.....	35
Figura 23. Ejemplo Ontología Génica de los Procesos biológicos.....	37
Figura 24. Ejemplo Ontología Génica de las Funciones moleculares.....	37
Figura 25. Ejemplo Ontología Génica de Componentes celulares.....	37
Figura 26. Vías KEGG cáncer de colon.....	38
Figura 27. Vías KEGG cáncer epidermoide.....	38
Figura 28. Vías KEGG cáncer de mama.....	39
Figura 29. Vías KEGG cáncer de próstata.....	39
Figura 30. Vías KEGG cáncer de pulmón no microcítico.....	40
Figura 31. Lista de genes de la vía “Pathways in cancer” en cáncer de colon.....	41
Figura 32. Lista de genes de la vía “Small cell lung cancer” en cancer de colon.....	41
Figura 33. Lista de genes de la vía “p53 signalling pathway” en cancer de colon.....	41
Figura 34. Mapa de la vía de señalización del cáncer de pulmón de células pequeñas..	43
Figura 35. Mapa de la vía de señalización del gen supresor tumoral p53.....	44
Figura 36. Mapa de las vías de señalización implicadas en el cáncer.....	45
Tabla 2. LogFC para cada gen en diferentes cánceres.....	46
Figura 37. Papel del supresor tumoral p53.....	49
Figura 38. Papel dual de p21 o <i>CDKN1A</i>	50

1. Introducción

1.1. Introducción

1.1.1. Definición de cáncer

El término cáncer engloba un grupo de enfermedades relacionadas que se caracterizan por el desarrollo de células anormales, que se dividen, crecen y se diseminan sin ningún tipo de control en cualquier parte del cuerpo ¹.

El proceso de división celular está regulado por una serie de mecanismos de control que indican cuándo la célula puede comenzar a dividirse y cuándo permanecer estática. Normalmente, las células humanas crecen y se dividen para formar nuevas células y cuando se produce un daño celular que no puede ser reparado se produce la autodestrucción celular que impide que el daño sea heredado por las células descendientes ². Sin embargo, la célula cancerosa “pierde” la capacidad para morir e inicia una división incontrolada (*Figura 1*). Esta multiplicación en el número de células llega a formar masas tumorales o neoplasias, que en su expansión pueden destruir y sustituir a los tejidos normales ^{1,3}. Cuando las células que constituyen el tumor no poseen la capacidad de invadir y destruir otros órganos, hablamos de tumores benignos, pero cuando adquieren la facultad de invadir tejidos y órganos de alrededor (infiltración), y de trasladarse y proliferar en otras partes del organismo (metástasis), hablamos de tumor maligno o cáncer ².

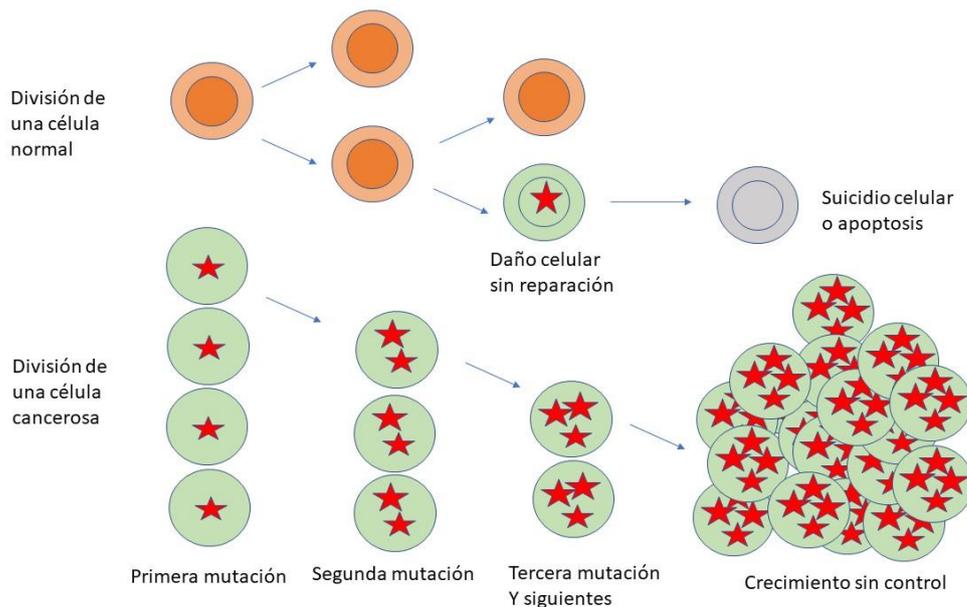


Figura 1. División celular en células normales y en células cancerosas.

1.1.2. Diferencias entre células cancerosas y normales.

La principal diferencia entre las células normales y las células cancerosas es que estas últimas tienen la capacidad de proliferar sin control, volviéndose invasivas. De esta forma, las células cancerosas pueden ignorar las señales celulares que conducen a las células a la muerte celular programada o apoptosis, la cual usa el cuerpo para deshacerse de las células que no le son necesarias^{3,4}.

Las células cancerosas pueden influir en las células normales, en las moléculas y en los vasos sanguíneos que rodean y alimentan al tumor (microambiente), por ejemplo, induciendo la angiogénesis o la formación de nuevos vasos sanguíneos que le suministren oxígeno y nutrientes necesarios para la progresión tumoral. Por otra parte, las células cancerosas son capaces de evadir el control del sistema inmunitario^{3,4}.

De esta forma podemos enumerar las principales capacidades adquiridas de una célula cancerosa:

- Potencial replicativo ilimitado.
- Inducción de la angiogénesis.
- Evasión de la apoptosis.
- Autosuficiencia en las señales de crecimiento.
- Insensibilidad hacia las señales de anti-crecimiento.
- Invasión tisular y metástasis.

1.1.3. Origen de las células tumorales

El proceso por el cual una célula normal se transforma en cancerosa se denomina carcinogénesis. El cáncer comienza con la transformación de una sola célula, que puede tener su origen en agentes externos y en factores genéticos heredados⁵. La transformación progresiva de células normales a derivados altamente malignos se origina en alteraciones en el material genético. Las mutaciones confieren a la célula la capacidad de dividirse a una tasa mayor que su cohorte y generar descendencia que conserva esta mutación (clones). Posteriormente, las células hijas acumulan subsecuentes y diversas mutaciones que permite generar distintos clones con mayores capacidades de supervivencia y crecimiento. Además, algunos clones pueden evadir los mecanismos de control del sistema inmunitario, lo que les confiere una ventaja proliferativa y permite el desarrollo de una neoplasia⁶.

De esta forma podemos decir que el cáncer es una enfermedad principalmente genética, causada por cambios en los genes que controlan la forma como funcionan nuestras células, especialmente la forma como crecen y se dividen. Los cambios genéticos que causan el cáncer pueden heredarse de los padres o pueden suceder también a lo largo de la vida de una persona, como resultado de errores que ocurren al dividirse las células o por el daño del ADN causado por algunas exposiciones del ambiente³.

Los cambios genéticos que contribuyen al cáncer tienden a afectar a tres tipos principales de genes:

- Proto-oncogenes. Los proto-oncogenes regulan el crecimiento y la diferenciación celular. Sus proteínas se expresan en diferentes momentos del ciclo y son imprescindibles para su regulación. Cambios estructurales y/o funcionales en los proto-oncogenes contribuyen a la malignación de la estirpe celular, convirtiéndolos en oncogenes. Estos oncogenes originarán proteínas con expresión y función alterada que favorecerán el crecimiento y/o la invasividad tumoral, reducción de la diferenciación celular e inhibición de la muerte celular. Las mutaciones en los proto-oncogenes son dominantes en la naturaleza, debido a que los oncogenes sólo precisan estar mutados en un alelo para que se produzca la sobreexpresión de una proteína dada y esta ejerza su acción promotora. Ejemplos de oncogenes son: *Kit*, *c-Raf*, *L-myc*, *N-myc*, *c-Myc*, *erb-B2*, *Bcl-1*, *Ret* o *Abl* ^{7,8}.
- Genes supresores de tumores. Los genes supresores de tumores actúan como factores de control que inhiben la proliferación celular excesiva. La función de las proteínas producto de estos genes es muy variada, pudiendo actuar en la regulación de la progresión del ciclo celular, control de la transmisión de señales, inducción de apoptosis, modulación de la transcripción o promoción de la adhesión celular. De esta forma, cuando se alteran o se inhiben los genes supresores tumorales, se permite la proliferación o crecimiento celular. Para que se manifieste el fenotipo tumoral, a diferencia de los proto-oncogenes que era suficiente con la alteración de uno solo de sus alelos, en los genes supresores de tumores es necesario la inactivación de sus dos copias genómicas. Ejemplos de genes supresores de tumores son: *p53*, *PTEN*, *Ras* o *TGF-β* ⁹.
- Genes reparadores del ADN. Los genes reparadores del ADN corrigen los errores que se producen espontáneamente en el ADN durante la división celular, o que son consecuencia de la exposición a productos químicos mutágenos o a la radiación. Entre estos genes encontramos *hMSH2* y *hMLH1*, que intervienen en la reparación del ADN desacoplado ¹⁰.

1.1.4. Epidemiología

La Organización Mundial de la Salud estima que el cáncer es la segunda causa de muerte en el mundo, ocasionando en el año 2015, 8.8 millones de defunciones. Aproximadamente un 30% de las muertes por cáncer se deben a cinco factores de riesgo comportamentales y alimentarios (índice de masa corporal elevado, consumo insuficiente de frutas y verduras, falta de actividad física y consumo de tabaco y alcohol) y, por tanto, pueden prevenirse ¹¹.

Según el informe anual editado por la Sociedad Española de Oncología Médica, “Las Cifras del Cáncer en España”, en 2017 los nuevos casos estimados de cáncer en España han sido de 228.482 y para el año 2035 se estima que habrá 315.413 nuevos casos de cáncer. Las estimaciones poblacionales indican que el número de casos nuevos probablemente aumenten en un 70% en las próximas décadas, alcanzando a nivel mundial los 24 millones de casos aproximadamente en el año 2035 ¹².

Los tumores más diagnosticados a nivel mundial son el cáncer colorrectal, próstata, pulmón, mama, vejiga y estómago. Sin embargo, si fragmentamos la incidencia de los distintos tumores por sexo, se estima que los tumores más frecuentemente diagnosticados en varones en España en 2017 fueron los de próstata, pulmón, colorrectal, vejiga, estómago, riñón, hígado, páncreas, linfoma no Hodgkin y leucemias. Por otra parte, en mujeres fueron los de mama, colon, útero, pulmón, ovario, páncreas, estómago, linfoma no Hodgkin, melanoma cutáneo y cérvix ¹².

De hecho, los tumores responsables del mayor número de fallecimientos en 2016 en España en la población general fueron el cáncer de pulmón (22.187 muertes) y el cáncer colorrectal (15.802 muertes), seguidos a una gran distancia del cáncer de páncreas (6.789 casos), el cáncer de mama (6.477 muertes), y de próstata (5.752 muertes) ¹².

1.1.5. Definición y función de *PTEN*.

El homólogo de la fosfatasa y la tensina (*PTEN*) es un potente supresor tumoral localizado en el cromosoma 10 (*Figura 2*), cuya pérdida de función se observa frecuentemente en una gran cantidad de cánceres ¹³. De hecho, *PTEN* es el segundo gen mutado con más frecuencia en el cáncer humano después del supresor tumoral *p53* ¹⁴.

PTEN tiene actividades dependientes e independientes de fosfatasa en la célula y regula gran variedad de procesos biológicos, incluido el mantenimiento de la estabilidad genómica, la supervivencia celular, la migración, la proliferación y el metabolismo. Incluso una disminución sutil en los niveles de *PTEN* y de su actividad resulta en la susceptibilidad al cáncer y favorece la progresión del tumor. Por lo tanto, la regulación de *PTEN* se ha convertido en un tema de intensa investigación en biología de tumores ¹⁵.

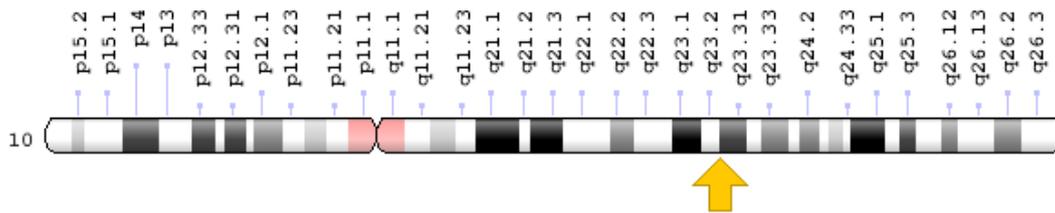


Figura 2. Localización molecular de *PTEN*. El gen se encuentra en el cromosoma 10 en la posición 23.31y comprende los pares de base 87.863.438 al 87.971.930 ¹⁶.

PTEN contiene dos dominios, el dominio de fosfatasa N-terminal y el dominio C-terminal. Aunque el papel de *PTEN* en la supresión tumoral se atribuya principalmente a su actividad de la fosfatasa lipídica N-terminal, más del 40% de las mutaciones tumorígenicas se producen en el dominio C-terminal ¹⁴.

PTEN es una fosfatasa con especificidad dual que puede desfosforilar proteínas y lípidos ¹⁷. La proteína codificada por este gen es una fosfatidilinositol-3,4,5-trisfosfato 3-fosfatasa, que contiene un dominio similar a la tensina, así como un dominio catalítico similar al de la proteína de especificidad dual tirosina fosfatasas. *PTEN* funciona como

un supresor de tumores mediante la regulación negativa de la vía de señalización PI3K-AKT/PKB¹⁸ (fosfoinosítida 3-quinasa-RAC-alfa serina / treonina-proteína quinasa) (Figura 3) a partir de la desfosforilación de fosfoinosítidos y por lo tanto de la modulación de la progresión del ciclo celular y la supervivencia celular.

La forma no fosforilada de la proteína PTEN coopera con AIP1 para suprimir la activación de AKT1. Además, desempeña un papel como modulador clave de la vía de señalización AKT-mTOR que controla el ritmo del proceso de integración de las neuronas recién nacidas durante la neurogénesis adulta, incluida la correcta colocación de las neuronas, el desarrollo dendrítico y la formación de sinapsis¹⁹.

PTEN puede ser un regulador negativo de la señalización de la insulina y del metabolismo de la glucosa en el tejido adiposo¹⁹. Por otra parte, el empalme alternativo y el uso de múltiples codones de inicio de la traducción dan como resultado múltiples variantes de transcripción que codifican diferentes isoformas, entre ellas, una isoforma de PTEN más larga que puede ayudar a regular el metabolismo energético en las mitocondrias¹⁸.

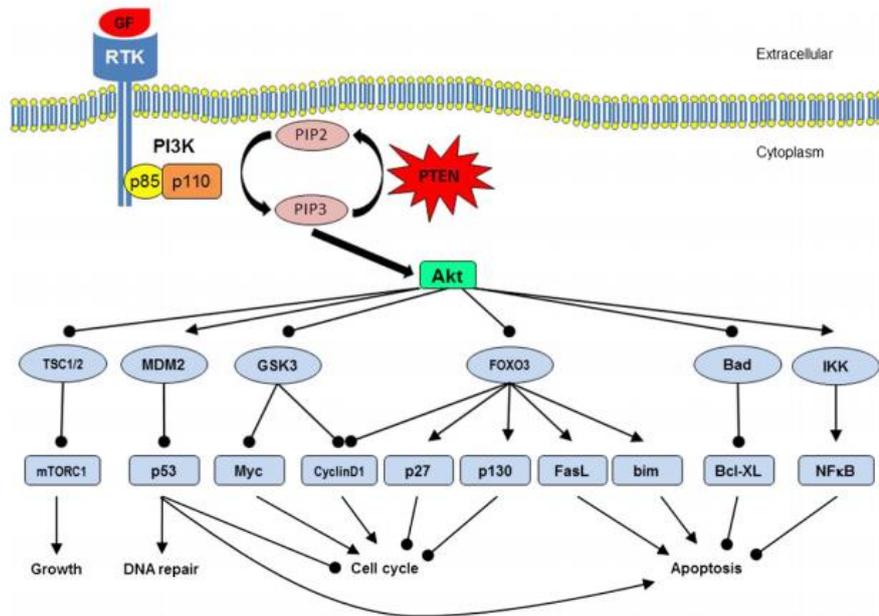


Figura 3. Vía de señalización de PTEN/PI3K/Akt²⁰.

PTEN también ejerce algunas actividades biológicas independientemente de su actividad catalítica. Por ejemplo, PTEN interactúa directamente con el supresor de tumores p53, mejorando su estabilidad y actividad transcripcional^{21,22}.

1.1.6. Mutaciones en *PTEN* y su papel en el desarrollo tumoral.

Las numerosas mutaciones somáticas de *PTEN* identificadas en los cánceres humanos afectan la estabilidad de *PTEN*, la localización subcelular y/o las actividades de la lípido

fosfatasa/lipídica y proteína fosfatasa²³. Las mutaciones de *PTEN* ocurren con más frecuencia en tres tipos de cáncer humano: glioblastoma, cáncer de endometrio y próstata²⁴.

Se han encontrado mutaciones en la línea germinal de *PTEN* en síndromes de susceptibilidad al cáncer, como el síndrome de Cowden, en el que más del 80% de los pacientes tienen mutaciones de *PTEN*. Las características distintivas de la enfermedad de Cowden son el sobrecrecimiento benigno de diversos tejidos, incluidos la piel, el seno, el intestino y el cerebro, así como una predisposición al desarrollo de tumores de mama y tiroides²⁴.

La eliminación homocigótica de *PTEN* causa letalidad embrionaria, lo que sugiere que *PTEN* es esencial para el desarrollo embrionario. Los ratones heterocigotos para *PTEN* desarrollan tumores espontáneos en una variedad de órganos comparables con el espectro de sus mutaciones en el cáncer humano¹⁴.

De esta forma, los estudios celulares revelan una gran cantidad de cambios fenotípicos en células deficientes en *PTEN*, como el aumento de la proliferación^{25,26}, la reducción de la apoptosis²⁷ y la mejora de la migración^{28,29}. En contraste, la sobreexpresión de *PTEN* en células cancerosas induce apoptosis y bloquea la progresión del ciclo celular, la formación de colonias y la migración celular^{28,30,31}.

1.1.7. Mutaciones en *PTEN* y resistencia a tratamientos.

La pérdida o la mutación de *PTEN* se asocia con la resistencia tumoral a ciertos tratamientos. *Eze N et. al. (2019)* demostró que la pérdida de la expresión de *PTEN* se asociaba con una baja eficiencia al tratamiento con cetuximab³² (anticuerpo IgG1 monoclonal quimérico cuya diana es el receptor del factor de crecimiento epidérmico) en el carcinoma de células escamosas de cabeza y cuello³³. *Fernández Garcia et. al.* mostró que la resistencia al tratamiento con trastuzumab en cáncer de mama se asociaba significativamente con la ausencia total de expresión de *PTEN*³⁴. Por otra parte, *Zhao et. al. (2019)* encontró mutaciones somáticas en *PTEN* que estaban asociadas con la resistencia a la inmunoterapia en pacientes con glioblastomas. Realizando un análisis estructural de la proteína PTEN encontraron mutaciones de pérdida de función múltiple, incluyendo D107Y, G132D, K147 y R173C en el dominio fosfatasa de PTEN, y R233, D252Y y L295fs dentro del dominio C2 de PTEN. Como consecuencia de estas mutaciones, PTEN sufre una pérdida de función y por tanto encontraron una mayor activación de la ruta PI3K-AKT en individuos que no respondían al tratamiento^{35,36}.

La radiación ionizante y la quimioterapia son tratamientos estándar para los pacientes con glioblastoma y ambos producen daño en el ADN, sin embargo, la eficacia clínica es limitada debido a la resistencia terapéutica. *Jianhui et. al. (2019)* identificaron un mecanismo de resistencia al tratamiento mediado por la fosforilación de PTEN en la tirosina 240 (pY240-PTEN) por FGFR2. pY240-PTEN se une a la cromatina a través de la interacción con Ki-67 en respuesta al tratamiento con radiación y facilita el reclutamiento de RAD51 para promover la reparación del ADN y por tanto la supervivencia de tumor al tratamiento. Estos resultados sugieren que el pY240-PTEN mediado por FGFR es un mecanismo clave de la resistencia a la radiación y es un objetivo viable para mejorar la eficacia de la radioterapia (Figura 4)³⁷.

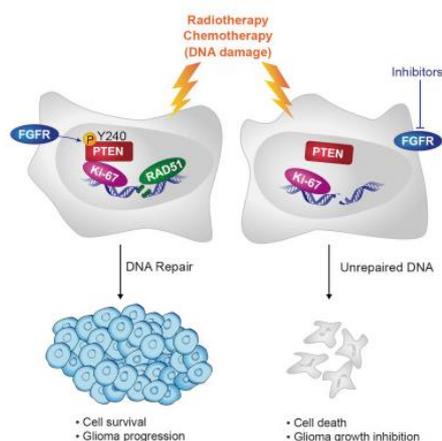


Figura 4. Fosforilación de PTEN en la tirosina 240 y posterior reparación del ADN³⁷.

1.1.8. Cáncer de próstata y *PTEN*

El cáncer de próstata es la segunda causa principal de muerte en los hombres de los Estados Unidos, después del cáncer de pulmón. Aproximadamente 1 de cada 41 hombres morirá por cáncer de próstata. El cáncer de próstata puede ser una enfermedad grave, aunque la mayoría de los hombres diagnosticados con este cáncer no muere a causa de esta enfermedad. De hecho, en los Estados Unidos, más de 2.9 millones de hombres que han sido diagnosticados con cáncer de próstata en algún momento, siguen vivos hoy en día³⁸.

La mutación del gen supresor tumoral *PTEN* es de las más comunes en el cáncer de próstata. La inactivación de *PTEN* por delección o mutación es identificada en aproximadamente el 20% de las muestras de tumores de próstata primarios y en hasta el 50% de los tumores resistentes a la castración. La pérdida de la función de *PTEN* conduce a la activación de la vía PI3K-AKT y está fuertemente asociada con resultados oncológicos adversos, haciendo de *PTEN* un marcador genómico potencialmente útil para distinguir la enfermedad indolente de la agresiva en pacientes con tumores clínicamente localizados³⁹.

Además, los compuestos terapéuticos dirigidos a los nodos en la vía de señalización PI3K-AKT-mTOR se están probando en ensayos clínicos para pacientes con cáncer de próstata metastásico resistente a la castración. El conocimiento del estado de *PTEN* podría ser útil para identificar a los pacientes que tienen más probabilidades de beneficiarse de estas terapias. Finalmente, los estudios han demostrado que *PTEN* podría tener funciones adicionales que son independientes de la vía PI3K-AKT, incluidas las que afectan el crecimiento del tumor a través de la modulación de la respuesta inmune y el microentorno del tumor³⁹.

1.1.9. Cáncer de colon y *PTEN*

El cáncer colorrectal es el que se origina en el colon o el recto. A estos cánceres también se les puede llamar cáncer de colon o cáncer de recto (rectal) dependiendo del lugar donde se originen⁴⁰. El cáncer colorrectal se produce cuando se forman tumores en el revestimiento del intestino grueso. Es común tanto en hombres como en mujeres y el riesgo de desarrollarlo aumenta después de los 50 años⁴¹.

La mutación de *PTEN* ocurre en el 5-14% de los cánceres colorrectales⁴²⁻⁴⁴. Además, en estudios retrospectivos, la delección de *PTEN* se ha asociado a una disminución de la sensibilidad del cáncer colorrectal al tratamiento de anticuerpos anti-EGFR (receptor del factor de crecimiento epidérmico)^{43,45}. Otros autores señalan que la pérdida de *PTEN* se relaciona con la resistencia al tratamiento con cetuximab en pacientes con cáncer de colon⁴⁵.

1.1.10. Carcinoma epidermoide y *PTEN*

El carcinoma epidermoide es un tipo de cáncer que empieza en las células escamosas de la piel que son células delgadas y planas que se parecen a las escamas de los peces. Estas células se encuentran en el tejido que forma la superficie de la piel, el revestimiento de los órganos huecos del cuerpo y el revestimiento de los aparatos respiratorio y digestivo. La mayoría de los cánceres de ano, cuello uterino, cabeza y cuello, y vagina son carcinomas epidermoides, también llamado carcinoma de células escamosas⁴⁶.

La pérdida de *PTEN*, tanto en el número de copias génicas como en la expresión proteica, parece relacionarse con tumores menos agresivos del carcinoma epidermoide del conducto auditivo externo⁴⁷. Además, como ocurre en otros cánceres, la pérdida de *PTEN* se relaciona con una resistencia al tratamiento con cetuximab en el carcinoma epidermoide del cuello³³.

1.1.11. Cáncer de pulmón no microcítico y *PTEN*

Existen dos tipos principales de cáncer de pulmón: cáncer de pulmón no microcítico (CPNM) o cáncer de pulmón de células no pequeñas que constituye alrededor del 80% al 85% de los cánceres de pulmón y cáncer de pulmón microcítico o de células pequeñas que constituye alrededor del 10% al 15% de los cánceres de pulmón⁴⁸.

Hay tres subtipos de cáncer de pulmón no microcítico, los cuales se originan de diferentes tipos de células pulmonares: adenocarcinoma, carcinoma de células escamosas o epidermoide y carcinoma de células grandes⁴⁸.

El cáncer de pulmón no microcítico es la principal causa de muerte por cáncer en todo el mundo. En los últimos años, la identificación de mutaciones que activan EGFR, confieren mayor sensibilidad y respuesta de la enfermedad a los inhibidores de la tirosina quinasa, cambiando la perspectiva de los pacientes con CPNM. Las alteraciones en la vía PI3K / AKT, principalmente la inactivación de *PTEN*, se han asociado con la

resistencia a la terapia con inhibidores de la EGFR-tirosina quinasa y una menor supervivencia en pacientes con CPNM⁴⁹.

1.1.12. Cáncer de mama y *PTEN*

El cáncer de mama es el tumor más frecuente en la mujer, siendo el grupo de edad con mayor prevalencia el comprendido entre los 65 y 70 años. La historia familiar es el factor de riesgo más importante, pero solo el 5-10% de las pacientes con cáncer de mama presentan una mutación heredada de uno de sus padres. Aunque las mutaciones germinales en los genes de alta susceptibilidad al cáncer de mama, *BRCA1* y *BRCA2* (*BRCA1/2*), son las más frecuentemente asociadas al síndrome de cáncer de mama/ovario hereditario, se han descrito otros genes de alta penetrancia asociados al cáncer de mama como *TP53* (síndrome de Li-Fraumeni), *PTEN* (síndrome de Cowden) y *STK11* (síndrome de Peutz-Jeghers)⁵⁰.

La probabilidad en las mujeres portadoras de la mutación en *PTEN* de presentar cáncer de mama se estima entre el 25-50% con una media de edad al diagnóstico de 38-50 años⁵¹. Trastuzumab, el anticuerpo monoclonal humanizado del receptor del factor de crecimiento epidérmico humano 2 (HER2), y lapatinib, un inhibidor del receptor del factor de crecimiento epidérmico / HER2 tirosina quinasa, son tratamientos establecidos para los cánceres de mama que sobreexpresan HER2⁵². Recientes estudios indican que la activación de la vía PI3K producida por la mutación en el gen *PTEN* se asocia a la resistencia al tratamiento con lapatinib y con trastuzumab en cáncer de mama⁵³.

1.2. Contexto y justificación del Trabajo

1.2.1. Descripción general

El cáncer es un proceso secuencial que implica la acumulación de sucesivas mutaciones en uno o varios genes de distintos tipos (oncogenes, genes supresores, genes de susceptibilidad). El proceso que conduce al desarrollo de un tumor maligno podría comenzar con una mutación individual en una sola célula del organismo que contribuye a incrementar la tasa de crecimiento de la célula portadora respecto a las adyacentes, favoreciendo la aparición y selección de un nuevo clon de células con una mayor capacidad de división⁵⁴.

PTEN es un gen supresor tumoral que se encuentra mutado en gran variedad de cánceres^{55,56}, sugiriendo un papel relevante en el desarrollo de la enfermedad. Este trabajo se centrará en el estudio del gen supresor tumoral *PTEN* y en las vías metabólicas implicadas en la mutación del mismo. De esta forma, encontrar una vía metabólica implicada en la mutación de *PTEN*, común en diversos cánceres, permitirá identificar una nueva diana terapéutica que podrá establecer la base para el desarrollo de nuevos fármacos.

1.2.2. Justificación del TFG

El cáncer es la segunda causa de muerte en el mundo. La aparición de mutaciones en ciertos genes es clave para el desarrollo tumoral, siendo una de las mutaciones más frecuentes la alteración del gen supresor tumoral *PTEN*⁵⁷. La función principal de un gen supresor tumoral consiste en inhibir la proliferación celular excesiva, por lo que la mutación en el gen *PTEN* facilitará el desarrollo del tumor. Además, existe evidencia de que la mutación en *PTEN* estaría relacionada con la resistencia tumoral a ciertos tratamientos^{53,58}. Es evidente la necesidad actual de encontrar una nueva diana terapéutica que pueda ser utilizada en diferentes cánceres que presenten mutación en el gen supresor tumoral *PTEN*.

Este trabajo desarrollado en el área de “genoma y enfermedades oncológicas” y centrado en el estudio de la mutación del gen supresor tumoral *PTEN* identificará las vías metabólicas implicadas con la alteración de *PTEN* y que son comunes en diferentes cánceres. Por tanto, el objetivo principal de este trabajo se centrará en encontrar una vía común en diferentes tipos de cánceres y los genes que la componen, susceptibles de ser una nueva diana terapéutica.

1.3. Objetivos del Trabajo

1.3.1. Objetivos generales

1. Identificar una vía metabólica o de señalización implicada en la alteración de *PTEN* y que sea común en diversos cánceres.
2. Identificar una nueva diana terapéutica.

1.3.2. Objetivos específicos

1. Describir la función del gen *PTEN*, su relación con el desarrollo del tumor y con el desarrollo de resistencias a tratamientos.
2. Identificar las vías metabólicas implicadas en la alteración de *PTEN* en diferentes tipos de cánceres.
3. Describir la función de una vía metabólica/señalización implicada en la alteración de *PTEN* y que sea común en diversos cánceres.
4. Identificar los genes que tengan un papel en la vía metabólica/señalización implicada en la alteración de *PTEN* y que sean comunes en diferentes tipos de cánceres.
5. Evaluar la posible diana terapéutica.

1.4. Enfoque y método seguido

El trabajo final, se ha enfocado en el estudio de la mutación del gen supresor tumoral *PTEN* en diferentes tipos de cáncer. Determinar cuáles son las vías metabólicas implicadas en la mutación, comunes en diversos cánceres, con el objetivo de encontrar una vía en común en diferentes cánceres y sus genes implicados, que sean susceptibles de ser diana terapéutica.

De esta forma se utilizan diferentes bases de datos de microarrays (GEO data) y de genes (RefSeq) con el objetivo de encontrar estudios experimentales que hayan utilizado la mutación de *PTEN* en cáncer. Se buscan estudios experimentales que utilicen modelos con *PTEN knock-out* y se estudian las vías metabólicas implicadas. Se analizan los estudios experimentales con diversos paquetes bioinformáticos del software R. Se consultan bases de datos como PubMed para realizar una revisión del tema principal.

Como producto se obtienen las vías metabólicas en común para todos los cánceres estudiados y la identificación de tres posibles dianas terapéuticas que servirán para futuras investigaciones y para el desarrollo de un nuevo fármaco.

Los condicionantes principales para la elección de este método de investigación son:

- Ausencia de un estudio experimental propio.
- Necesidad de estudios con distintos tipos de cáncer.
- Presupuesto económico: 0 €
- Tiempo disponible para la realización: 2-3 meses.

Debido a los condicionantes mencionados, resultaría imposible realizar un estudio experimental con muestras de diferentes cánceres *vs.* muestras de diferentes cánceres, pero con el gen supresor tumoral *PTEN* mutado. Esta sería la mejor forma de identificar las vías metabólicas implicadas en la mutación de *PTEN* y comunes en diferentes cánceres ya que todas las muestras tendrían las mismas condiciones y se estarían evitando sesgos. Debido a una limitación en el tiempo y a la ausencia de presupuesto es imposible realizar un estudio experimental, por lo que se debe optar por la opción de crear la base de datos a partir de estudios experimentales ya realizados y que se puedan encontrar en bases de datos públicas.

De esta forma, los materiales necesarios para llevar a cabo este estudio son:

- Portátil ASUS Intel Core i7-6500U, up to 3.1 GHz. Windows 10. 12GB de memoria RAM.
- Programa R Studio.
- Bases de datos de microarrays (GEO data) y de genes (RefSeq), así como Pubmed.
- Programas depositados en Bioconductor que funcionan con R: affy, limma, DESeq2, edgeR, etc.
- Microsoft Word y Power Point.

1.5. Planificación del Trabajo

1.5.1. Tareas

Tabla 1. Tareas

<i>Tareas</i>	Fecha de inicio	Fecha de finalización
1. Entrega PEC 1: Plan de Trabajo	05/03/2019	18/03/2019
2. Desarrollo del trabajo Fase 1		
2.1. <i>Búsqueda general de información</i>	19/03/2019	24/03/2019
2.2. <i>Función PTEN y relación con el cáncer</i>	25/03/2019	31/03/2019
2.3. <i>Mutaciones en PTEN y resistencia al tratamiento</i>	01/04/2019	03/04/2019
2.4. <i>Búsqueda de estudios</i>	04/04/2019	10/04/2019
2.5. <i>Descripción tipos de cánceres a estudiar.</i>	11/04/2019	15/04/2019
2.6. <i>Descripción de la bibliografía actual sobre la relación del gen PTEN y cada cáncer a estudiar, frecuencia de la mutación y su implicación.</i>	16/04/2019	20/04/2019
2.7. <i>Revisión final del documento PEC2</i>	21/04/2019	23/04/2019
2.8. <i>Entrega PEC2</i>	24/04/2019	24/04/2019
3. Desarrollo del trabajo Fase 2		
3.1. <i>Análisis bioinformático</i>	25/04/2019	05/05/2019
3.2. <i>Identificación vías metabólicas implicadas en alteración PTEN</i>	06/05/2019	07/05/2019
3.3. <i>Identificación de una vía metabólica común</i>	08/05/2019	09/05/2019
3.4. <i>Identificación de una posible diana terapéutica</i>	10/05/2019	11/05/2019
3.5. <i>Búsqueda bibliográfica</i>	12/05/2019	16/05/2019
3.5. <i>Revisión final del documento PEC3</i>	17/05/2019	19/05/2019
3.6. <i>Entrega PEC3</i>	20/05/2019	20/05/2019
4. Elaboración de la Memoria y la presentación.		
4.1. <i>Elaboración de la memoria.</i>	21/05/2019	26/05/2019
4.2. <i>Corrección y revisión de la memoria.</i>	27/05/2019	03/06/2019
4.3. <i>Entrega PEC4: Cierre de la memoria</i>	04/06/2019	04/06/2019
4.4. <i>Elaboración de la presentación.</i>	05/06/2019	09/06/2019
4.5. <i>Corrección y revisión de la presentación.</i>	10/06/2019	11/06/2019
4.6. <i>Entrega PEC5a: Elaboración de la Presentación.</i>	12/06/2019	12/06/2019
4.7. <i>Entrega PEC5B: Defensa pública.</i>	13/06/2019	25/06/2019

1.5.2. Calendario

El calendario de las tareas se ha planificado teniendo en cuenta diversas situaciones, por lo que se ha asignado más tiempo a aquellas tareas que coinciden en vacaciones, días festivos, fin de semana o días en los que no se tiene disponibilidad por diversos motivos. Se tienen en cuenta las fechas de entrega de las diferentes PEC.

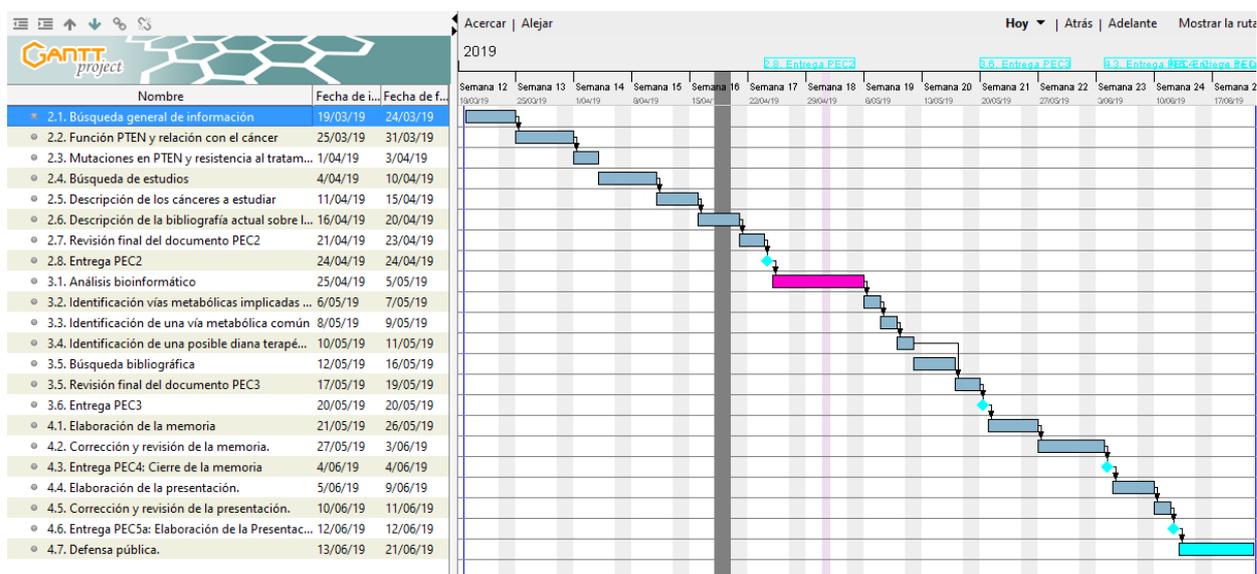


Figura 5. Diagrama de Gantt. En color se observa el tiempo dedicado al análisis bioinformático. Las columnas grises representan fines de semana y festivos.

1.5.3. Hitos

- Entrega PEC 1: Plan de Trabajo (18/03/2019)
- Entrega PEC 2: Desarrollo del Trabajo Fase 1 (24/04/2019)

La PEC2 contendrá los apartados 2.1 a 2.6 especificados en la planificación de tareas.

- Entrega PEC 3: Desarrollo del Trabajo Fase 2 (20/05/2019)

La PEC3 contendrá los apartados 3.1 a 3.5 especificados en la planificación de tareas.

- Entrega PEC 4: Cierre de la Memoria (04/06/2019)
- Entrega PEC 5a: Elaboración de la Presentación (12/06/2019)
- Entrega PEC 5b: Defensa pública (13/06/2019)

1.5.4. Análisis de riesgos

Los factores que pueden repercutir negativamente en el seguimiento del plan de trabajo y en la consecución del proyecto son:

1. Factor tiempo.
2. Encontrar los estudios adecuados para poder realizar el análisis bioinformático.

3. Viaje o circunstancia inesperada, como podría ser una enfermedad que pueda retrasar alguna tarea.

1.5.5. Acciones de mitigación:

Siendo el factor tiempo el principal problema, se dispone de una lista de tareas bien planificada para repartir todas las tareas. En esta lista de tareas no se incluyen festivos o fines de semana. Si un viaje inesperado o simplemente el tiempo fuera realmente un problema se utilizarían los días festivos para finalizar el proyecto con el objetivo de disponer de más tiempo para la realización de este. Además, se podría redistribuir las tareas, quitando tiempo a ciertas tareas y añadiéndoselo a otras más necesarias en función de las necesidades del momento.

Respecto al punto 2 sobre encontrar los estudios necesarios se podría modificar ligeramente el método a utilizar buscando en otras bases de datos y añadiendo más tiempo a ciertas tareas.

1.6. Breve resumen de productos obtenidos

Este estudio permite obtener las vías metabólicas estadísticamente significativas entre diferentes cánceres con el gen supresor tumoral *PTEN* inhibido comparado con los mismos cánceres controles. De esta forma se obtiene una lista de las vías metabólicas que desempeñan un papel diferencial cuando *PTEN* está inhibido. A partir de las vías metabólicas se obtienen los genes que las constituyen y que son comunes en el mayor número de cánceres. En total se obtienen 3 posibles dianas terapéuticas.

1.7. Breve descripción de los otros capítulos de la memoria

En el apartado 2 (Metodología) se describe en detalle los datos usados para completar los objetivos del Trabajo de Fin de Máster. Se explica resumidamente los diferentes pasos utilizados con R para analizar los datos. Además, se muestran resultados parciales de cada etapa del análisis.

En el apartado 3 (Resultados y discusión) se detallan los resultados finales obtenidos y se relacionan con la bibliografía actual. Se muestran las vías metabólicas implicadas en la mutación de *PTEN* y las posibles dianas terapéuticas y sus implicaciones.

En el apartado 4 (Limitaciones) se detallan las limitaciones del estudio.

En el apartado 5 (Conclusiones) se sintetizan los resultados clave del estudio, así como las posibles implicaciones. Se realiza una reflexión y un análisis crítico. Además, se detallan las líneas futuras de trabajo.

2. Metodología

Se realiza una búsqueda en bases de datos de *microarrays* (GEO data) y de genes (RefSeq) con el objetivo de encontrar estudios experimentales en diferentes cánceres que hayan utilizado la delección del gen *PTEN* y que cuenten con una muestra control con la que poder realizar la comparación de la diferente expresión génica. El reducido número de estos estudios es una limitación propia de este trabajo.

Criterios de inclusión:

- Las muestras deben ser de algún tipo de cáncer.
- Las muestras deben contener la delección del gen *PTEN* (*PTEN* null, *PTEN* loss, *PTEN* -/-, *PTEN* knock-out).
- Se debe incluir una muestra control con la que poder comparar la muestra *PTEN* loss.

Se encuentran 5 tipos de estudios experimentales que cumplen con estos criterios de inclusión:

- Cáncer de colon:

Puede consultarse la información y descargarse los datos del cáncer de colon en <https://www.ncbi.nlm.nih.gov/sites/GDSbrowser?acc=GDS2446#details>. Se trata de un estudio experimental en el que se pretendía estudiar el papel de *PTEN* en la patogénesis del cáncer. Se utilizaron 5 arrays, 2 arrays de cáncer de colon control y 3 arrays de cáncer de colon en la que se realizó la delección de *PTEN* a partir de “gene targeting”. Este estudio experimental se realizó en *Homo sapiens* con Affymetrix Human Genome U133A Array.

- Cáncer de mama.

Puede consultarse la información y descargarse los datos del cáncer de mama en <https://www.ncbi.nlm.nih.gov/sites/GDSbrowser?acc=GDS2958#details>. Se trata de un estudio experimental en el que se suprimió *PTEN* con el objetivo de profundizar en las bases moleculares del fenotipo molecular resultante de la inactivación de *PTEN*. Se utilizaron 4 arrays, 2 arrays control y 2 arrays de cáncer de mama en la que se realizó la delección de *PTEN*. Este estudio experimental se realizó en *Homo sapiens* con Affymetrix Human Genome U133 Plus 2.0 Array.

- Cáncer epidermoide.

Puede consultarse la información y descargarse los datos del cáncer epidermoide en <https://www.ncbi.nlm.nih.gov/sites/GDSbrowser?acc=GDS2958#details>. Se trata de un estudio experimental en el que se suprimió *PTEN* con el objetivo de profundizar en las bases moleculares del fenotipo molecular resultante de la inactivación de *PTEN*. Se utilizaron 4 arrays, 2 arrays control y 2 arrays de cáncer epidermoide en la que se

realizó la delección de *PTEN*. Este estudio experimental se realizó en *Homo sapiens* con Affymetrix Human Genome U133 Plus 2.0 Array.

- Cáncer de pulmón no microcítico.

Puede consultarse la información y descargarse los datos del cáncer de pulmón no microcítico en <https://www.ncbi.nlm.nih.gov/sites/GDSbrowser?acc=GDS2958#details>. Se trata de un estudio experimental en el que se suprimió *PTEN* con el objetivo de profundizar en las bases moleculares del fenotipo molecular resultante de la inactivación de *PTEN*. Se utilizaron 4 arrays, 2 arrays control y 2 arrays de cáncer de pulmón no microcítico en la que se realizó la delección de *PTEN*. Este estudio experimental se realizó en *Homo sapiens* con Affymetrix Human Genome U133 Plus 2.0 Array.

- Cáncer de próstata.

Puede consultarse la información y descargarse los datos del cáncer de próstata en <https://www.ncbi.nlm.nih.gov/sites/GDSbrowser?acc=GDS4125#details>. Se trata de un estudio experimental en el que se suprimió *PTEN* y se activó *K-ras* con el objetivo de establecer la relación entre *K-ras* y *PTEN* y su relación con la promoción del cáncer de próstata y la metástasis. Es un estudio de 6 arrays (3 arrays controles y 3 arrays con *PTEN*null/*Kras*+) que se realizó en *Mus musculus* con Affymetrix Mouse Genome 430 2.0 Array. Debido a la distinta naturaleza de este cáncer, ya que se trata de muestras en ratón y no en humanos como los otros 4 cánceres a estudiar, y además, debido a que se activa *K-ras* lo cual puede influir la expresión génica e incluir sesgos en el análisis, esta muestra de cáncer de próstata se utilizará como complementaria. Es decir, la muestra de cáncer de próstata debido a las diferencias descritas se procesará como las demás, pero únicamente complementará los resultados, de forma que si se encuentran vías metabólicas diferenciadas que son comunes con los otros cánceres, nos validará nuestro análisis, pero nunca se tomará como resultado base.

Se descargan los arrays correspondientes de cada tipo de cáncer.

2.1. Lectura de los datos

La lectura de los datos se realizó a partir de los paquetes *Biobase* y *affy* de Bioconductor. Se ha creado un archivo “target.txt” para cada tipo de cáncer (cáncer de colon, mama, próstata, epidermoide y de pulmón no microcítico) con la información de las condiciones experimentales y sus *archivos* .*CEL* asociados. El contenido del archivo target.txt se utilizó en la lectura de los datos y la posterior creación del objeto *rawData* de la clase *affybatch* con las intensidades en crudo de cada archivo.CEL.

Los objetos *rawData* serán la base para el posterior análisis:

```
show(rawDataColon)

## AffyBatch object
## size of arrays=712x712 features (20 kb)
## cdf=HG-U133A (22283 affyids)
## number of samples=5
```

```
## number of genes=22283
## annotation=hgu133a
## notes=
```

Se obtiene que para el estudio con cáncer de colon los datos pertenecen a arrays humano HG-U133A con 22283 transcritos. Se han utilizado 5 arrays (2 para control y 3 para PTEN-null).

```
show(rawDataepidermoid)
```

```
## AffyBatch object
## size of arrays=1164x1164 features (20 kb)
## cdf=HG-U133_Plus_2 (54675 affyids)
## number of samples=4
## number of genes=54675
## annotation=hgu133plus2
## notes=
```

Se obtiene que para el cáncer epidermoide los datos pertenecen a arrays humano HG-U133plus2 con 54675 transcritos. Se han utilizado 4 arrays (2 para control y 2 para PTEN-null).

```
show(rawDatamama)
```

```
## AffyBatch object
## size of arrays=1164x1164 features (20 kb)
## cdf=HG-U133_Plus_2 (54675 affyids)
## number of samples=4
## number of genes=54675
## annotation=hgu133plus2
## notes=
```

Se obtiene que para el cáncer de mama los datos pertenecen a arrays humano HG-U133Aplus2 con 54675 transcritos. Se han utilizado 4 arrays (2 para control y 2 para PTEN-null).

```
show(rawDatalung)
```

```
## AffyBatch object
## size of arrays=1164x1164 features (20 kb)
## cdf=HG-U133_Plus_2 (54675 affyids)
## number of samples=4
## number of genes=54675
## annotation=hgu133plus2
## notes=
```

Se obtienen que para el carcinoma pulmonar no microcítico los datos pertenecen a arrays humano HG-U133Aplus 2 con 54675 transcritos. Se han utilizado 4 arrays (2 para control y 4 para PTEN-null).

```
## AffyBatch object
## size of arrays=1002x1002 features (21 kb)
## cdf=Mouse430_2 (45101 affyids)
## number of samples=6
## number of genes=45101
```

```
## annotation=mouse4302
## notes=
```

Se obtienen que para el cáncer de próstata los datos pertenecen a arrays de ratón mouse4302 con 45101 transcritos. Se han utilizado 6 arrays (3 para control y 3 para PTEN-null).

2.2. Preprocesado: Exploración, Control de Calidad y Normalización

Los datos que proceden de la lectura de los microarrays son datos en “crudo” y deben ser pre-procesados de diversas formas antes de analizarlos. El pre-procesado engloba:

- Exploración y calidad de los datos.
- Normalización y resumen de los valores de las sondas de cada grupo.
- Filtrado no específico para eliminar el efecto de genes que no se expresen o bien no se expresan de forma distinta entre los grupos que debemos comparar.

2.2.1. Exploración de los datos

La exploración y el control de calidad se dividirán en exploraciones estadísticas estándar y en técnicas de control de calidad desarrolladas específicamente para datos de microarrays.

La exploración de los datos se suele basar en técnicas univariantes como histogramas o diagramas de cajas o en técnicas multivariantes como los análisis de conglomerados con la finalidad de obtener a grandes rasgos una idea general de cómo ha resultado el experimento⁵⁹.

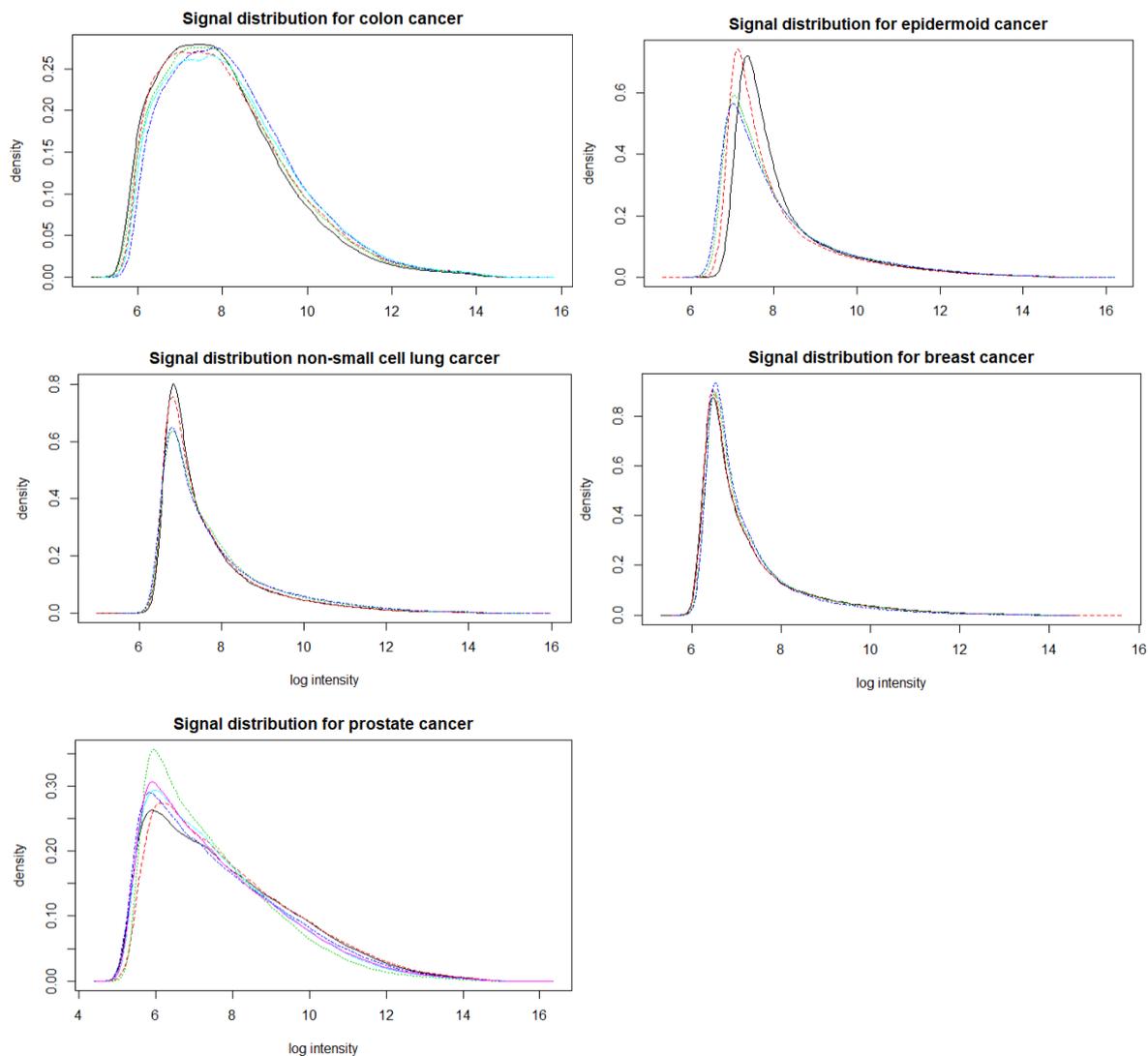


Figura 6. Gráficos de densidad. Gráficos de densidad para los *arrays* en función de cada tipo de cáncer.

Los gráficos de densidad permiten hacerse una idea de si las distribuciones de los distintos *arrays* son similares en forma y posición. Aunque los gráficos muestran cierta heterogeneidad entre *arrays*, no sugiere un problema en los datos.

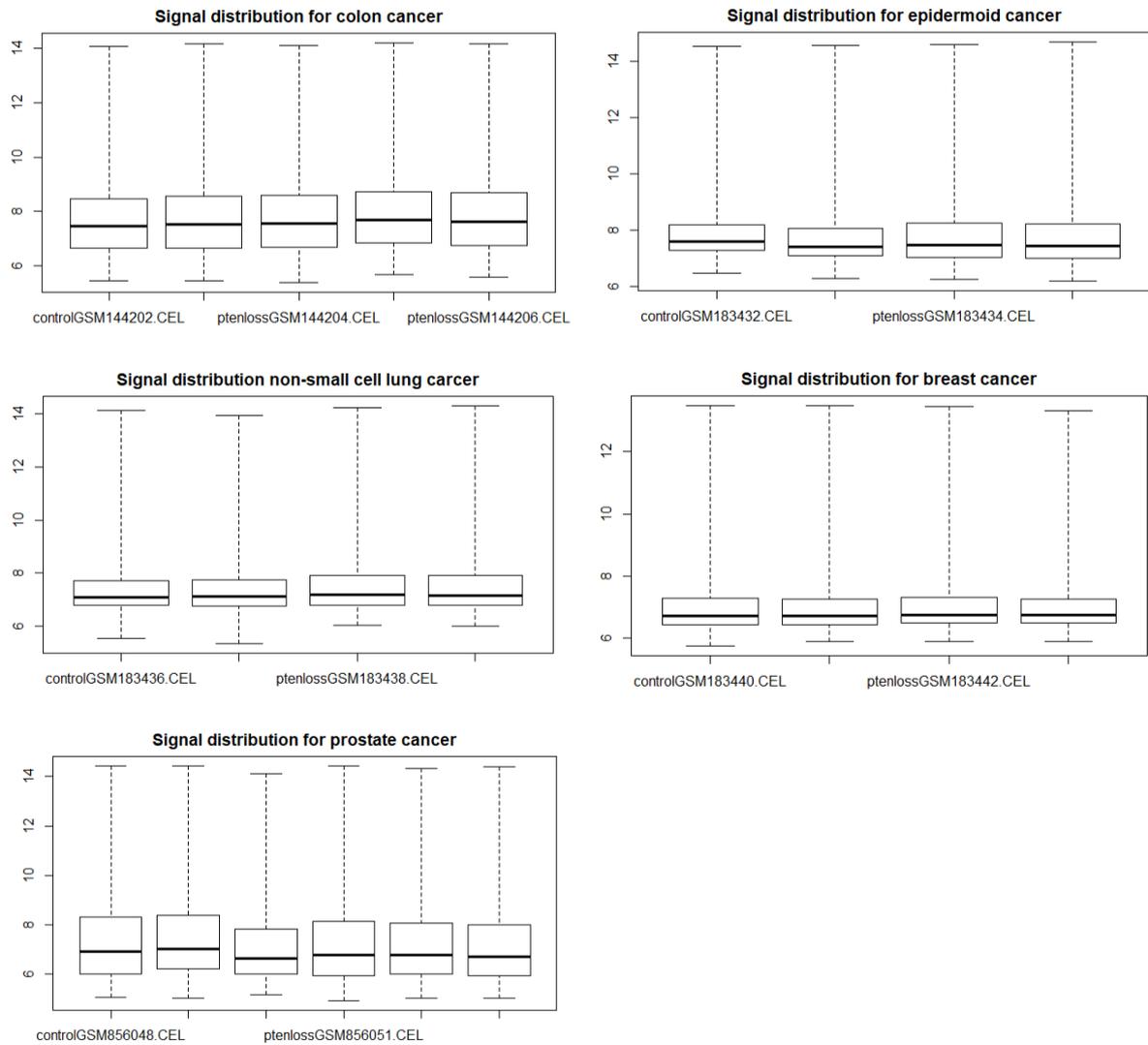


Figura 7. Boxplots o diagramas de cajas. Boxplots correspondientes a cada array de los diferentes cánceres.

Tanto el boxplot como el histograma o gráfico de densidad muestran la distribución de las señales de los datos originales, sin normalizar. Se observa que la distribución y posición siguen un mismo patrón, por lo que son similares y aunque existe heterogeneidad, lo que sugiere que es conveniente normalizar la muestra, no se muestra ninguna diferencia sistémica, por lo que no parece que existan arrays problemáticos en ninguno de los cánceres estudiados.

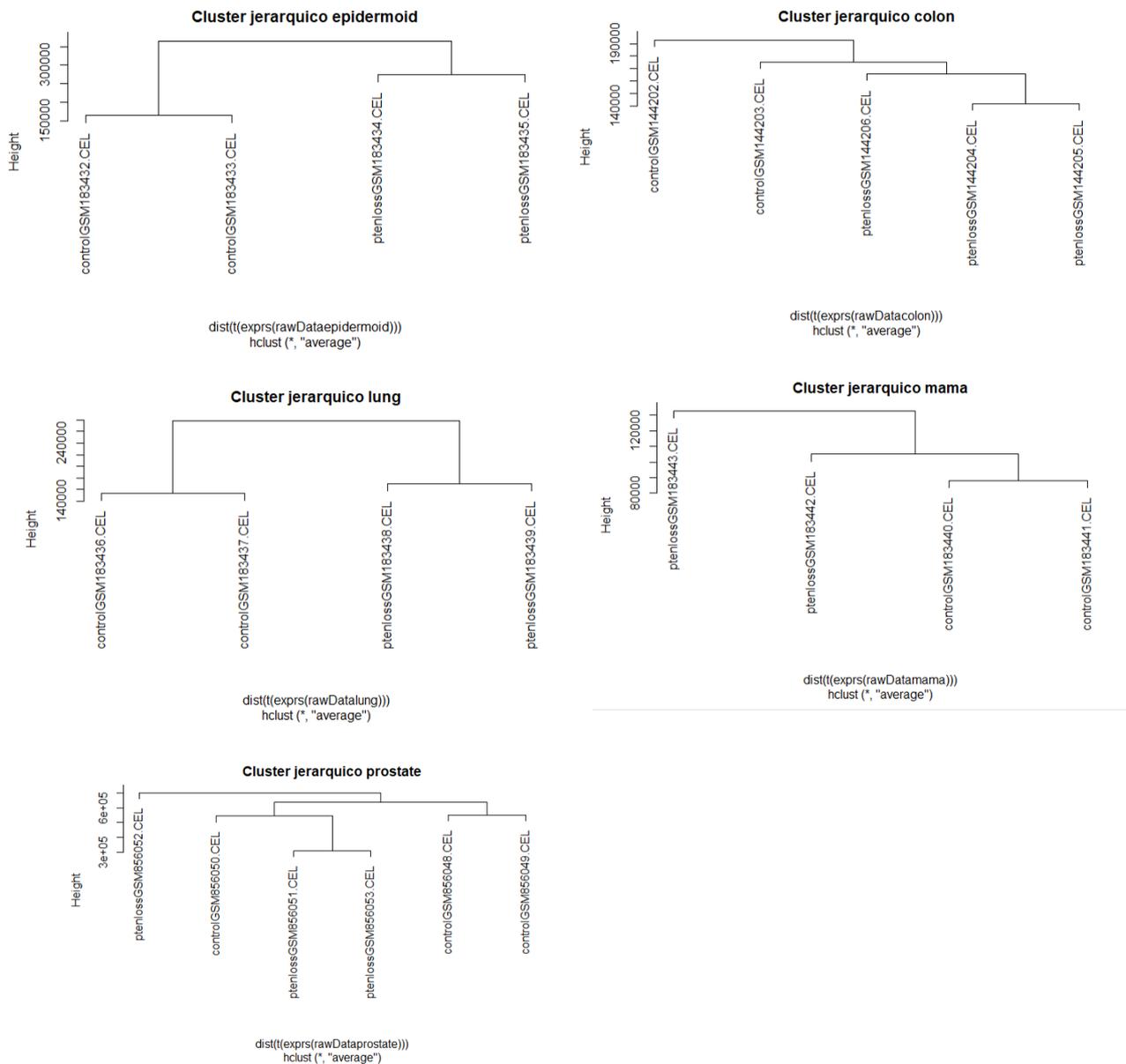


Figura 8. Dendogramas. Dendogramas resultante de realizar un clúster jerárquico de las muestras, basado en la información de todos los genes.

El dendograma o clúster jerárquico indica como se agrupan las diferentes condiciones experimentales, con el fin de evaluar de forma descriptiva si hay alguna o varias de ellas con un comportamiento o agrupación distinta de la esperada⁵⁹. En este caso, observamos que los arrays de las muestras de cáncer de colon, epidermoide, pulmón no microcítico y mama se agrupan por condiciones experimentales (control y *PTEN* loss). Los arrays de la muestra de cáncer de próstata se agrupan por condiciones experimentales, excepto el array GSM856052 (*PTEN* loss). El caso de que un array no se agrupe por condiciones experimentales, no es necesariamente indicador de problemas, puede deberse a la presencia de otra fuente de variación o bien al hecho de que se trata de un gráfico basado en todos los datos y las condiciones experimentales

pueden haber afectado un pequeño número de genes. Al comparar este dato con el diagrama de cajas o histograma anteriores, no parece que exista ningún array defectuoso.

2.2.2. Control de calidad

Las exploraciones anteriores proporcionan una idea general acerca de la distribución de los datos y la posible presencia de grupos naturales (por ejemplo, del tratamiento control o *PTEN* loss) o artificiales (por ejemplo, los días de procesado).

Entre los controles de calidad de *Affymetrix* se encuentra el gráfico de degradación que permite hacerse una idea del proceso de hibridación de las muestras.

La calidad de la hibridación del ARN a lo largo de los conjuntos de sondas puede variar si en un grupo de arrays el nivel de degradación es desigual y suele considerarse que su calidad es baja⁵⁹.

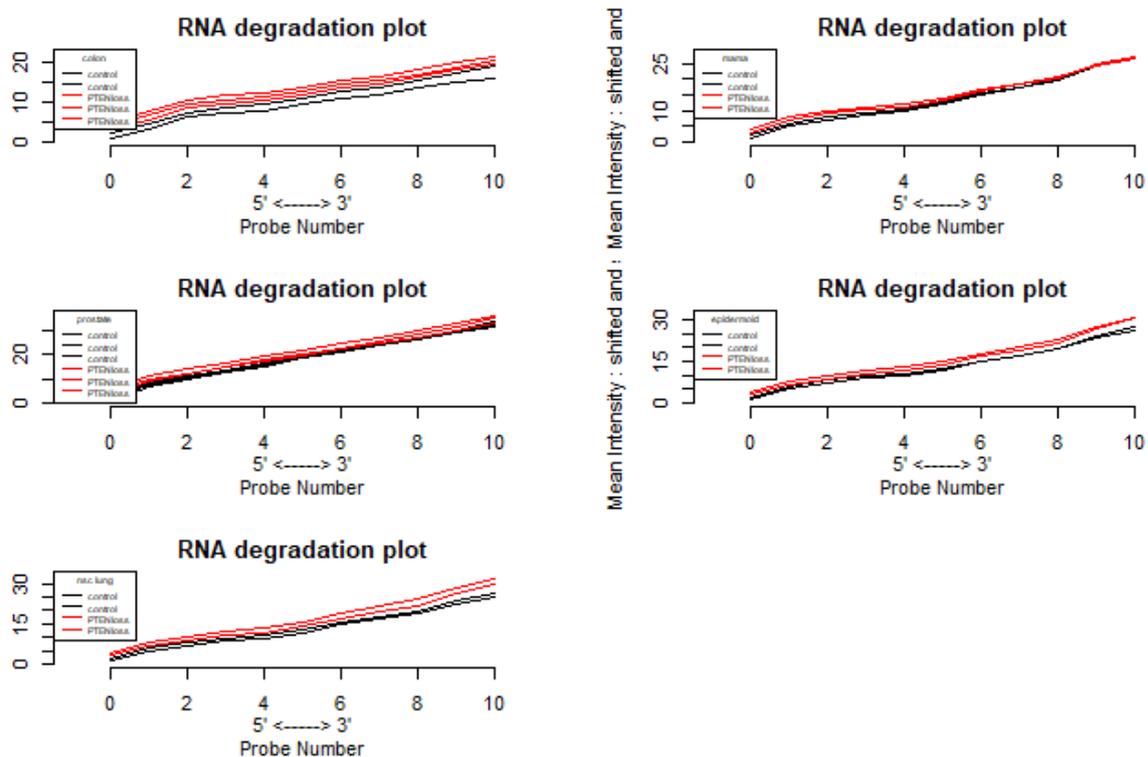


Figura 9. Gráficos de degradación de cada array. Las líneas son ascendentes y mayormente paralelas por lo que sugiere una calidad similar en todos los arrays.

El paquete *affyPLM* realiza un control de calidad basado en modelos a nivel de sonda (PLM). Los valores estimados por este modelo se comparan con los valores reales y se obtienen los errores o residuos del ajuste. Con los valores ajustados del modelo se calcula la expresión relativa en escala logarítmica (relative log expresión RLE), que es una medida estandarizada de la expresión y debería presentar una distribución similar

en todos los arrays, y el error no estandarizados y normalizado o NUSE que es el más informativo, ya que presenta la distribución de los residuos. Si un array es problemático la caja correspondiente en el boxplot es desplazada hacia arriba o hacia debajo de las demás⁵⁹.

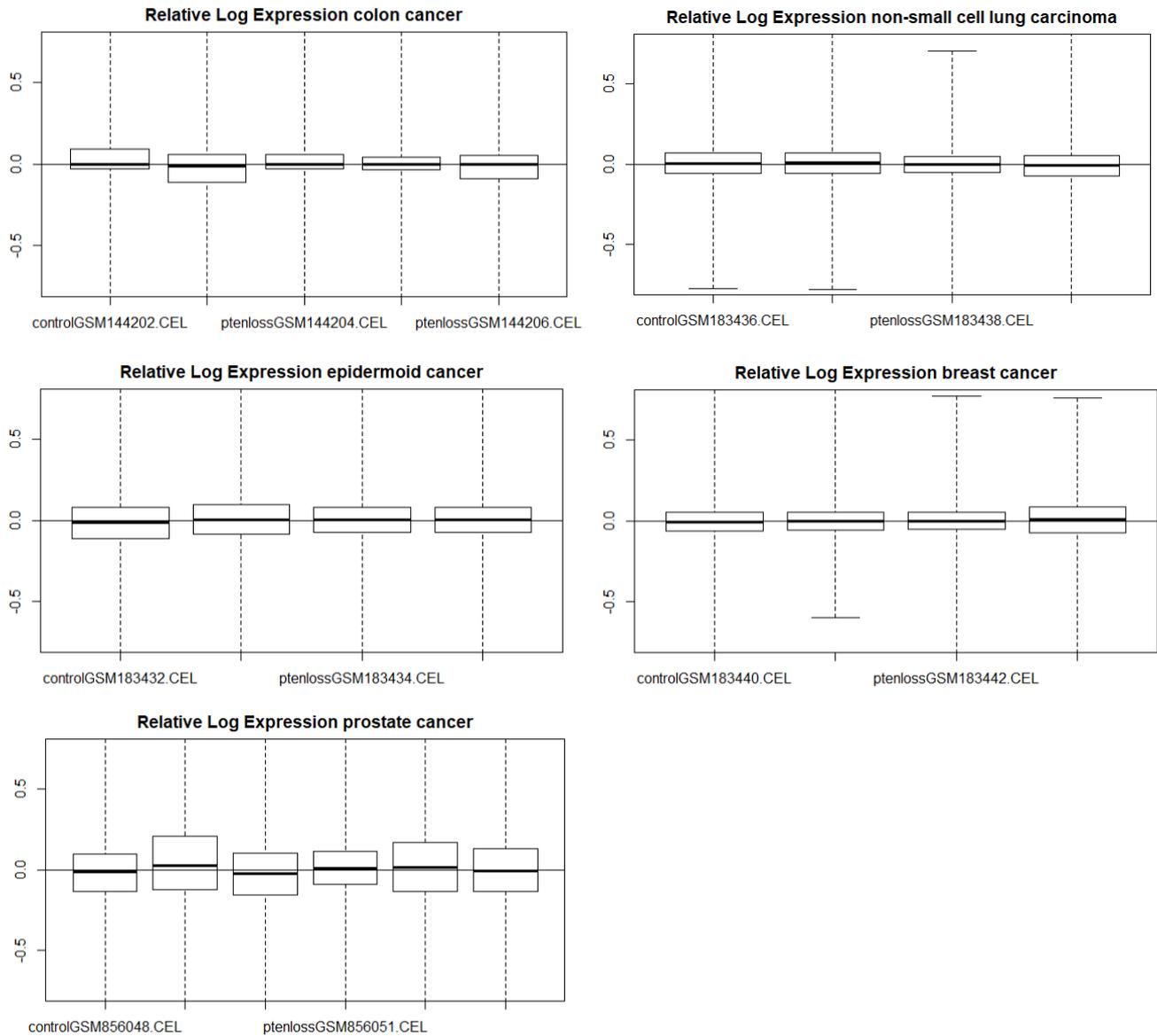


Figura 10. Gráficos de expresiones relativas. Las muestras son simétricas lo que sugiere una calidad aceptable de los datos.

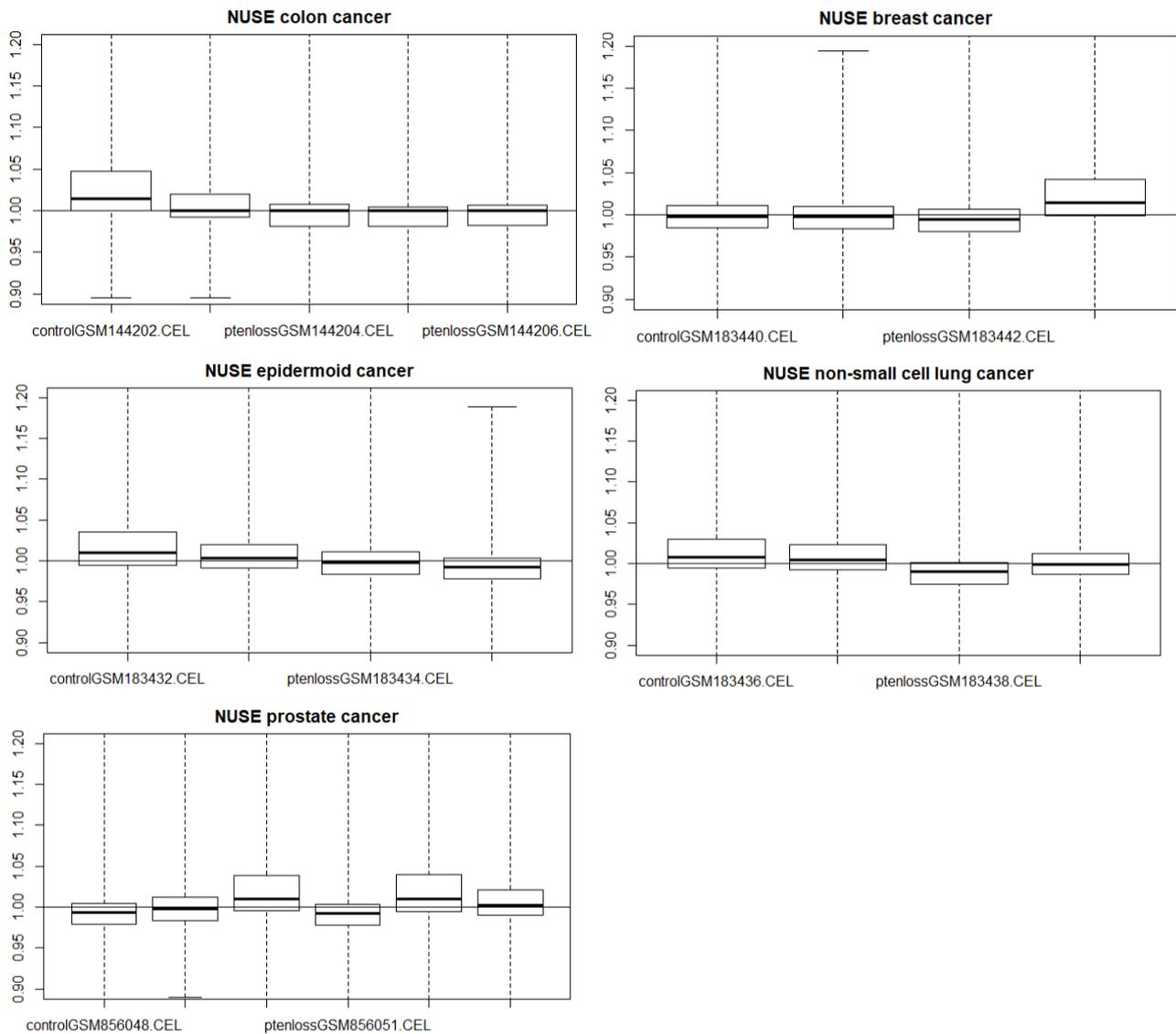


Figura 11. Gráficos del error no estandarizado y normalizado o NUSE para cada array de los diferentes cánceres. Aunque existe heterogeneidad, ningún array presenta un gran desplazamiento hacia arriba o hacia debajo de los demás, por lo que la mayor parte son simétricos y no se observan arrays problemáticos.

2.2.3. Normalización

La normalización describe las técnicas utilizadas para transformar adecuadamente los datos antes de que sean analizados con la finalidad de corregir diferencias sistemáticas⁵⁹. De esta manera, la normalización permite comparar mediciones de diferentes arrays, eliminando sesgos sistemáticos.

Existen muchos métodos de normalización, entre ellos el RMA (robust multiarray average) que se basa en los distintos valores de la misma sonda entre todos los arrays disponibles.

El método RMA ajusta el ruido de fondo (*background*), toma logaritmos en base 2 de cada intensidad ajustada por el *background*, realiza la normalización por cuantiles de los valores del paso anterior y por último, estima las intensidades de cada gen separadamente para cada conjunto de sondas. Este método se ha convertido en el estándar para muchos usuarios de Bioconductor y es el utilizado en este análisis^{59,60}.

Para realizar la normalización se utiliza la función *rma()*. Puede observarse el código entero en el Anexo: Pipeline.

2.2.4. Exploración de los datos normalizados

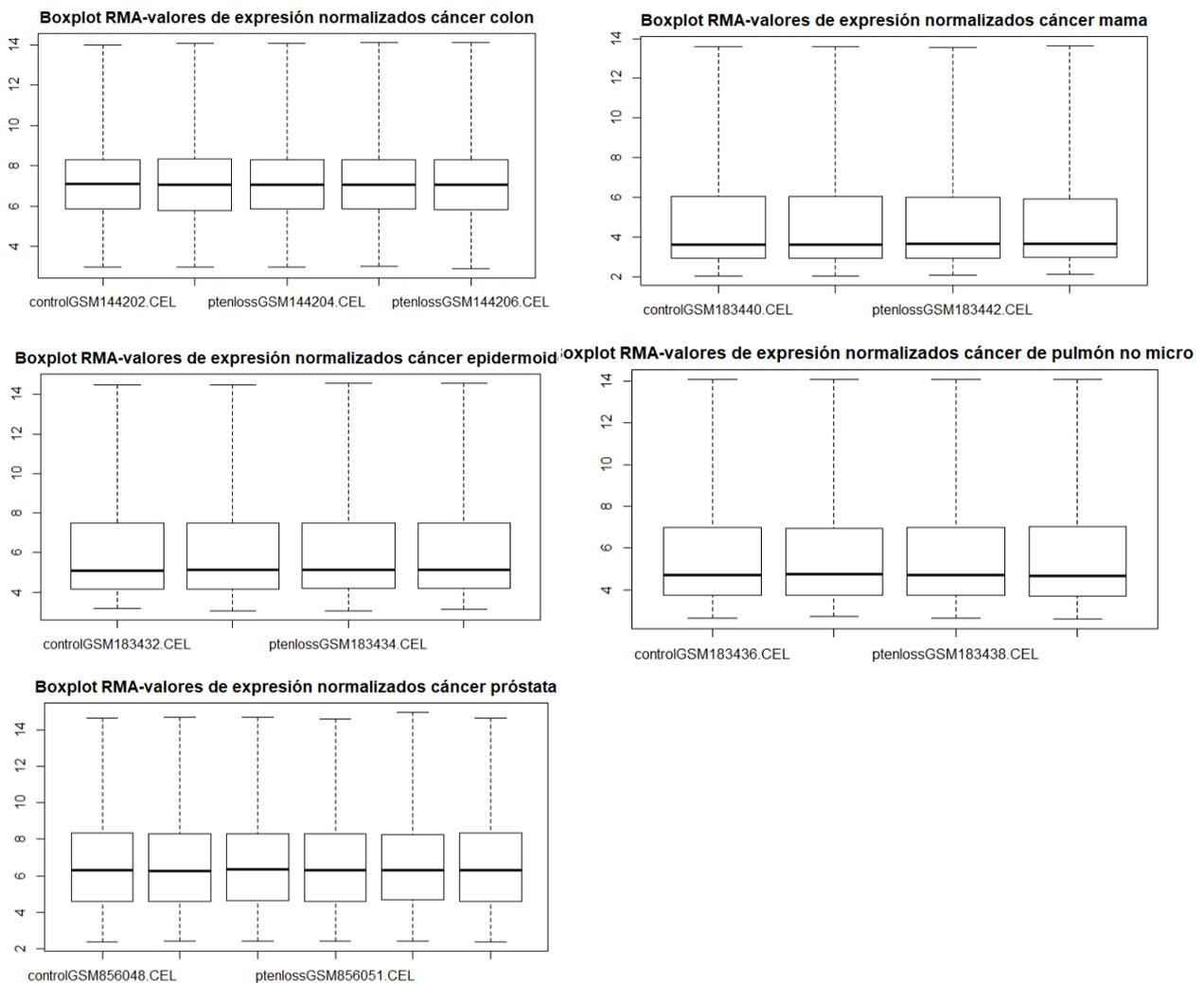


Figura 12. Diagrama de cajas. Diagrama de cajas tras la normalización de los datos. El diagrama de cajas muestra como después de la normalización la distribución de los

arrays coincide por lo que se ha transformado los valores a una escala completamente comparable.

2.2.5. Filtrado no específico

El filtrado no específico es recomendable para eliminar el ruido de fondo y limitar los ajustes posteriores a los necesarios. De esta forma pueden eliminarse genes que no presenten una variación significativa en su señal, entre distintas condiciones experimentales^{59,60}.

Para realizar el filtrado no específico se utiliza la función *nsFilter()*. Puede observarse el código entero en el Anexo: Pipeline.

De esta forma se obtiene un objeto *expressionSet* para cada tipo de cáncer almacenado en un archivo binario *tipodecancer.Rda* que será la base para el posterior estudio.

2.3. Selección de genes diferencialmente expresados

Para realizar la selección de genes se crea la matriz de diseño y los contrastes, entre la muestra control y *PTEN* loss.

Una vez definidos los contrastes se realiza el ajuste del modelo que implementa *limma*⁶¹. En este caso se utiliza el método empírico de Bayes, mejor que el tradicional análisis *t-test* ya que permite obtener estimaciones más precisas para la significancia estadística de los genes. El ajuste de las comparaciones múltiples se realiza por el método Benjamini y Hochberg^{59,60}. La función *topTable()* genera para cada contraste una lista de genes ordenados de más a menos diferencialmente expresados.

2.4. Anotación de resultados

La identificación de los genes resulta más sencilla si se utilizan nombres estándar como el símbolo del gen o *gene symbol*. Debido a que los microarrays son un producto comercial, para poder indicar el nombre de cada gen seleccionado, es preciso disponer de una tabla de anotaciones que relacione cada sonda con los genes a los que apunta. Con la finalidad de valorar los genes diferencialmente expresados se obtiene una lista con los nombres de cada gen en la columna *gene.symbols*. Los genes aparecen en orden de menor *p-valor* a mayor y por tanto de más significativo a menos.

La columna logFC permite intuir si el gen se encuentra con una regulación negativa (valor negativo) o sobreexpresado (positivo). El valor adecuado para indicar la diferencia significativa entre las dos condiciones es el valor de “adj.p.value”, ya que al haberse realizado múltiples comparaciones y se ha ajustado por el método de Benjamini y Hochberg. Además, cuanto mayor sea el valor absoluto del estadístico t mayor es la probabilidad de que el gen esté diferencialmente expresado⁶⁰.

	GeneSymbol <fctr>	GeneName <fctr>	logFC <dbl>	AveExpr <dbl>	t <dbl>	P.Value <dbl>	adj.P.Val <dbl>
219850_s_at	SEM1	SEM1, 26S proteasome complex subunit	2.1284010	6.880886	7.156885	1.101429e-07	0.0004026113
211711_s_at	LARGE1	LARGE xylosyl- and glucuronyltransferase	-1.2203092	5.973048	-7.085522	1.317542e-07	0.0004026113
204036_at	JUN	Jun proto-oncogene, AP-1 transcription f	1.1855056	8.082167	6.930635	1.947805e-07	0.0004026113
211668_s_at	TRIP4	thyroid hormone receptor interactor 4	1.2701582	9.043244	6.697772	3.524247e-07	0.0005440235
202284_s_at	HDLBP	high density lipoprotein binding protein	1.1208819	11.374354	6.612410	4.386579e-07	0.0005440235
221577_x_at	EPYC	epiphycan	0.9827011	11.304636	5.922844	2.640703e-06	0.0027291666

Figura 13. Genes diferencialmente expresados en cáncer de colon. Lista de los 6 primeros genes diferencialmente expresados entre *PTEN* loss y control en el cáncer de colon.

	GeneSymbol <fctr>	GeneName <fctr>	logFC <dbl>	AveExpr <dbl>	t <dbl>	P.Value <dbl>	adj.P.Val <dbl>
213711_at	NA	NA	-6.333746	9.686074	-67.45628	3.368451e-09	2.941982e-05
204614_at	PPP4R1L	protein phosphatase 4 regulatory subunit	-6.911917	7.900628	-61.01136	5.829747e-09	2.941982e-05
228038_at	LOC101929538	uncharacterized LOC101929538	5.251163	6.854119	51.76178	1.430629e-08	4.813112e-05
231771_at	LINC00304	long intergenic non-protein coding RNA 3	-4.505466	8.948839	-45.89404	2.758712e-08	6.712397e-05
211990_at	KDSR	3-ketodihydrosphingosine reductase	-4.881669	7.946787	-42.16736	4.378732e-08	6.712397e-05
208885_at	NA	NA	-3.799959	9.791446	-41.67329	4.669441e-08	6.712397e-05

Figura 14. Genes diferencialmente expresados en cáncer epidermoide. Lista de los 6 primeros genes diferencialmente expresados entre *PTEN* loss y control en el cáncer epidermoide.

	GeneSymbol <fctr>	GeneName <fctr>	logFC <dbl>	AveExpr <dbl>	t <dbl>	P.Value <dbl>	adj.P.Val <dbl>
204053_x_at	LOC285593	uncharacterized LOC285593	-3.160925	5.349769	-18.657559	1.985249e-14	2.003712e-10
204664_at	DLEU2	deleted in lymphocytic leukemia 2	-1.616885	5.284307	-10.598380	8.057523e-10	4.066229e-06
223315_at	NA	NA	-1.680513	4.383153	-10.263818	1.414507e-09	4.758874e-06
202627_s_at	LOC101929552	uncharacterized LOC101929552	-1.519972	5.615900	-9.758295	3.392094e-09	8.559100e-06
202718_at	LOC101928504	uncharacterized LOC101928504	1.431007	6.549679	8.826434	1.844017e-08	3.229060e-05
202391_at	MRGPRX3	MAS related GPR family member X3	-1.311167	7.114044	-8.805033	1.919584e-08	3.229060e-05

Figura 15. Genes diferencialmente expresados en cáncer de mama. Lista de los 6 primeros genes diferencialmente expresados entre *PTEN* loss y control en el cáncer de mama.

	GeneSymbol <fctr>	GeneName <fctr>	logFC <dbl>	AveExpr <dbl>	t <dbl>	P.Value <dbl>	adj.P.Val <dbl>
1417256_at	Hunk	hormonally upregulated Neu-associated ki	-6.103436	6.856694	-19.98642	2.621939e-07	0.00228245
1425002_at	Lcn5	lipocalin 5	4.532662	6.089868	18.47759	4.440132e-07	0.00228245
1436790_a_at	Serpine1	serine (or cysteine) peptidase inhibitor	-4.426665	5.685709	-14.15595	2.621873e-06	0.00898516
1418158_at	D15Ert55e	DNA segment, Chr 15, ERATO Doi 55, expre	3.952427	6.385902	12.01535	7.719421e-06	0.01853826
1449031_at	Dlg3	discs large MACUK scaffold protein 3	4.214363	6.196528	11.73356	9.015784e-06	0.01853826
1449254_at	Klk11	kallikrein related-peptidase 11	-2.969214	12.365677	-11.08584	1.305475e-05	0.01912992

Figura 16. Genes diferencialmente expresados en cáncer de próstata. Lista de los 6 primeros genes diferencialmente expresados entre *PTEN* loss y control en el cáncer de próstata.

	GeneSymbol <fctr>	GeneName <fctr>	logFC <dbl>	AveExpr <dbl>	t <dbl>	P.Value <dbl>	adj.P.Val <dbl>
204053_x_at	LOC101929319	uncharacterized LOC101929319	-2.954882	7.359783	-18.519620	6.109007e-76	6.165821e-72
211756_at	LINC01533	long intergenic non-protein coding RNA 1	1.465504	5.383686	9.184993	4.502992e-20	2.272435e-16
221760_at	LINC02223	long intergenic non-protein coding RNA 2	1.388564	6.940303	8.702776	3.482934e-18	1.171775e-14
209318_x_at	NA	NA	1.342009	4.049666	8.410995	4.332350e-17	1.093160e-13
213428_s_at	JPX	JPX transcript, XIST activator	1.203582	5.681046	7.543406	4.772185e-14	9.633132e-11
206002_at	VWA5B1	von Willebrand factor A domain containin	1.194569	4.718262	7.486918	7.340351e-14	1.234769e-10

Figura 17. Genes diferencialmente expresados en cáncer de pulmón no microcítico. Lista de los 6 primeros genes diferencialmente expresados entre *PTEN* loss y control en el cáncer de pulmón no microcítico.

Se observa que para el cáncer de colon el gen que presenta una mayor diferencia significativa entre la muestra control y *PTEN* loss es *SEMI* y se encuentra sobreexpresado ($\log_{2}FC > 0$). Para el cáncer epidermoide el gen cuyo nombre ha sido identificado que presenta una mayor diferencia significativa entre tratamiento y control es *PPP4R1L* y se encuentra inhibido ($\log_{2}FC < 0$). Para el cáncer de mama el gen cuyo nombre ha sido identificado es *DLEU2* (deleted in lymphocytic leukemia 2) y se encuentra inhibido ($\log_{2}FC < 0$). Para el cáncer de próstata el gen con mayor diferencia significativa entre las 2 condiciones es *Hunk* (hormonally upregulated Neu-associated kinase) y se encuentra inhibido ($\log_{2}FC < 0$) y para el cáncer de pulmón no microcítico el gen es *LINC* (long intergenic non-protein coding RNA) y se encuentra inhibido ($\log_{2}FC < 0$).

2.5. Genes sobreexpresados y downregulados o inhibidos.

Como curiosidad y con la finalidad de saber cuántos genes obtenidos del apartado anterior se encuentran sobreexpresados y cuántos downregulados se procede a utilizar la función *decideTests()*. Se tiene en cuenta el ajuste de Benjamini y Hochberg con un p-valor < 0.05 :

```
#cáncer de colon
require(annotate)
fit.Symbolsc <- getSYMBOL (rownames(fitcolon), "hgu133a.db")
resc<-decideTests(fitcolon, method="separate",
adjust.method="BH", p.value=0.05)
sum.res.rowsc<-apply(abs(resc), 1, sum)
res.selectedc<-resc[sum.res.rowsc!=0, ]
print(summary(resc))

##          (Intercept) groupscT
## Down              0          7
## NotSig            0         6178
## Up                6201         16
```

En el cáncer de colon se obtiene en el grupo *PTEN* loss comparado con el control, 16 genes sobreexpresados y 7 con regulación negativa con un p-valor < 0.05 .

```
#cáncer de mama
require(annotate)
fit.Symbolsm <- getSYMBOL (rownames(fitmama), "hgu133plus2.db")
resm<-decideTests(fitmama, method="separate",
adjust.method="BH", p.value=0.05)
sum.res.rowsm<-apply(abs(resm), 1, sum)
res.selectedm<-resm[sum.res.rowsm!=0, ]
print(summary(resm))

##          (Intercept) groupsmT
## Down              0          446
## NotSig            0         9554
## Up                10093         93
```

En el cáncer de mama se obtienen 446 genes inhibidos y 93 sobreexpresados al comparar *PTEN* loss con control, con un p-valor<0.05.

```
#cáncer de próstata
require(annotate)
fit.Symbolsp <- getSYMBOL (rownames(fitprostate), "mouse4302.db")
res<-decideTests(fitprostate, method="separate",
adjust.method="BH",p.value=0.05)
sum.res.rowsp<-apply(abs(res),1,sum)
res.selectedp<-res[sum.res.rowsp!=0,]
print(summary(res))

##          (Intercept) groupspT
## Down              0         97
## NotSig            17       10141
## Up                10264        43
```

En el cáncer de próstata se obtienen con un p-valor <0.05, 97 genes inhibidos y 43 sobreexpresados al comparar el grupo *PTEN* loss con control.

```
#cáncer epidermoide
require(annotate)
fit.Symbolse <- getSYMBOL (rownames(fitepidermoid), "hgu133plus2.db")
rese<-decideTests(fitepidermoid, method="separate",
adjust.method="BH",p.value=0.01)
sum.res.rowse<-apply(abs(rese),1,sum)
res.selectede<-rese[sum.res.rowse!=0,]
print(summary(rese))

##          (Intercept) groupseT
## Down              0         3256
## NotSig            0         3440
## Up                10093        3397
```

En el cáncer epidermoide se obtienen al comparar el grupo con la delección de *PTEN* con el grupo control, 3256 genes inhibidos y 3397 genes sobreexpresados. Se observa en comparación con los demás cánceres un mayor número tanto de genes sobreexpresados como inhibidos.

```
#cáncer de pulmón no microcítico
require(annotate)
fit.Symbolsl <- getSYMBOL (rownames(fitlung), "hgu133plus2.db")
resl<-decideTests(fitlung, method="separate",
adjust.method="BH",p.value=0.01)
sum.res.rowsl<-apply(abs(resl),1,sum)
res.selectedl<-resl[sum.res.rowsl!=0,]
print(summary(resl))

##          (Intercept) groupslT
## Down              0         106
## NotSig            0         9739
## Up                10093        248
```

En el cáncer de pulmón no microcítico se obtienen con un p-valor <0.05 , 106 genes inhibidos y 248 genes sobreexpresados en el grupo *PTEN* loss comparado con el grupo control.

2.6. Visualización de los perfiles de expresión.

Tras la selección de los genes diferencialmente expresados, puede visualizarse las expresiones de cada gen agrupándose para destacar los genes que se encuentran sobreexpresados o inhibidos simultáneamente, constituyendo perfiles de expresión.

Los mapas de color o Heatmaps convierten los valores numéricos de una matriz de datos en tonos de color, de manera que sea posible obtener una visión de conjunto de cómo varían los distintos valores entre muestras y entre genes. Además, se suele hacer una agrupación jerárquica de las columnas (muestras), las filas (genes) o ambos a la vez, con lo cual, los valores parecidos tenderán a estar agrupados, lo que ayudará a visualizar la variación conjunta⁵⁹.

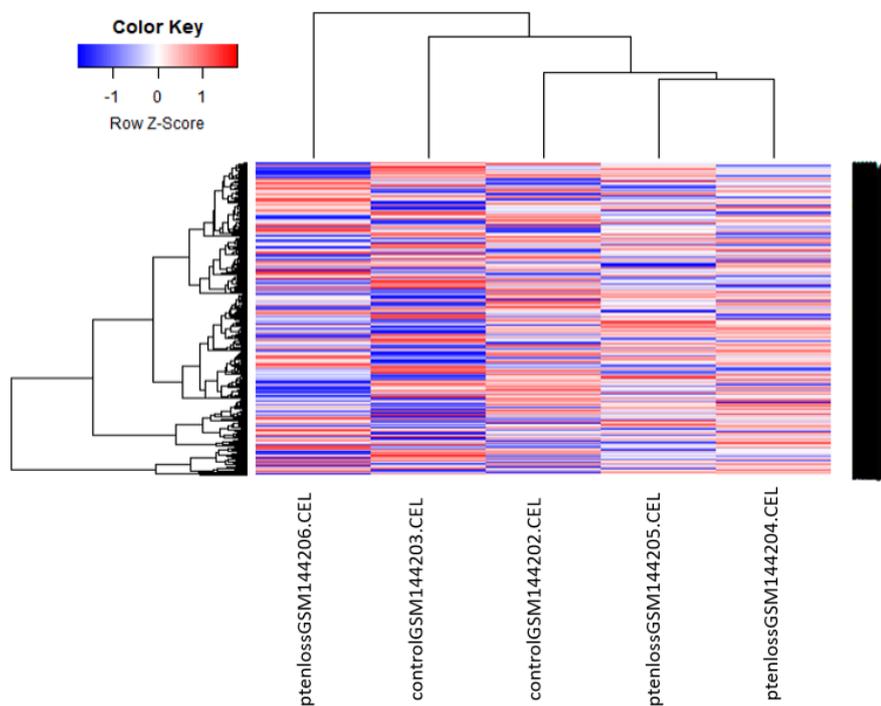


Figura 18. HeatMap para los genes seleccionados como diferencialmente expresados en el cáncer de colon. Se observa que las muestras de *PTEN* loss, excepto la GSM144206, tienen perfiles de expresión más parecidos entre ellos que con las muestras control. Se observa que el perfil de expresión de *PTEN*lossGSM144206 es parecido a la muestra control GSM144203.

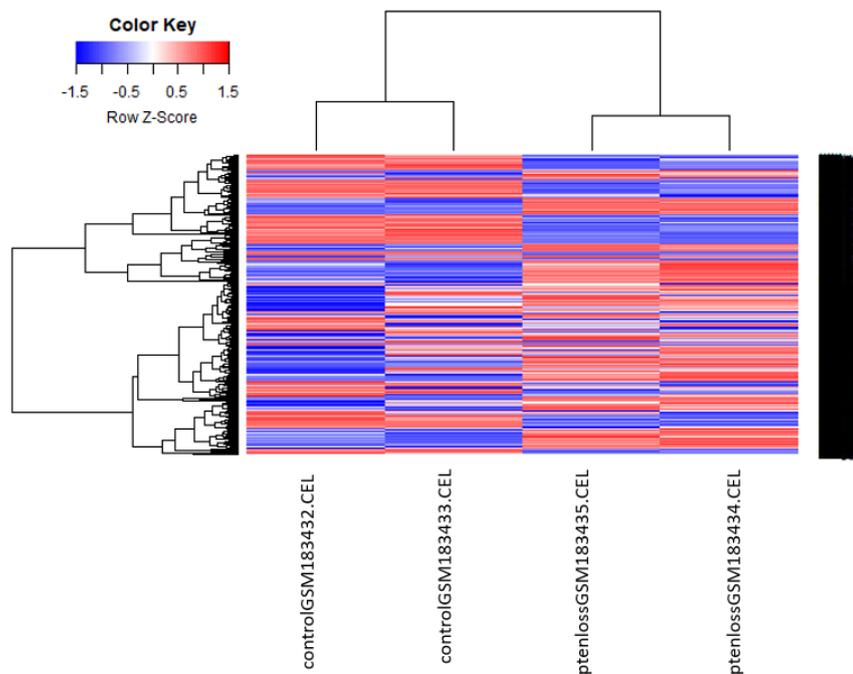


Figura 19. HeatMap para los genes seleccionados como diferencialmente expresados en el carcinoma epidermoide. Se observa que los perfiles de expresión para las muestras de cáncer epidermoide con *PTEN* loss son similares entre sí y difieren de las muestras control, las cuales también tienen perfiles de expresión parecidos.

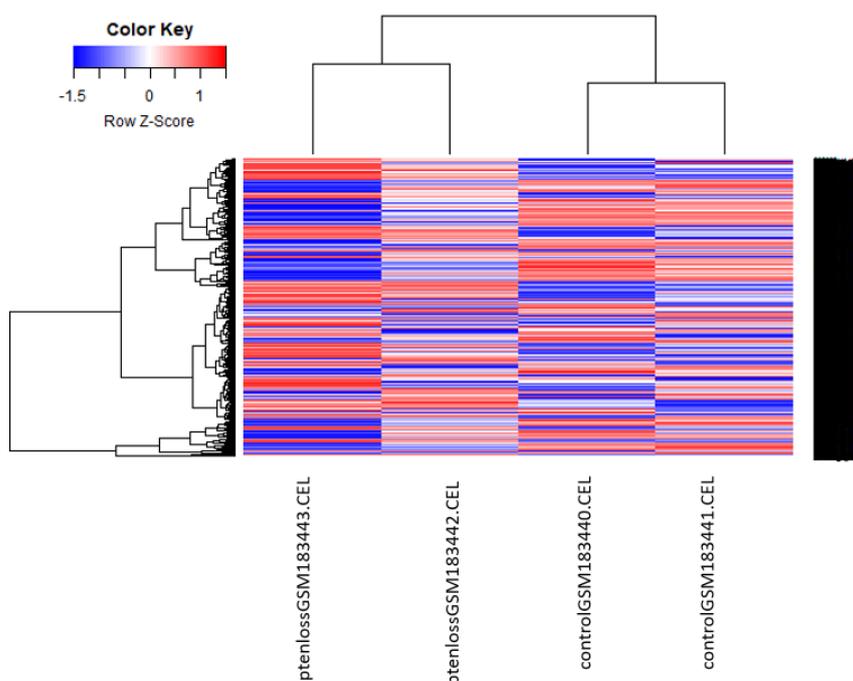


Figura 20. HeatMap para los genes seleccionados como diferencialmente expresados en el cáncer de mama. Se observa que las muestras control tienen perfiles de expresión similares y diferentes a las muestras *PTEN* loss.

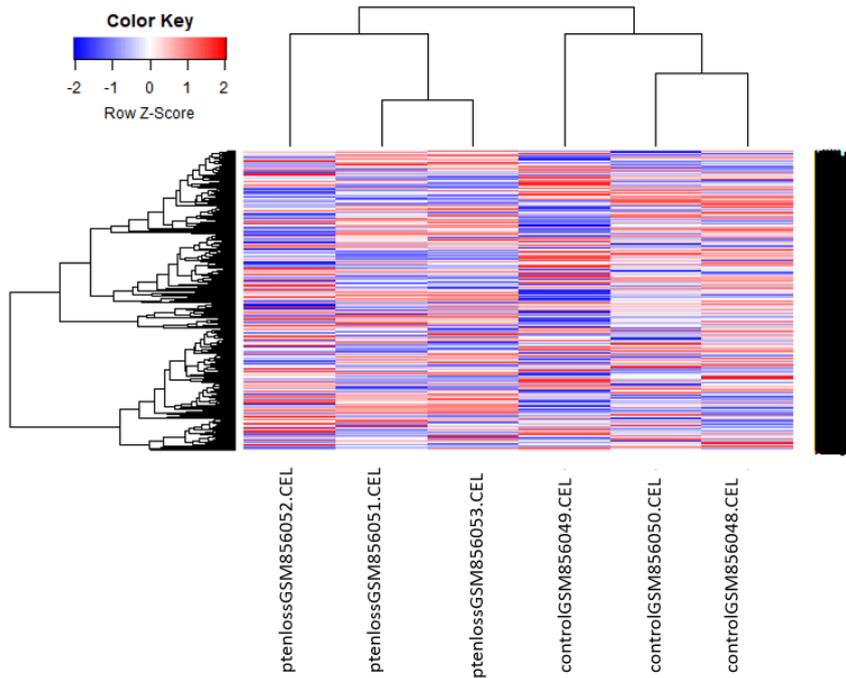


Figura 21. HeatMap para los genes seleccionados como diferencialmente expresados en el cáncer de próstata. Se observa que los perfiles de expresión para las muestras de cáncer de próstata con *PTEN* loss son similares entre sí y difieren de las muestras control, las cuales también tienen perfiles de expresión parecidos.

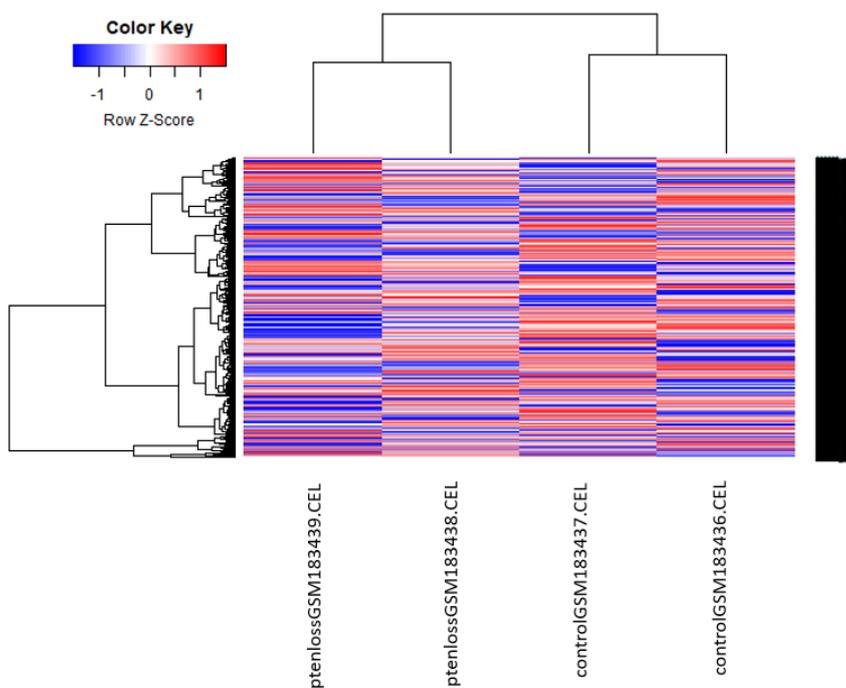


Figura 22. HeatMap para los genes seleccionados como diferencialmente expresados en el cáncer de pulmón no microcítico. Se observa que las muestras de cáncer con *PTEN*

loss tienen perfiles de expresión parecidos y difieren de los controles, las cuales también presentan entre ellas perfiles de expresión parecidos.

2.7. Análisis de significación biológica.

Se estudiaron las funciones de los genes buscando sus anotaciones en bases de datos de anotación funcional como la Gene Ontology (GO)⁶² o la Kyoto Encyclopedia of Genes and Genomes (KEGG)⁶³.

2.7.1. Análisis de enriquecimiento

El objetivo del análisis de enriquecimiento es establecer si una determinada categoría que representa un proceso biológico (GO) o una vía metabólica o *pathway* (KEGG) aparece con mayor o menor frecuencia en la lista de genes seleccionados, por lo que será probable que tenga algo que ver con el proceso en base al cual se ha seleccionado la lista que se estudia^{59,60}.

2.7.2. Ontología Génica (GO)

Puede observarse el análisis realizado en el *Anexo: Pipeline*. La librería GOSTATS tiene algunas herramientas para realizar el análisis de enriquecimiento. Antes del análisis, los ID de las sondas del array deben convertirse a EntrezIDs que son identificadores específicos de genes usados por NCBI para entrelazar diferentes bases de datos.

La ontología genética (GO) describe nuestro conocimiento del dominio biológico con respecto a tres aspectos:

- Procesos biológicos (BP): Los procesos más grandes, o “programas biológicos” realizados por múltiples actividades moleculares. Debe tenerse en cuenta que un proceso biológico no es equivalente a una vía⁶⁴.
- Funciones moleculares (MF): Actividades a nivel molecular realizadas por productos genéticos. Los términos de función molecular describen actividades que ocurren a nivel molecular, como "catálisis" o "transporte". Los términos de función molecular de GO representan actividades en lugar de las moléculas o complejos que realizan las acciones, y no especifican dónde, cuándo o en qué contexto tiene lugar la acción. Las funciones moleculares generalmente corresponden a actividades que pueden realizarse por productos génicos individuales (es decir, una proteína o ARN), pero algunas actividades se realizan por complejos moleculares compuestos de múltiples productos génicos⁶⁴.
- Componentes celulares (CC): Las ubicaciones relativas a las estructuras celulares en las que un producto génico desempeña una función, ya sea compartimentos celulares (por ejemplo, mitocondria), o complejos macromoleculares estables de los que forman parte (por ejemplo, el ribosoma).

A diferencia de los otros aspectos de GO, las clases de componentes celulares se refieren no a procesos, sino a una anatomía celular⁶⁴.

Gene to GO BP test for over-representation

GOBPID	Pvalue	OddsRatio	ExpCount	Count	Size	Term
GO:0070482	0.000	16.270	1	7	305	response to oxygen levels
GO:0007626	0.000	19.603	0	5	164	locomotory behavior
GO:0014003	0.000	53.194	0	3	35	oligodendrocyte development
GO:0042981	0.000	6.388	2	10	1233	regulation of apoptotic process
GO:0043067	0.000	6.307	3	10	1247	regulation of programmed cell death
GO:0007420	0.000	8.206	1	7	583	brain development
GO:0010941	0.000	5.793	3	10	1344	regulation of cell death
GO:0060322	0.000	7.671	1	7	621	head development
GO:0045475	0.000	135.381	0	2	10	locomotor rhythm

Figura 23. Ejemplo Ontología Génica de los Procesos biológicos. Ejemplo de la lista de Ontología Génica de los Procesos biológicos (BP) en el cáncer de colon al comparar los *arrays* con delección de *PTEN* con control. Se observan procesos de la regulación de la muerte celular o apoptosis.

Gene to GO MF test for over-representation

GOMFID	Pvalue	OddsRatio	ExpCount	Count	Size	Term
GO:0001965	0.001	63.206	0	2	20	G-protein alpha-subunit binding
GO:0004083	0.002	Inf	0	1	1	bisphosphoglycerate 2-phosphatase activity
GO:0019912	0.002	Inf	0	1	1	cyclin-dependent protein kinase activating kinase activity
GO:0051800	0.002	Inf	0	1	1	phosphatidylinositol-3,4-bisphosphate 3-phosphatase activity
GO:0016314	0.004	542.571	0	1	2	phosphatidylinositol-3,4,5-trisphosphate 3-phosphatase activity
GO:0034416	0.004	542.571	0	1	2	bisphosphoglycerate phosphatase activity
GO:0051717	0.004	542.571	0	1	2	inositol-1,3,4,5-tetrakisphosphate 3-phosphatase activity
GO:0003960	0.006	271.262	0	1	3	NADPH:quinone reductase activity
GO:0106017	0.006	271.262	0	1	3	phosphatidylinositol-3,4-bisphosphate phosphatase activity
GO:0042803	0.007	4.815	1	5	661	protein homodimerization activity

Figura 24. Ejemplo Ontología Génica de las Funciones moleculares. Ejemplo de la lista de Ontología Génica de las Funciones moleculares (MF) en el cáncer de colon al comparar los *arrays* con *PTEN* loss con control.

Gene to GO CC test for over-representation

GOCCID	Pvalue	OddsRatio	ExpCount	Count	Size	Term
GO:0031968	0.000	15.311	0	4	163	organelle outer membrane
GO:0019867	0.000	15.119	0	4	165	outer membrane
GO:0005741	0.003	12.410	0	3	143	mitochondrial outer membrane
GO:0070557	0.004	532.818	0	1	2	PCNA-p21 complex
GO:0035749	0.012	106.527	0	1	6	myelin sheath adaxonal region
GO:0043220	0.017	66.562	0	1	9	Schmidt-Lanterman incisure
GO:0043218	0.023	48.397	0	1	12	compact myelin
GO:0009898	0.024	9.132	0	2	123	cytoplasmic side of plasma membrane
GO:0031089	0.025	44.360	0	1	13	platelet dense granule lumen
GO:0099524	0.029	38.016	0	1	15	postsynaptic cytosol

Figura 25. Ejemplo Ontología Génica de Componentes celulares. Ejemplo de la lista de Ontología Génica de Componentes celulares (CC) en el cáncer de colon al comparar los *arrays* con delección de *PTEN* con controles.

2.7.3. Vías de señalización o *pathways* (KEGG).

Tras analizar los procesos biológicos que son estadísticamente significativos en los *arrays* en los que existe la delección de *PTEN* se analizaron las vías de señalización implicadas con la finalidad de encontrar alguna vía que sea común en los diferentes tipos de cánceres, de acuerdo con el objetivo del trabajo.

Gene to KEGG test for over-representation

KEGGID	Pvalue	OddsRatio	ExpCount	Count	Size	Term
04115	0.000	51.016	0	5	61	p53 signaling pathway
05214	0.011	14.577	0	2	59	Glioma
05218	0.014	12.761	0	2	67	Melanoma
05215	0.023	9.954	0	2	85	Prostate cancer
05222	0.023	9.954	0	2	85	Small cell lung cancer
00360	0.041	27.458	0	1	15	Phenylalanine metabolism
05200	0.050	4.281	1	3	306	Pathways in cancer

Figura 26. Vías KEGG cáncer de colon. Vías implicadas en el cáncer de colon con delección en el gen *PTEN* al comparar con muestras control con un p-valor<0.005.

Gene to KEGG test for over-representation

KEGGID	Pvalue	OddsRatio	ExpCount	Count	Size	Term
03008	0.000	3.573	25	46	68	Ribosome biogenesis in eukaryotes
04144	0.000	1.830	72	99	192	Endocytosis
04115	0.000	2.684	25	41	67	p53 signaling pathway
05222	0.000	2.319	32	49	85	Small cell lung cancer
05200	0.000	1.513	120	150	321	Pathways in cancer
04210	0.001	1.998	32	47	87	Apoptosis
03013	0.002	1.704	51	68	136	RNA transport
04120	0.002	1.703	49	65	130	Ubiquitin mediated proteolysis
05211	0.004	1.960	26	37	69	Renal cell carcinoma
05219	0.005	2.389	15	24	41	Bladder cancer
04141	0.005	1.541	60	76	160	Protein processing in endoplasmic reticulum
05212	0.008	1.848	26	36	69	Pancreatic cancer
00270	0.008	2.366	13	21	36	Cysteine and methionine metabolism
04142	0.010	1.585	43	56	116	Lysosome
04130	0.013	2.253	13	20	35	SNARE interactions in vesicular transport
00520	0.013	2.013	17	25	46	Amino sugar and nucleotide sugar metabolism
04320	0.018	2.625	9	14	23	Dorso-ventral axis formation
04330	0.019	1.921	18	25	47	Notch signaling pathway
00471	0.019	Inf	1	4	4	D-Glutamine and D-glutamate metabolism
04110	0.020	1.487	46	57	122	Cell cycle
00310	0.022	1.943	16	23	43	Lysine degradation
00330	0.024	1.831	19	26	50	Arginine and proline metabolism
01100	0.026	1.150	394	422	1056	Metabolic pathways
03018	0.027	1.643	26	34	69	RNA degradation
00051	0.030	2.005	13	19	35	Fructose and mannose metabolism
05220	0.032	1.600	27	35	72	Chronic myeloid leukemia
05145	0.042	1.407	44	54	119	Toxoplasmosis
00561	0.045	1.689	19	25	50	Glycerolipid metabolism
00071	0.048	1.772	15	21	41	Fatty acid degradation
00564	0.049	1.517	28	35	74	Glycerophospholipid metabolism
03022	0.049	1.912	12	17	32	Basal transcription factors

Figura 27. Vías KEGG cáncer epidermoide. Vías implicadas en el cáncer epidermoide con delección en el gen *PTEN* al comparar con muestras control con un p-valor<0.05.

Gene to KEGG test for over-representation

KEGGID	Pvalue	OddsRatio	ExpCount	Count	Size	Term
04115	0.001	4.377	2	9	67	p53 signaling pathway
05222	0.001	3.769	3	10	85	Small cell lung cancer
05200	0.002	2.132	11	22	321	Pathways in cancer
02010	0.004	4.517	2	6	43	ABC transporters
04142	0.008	2.651	4	10	116	Lysosome
04510	0.021	1.999	7	13	196	Focal adhesion
03320	0.031	2.727	2	6	67	PPAR signaling pathway
04710	0.032	4.855	1	3	20	Circadian rhythm
00600	0.034	3.559	1	4	35	Sphingolipid metabolism
05100	0.035	2.639	2	6	69	Bacterial invasion of epithelial cells
04370	0.039	2.557	3	6	71	VEGF signaling pathway
04666	0.040	2.341	3	7	90	Fc gamma R-mediated phagocytosis
00910	0.041	4.342	1	3	22	Nitrogen metabolism
05223	0.042	2.820	2	5	54	Non-small cell lung cancer

Figura 28. Vías KEGG cáncer de mama. Vías implicadas en el cáncer de mama con delección en el gen *PTEN* al comparar con muestras control con un p-valor<0.05.

Gene to KEGG test for over-representation

KEGGID	Pvalue	OddsRatio	ExpCount	Count	Size	Term
04512	0.000	11.513	1	7	83	ECM-receptor interaction
04510	0.000	5.363	2	8	195	Focal adhesion
04350	0.001	7.832	1	5	81	TGF-beta signaling pathway
00980	0.002	8.339	1	4	60	Metabolism of xenobiotics by cytochrome P450
05146	0.003	5.805	1	5	107	Amoebiasis
00982	0.003	7.402	1	4	67	Drug metabolism - cytochrome P450
04060	0.004	3.976	2	7	221	Cytokine-cytokine receptor interaction
04974	0.005	6.465	1	4	76	Protein digestion and absorption
05144	0.007	8.373	0	3	44	Malaria
00532	0.015	11.829	0	2	21	Glycosaminoglycan biosynthesis - chondroitin sulfate / dermatan sulfate
04145	0.041	3.304	1	4	143	Phagosome
05219	0.054	5.740	0	2	41	Bladder cancer
00480	0.079	4.560	0	2	51	Glutathione metabolism
00920	0.088	12.239	0	1	10	Sulfur metabolism
04612	0.093	4.133	1	2	56	Antigen processing and presentation
04115	0.122	3.481	1	2	66	p53 signaling pathway
04610	0.128	3.374	1	2	68	Complement and coagulation cascades
00770	0.129	7.860	0	1	15	Pantothenate and CoA biosynthesis

Figura 29. Vías KEGG cáncer de próstata. Vías implicadas en el cáncer de próstata con delección en el gen *PTEN* al comparar con muestras control con un p-valor<0.13. Se observa que para un p-valor<0.05 no se obtiene ninguna de las vías de señalización que se encuentran en el cáncer de mama, colon y epidermoide, pero con un p-valor=0.122 se obtiene la vía de señalización de p53.

Gene to KEGG test for over-representation

KEGGID	Pvalue	OddsRatio	ExpCount	Count	Size	Term
04940	0.011	5.035	1	4	39	Type I diabetes mellitus
05323	0.012	3.420	2	6	84	Rheumatoid arthritis
04964	0.015	6.572	1	3	23	Proximal tubule bicarbonate reclamation
04672	0.017	4.402	1	4	44	Intestinal immune network for IgA production
05212	0.020	3.452	2	5	69	Pancreatic cancer
05310	0.020	5.711	1	3	26	Asthma
04520	0.022	3.347	2	5	71	Adherens junction
05150	0.023	3.998	1	4	48	Staphylococcus aureus infection
04060	0.026	2.072	6	11	250	Cytokine-cytokine receptor interaction
04974	0.027	3.153	2	5	75	Protein digestion and absorption
03013	0.035	2.409	3	7	136	RNA transport
05330	0.038	4.373	1	3	33	Allograft rejection
05332	0.045	4.098	1	3	35	Graft-versus-host disease
00900	0.045	6.692	0	2	15	Terpenoid backbone biosynthesis
05214	0.057	2.923	1	4	64	Glioma
04612	0.057	2.923	1	4	64	Antigen processing and presentation
05416	0.060	2.875	1	4	65	Viral myocarditis
00100	0.062	5.434	0	2	18	Steroid biosynthesis
05140	0.063	2.828	2	4	66	Leishmaniasis
04610	0.066	2.782	2	4	67	Complement and coagulation cascades
04144	0.071	1.926	4	8	192	Endocytosis
04514	0.071	2.167	3	6	128	Cell adhesion molecules (CAMs)
04510	0.078	1.884	4	8	196	Focal adhesion
05320	0.086	3.043	1	3	46	Autoimmune thyroid disease
00471	0.088	14.406	0	1	4	D-Glutamine and D-glutamate metabolism
04350	0.109	2.300	2	4	80	TGF-beta signaling pathway
04512	0.116	2.240	2	4	82	ECM-receptor interaction
05223	0.124	2.562	1	3	54	Non-small cell lung cancer
05222	0.128	2.156	2	4	85	Small cell lung cancer
05145	0.134	1.919	3	5	119	Toxoplasmosis
04012	0.136	2.103	2	4	87	ErbB signaling pathway
00232	0.149	7.199	0	1	7	Caffeine metabolism

Figura 30. Vías KEGG en el cáncer de pulmón no microcítico. Vías implicadas en el cáncer de pulmón no microcítico con delección en el gen *PTEN* al comparar con muestras control con un p -valor <0.15 . Se observa que para un p -valor <0.05 no se obtiene ninguna de las vías que se encuentran en el cáncer de mama, colon y epidermoide, pero con un p -valor $=0.128$ se obtiene la vía de señalización “small cell lung cancer”.

De esta forma se obtienen 3 vías de señalización en común en los cánceres de mama, colon y epidermoide:

- KEGGID: 05222. “Small cell lung cancer” o vía metabólica del cáncer de pulmón de células pequeñas.
- KEGGID 04115. “p53 signalling pathway” o vía de señalización de p53.
- KEGGID 05200. “Pathways in cancer” o vías metabólicas implicadas en el cáncer.

Debe tenerse en cuenta que para el cáncer de pulmón no microcítico y el cáncer de próstata no se ha encontrado una vía en común con los otros cánceres a un p -valor significativo (p -valor <0.05).

Se procede a encontrar los genes implicados en estas 3 vías que son las comunes en el mayor número de cánceres estudiados.

2.8. Genes implicados en las vías metabólicas seleccionadas

Para la búsqueda de los genes implicados en las vías metabólicas seleccionadas se utiliza la librería *globaltest* y la función *gtKEGG()* indicando el código de la vía de la que queremos extraer los genes en cuestión y el objeto *expressionSet* para los datos normalizados:

	alias <chr>	p-value <dbl>	Statistic <dbl>	Expected <dbl>	Std.dev <dbl>	#Cov <dbl>
201731_s_at	TPR	0.001289021	97.91158	25	30.61862	1
216252_x_at	FAS	0.001397613	97.79639	25	30.61862	1
202284_s_at	CDKN1A	0.002506561	96.75408	25	30.61862	1
211711_s_at	PTEN	0.003847820	95.69001	25	30.61862	1
202871_at	TRAF4	0.004032833	95.55418	25	30.61862	1
211899_s_at	TRAF4	0.004318220	95.34882	25	30.61862	1
204053_x_at	PTEN	0.005986829	94.23041	25	30.61862	1
211665_s_at	SOS2	0.006145086	94.13040	25	30.61862	1
202845_s_at	RALBP1	0.007285398	93.43466	25	30.61862	1
204780_s_at	FAS	0.010750622	91.52482	25	30.61862	1

Figura 31. Ejemplo de la lista de genes en cáncer de colon que pertenecen a la vía de señalización “Pathways in cancer”.

	alias <chr>	p-value <dbl>	Statistic <dbl>	Expected <dbl>	Std.dev <dbl>	#Cov <dbl>
211711_s_at	PTEN	0.003847820	95.69001	25	30.61862	1
202871_at	TRAF4	0.004032833	95.55418	25	30.61862	1
211899_s_at	TRAF4	0.004318220	95.34882	25	30.61862	1
204053_x_at	PTEN	0.005986829	94.23041	25	30.61862	1
204054_at	PTEN	0.013060204	90.37460	25	30.61862	1
209112_at	CDKN1B	0.017589152	88.31406	25	30.61862	1
208905_at	CYCS	0.025619288	85.09189	25	30.61862	1
202449_s_at	RXRA	0.026839334	84.63808	25	30.61862	1
215099_s_at	RXRB	0.033190635	82.39231	25	30.61862	1
213992_at	COL4A6	0.044829147	78.67346	25	30.61862	1

Figura 32. Ejemplo de la lista de genes en cáncer de colon que pertenecen a la vía de señalización “Small cell lung cancer”.

	alias <chr>	p-value <dbl>	Statistic <dbl>	Expected <dbl>	Std.dev <dbl>	#Cov <dbl>
216252_x_at	FAS	0.001397613	97.79639	25	30.61862	1
202284_s_at	CDKN1A	0.002506561	96.75408	25	30.61862	1
211711_s_at	PTEN	0.003847820	95.69001	25	30.61862	1
204053_x_at	PTEN	0.005986829	94.23041	25	30.61862	1
200952_s_at	CCND2	0.006310147	94.02705	25	30.61862	1
212743_at	RCHY1	0.007955985	93.04348	25	30.61862	1
204780_s_at	FAS	0.010750622	91.52482	25	30.61862	1
210609_s_at	TP53I3	0.011870519	90.95704	25	30.61862	1
204054_at	PTEN	0.013060204	90.37460	25	30.61862	1
209903_s_at	ATR	0.018772698	87.80913	25	30.61862	1

Figura 33. Ejemplo de la lista de genes en cáncer de colon que pertenecen a la vía de señalización “p53 signalling pathway”.

El procedimiento se ha aplicado en todos los cánceres estudiados. Aunque las 3 vías seleccionadas no aparecen como estadísticamente significativas en el cáncer epidermoide y de próstata, no quiere decir que en estos cánceres no haya genes que pertenezcan a esas vías. De esta manera se ha procedido a buscar los genes que pertenecen a estas vías en todos los tipos de cánceres con la finalidad de encontrar un gen en común para todos los cánceres.

3. Resultados y discusión

A partir de la exploración de los datos, se observa que la distribución y posición de los arrays de los diferentes cánceres siguen un mismo patrón, por lo que no parece que existan arrays problemáticos en ninguno de los cánceres estudiados (*Figura 6 y 7*). Además, los controles de calidad realizados sugieren una calidad similar en todos los arrays (*Figura 9 y 10*).

3.1. Vías de señalización implicadas en la mutación de *PTEN*.

Se ha realizado el análisis de significación biológica para encontrar las vías de señalización implicadas. El análisis de significación biológica se basa en analizar los genes seleccionados por estar diferencialmente expresados entre las dos condiciones (*PTEN* loss-control) y observar a qué vía corresponden, con la premisa de que los genes que los caracterizan aparecerán en esta lista con mayor frecuencia que el resto de genes analizados. De esta forma se obtienen las vías de señalización implicadas en la mutación de *PTEN*. Los análisis de este tipo necesitan un número mínimo de genes para resultar fiables por lo que se incluyen todos los genes con *p-valores* ajustados inferiores a 0.05. Tras el análisis se identificaron 3 vías de señalización predominantes entre los diferentes cánceres:

- KEGGID: 05222. “Small cell lung cancer” o vía de señalización del cáncer de pulmón de células pequeñas.
- KEGGID 04115. “p53 signalling pathway” o vía de señalización de p53.
- KEGGID 05200. “Pathways in cancer” o vías de señalización implicadas en el cáncer.

Debe tenerse en cuenta que estas vías de señalización se identifican de forma estadísticamente significativa en cáncer de mama, epidermoide y colon. En el cáncer de pulmón no microcítico y de próstata no se obtienen vías metabólicas en común con el resto de cánceres a un p -valor < 0.05.

3.1.1 Vía de señalización del cáncer de pulmón de células pequeñas.

La vía de señalización del cáncer de pulmón parece tener un papel significativo en el cáncer de mama (p -value=0.001), colon (p -value=0.023) y epidermoide (p -value=0.000). En el cáncer de próstata y en el cáncer de pulmón no microcítico, esta vía no es estadísticamente significativa, con un p -valor= 0.546 y 0.128, respectivamente.

El cáncer de pulmón es una de las principales causas de muerte por cáncer entre hombres y mujeres en los países industrializados. El carcinoma de pulmón de células pequeñas (CPCP) es una neoplasia muy agresiva, que representa aproximadamente el

25% de todos los casos de cáncer de pulmón⁶⁵. Se trata de un cáncer muy poco común en personas que nunca han fumado⁶⁶. Los mecanismos moleculares alterados en el CPCP incluyen la expresión inducida del oncogén *MYC* y la pérdida de genes supresores tumorales, como *p53*, *PTEN*, *RB* y *FHIT*. La sobreexpresión de proteínas *MYC* en el CPCP conduce a una proliferación tumoral más rápida y pérdida de diferenciación terminal. La mutación o eliminación de *p53* o *PTEN* puede llevar a una proliferación más rápida y reducir la apoptosis. El gen del retinoblastoma *RBI* codifica una fosfoproteína nuclear que ayuda a regular la progresión del ciclo celular. El gen *FHIT* codifica la enzima diadenosina trifosfato hidrolasa, que se cree que tiene un papel indirecto en la proapoptosis y el control del ciclo celular⁶⁵.

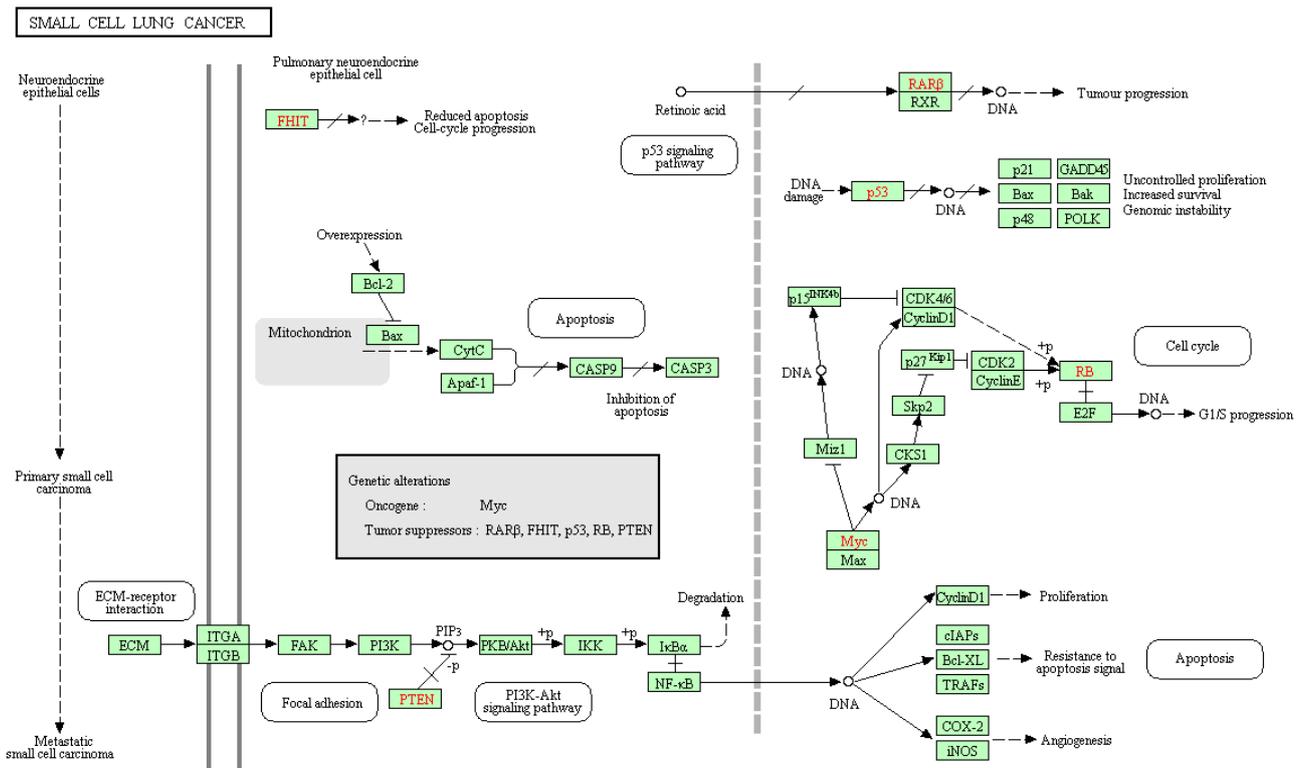


Figura 34. Mapa de la vía de señalización del cáncer de pulmón de células pequeñas⁶⁵.

3.1.2 Vía de señalización de *p53*.

La vía de señalización de *p53* parece tener un papel significativo en el cáncer de colon (p -value=0.000), mama (p -value=0.001) y epidermoide (p -value=0.000). En el cáncer de próstata y en el cáncer de pulmón no microcítico, esta vía no es estadísticamente significativa, con un p -valor=0.122 y 0.454, respectivamente.

El gen supresor tumoral *p53* y sus genes diana forman una vía de señalización compleja de *p53* para regular una amplia variedad de procesos biológicos con la finalidad de prevenir la carcinogénesis⁶⁷. La activación de *p53* es inducida por una serie de señales

- Vía de señalización HIF-1.
- Vía de señalización Notch.
- Vía de señalización Hedgehog.
- Vía de señalización VEGF.
- Vía de señalización mTOR.
- Vía de señalización Wnt.

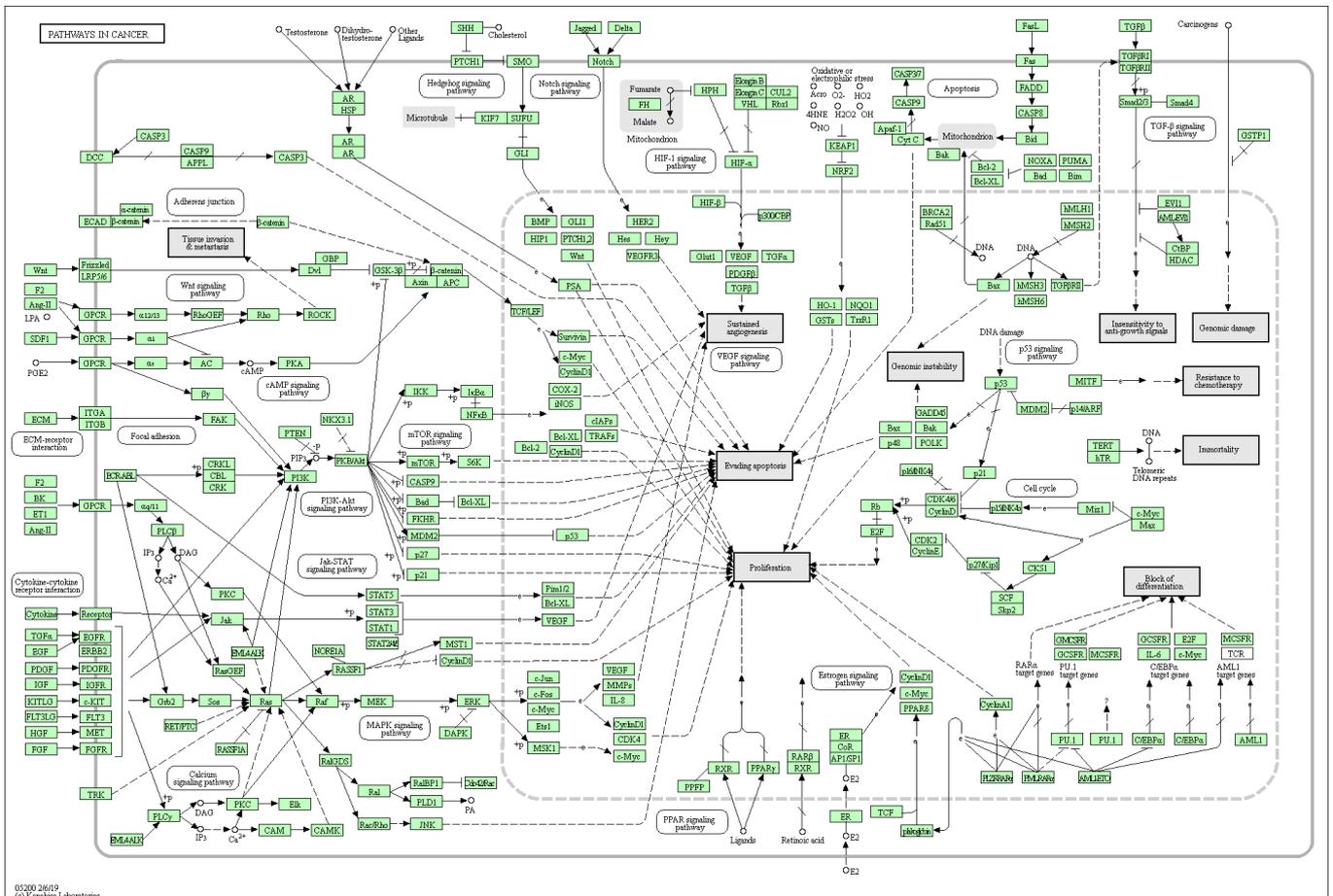


Figura 36. Mapa de las vías de señalización implicadas en el cáncer⁷¹.

3.2. Genes en común en las diferentes vías de señalización seleccionadas.

Después de seleccionar las vías de señalización comunes en el mayor número de cánceres estudiados, se procede a analizar los genes que comprenden estas vías con el objetivo de encontrar genes comunes en diferentes cánceres que sean susceptibles de ser diana terapéutica.

Se intenta encontrar un gen que esté presente en un mayor número de cánceres por lo que tras el análisis se obtiene que el gen *TRAF4* (TNF receptor-associated factor 4) es el único en común en los 4 cánceres humanos (cáncer de colon, mama, pulmón no

microcítico y epidermoide) para las 3 vías seleccionadas y aunque no pueda validarse con el cáncer de próstata en ratón, se considera que puede ser candidato a una posible diana terapéutica.

Además del gen *TRAF4*, en cáncer de mama, epidermoide y pulmón no microcítico se encuentran en común el gen *TP53* y *CDKN1A* (inhibidor de la quinasa dependiente de ciclina 1A) y se valida esta observación al buscar en el cáncer de próstata en ratón en el que también aparecen significativos. De esta forma se obtienen 2 candidatos más a una posible diana terapéutica.

De esta forma se obtienen 3 posibles dianas terapéuticas:

- *TRAF4*
- *TP53*
- *CDKN1A*

Se realiza una búsqueda de cada gen en los resultados obtenidos en el apartado “2.3. Selección de genes diferencialmente expresados” y se encuentra para cada gen su *logFC* (*logarithm fold change*) con la finalidad de saber si se encuentran sobreexpresados o inhibidos (Tabla3).

Tabla 2. LogFC para cada gen en diferentes cánceres.

	Cáncer de colon	Cáncer epidermoide	Cáncer de mama	Cáncer de pulmón no microcítico	Cáncer de próstata
TRAF4	0.248	-0.359	0.274	0.114	-
TP53	-	1.025	0.192	0.177	-0.463
CDKN1A	-	0.461	0.312	0.034	-0.750

De esta forma se obtiene que *TRAF4* se encuentra sobreexpresado en cáncer de colon, mama y pulmón no microcítico e inhibido en carcinoma epidermoide. El gen *TP53* se encuentra sobreexpresado en cáncer epidermoide, mama y pulmón microcítico y se encuentra inhibido en cáncer de próstata (*Mus musculus*). De la misma forma sucede con el gen *CDKN1A* o *p21* que se encuentra sobreexpresado en el cáncer epidermoide, mama y pulmón no microcítico y se encuentra inhibido en el cáncer de próstata. Esta falta de concordancia entre la expresión de *TP53* y *CDKN1A* en el cáncer de próstata y los demás cánceres indica que no es posible validar los resultados de los genes encontrados en *Mus musculus*, debido a que como puede apreciarse no concuerda la expresión. Además, debe recordarse que no solo la especie puede estar variando la expresión génica en el cáncer de próstata, sino también la sobreexpresión de *K-ras*.

3.3. Posibles dianas terapéuticas.

3.3.1. Factor 4 asociado a receptor de TNF (*TRAF4*).

El gen *TRAF4* se encuentra en el cromosoma 17q11-q12⁷², presenta un tamaño de aproximadamente 5,5 kb y contiene 7 exones⁷³. El gen *TRAF4* codifica a un miembro de la familia de factores asociados a los receptores de los factores de necrosis tumoral

(TNF). Las proteínas TRAF están asociadas y median la transducción de señales de los miembros de la superfamilia de receptores de TNF. Se ha demostrado que la proteína codificada interactúa con el receptor de neurotrofina, p75 (NTR / NTSR1), y regula negativamente la muerte celular inducida por NTR y la activación de NF-kappa B. Se ha encontrado que esta proteína se une a p47phox, un factor regulador citosólico incluido en un complejo multiproteico conocido como NAD(P)H oxidasa. Por lo tanto, se cree que esta proteína está involucrada en la activación oxidativa de MAPK8/JNK⁷⁴.

Para investigar la función biológica de *TRAF4*, Regnier *et al.* (2002) generaron ratones deficientes en *TRAF4* por disrupción génica. La mutación fue letal en el embrión pero con una gran variación individual, ya que aproximadamente un tercio de los embriones homocigotos murieron en el útero alrededor del día 14 embrionario, mientras que los otros alcanzaron la edad adulta. Los ratones mutantes supervivientes manifestaron numerosas anomalías en el desarrollo. En particular, el 100% de los ratones mutantes homocigotos sufrieron trastornos respiratorios causados por la interrupción del anillo traqueal. Las malformaciones adicionales fueron principalmente en el esqueleto axial, como las costillas, el esternón, la cola y los arcos vertebrales, con diversos grados de penetrancia. Los ratones deficientes en *TRAF4* también mostraron una alta incidencia de espina bífida, un defecto relacionado con defectos del tubo neural que son malformaciones congénitas comunes en los seres humanos. Los resultados demostraron que se requiere *TRAF4* durante la embriogénesis en procesos biológicos clave que incluyen la formación de la tráquea, el desarrollo del esqueleto axial y el cierre del tubo neural⁷⁵.

Un reciente estudio indica que la sobreexpresión de *TRAF4* incrementa la proliferación celular y la migración, además la inhibición de *TRAF4* parece estar relacionada con la reducción negativa de la vía de señalización PI3K/AKT en cáncer de endometrio⁷⁶. *TRAF4* se encuentra sobreexpresado en el 17.5% de los cánceres de mama, lo que sugiere que este gen puede tener un papel importante en la formación y progresión tumoral y en la metástasis⁷⁷. Por el contrario, otro estudio indica que *TRAF4* se encuentra inhibido en cáncer de mama, con solo un 10% de sobreexpresión del gen⁷⁸. Las discrepancias entre diferentes estudios parecen ser debidas a la diferencia en el grado de diferenciación del tumor estudiado⁷⁷.

De esta forma, y basándonos en el estudio anterior⁷⁷, la diferente expresión de *TRAF4* en el cáncer epidermoide, comparado con los otros cánceres, podría ser debida a una diferencia en el grado de diferenciación del tumor. Además, se ha de tener en cuenta la mayor diversidad del cáncer epidermoide en comparación con los otros tipos de cánceres. Serían necesarios experimentos posteriores en los que se estudiara la expresión del gen en diferentes etapas del tumor.

Por otra parte, *TRAF4* parece ser una buena diana terapéutica debido a su papel importante en la tumorigénesis a través de la interacción directa y la activación de *Akt*⁷⁹. De esta forma, si al hecho de que *PTEN* funcione como un supresor de tumores mediante la antagonización de la vía de señalización PI3K-AKT/PKB¹⁸ y en numerosos cánceres se encuentra mutado o inhibido, se le añade el hecho de que *TRAF4* está sobreexpresado, la vía de señalización PI3K/AKT estará sobreexpresada, por lo que un

inhibidor de *TRAF4* reduciría la expresión de la vía PI3K/AKT resultando en un aumento de la muerte celular. De esta forma se postula que *TRAF4* podría ser una excelente diana terapéutica en los cánceres de mama, colon, epidermoide y pulmón no microcítico en los que *PTEN* se encuentre inhibido.

De hecho, un estudio de *Yang et. al.* (2018) demostró que un microRNA llamado miR-302c-3p regula negativamente a *TRAF4* ya que se dirige directamente a la región 3'-UTR del ARNm de *TRAF4*, reduciendo la migración y la invasión tumoral y mejorando el pronóstico del cáncer. De esta forma, el nivel de mRNA de *TRAF4* se correlacionó inversamente miR-302c-3p⁸⁰. Este estudio conduce a pensar que *TRAF4* puede ser una buena diana terapéutica en multitud de cánceres.

3.3.2. Proteína tumoral 53 (TP53).

El gen *TP53* proporciona instrucciones para producir una proteína llamada proteína tumoral p53 (o p53). Esta proteína actúa como un supresor de tumores, lo que significa que regula la división celular evitando que las células proliferen demasiado rápido o de manera incontrolada. La proteína p53 se encuentra en el núcleo de las células de todo el cuerpo, donde se une directamente al ADN. Cuando el ADN de una célula se daña por agentes como los químicos tóxicos, la radiación o los rayos ultravioleta de la luz solar, esta proteína juega un papel fundamental en la determinación de si el ADN se reparará o la célula dañada resultará en apoptosis. Si el ADN se puede reparar, p53 activa otros genes para reparar el daño. Si el ADN no se puede reparar, esta proteína evita que la célula se divida y comienza el proceso de suicidio celular o apoptosis. Al evitar que las células con ADN mutado o dañado se dividan, p53 ayuda a prevenir el desarrollo de tumores. Debido a que p53 es esencial para regular la división celular y prevenir la formación de tumores, se le ha denominado el "guardián del genoma"⁸¹.

TP53 (p53) es el gen individual más frecuentemente alterado en los cánceres humanos, con mutaciones presentes en aproximadamente el 50% de todos los tumores invasivos. Sin embargo, en algunos de los cánceres más difíciles de tratar, como los cánceres de ovario seroso de alto grado, los cánceres de mama triple negativos, los cánceres de esófago, los cánceres de pulmón de células pequeñas y los cánceres de pulmón de células escamosas, p53 está mutado en al menos el 80% de muestras⁸². De hecho, la proteína p53 mutada es un biomarcador potencial y un objetivo terapéutico para pacientes con cáncer de mama⁸³. Existe una creciente evidencia de que estos p53 mutantes han perdido la actividad supresora de tumores y han ganado funciones que ayudan a contribuir a la progresión maligna⁸⁴. Por lo tanto, claramente, la proteína p53 mutante es un objetivo candidato importante contra el cual se podrían desarrollar nuevos tratamientos contra el cáncer⁸².

En la última década, se han puesto a disposición varios compuestos que pueden reactivar la proteína p53 mutante y convertirla en una conformación con propiedades de p53 no mutado. Se ha encontrado que algunos de estos compuestos, especialmente PRIMA-1, APR-246 PK11007 y COTI-2, exhiben actividad anticancerígena en modelos

preclínicos⁸³. Hasta la fecha, dos de estos compuestos, es decir, APR-246 y COTI-2 han progresado a ensayos clínicos. Un ensayo clínico de fase I/IIa con APR-246 no informó efectos adversos importantes. Actualmente, el APR-246 se encuentra en un ensayo de fase Ib/II en pacientes con cáncer de ovario seroso avanzado, mientras que el COTI-2 se está evaluando en un ensayo de fase I en pacientes con cánceres ginecológicos avanzados⁸².

De esta forma, los datos obtenidos en este estudio coinciden con la investigación actual, en la cual *TP53* es una diana terapéutica potencial y ya se está investigando su utilización en gran variedad de cánceres⁸²⁻⁸⁴. Encontrar compuestos que puedan sobreexpresar *TP53* sería una buena diana terapéutica para los cánceres de mama, epidermoide y pulmón no microcítico con delección en *PTEN*.

Debe tenerse en cuenta que, en el cáncer de mama, epidermoide y pulmón no microcítico, *TP53* se encuentra sobreexpresado. Una posible explicación es su papel como supresor tumoral, ya que al estar inhibido *PTEN*, *TP53* se activa en respuesta a la activación oncogénica con la finalidad de inhibir el crecimiento y división tumoral que está ocurriendo en cada cáncer (Figura 26). De igual forma, debido a que p53 activa directamente a p21 o *CDNK1A*, este se encuentra también sobreexpresado en los cánceres de mama, epidermoide y pulmón no microcítico.

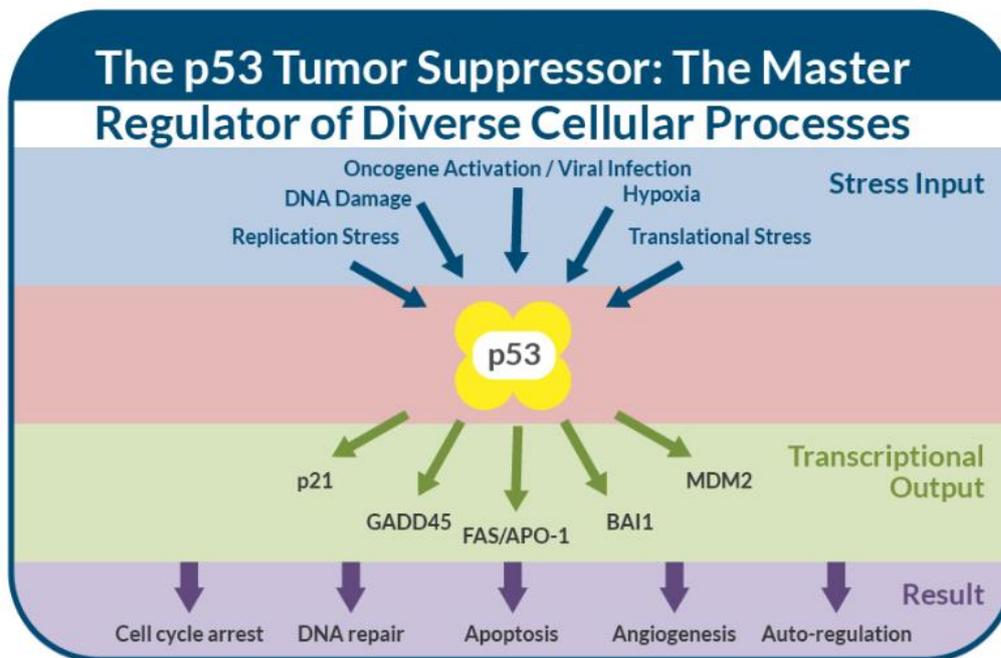


Figura 37. Papel del supresor tumoral p53⁸⁵.

3.3.3. Inhibidor de la quinasa dependiente de ciclina (CDKN1A)

El gen *CDKN1A* o p21 codifica un potente inhibidor de la quinasa dependiente de ciclina. La proteína codificada se une e inhibe la actividad de los complejos de quinasa 2 ciclina dependientes o quinasa 4 ciclina dependientes, y funciona por lo tanto, como

un regulador de la progresión del ciclo celular en G1. La expresión de este gen está estrechamente controlada por la proteína supresora tumoral p53, a través de la cual esta proteína media la detención de la fase G1 del ciclo celular dependiente de p53 en respuesta a una variedad de estímulos de estrés. Esta proteína puede interactuar con el antígeno nuclear de células en proliferación, un factor accesorio de la ADN polimerasa, y desempeña un papel regulador en la replicación del ADN en la fase S y en la reparación del daño en el ADN⁸⁶. p21 facilita la detención del ciclo celular al interactuar con diferentes estímulos como p53, procesos de reparación de ADN, CDK, E2F1, MYC, PCNA, STAT3 AP4, proteasomas, K1F, CDX2 y ER α . p21 actúa tanto como un gen supresor de tumores como un inhibidor de la apoptosis al interactuar con varias moléculas y factores de transición⁸⁷. Además, esta proteína interacciona con caspasas similares a CASP3, lo que conduce a una activación de la quinasa 2 dependiente de ciclina, y puede ser clave en la ejecución de la apoptosis después de la activación de la caspasa. Curiosamente, los ratones que carecen de este gen tienen la capacidad de regenerar tejido dañado o perdido⁸⁶.

La alteración de p21 o *CDKN1A* no es común en humanos. Las enfermedades asociadas con *CDKN1A* incluyen cáncer de lengua y tumor neuroendocrino gastrointestinal⁸⁶.

El estudio de *Romanov, et. al. (2016)* reveló un papel dual de p21 en el cáncer. Demostró que p21 podía inducir la proliferación tumoral en aquellas células que tenían mutado *p53*, mientras que en aquellas en las que el papel de *p53* seguía intacto, la expresión de p21 induce suicidio celular o apoptosis^{87,88} (*Figura 27*). Este hecho debe tenerse en cuenta al considerar a *CDKN1A* o p21 como posible diana terapéutica, debido a que, si el cáncer a tratar presenta una función mutada de *p53*, en caso de sobreexpresar p21 se podría estar favoreciendo la proliferación tumoral.

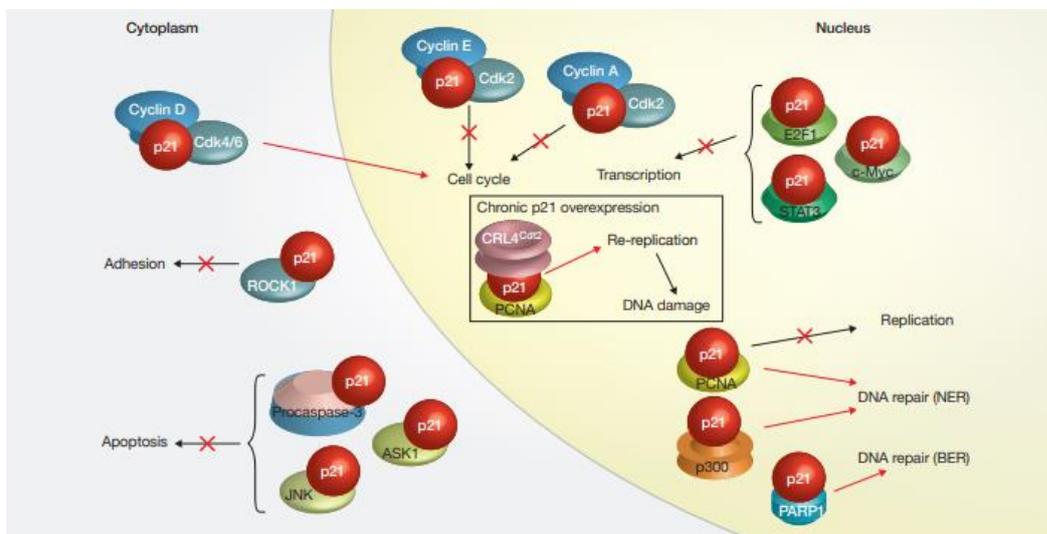


Figura 38. Papel dual de p21 o *CDKN1A*⁸⁸.

Algunos estudios indican que la actividad supresora de tumores de p21 podría tener un papel relevante en una nueva estrategia de tratamiento, que puede llevar al desarrollo de agentes anticancerosos nuevos y selectivos para la terapia dirigida de los cánceres^{87,89}.

3.4. Implicaciones.

De esta forma y tras la búsqueda bibliográfica, se considera que *TRAF4* puede ser una buena diana terapéutica en el mayor número de cánceres estudiados con el gen supresor tumoral *PTEN* inhibido (colon, mama, epidermoide y pulmón no microcítico). De hecho, *PTEN* regula negativamente la vía de señalización PI3K/AKT y cuando *PTEN* está inhibido esta vía se encuentra sobreexpresada. Si a este hecho se le añade una sobreexpresión de *TRAF4*, que presenta un importante papel en la tumorigénesis a través de la interacción directa y la activación de Akt⁷⁹, esto resultaría en que la vía PI3K/AKT, aparte de no tener un gen supresor tumoral que la inhiba, estaría sobreexpresada por *TRAF4*. Por este motivo, una inhibición de *TRAF4* regularía a la baja la vía PI3K/AKT, resultando en un aumento de la muerte celular. Esto podría tener un gran impacto en el tratamiento para ciertos tipos de cánceres.

En relación a *TP53*, son numerosos los estudios que consideran a p53 una buena diana terapéutica⁸²⁻⁸⁴ debido a su papel de supresor tumoral, por el cual se ha ganado el nombre de “guardián del genoma”. De hecho, actualmente ya existen estudios clínicos en los que se testan varios compuestos con actividad anticancerígena⁸³.

Por otra parte, debido al papel dual que recientemente se ha descubierto en p21 o *CDKN1A*^{87,88}, este gen puede ser una buena diana terapéutica en cánceres con *PTEN* mutado o inhibido pero que conserven la función no mutada de *p53*. En caso contrario en los que *p53* se encuentre mutado, la activación de p21 conduciría a la progresión tumoral.

Debe tenerse en cuenta que la pérdida o la mutación de *PTEN* se asocia con la resistencia tumoral a ciertos tratamientos. De hecho, la delección de *PTEN* se ha asociado a una disminución de la sensibilidad del cáncer colorrectal al tratamiento de anticuerpos anti-EGFR^{43,45}. En cáncer epidermoide y de colon, la pérdida de *PTEN* se relaciona con una resistencia al tratamiento con cetuximab^{45,33}. En cáncer de pulmón no microcítico, la inactivación de *PTEN*, se ha asociado con la resistencia a la terapia con inhibidores de la EGFR-tirosina quinasa y una menor supervivencia⁴⁹. Por último, en cáncer de mama la inhibición en *PTEN* se asocia a la resistencia al tratamiento con lapatinib y con trastuzumab⁵³.

De esta forma, estas tres dianas terapéuticas implican un avance en la investigación hacia el desarrollo de nuevos fármacos. De hecho, si bien es cierto que con el gen *p53* ya se están realizando ensayos clínicos, para *TRAF4* y *CDKN1A* aún no se están realizando ensayos clínicos, pero podrían ser una vía de estudio interesante para el tratamiento de cánceres con delección de *PTEN*.

4. Limitaciones

La principal limitación es la reducida disponibilidad de estudios experimentales que hayan utilizado la delección de *PTEN* en diferentes cánceres y que presenten una muestra control con la que compararla. De hecho, se encontraron dos estudios, en cáncer de ovarios⁹⁰ y en cáncer de hígado⁹¹, que no han podido incluirse en el estudio ya que a pesar de que presentaban la delección en *PTEN*, no pertenecían a un cáncer, sino que se realizaba la delección del gen supresor tumoral *PTEN* en una célula normal que con el paso del tiempo desarrollaba un tumor. Por este motivo, con la finalidad de homogeneizar la muestra y evitar sesgos, se han incluido únicamente aquellos estudios experimentales que trabajaban desde el inicio con una muestra de cáncer. Además, se ha de mencionar que no pudieron considerarse aquellos estudios que no presentaban una muestra control con la que comparar la muestra con *PTEN* loss. Todo esto reduce la cantidad estudios experimentales que incluir en el trabajo.

Otra limitación ha sido la especie en la cual se realizaba el experimento. Se han encontrado estudios en *Homo sapiens* pero también en *Mus musculus*, y aunque la especie no fuera un criterio de exclusión, se ha comprobado que los resultados de la muestra de cáncer perteneciente a *Mus musculus* no era comparable con los resultados obtenidos en los diferentes cánceres en *Homo sapiens*.

La limitación en la cantidad de estudios experimentales ya realizados, reduce el poder estadístico del estudio.

5. Conclusiones

5.1. Conclusiones

PTEN es un potente supresor tumoral cuya pérdida de función se observa en gran cantidad de cánceres, siendo el segundo gen mutado con más frecuencia en el cáncer humano después del supresor tumoral *p53*.

PTEN funciona como supresor de tumores mediante la regulación negativa de la vía de señalización PI3K-AKT/PKB. Además, regula gran variedad de procesos biológicos, incluido el mantenimiento de la estabilidad genómica, la supervivencia celular, la migración, la proliferación y el metabolismo.

La pérdida o la mutación de *PTEN* se asocia con la resistencia tumoral a ciertos tratamientos. Además, incluso una disminución sutil de los niveles de *PTEN* resulta en la susceptibilidad al cáncer y favorece la progresión tumoral.

La vía de señalización del cáncer de pulmón de células pequeñas (KEGGID:05222), la vía de señalización de *p53* (KEGGID:04115) y las vías de señalización implicadas en el cáncer (KEGGID:05200) son las vías de señalización comunes en el cáncer de colon, epidermoide y mama y estadísticamente significativas al comparar un cáncer con delección del supresor tumoral *PTEN* con su control.

El gen *TRAF4* puede ser una buena diana terapéutica en el mayor número de cánceres estudiados con el gen supresor tumoral *PTEN* inhibido (colon, mama, epidermoide y pulmón no microcítico), debido a que su inhibición regularía negativamente la vía PI3K/AKT, resultando en un aumento de la muerte celular.

TP53 es una buena diana terapéutica debido a su papel como supresor tumoral. Actualmente ya se están realizando estudios clínicos en los que se utilizan compuestos que pueden reactivar la proteína *p53* mutante y convertirla en una conformación con propiedades de tipo no mutado. Encontrar compuestos que puedan sobreexpresar *TP53* sería una buena diana terapéutica para los cánceres de mama, epidermoide y pulmón no microcítico con delección en *PTEN*.

El gen *CDKN1A* o *p21* puede ser una buena diana terapéutica en cánceres de mama, pulmón no microcítico o epidermoide con *PTEN* mutado o inhibido, pero que conserven la función no mutada de *p53*. Debido a la función dual de *CDKN1A*, si *p53* se encuentra mutado, la activación de *p21* conduciría a la progresión tumoral.

5.2. Reflexión crítica

Todos los objetivos principales y secundarios se han logrado. Se pretendía encontrar una posible diana terapéutica y se han obtenido tres posibles dianas terapéuticas (*TRAF4*, *CDKN1A* y *TP53*) para cánceres en los que exista una delección en el gen supresor tumoral *PTEN*.

Es importante mencionar que el primer objetivo, en el que se basaba el trabajo, consistía en encontrar una posible diana terapéutica en el mayor número de cánceres con delección de *PTEN*. En un primer momento, se pretendió buscar aquellos cánceres en los que la mutación de *PTEN* fuera más frecuente para que así los resultados tuvieran una mayor relevancia, pero la cantidad de estudios experimentales que utilizan la delección de *PTEN* en cáncer es limitada. Sin embargo, los cánceres estudiados presentan una gran relevancia debido a que el cáncer de mama es el tumor más frecuente en la mujer⁵⁰ y el cáncer de pulmón no microcítico es la principal causa de muerte por cáncer en todo el mundo⁴⁹.

5.3. Análisis crítico

Se han producido pequeñas desviaciones de días sobre el cumplimiento de los plazos establecidos al inicio para las entregas parciales del trabajo. No obstante, se han entregado todas las entregas parciales a tiempo y con el contenido previsto.

La metodología utilizada ha sido la adecuada con los medios disponibles. Debido a una limitación en el tiempo y a la ausencia de presupuesto es imposible realizar un estudio experimental en los que se estudien diferentes cánceres con o sin delección en *PTEN*. Por este motivo, la mejor opción ha sido encontrar estudios experimentales ya realizados y cuyos datos se encuentren en bases de datos públicas, con *PTEN knock-out* y con su control para poder compararlos. Establecer la especie de estudio hubiera sido un factor a considerar, debido a que como se ha observado, no se han podido obtener resultados comparables entre *Mus musculus* y *Homo sapiens*.

5.4. Líneas futuras de trabajo

Las nuevas líneas de trabajo deberían abordar el estudio de estas tres dianas terapéuticas. Se debería de inhibir *TRAF4*, sobreexpresar *TP53* o sobreexpresar *CDKN1A* en diferentes estudios con cánceres con delección de *PTEN* y observar la progresión del cáncer.

Por otra parte, sería interesante realizar este mismo estudio, pero con los cánceres en los que *PTEN* presente una mayor tasa de mutación (glioblastoma, cáncer de endometrio y próstata). De esta forma, se realizaría un estudio experimental en condiciones iguales para todos los cánceres y se inhibiría *PTEN*. De esta forma se está homogeneizando la muestra y evitando sesgos.

6. Glosario

CDKN1A: Inhibidor de la quinasa dependiente de ciclina 1^a

CPNM: Cáncer de pulmón no microcítico.

GO: Ontología génica.

KEGG: Enciclopedia de Kyoto de Genes y Genoma.

mTOR: Diana de rapamicina en células de mamífero.

PI3K: Fosfoinositol-3-quinasas.

PKB: Proteína quinasa B.

PTEN: Fosfatidilinositol-3,4,5-trisfosfato 3-fosfatasa.

TP53 (p53): Proteína tumoral 53

TRAF4: Factor 4 asociado a los receptores de los factores de necrosis tumoral.

7. Bibliografía

1. Javier GP, de Velasco. ¿Qué es el cáncer y cómo se desarrolla? - SEOM: Sociedad Española de Oncología Médica © 2019. 2017. <https://seom.org/informacion-sobre-el-cancer/que-es-el-cancer-y-como-se-desarrolla>. Accessed April 20, 2019.
2. Asociación Española Contra el Cáncer. ¿Qué es el Cáncer? ¿Cómo se Desarrolla? | AECC. <https://www.aecc.es/es/todo-sobre-cancer/que-es-cancer>. Accessed April 21, 2019.
3. Instituto Nacional del Cáncer. ¿Qué es el cáncer? - National Cancer Institute. <https://www.cancer.gov/espanol/cancer/naturaleza/que-es>. Accessed April 20, 2019.
4. Hanahan D, Weinberg RA. The hallmarks of cancer. *Cell*. 2000;100(1):57-70. doi:10.1016/S0092-8674(00)81683-9
5. OMS | Datos y cifras sobre el cáncer. WHO. 2014. <https://www.who.int/cancer/about/facts/es/>. Accessed April 21, 2019.
6. Catherine Sánchez N. Conociendo y comprendiendo la célula cancerosa: Fisiopatología del cáncer. *Rev Médica Clínica Las Condes*. 2013;24(4):553-562. doi:10.1016/S0716-8640(13)70659-X
7. 2.1 Proto-oncogenes Y Oncogenes. <http://www.biocancer.com/journal/1321/21-proto-oncogenes-y-oncogenes>. Accessed April 21, 2019.
8. Chial H. Proto-oncogenes to Oncogenes to Cancer. *Nat Educ*. 2008;1(1):33.
9. Santos E. *BIOLOGÍA MOLECULAR DEL CÁNCER: DEL LABORATORIO A LA CLÍNICA*. <http://www.cicancer.org/es/el-cancer>.
10. Genes que regulan la reparación del DNA. http://www10.uniovi.es/anatopatodon/modulo10/tema02_biologia/07dna.htm. Accessed April 21, 2019.
11. Plummer M, de Martel C, Vignat J, Ferlay J, Bray F, Franceschi S. Global burden of cancers attributable to infections in 2012: a synthetic analysis. *Lancet Glob Heal*. 2016;4(9):e609-e616. doi:10.1016/S2214-109X(16)30143-7
12. *Las Cifras Del Cáncer En España.*; 2018. https://seom.org/seomcms/images/stories/recursos/Las_Cifras_del_cancer_en_Espana2018.pdf. Accessed April 21, 2019.
13. Song MS, Salmena L, Pandolfi PP. The functions and regulation of the PTEN tumour suppressor. *Nat Rev Mol Cell Biol*. 2012;13(5):283-296. doi:10.1038/nrm3330
14. Yin Y, Shen WH. PTEN: a new guardian of the genome. *Oncogene*. 2008;27(41):5443-5453. doi:10.1038/onc.2008.241
15. Lee Y-R, Chen M, Pandolfi PP. The functions and regulation of the PTEN tumour suppressor: new modes and prospects. *Nat Rev Mol Cell Biol*. 2018;19(9):547-562. doi:10.1038/s41580-018-0015-0
16. PTEN gene phosphatase and tensin homolog. <https://ghr.nlm.nih.gov/gene/PTEN#location>. Accessed April 21, 2019.
17. Tan M-H, Mester JL, Ngeow J, Rybicki LA, Orloff MS, Eng C. Lifetime Cancer Risks in Individuals with Germline PTEN Mutations. *Clin Cancer Res*. 2012;18(2):400-407. doi:10.1158/1078-0432.CCR-11-2283
18. PTEN phosphatase and tensin homolog [Homo sapiens (human)]. <https://www.ncbi.nlm.nih.gov/gene/5728>. Accessed April 10, 2019.

19. UniProtKB - P60484 (PTEN_HUMAN).
<https://www.uniprot.org/uniprot/P60484>.
20. Govender D, Chetty R. Gene of the month: PTEN. *J Clin Pathol*. 2012;65(7):601-603. doi:10.1136/jclinpath-2012-200711
21. Freeman DJ, Li AG, Wei G, et al. PTEN tumor suppressor regulates p53 protein levels and activity through phosphatase-dependent and -independent mechanisms. *Cancer Cell*. 2003;3(2):117-130. doi:10.1016/S1535-6108(03)00021-7
22. Tang Y, Eng C. PTEN Autoregulates Its Expression by Stabilization of p53 in a Phosphatase-Independent Manner. *Cancer Res*. 2006;66(2):736-742. doi:10.1158/0008-5472.CAN-05-1557
23. Kotelevets L, Scott MGH, Chastre E. Targeting PTEN in Colorectal Cancers. In: *Advances in Experimental Medicine and Biology*. Vol 1110. ; 2018:55-73. doi:10.1007/978-3-030-02771-1_5
24. Liaw D, Marsh DJ, Li J, et al. Germline mutations of the PTEN gene in Cowden disease, an inherited breast and thyroid cancer syndrome. *Nat Genet*. 1997;16(1):64-67. doi:10.1038/ng0597-64
25. Sun H, Lesche R, Li DM, et al. PTEN modulates cell cycle progression and cell survival by regulating phosphatidylinositol 3,4,5,-trisphosphate and Akt/protein kinase B signaling pathway. *Proc Natl Acad Sci U S A*. 1999;96(11):6199-6204. <http://www.ncbi.nlm.nih.gov/pubmed/10339565>. Accessed April 22, 2019.
26. Backman SA, Stambolic V, Suzuki A, et al. Deletion of Pten in mouse brain causes seizures, ataxia and defects in soma size resembling Lhermitte-Duclos disease. *Nat Genet*. 2001;29(4):396-403. doi:10.1038/ng782
27. Stambolic V, Suzuki A, de la Pompa JL, et al. Negative regulation of PKB/Akt-dependent cell survival by the tumor suppressor PTEN. *Cell*. 1998;95(1):29-39. <http://www.ncbi.nlm.nih.gov/pubmed/9778245>. Accessed April 22, 2019.
28. Tamura M, Gu J, Matsumoto K, Aota S, Parsons R, Yamada KM. Inhibition of cell migration, spreading, and focal adhesions by tumor suppressor PTEN. *Science*. 1998;280(5369):1614-1617. <http://www.ncbi.nlm.nih.gov/pubmed/9616126>. Accessed April 22, 2019.
29. Liliental J, Moon SY, Lesche R, et al. Genetic deletion of the Pten tumor suppressor gene promotes cell motility by activation of Rac1 and Cdc42 GTPases. *Curr Biol*. 2000;10(7):401-404. <http://www.ncbi.nlm.nih.gov/pubmed/10753747>. Accessed April 22, 2019.
30. Li J, Simpson L, Takahashi M, et al. The PTEN/MMAC1 tumor suppressor induces cell death that is rescued by the AKT/protein kinase B oncogene. *Cancer Res*. 1998;58(24):5667-5672. <http://www.ncbi.nlm.nih.gov/pubmed/9865719>. Accessed April 22, 2019.
31. Weng L, Brown J, Eng C. PTEN induces apoptosis and cell cycle arrest through phosphoinositol-3-kinase/Akt-dependent and -independent pathways. *Hum Mol Genet*. 2001;10(3):237-242. <http://www.ncbi.nlm.nih.gov/pubmed/11159942>. Accessed April 22, 2019.
32. Cetuximab. <https://www.vademecum.es/principios-activos-cetuximab-101xc06>. Accessed June 2, 2019.
33. Eze N, Lee J-W, Yang D-H, et al. PTEN loss is associated with resistance to cetuximab in patients with head and neck squamous cell carcinoma. *Oral Oncol*. 2019;91:69-78. doi:10.1016/j.oraloncology.2019.02.026
34. Fernández García MS, Ablanado Ablanado P, Fuentes Martínez N, Vivanco Allende B, Touris Noriega J, García Varona A FFM. ACTIVIDAD DE PTEN Y

- RESPUESTA AL TRATAMIENTO CON TRASTUZUMAB EN TUMORES DE MAMA CON SOBREEXPRESION DE HER2 O AMPLIFICACION DE SU GEN - Pósteres - Sociedad Española de Anatomía Patológica. https://www.seap.es/posters4/-/asset_publisher/6JmC/content/id/92766?inheritRedirect=false. Accessed April 22, 2019.
35. Cheng F, Eng C. PTEN Mutations Trigger Resistance to Immunotherapy. *Trends Mol Med*. March 2019. doi:10.1016/j.molmed.2019.03.003
 36. Zhao J, Chen AX, Gartrell RD, et al. Immune and genomic correlates of response to anti-PD-1 immunotherapy in glioblastoma. *Nat Med*. 2019;25(3):462-469. doi:10.1038/s41591-019-0349-y
 37. Ma J, Benitez JA, Li J, et al. Inhibition of Nuclear PTEN Tyrosine Phosphorylation Enhances Glioma Radiation Sensitivity through Attenuated DNA Repair. *Cancer Cell*. 2019;35(3):504-518.e7. doi:10.1016/j.ccell.2019.01.020
 38. Estadísticas importantes sobre el cáncer de próstata. <https://www.cancer.org/es/cancer/cancer-de-prostata/acerca/estadisticas-clave.html>. Accessed April 22, 2019.
 39. Jamaspishvili T, Berman DM, Ross AE, et al. Clinical implications of PTEN loss in prostate cancer. *Nat Rev Urol*. 2018;15(4):222-234. doi:10.1038/nrurol.2018.9
 40. ¿Qué es el cáncer colorrectal? <https://www.cancer.org/es/cancer/cancer-de-colon-o-recto/acerca/que-es-cancer-de-colon-o-recto.html>. Accessed April 22, 2019.
 41. Cáncer del colon y del recto. <https://medlineplus.gov/spanish/colorectalcancer.html>. Accessed April 22, 2019.
 42. Berg M, Danielsen SA, Ahlquist T, et al. DNA Sequence Profiles of the Colorectal Cancer Critical Gene Set KRAS-BRAF-PIK3CA-PTEN-TP53 Related to Age at Disease Onset. Chan KYK, ed. *PLoS One*. 2010;5(11):e13978. doi:10.1371/journal.pone.0013978
 43. Roock W De, Vriendt V De, Normanno N, Ciardiello F, Tejpar S. KRAS, BRAF, PIK3CA, and PTEN mutations: implications for targeted therapies in metastatic colorectal cancer. *Lancet Oncol*. 2011;12(6):594-603. doi:10.1016/S1470-2045(10)70209-6
 44. Dicuonzo G, Angeletti S, Garcia-Foncillas J, et al. Colorectal carcinomas and PTEN/MMAC1 gene mutations. *Clin Cancer Res*. 2001;7(12):4049-4053. <http://www.ncbi.nlm.nih.gov/pubmed/11751500>. Accessed April 22, 2019.
 45. Loupakis F, Pollina L, Stasi I, et al. PTEN Expression and KRAS Mutations on Primary Tumors and Metastases in the Prediction of Benefit From Cetuximab Plus Irinotecan for Patients With Metastatic Colorectal Cancer. *J Clin Oncol*. 2009;27(16):2622-2629. doi:10.1200/JCO.2008.20.2796
 46. Definición de carcinoma epidermoide - Diccionario de cáncer - National Cancer Institute. <https://www.cancer.gov/espanol/publicaciones/diccionario/def/carcinoma-epidermoide>. Accessed April 22, 2019.
 47. Sánchez Fernández R. *Alteraciones Genético-Moleculares En El Carcinoma Epidermoide de Oído Externo*. http://digibuo.uniovi.es/dspace/bitstream/10651/37416/1/TD_RafaelSanchezFernandez.pdf. Accessed April 22, 2019.
 48. ¿Qué es el cáncer de pulmón no microcítico? <https://www.cancer.org/es/cancer/cancer-de-pulmon-no-microcitico/acerca/que-es-cancer-de-pulmon-no-microcitico.html>. Accessed April 22, 2019.

49. Pérez-Ramírez C, Cañadas-Garre M, Molina MÁ, Faus-Dáder MJ, Calleja-Hernández MÁ. PTEN and PI3K/AKT in non-small-cell lung cancer. *Pharmacogenomics*. 2015;16(16):1843-1862. doi:10.2217/pgs.15.122
50. *Cáncer de Mama Hereditario Más Allá de BRCA1/BRCA2*. <https://revistageneticamedica.com/wp-content/uploads/2018/04/GMG02-67-77-Chirivella-2018-web.pdf>. Accessed April 22, 2019.
51. Ngeow J, Sesock K, Eng C. Breast cancer risk and clinical implications for germline PTEN mutation carriers. *Breast Cancer Res Treat*. 2017;165(1):1-8. doi:10.1007/s10549-015-3665-z
52. Dave B, Migliaccio I, Gutierrez MC, et al. Loss of Phosphatase and Tensin Homolog or Phosphoinositol-3 Kinase Activation and Response to Trastuzumab or Lapatinib in Human Epidermal Growth Factor Receptor 2–Overexpressing Locally Advanced Breast Cancers. *J Clin Oncol*. 2011;29(2):166-173. doi:10.1200/JCO.2009.27.7814
53. Rimawi MF, De Angelis C, Contreras A, et al. Low PTEN levels and PIK3CA mutations predict resistance to neoadjuvant lapatinib and trastuzumab without chemotherapy in patients with HER2 over-expressing breast cancer. *Breast Cancer Res Treat*. 2018;167(3):731-740. doi:10.1007/s10549-017-4533-9
54. El desarrollo tumoral: acumulación de mutaciones y selección clonal. :: Centro de investigación del Cáncer - Comprehensive Cancer Center Research. <http://www.cicancer.org/es/el-desarrollo-tumoral-acumulacion-de-mutaciones-y-seleccion-clonal>. Accessed March 13, 2019.
55. Naderali E, Khaki AA, Rad JS, Ali-Hemmati A, Rahmati M, Charoudeh HN. Regulation and modulation of PTEN activity. *Mol Biol Rep*. 2018;45(6):2869-2881. doi:10.1007/s11033-018-4321-6
56. Álvarez-García V, Tawil Y, Wise HM, Leslie NR. Mechanisms of PTEN loss in cancer: It's all about diversity. *Semin Cancer Biol*. February 2019. doi:10.1016/j.semcancer.2019.02.001
57. Raffone A, Travaglino A, Saccone G, et al. Loss of PTEN expression as diagnostic marker of endometrial precancer: A systematic review and meta-analysis. *Acta Obstet Gynecol Scand*. 2019;98(3):275-286. doi:10.1111/aogs.13513
58. Cowell JK, Qin H, Hu T, Wu Q, Bhole A, Ren M. Mutation in the FGFR1 tyrosine kinase domain or inactivation of PTEN is associated with acquired resistance to FGFR inhibitors in FGFR1-driven leukemia/lymphomas. *Int J Cancer*. 2017;141(9):1822-1829. doi:10.1002/ijc.30848
59. M. Carmen Ruíz de Villa AS-P. *Análisis de Datos de Microarrays*.; 2018.
60. Jarno Tuimala. *DNA Microarray Data Analysis Using Bioconductor*. CSC, the Finnish IT center for Science; 2008.
61. Smyth GK, Ritchie M, Thorne N, Wettenhall J, Shi W, Hu Y. *Limma: Linear Models for Microarray and RNA-Seq Data User's Guide*. Melbourne, Australia: Bioinformatics Division, The Walter and Eliza Hall Institute of Medical Research; 2002.
62. Gene Ontology Resource. <http://geneontology.org/>. Accessed May 21, 2019.
63. Kanehisa M, Sato Y, Furumichi M, Morishima K, Tanabe M. New approach for understanding genome variations in KEGG. *Nucleic Acids Res*. 2019;47(D1):D590-D595. doi:10.1093/nar/gky962
64. Gene Ontology overview. <http://geneontology.org/docs/ontology-documentation/>. Accessed May 21, 2019.
65. KEGG PATHWAY: Small cell lung cancer - Homo sapiens (human).

- https://www.genome.jp/kegg-bin/show_pathway?map=hsa05222&show_description=show. Accessed May 26, 2019.
66. Cáncer pulmonar de células pequeñas: MedlinePlus enciclopedia médica. <https://medlineplus.gov/spanish/ency/article/000122.htm>. Accessed May 26, 2019.
 67. Liu J, Zhang C, Feng Z. Tumor suppressor p53 and its gain-of-function mutants in cancer. *Acta Biochim Biophys Sin (Shanghai)*. 2014;46(3):170-179. doi:10.1093/abbs/gmt144
 68. KEGG PATHWAY: p53 signaling pathway - Homo sapiens (human). https://www.genome.jp/kegg-bin/show_pathway?hsa04115. Accessed May 26, 2019.
 69. Clarke AR, Purdie CA, Harrison DJ, et al. Thymocyte apoptosis induced by p53-dependent and independent pathways. *Nature*. 1993;362(6423):849-852. doi:10.1038/362849a0
 70. Lowe SW, Schmitt EM, Smith SW, Osborne BA, Jacks T. p53 is required for radiation-induced apoptosis in mouse thymocytes. *Nature*. 1993;362(6423):847-849. doi:10.1038/362847a0
 71. KEGG PATHWAY: Pathways in cancer - Homo sapiens (human). https://www.genome.jp/kegg-bin/show_pathway?hsa05200. Accessed May 26, 2019.
 72. Tomasetto C, Régnier C, Moog-Lutz C, et al. Identification of Four Novel Human Genes Amplified and Overexpressed in Breast Carcinoma and Localized to the q11-q21.3 Region of Chromosome 17. *Genomics*. 1995;28(3):367-376. doi:10.1006/geno.1995.1163
 73. Regnier CH, Tomasetto C, Moog-Lutz C, et al. Presence of a New Conserved Domain in CART1, a Novel Member of the Tumor Necrosis Factor Receptor-associated Protein Family, Which Is Expressed in Breast Carcinoma. *J Biol Chem*. 1995;270(43):25715-25721. doi:10.1074/jbc.270.43.25715
 74. TRAF4 Gene. TNF Receptor Associated Factor 4. <https://www.genecards.org/cgi-bin/carddisp.pl?gene=TRAF4>. Accessed May 24, 2019.
 75. Regnier CH, Masson R, Kedinger V, et al. Impaired neural tube closure, axial skeleton malformations, and tracheal ring disruption in TRAF4-deficient mice. *Proc Natl Acad Sci*. 2002;99(8):5585-5590. doi:10.1073/pnas.052124799
 76. Xie P, Wang X, Kong M, Bai X, Jiang T. TRAF4 promotes endometrial cancer cell growth and migration by activation of PI3K/AKT/Oct4 signaling. *Exp Mol Pathol*. 2019;108:9-16. doi:10.1016/j.yexmp.2019.03.003
 77. Wu H. *TNF Receptor Associated Factors (TRAFs)*. Springer Science+Business Media; 2007. https://books.google.es/books?id=gsbhUG77xXYC&pg=PA68&lpg=PA68&dq=traf4+funcion&source=bl&ots=zOdqRE-dXd&sig=ACfU3U2ZtPkYKEpNPxxpD_HVFn4hek1AVQ&hl=es&sa=X&ved=2ahUKewiQ7tmJn7riAhV68OAKHZPcC9cQ6AEwDnoECAgQAQ#v=onepage&q=traf4+funcion&f=false. Accessed May 27, 2019.
 78. Krajewska M, Krajewski S, Zapata JM, et al. TRAF-4 expression in epithelial progenitor cells. Analysis in normal adult, fetal, and tumor tissues. *Am J Pathol*. 1998;152(6):1549-1561. <http://www.ncbi.nlm.nih.gov/pubmed/9626059>. Accessed May 27, 2019.
 79. ZHANG J, LI X, YANG W, JIANG X, LI N. TRAF4 promotes tumorigenesis of

- breast cancer through activation of Akt. *Oncol Rep.* 2014;32(3):1312-1318. doi:10.3892/or.2014.3304
80. Yang L, Guo Y, Liu X, et al. The tumor suppressive miR-302c-3p inhibits migration and invasion of hepatocellular carcinoma cells by targeting TRAF4. *J Cancer.* 2018;9(15):2693-2701. doi:10.7150/jca.25569
 81. Reference GH. TP53 gene. <https://ghr.nlm.nih.gov/gene/TP53#normalfunction>. Accessed May 25, 2019.
 82. Duffy MJ, Synnott NC, Crown J. Mutant p53 as a target for cancer treatment. *Eur J Cancer.* 2017;83:258-265. doi:10.1016/j.ejca.2017.06.023
 83. Duffy MJ, Synnott NC, Crown J. Mutant p53 in breast cancer: potential as a therapeutic target and biomarker. *Breast Cancer Res Treat.* 2018;170(2):213-219. doi:10.1007/s10549-018-4753-7
 84. Muller PAJ, Vousden KH. Mutant p53 in Cancer: New Functions and Therapeutic Opportunities. *Cancer Cell.* 2014;25(3):304-317. doi:10.1016/j.ccr.2014.01.021
 85. Jockers D. The p53 Gene and Cancer Development. <https://drjockers.com/p53-gene-cancer-development/>. Accessed May 26, 2019.
 86. CDKN1A. Cyclin Dependent Kinase Inhibitor 1A. <https://www.genecards.org/cgi-bin/carddisp.pl?gene=CDKN1A>. Accessed May 26, 2019.
 87. Parveen A, Akash MSH, Rehman K, Kyunn WW. Dual Role of p21 in the Progression of Cancer and Its Treatment. *Crit Rev Eukaryot Gene Expr.* 2016;26(1):49-62. doi:10.1615/CritRevEukaryotGeneExpr.v26.i1.60
 88. Romanov VS, Rudolph KL. p21 shapes cancer evolution. *Nat Cell Biol.* 2016;18(7):722-724. doi:10.1038/ncb3382
 89. El-Deiry WS. p21(WAF1) Mediates Cell-Cycle Inhibition, Relevant to Cancer Suppression and Therapy. *Cancer Res.* 2016;76(18):5189-5191. doi:10.1158/0008-5472.CAN-16-2055
 90. Russo A, Czarnecki AA, Dean M, et al. PTEN loss in the fallopian tube induces hyperplasia and ovarian tumor formation. *Oncogene.* 2018;37(15):1976-1990. doi:10.1038/s41388-017-0097-8
 91. Wang G, Luo X FG. Temporal gene expression patterns (Next Generation Sequencing) in liver tumourigenesis of hepatocyte-specific deletion of shp2 and/or pten. Dec 07. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE123427>. Published 2018. Accessed May 29, 2019.

Anexo: Pipeline

1. Directorio y opciones de trabajo

```
setwd("C:/Users/laura/Desktop/Datos")
workingDir <- getwd() #indicamos el directorio de trabajo
dataDir <- file.path(workingDir, "data")
resultsDir <- file.path(workingDir, "results")
celfilesDircolon <- file.path(workingDir, "celfilescolon")
setwd(workingDir)
```

2. Lectura de los datos

```
require(affy)
#colon
sampleInfocolon <-
read.AnnotatedDataFrame(file.path(dataDir, "targetcolon.txt"), header =
TRUE, row.names=1, sep="\t")
fileNamescolon <- rownames(pData(sampleInfocolon))
rawDatacolon <- read.affybatch(fileNames=file.path(celfilesDircolon,
fileNamescolon), phenoData=sampleInfocolon)
```

```
#mama
celfilesDirmama <- file.path(workingDir, "celfilesmama")
sampleInfomama <-
read.AnnotatedDataFrame(file.path(dataDir, "targetmama.txt"), header =
TRUE, row.names=1, sep="\t")
fileNamesmama <- rownames(pData(sampleInfomama))
rawDatamama <- read.affybatch(fileNames=file.path(celfilesDirmama,
fileNamesmama), phenoData=sampleInfomama)
```

```
#epidermoide
celfilesDirepidermoid <- file.path(workingDir, "celfilesepidermoid")
sampleInfoepidermoid <-
read.AnnotatedDataFrame(file.path(dataDir, "targetepidermoid.txt"),
header = TRUE, row.names=1, sep="\t")
fileNamesepidermoid <- rownames(pData(sampleInfoepidermoid))
rawDataepidermoid <-
read.affybatch(fileNames=file.path(celfilesDirepidermoid,
fileNamesepidermoid), phenoData=sampleInfoepidermoid)
```

```
#cancer de pulmón no microcítico
celfilesDirlung <- file.path(workingDir, "celfileslung")
sampleInfolung <-
read.AnnotatedDataFrame(file.path(dataDir, "targetlung.txt"), header =
TRUE, row.names=1, sep="\t")
fileNameslung <- rownames(pData(sampleInfolung))
rawDatalung <- read.affybatch(fileNames=file.path(celfilesDirlung,
fileNameslung), phenoData=sampleInfolung)
```

```
#próstata
celfilesDirprostate <- file.path(workingDir, "celfilesprostate")
```

```

sampleInfoprostate <-
read.AnnotatedDataFrame(file.path(dataDir,"targetp.txt"), header =
TRUE, row.names=1, sep="\t")
fileNamesprostate <- rownames(pData(sampleInfoprostate))
rawDataprostate <-
read.affybatch(fileNames=file.path(celfilesDirprostate,
fileNamesprostate), phenoData=sampleInfoprostate)

```

3. Preprocesado: Exploración, Control de calidad y Normalización.

3.1. Exploración de los Datos

```

#Histograma
par(mfrow=c(3,2))
hist(rawDatacolon, main="Signal distribution colon")
hist(rawDataepidermoid, main="Signal distribution epidermoid")
hist(rawDataLung, main="Signal distribution lung")
hist(rawDatamama, main="Signal distribution mama")
hist(rawDataprostate, main="Signal distribution prostate")

#Boxplot
par(mfrow=c(3,2))
boxplot(rawDatacolon, main="Signal distribution colon")
boxplot(rawDataepidermoid, main="Signal distribution epidermoid")
boxplot(rawDataLung, main="Signal distribution lung")
boxplot(rawDatamama, main="Signal distribution mama")
boxplot(rawDataprostate, main="Signal distribution prostate")

#Clúster jerárquico
par(mfrow=c(3,2))
plot(hclust(dist(t(exprs(rawDatacolon))),method="average"),
main="Cluster jerarquico colon")
plot(hclust(dist(t(exprs(rawDataepidermoid))),method="average"),
main="Cluster jerarquico epidermoid")
plot(hclust(dist(t(exprs(rawDataLung))),method="average"),
main="Cluster jerarquico lung")
plot(hclust(dist(t(exprs(rawDatamama))),method="average"),
main="Cluster jerarquico mama")
plot(hclust(dist(t(exprs(rawDataprostate))),method="average"),
main="Cluster jerarquico prostate")

```

3.2. Control de calidad

```

#Gráfico de degradación

par(mfrow=c(3,2))
infoc <- data.frame(grupo=c("control", "control", "PTENloss",
"PTENloss", "PTENloss"))
deg <- AffyRNAdeg(rawDatacolon, log.it=F)
summaryAffyRNAdeg(deg)

plotAffyRNAdeg(deg, col=infoc$grupo)
legend("topleft", legend=infoc$grupo, col=infoc$grupo,
lty=1:ncol(infoc), cex=0.5, title="colon")

infom <- data.frame(grupo=c("control", "control", "PTENloss",

```

```

"PTENloss"))
degm <- AffyRNAdeg(rawDatamama, log.it=F)
summaryAffyRNAdeg(degm)

plotAffyRNAdeg(degm, col=infom$grupo)
legend("topleft", legend=infom$grupo, col=infom$grupo,
lty=1:ncol(infom), cex=0.5, title="mama")

infop <- data.frame(grupo=c("control", "control", "control",
"PTENloss", "PTENloss", "PTENloss"))
degp <- AffyRNAdeg(rawDataprostate, log.it=F)
summaryAffyRNAdeg(degpp)

plotAffyRNAdeg(degpp, col=infop$grupo)
legend("topleft", legend=infop$grupo, col=infop$grupo,
lty=1:ncol(infop), cex=0.5, title="prostate")

infoe <- data.frame(grupo=c("control", "control", "PTENloss",
"PTENloss"))
dege <- AffyRNAdeg(rawDataepidermoid, log.it=F)
summaryAffyRNAdeg(dege)

plotAffyRNAdeg(dege, col=infoe$grupo)
legend("topleft", legend=infoe$grupo, col=infoe$grupo,
lty=1:ncol(infoe), cex=0.5, title="epidermoid")

infol <- data.frame(grupo=c("control", "control", "PTENloss",
"PTENloss"))
degl <- AffyRNAdeg(rawDatalung, log.it=F)
summaryAffyRNAdeg(degl)

plotAffyRNAdeg(degl, col=infol$grupo)
legend("topleft", legend=infol$grupo, col=infol$grupo,
lty=1:ncol(infol), cex=0.5, title="nsc lung")

#Gráfico de expresiones relativas

require(affyPLM)
stopifnot(require(affyPLM))
Psetc<- fitPLM(rawDatacolon)
Psetm<- fitPLM(rawDatamama)
Psetp<- fitPLM(rawDataprostate)
Psetl<- fitPLM(rawDatalung)
Psete<- fitPLM(rawDataepidermoid)
par(mfrow=c(3,2))
RLE(Psetc, main = "Relative Log Expression colon")
RLE(Psetm, main = "Relative Log Expression mama")
RLE(Psetp, main = "Relative Log Expression prostate")
RLE(Psetl, main = "Relative Log Expression non-small cell lung
carcinoma")
RLE(Psete, main = "Relative Log Expression epidermoid")

#Gráfico de residuos centrados y normalizados o NUSE
par(mfrow=c(3,2))
NUSE(Psetc, main = "NUSE colon")

```

```

NUSE(Psetm, main = "NUSE mama")
NUSE(Psetp, main = "NUSE prostate")
NUSE(Psetl, main = "NUSE lung ")
NUSE(Psete, main = "NUSE epidermoid")

```

3.3. Normalización

```
stopifnot(require(affy))
```

```
#colon
```

```

normalize <- T
if(normalize){
  eset_rmacolon <- rma(rawDatacolon) #RMA preprocessing
  save(eset_rmacolon, file=file.path(dataDir,"colon.Rda"))
}else{
  load (file=file.path(dataDir,"colon.Rda"))
}

```

```
#mama
```

```

normalize <- T
if(normalize){
  eset_rmamama <- rma(rawDatamama) #RMA preprocessing
  save(eset_rmamama, file=file.path(dataDir,"mama.Rda"))
}else{
  load (file=file.path(dataDir,"mama.Rda"))
}

```

```
#prostate
```

```

normalize <- T
if(normalize){
  eset_rmaprostate <- rma(rawDataprostate) #RMA preprocessing
  save(eset_rmaprostate, file=file.path(dataDir,"prostate.Rda"))
}else{
  load (file=file.path(dataDir,"prostate.Rda"))
}

```

```
#Lung
```

```

normalize <- T
if(normalize){
  eset_rmalung <- rma(rawDataLung) #RMA preprocessing
  save(eset_rmalung, file=file.path(dataDir,"lung.Rda"))
}else{
  load (file=file.path(dataDir,"lung.Rda"))
}

```

```
#epidermoid
```

```

normalize <- T
if(normalize){
  eset_rmaepidermoid <- rma(rawDataepidermoid) #RMA preprocessing
  save(eset_rmaepidermoid, file=file.path(dataDir,"epidermoid.Rda"))
}else{
  load (file=file.path(dataDir,"epidermoid.Rda"))
}

```

3.4. Exploración de los datos normalizados

```
#Boxplot de Los datos normalizados
```

```
par(mfrow=c(3,2))
boxplot(eset_rmacolon, main="Boxplot RMA-valores de expresion
normalizados colon")
boxplot(eset_rmamama, main="Boxplot RMA-valores de expresion
normalizados mama")
boxplot(eset_rmaepidermoid, main="Boxplot RMA-valores de expresion
normalizados epidermoid")
boxplot(eset_rmalung, main="Boxplot RMA-valores de expresion
normalizados lung")
boxplot(eset_rmaprostate, main="Boxplot RMA-valores de expresion
normalizados prostate")
```

3.5. Filtraje

```
require(genefilter)
```

```
#colon
```

```
filteredcolon<-nsFilter(eset_rmacolon, require.entrez=TRUE,
                        remove.dupEntrez=TRUE, var.func=IQR,
                        var.cutoff=0.5, var.filter=TRUE,
                        filterByQuantile=TRUE, feature.exclude="^AFFX")
```

```
eset_filteredcolon <-filteredcolon$eset
save(eset_rmacolon, eset_filteredcolon, file=file.path(resultsDir,
"colon.Rda"))
```

```
#mama
```

```
filteredmama<-nsFilter(eset_rmamama, require.entrez=TRUE,
                       remove.dupEntrez=TRUE, var.func=IQR,
                       var.cutoff=0.5, var.filter=TRUE,
                       filterByQuantile=TRUE, feature.exclude="^AFFX")
```

```
eset_filteredmama <-filteredmama$eset
save(eset_rmamama, eset_filteredmama, file=file.path(resultsDir,
"mama.Rda"))
```

```
#epidermoid
```

```
filteredeepidermoid<-nsFilter(eset_rmaepidermoid, require.entrez=TRUE,
                               remove.dupEntrez=TRUE, var.func=IQR,
                               var.cutoff=0.5, var.filter=TRUE,
                               filterByQuantile=TRUE, feature.exclude="^AFFX")
```

```
eset_filteredeepidermoid <-filteredeepidermoid$eset
save(eset_rmaepidermoid, eset_filteredeepidermoid,
file=file.path(resultsDir, "epidermoid.Rda"))
```

```
#Lung
```

```
filteredlung<-nsFilter(eset_rmalung, require.entrez=TRUE,
                       remove.dupEntrez=TRUE, var.func=IQR,
                       var.cutoff=0.5, var.filter=TRUE,
                       filterByQuantile=TRUE, feature.exclude="^AFFX")
```

```

eset_filteredlung <-filteredlung$eset
save(eset_rmalung, eset_filteredlung, file=file.path(resultsDir,
"lung.Rda"))

#prostate
filteredprostate<-nsFilter(eset_rmaprostate, require.entrez=TRUE,
                           remove.dupEntrez=TRUE, var.func=IQR,
                           var.cutoff=0.5, var.filter=TRUE,
                           filterByQuantile=TRUE, feature.exclude="^AFFX")

eset_filteredprostate <-filteredprostate$eset
save(eset_rmaprostate, eset_filteredprostate,
file=file.path(resultsDir, "prostate.Rda"))

```

4. Selección de genes diferencialmente expresados

```

#cancer colon
groupsc<-c("C", "C", "T", "T", "T")
groupsc<-as.factor(groupsc)
designcolon<-model.matrix(~groupsc)
designcolon

#cancer de mama
groupsm<-c("C", "C", "T", "T")
groupsm<-as.factor(groupsm)
designmama<-model.matrix(~groupsm)
designmama

#carcinoma epidermoide
groupse<-c("C", "C", "T", "T")
groupse<-as.factor(groupse)
designepidermoid<-model.matrix(~groupse)
designepidermoid

#non-small cell lung carcinoma
groupsl<-c("C", "C", "T", "T")
groupsl<-as.factor(groupsl)
designlung<-model.matrix(~groupsl)
designlung

#próstata
groupsp<-c("C", "C", "C", "T", "T", "T")
groupsp<-as.factor(groupsp)
designprostate<-model.matrix(~groupsp)
designprostate

```

5. Análisis usando el modelo lineal

```

#colon
fitcolon<-lmFit(eset_filteredcolon, designcolon)
fitcolon<-eBayes(fitcolon) #el objeto fit contiene el resultado del
#analisis
toptablecolon<-topTable(fitcolon, coef=2, number=nrow(fitcolon)) #Los
#resultados son extraidos usando el comando toptable
head(toptablecolon)

```

```

#mama
fitmama<-lmFit(eset_filteredmama,designmama)
fitmama<-eBayes(fitmama) #el objeto fit contiene el resultado del
analisis
toptablemama<-topTable(fitmama, coef=2, number=nrow(fitmama)) #Los
resultados son extraidos usando el comando toptable
head(toptablemama)

#Lung
fitlung<-lmFit(eset_filteredlung,designlung)
fitlung<-eBayes(fitlung) #el objeto fit contiene el resultado del
analisis
toptablelung<-topTable(fitlung, coef=2, number=nrow(fitlung)) #Los
resultados son extraidos usando el comando toptable
head(toptablelung)

#prostate
fitprostate<-lmFit(eset_filteredprostate,designprostate)
fitprostate<-eBayes(fitprostate) #el objeto fit contiene el resultado
del analisis
toptableprostate<-topTable(fitprostate, coef=2,
number=nrow(fitprostate)) #Los resultados son extraidos usando el
comando toptable
head(toptableprostate)

#epidermoid
fitepidermoid<-lmFit(eset_filteredepidermoid,designepidermoid)
fitepidermoid<-eBayes(fitepidermoid) #el objeto fit contiene el
resultado del analisis
toptableepidermoid<-topTable(fitepidermoid, coef=2,
number=nrow(fitepidermoid)) #Los resultados son extraidos usando el
comando toptable
head(toptableepidermoid)

```

6. Anotación de resultados

```

require(hgu133a.db)
require(hgu133plus2.db)
require(mouse4302.db)
hgu133a()
hgu133plus2()
mouse4302()
#colon
require(annaffy)
geneSymbolsc<-unlist(as.list(hgu133aSYMBOL))
geneNamesc<-unlist(as.list(hgu133aGENENAME))
geneNamesc<-substring(geneNamesc,1,40)
genelistc<-data.frame(GeneSymbolc=geneSymbolsc,GeneNameec=geneNamesc)
topTabc <- topTable(fitcolon, number=nrow(fitcolon), coef=2,
adjust="BH", genelist=genelistc)
head(topTabc)

#epidermoid
geneSymbolse<-unlist(as.list(hgu133plus2SYMBOL))
geneNameese<-unlist(as.list(hgu133plus2GENENAME))

```

```

geneNamese<-substring(geneNamese,1,40)
geneliste<-data.frame(GeneSymbole=geneSymbolse, GeneNamee=geneNamese)
topTabe <- topTable(fitepidermoid, number=nrow(fitepidermoid), coef=2,
adjust="BH", geneliste=geneliste)
head(topTabe)

```

#mama

```

geneSymbolsm<-unlist(as.list(hgu133plus2SYMBOL))
geneNamesm<-unlist(as.list(hgu133plus2GENENAME))
geneNamesm<-substring(geneNamesm,1,40)
genelistm<-data.frame(GeneSymbolm=geneSymbolsm, GeneNameem=geneNamesm)
topTabm <- topTable(fitmama, number=nrow(fitmama), coef=2,
adjust="BH", geneliste=genelistm)
head(topTabm)

```

#prostate

```

geneSymbolsp<-unlist(as.list(mouse4302SYMBOL))
geneNamesp<-unlist(as.list(mouse4302GENENAME))
geneNamesp<-substring(geneNamesp,1,40)
genelistp<-data.frame(GeneSymbolp=geneSymbolsp, GeneNameep=geneNamesp)
topTabp <- topTable(fitprostate, number=nrow(fitprostate), coef=2,
adjust="BH", geneliste=genelistp)
head(topTabp)

```

#Lung

```

geneSymbolsl<-unlist(as.list(hgu133plus2SYMBOL))
geneNamesl<-unlist(as.list(hgu133plus2GENENAME))
geneNamesl<-substring(geneNamesl,1,40)
genelistl<-data.frame(GeneSymboll=geneSymbolsl, GeneNameel=geneNamesl)
topTabl <- topTable(fitlung, number=nrow(fitlung), coef=2,
adjust="BH", geneliste=genelistl)
head(topTabl)

```

7. Genes sobreexpresados o inhibidos o con regulación negativa.

#CoLon

```

require(annotate)
fit.Symbolsc <- getSYMBOL (rownames(fitcolon), "hgu133a.db")
resc<-decideTests(fitcolon, method="separate",
adjust.method="BH", p.value=0.01)
sum.res.rowsc<-apply(abs(resc),1,sum)
res.selectedc<-resc[sum.res.rowsc!=0,]
print(summary(resc))

```

#mama

```

require(annotate)
fit.Symbolsm <- getSYMBOL (rownames(fitmama), "hgu133plus2.db")
resm<-decideTests(fitmama, method="separate",
adjust.method="BH", p.value=0.01)
sum.res.rowsm<-apply(abs(resm),1,sum)
res.selectedm<-resm[sum.res.rowsm!=0,]
print(summary(resm))

```

#prostate

```

require(annotate)
fit.Symbolsp <- getSYMBOL (rownames(fitprostate), "mouse4302.db")

```

```

resp<-decideTests(fitprostate, method="separate",
adjust.method="BH",p.value=0.01)
sum.res.rowsp<-apply(abs(resp),1,sum)
res.selectedp<-resp[sum.res.rowsp!=0,]
print(summary(resp))

#epidermoid
require(annotate)
fit.Symbolse <- getSYMBOL (rownames(fitepidermoid), "hgu133plus2.db")
rese<-decideTests(fitepidermoid, method="separate",
adjust.method="BH",p.value=0.01)
sum.res.rowse<-apply(abs(rese),1,sum)
res.selectede<-rese[sum.res.rowse!=0,]
print(summary(rese))

#Lung
require(annotate)
fit.Symbolsl <- getSYMBOL (rownames(fitlung), "hgu133plus2.db")
resl<-decideTests(fitlung, method="separate",
adjust.method="BH",p.value=0.01)
sum.res.rowsl<-apply(abs(resl),1,sum)
res.selectedl<-resl[sum.res.rowsl!=0,]
print(summary(resl))

```

8. Visualización de los perfiles de expresión.

```

require(gplots)
#coLon
probeNamesc<-rownames(resc)
probeNames.selectedc<-probeNamesc[sum.res.rowsc!=0]
exprs2clusterc <-exprs(eset_rmacolon)[probeNames.selectedc,]
heatmap.2(exprs2clusterc, col=bluered(75), scale="row", key=TRUE,
symkey=FALSE, density.info="none", trace="none", cexCol=1)

#epidermoid
probeNamese<-rownames(rese)
probeNames.selectede<-probeNamese[sum.res.rowse!=0]
exprs2clustere <-exprs(eset_rmaepidermoid)[probeNames.selectede,]
heatmap.2(exprs2clustere, col=bluered(75), scale="row", key=TRUE,
symkey=FALSE, density.info="none", trace="none", cexCol=1)

#mama
probeNamesm<-rownames(resm)
probeNames.selectedm<-probeNamesm[sum.res.rowsm!=0]
exprs2clusterm <-exprs(eset_rmamama)[probeNames.selectedm,]
heatmap.2(exprs2clusterm, col=bluered(75), scale="row", key=TRUE,
symkey=FALSE, density.info="none", trace="none", cexCol=1)

#prostate
probeNamesp<-rownames(resp)
probeNames.selectedp<-probeNamesp[sum.res.rowsp!=0]
exprs2clusterp <-exprs(eset_rmaprostate)[probeNames.selectedp,]
heatmap.2(exprs2clusterp, col=bluered(75), scale="row", key=TRUE,
symkey=FALSE, density.info="none", trace="none", cexCol=1)

```

```

#Lung
probeNames1<-rownames(res1)
probeNames.selected1<-probeNames1[sum.res.rows1!=0]
exprs2cluster1 <-exprs(eset_rmalung)[probeNames.selected1,]
heatmap.2(exprs2cluster1, col=bluered(75), scale="row", key=TRUE,
symkey=FALSE, density.info="none", trace="none", cexCol=1)

```

9. Análisis de significación biológica (“Gene Enrichment Analysis”)

9.1. Gene ontology

```

#colOn
require(annotate)
#BIOLOGICAL PROCESS
whichGenesc <- topTabc["adj.P.Val"]<0.05
topTabID <- row.names(topTabc)[whichGenesc]
genes.totalidc<-row.names(eset_rmacolon)
# Escogemos los grupos de sondas a incluir en el análisis
geneIdsc<-
unique(getEG(as.character(topTabID[whichGenesc]),"hgu133a.db"))
geneIdsc<-geneIdsc[!is.na(geneIdsc)]
# Definimos el universo de genes: todos los que se han incluido en el
análisis
entrezUniversec<-unique(getEG(as.character(genes.totalidc),
"hgu133a.db"))
entrezUniversec<-entrezUniversec[!is.na(entrezUniversec)]
# Ejecutamos los análisis
paramsc<-new("GOHyperGParams", geneIds=geneIdsc,
universeGeneIds=entrezUniversec, annotation=c("hgu133a.db"),
ontology="BP", pvalueCutoff=0.05, conditional=FALSE,
testDirection="over")
resultBPcBP<-hyperGTest(paramsc)
# Creamos un informe html con los resultados
GOfilenamec <- file.path(resultsDir, paste("GOResultscolonBP.html",
sep=""))
htmlReport(resultBPcBP, file=GOfilenamec,
summary.args=list("htmlLinks"=TRUE))

#Molecular function
paramscMF<-new("GOHyperGParams", geneIds=geneIdsc,
universeGeneIds=entrezUniversec, annotation=c("hgu133a.db"),
ontology="MF", pvalueCutoff=0.05, conditional=FALSE,
testDirection="over")
resultBPcMF<-hyperGTest(paramscMF)
GOfilenamecMF <- file.path(resultsDir, paste("GOResultscolonMF.html",
sep=""))
htmlReport(resultBPcMF, file=GOfilenamecMF,
summary.args=list("htmlLinks"=TRUE))
resultBPcMF

#CELLULAR COMPONENT
paramscCC<-new("GOHyperGParams", geneIds=geneIdsc,
universeGeneIds=entrezUniversec, annotation=c("hgu133a.db"),
ontology="CC", pvalueCutoff=0.05, conditional=FALSE,
testDirection="over")

```

```

resultBPcCC<-hyperGTest(paramsCC)
GOfilenamecCC <- file.path(resultsDir, paste("GOResultscolonCC.html",
sep=""))
htmlReport(resultBPcCC, file=GOfilenamecCC,
summary.args=list("htmlLinks"=TRUE))
resultBPcCC

#epidermoide
#BIOLOGICAL PROCESS
whichGenese <- topTabe["adj.P.Val"]<0.05
topTabIDE <- row.names(topTabe)[whichGenese]
genes.totalide<-row.names(eset_rmaepidermoid)
geneIdse<-
unique(getEG(as.character(topTabIDE[whichGenese]), "hgu133plus2.db"))
geneIdse<-geneIdse[!is.na(geneIdse)]
entrezUniversee<-unique(getEG(as.character(genes.totalide),
"hgu133plus2.db"))
entrezUniversee<-entrezUniversee[!is.na(entrezUniversee)]
paramse<-new("GOHyperGParams", geneIds=geneIdse,
universeGeneIds=entrezUniversee, annotation=c("hgu133plus2.db"),
ontology="BP", pvalueCutoff=0.05, conditional=FALSE,
testDirection="over")
resultBPe<-hyperGTest(paramse)
GOfilenameee <- file.path(resultsDir,
paste("GOResultsepidermoidBP.html", sep=""))
htmlReport(resultBPe, file=GOfilenameee,
summary.args=list("htmlLinks"=TRUE))
resultBPe

#MOLECULAR FUNCTION
paramseMF<-new("GOHyperGParams", geneIds=geneIdse,
universeGeneIds=entrezUniversee, annotation=c("hgu133plus2.db"),
ontology="MF", pvalueCutoff=0.05, conditional=FALSE,
testDirection="over")
resultBPeMF<-hyperGTest(paramseMF)
GOfilenameeeMF <- file.path(resultsDir,
paste("GOResultsepidermoidMF.html", sep=""))
htmlReport(resultBPeMF, file=GOfilenameeeMF,
summary.args=list("htmlLinks"=TRUE))
resultBPeMF

#CELULULAR COMPONENT
paramseCC<-new("GOHyperGParams", geneIds=geneIdse,
universeGeneIds=entrezUniversee, annotation=c("hgu133plus2.db"),
ontology="CC", pvalueCutoff=0.05, conditional=FALSE,
testDirection="over")
resultBPeCC<-hyperGTest(paramseCC)
GOfilenameeeCC <- file.path(resultsDir,
paste("GOResultsepidermoidCC.html", sep=""))
htmlReport(resultBPeCC, file=GOfilenameeeCC,
summary.args=list("htmlLinks"=TRUE))
resultBPeCC

#mama
#BIOLOGICAL PROCESS

```

```

whichGenesm <- topTabm["adj.P.Val"]<0.05
topTabIDm <- row.names(topTabm)[whichGenesm]
genes.totalidm<-row.names(eset_rmamama)
geneIdsm<-
unique(getEG(as.character(topTabIDm[whichGenesm]),"hgu133plus2.db"))
geneIdsm<-geneIdsm[!is.na(geneIdsm)]
entrezUniversem<-unique(getEG(as.character(genes.totalidm),
"hgu133plus2.db"))
entrezUniversem<-entrezUniversem[!is.na(entrezUniversem)]
paramsm<-new("GOHyperGParams", geneIds=geneIdsm,
universeGeneIds=entrezUniversem, annotation=c("hgu133plus2.db"),
ontology="BP", pvalueCutoff=0.05, conditional=FALSE,
testDirection="over")
resultBPm<-hyperGTest(paramsm)
GOfilenamem <- file.path(resultsDir, paste("GOResultsmamaBP", sep=""))
htmlReport(resultBPm, file=GOfilenamem,
summary.args=list("htmlLinks"=TRUE))
resultBPm

#MOLECULAR FUNCTION
paramsmMF<-new("GOHyperGParams", geneIds=geneIdsm,
universeGeneIds=entrezUniversem, annotation=c("hgu133plus2.db"),
ontology="MF", pvalueCutoff=0.05, conditional=FALSE,
testDirection="over")
resultBPmMF<-hyperGTest(paramsmMF)
GOfilenamemMF <- file.path(resultsDir, paste("GOResultsmamaMF.html",
sep=""))
htmlReport(resultBPmMF, file=GOfilenamemMF,
summary.args=list("htmlLinks"=TRUE))
resultBPmMF

#CELLULAR COMPONENT
paramsmCC<-new("GOHyperGParams", geneIds=geneIdsm,
universeGeneIds=entrezUniversem, annotation=c("hgu133plus2.db"),
ontology="CC", pvalueCutoff=0.05, conditional=FALSE,
testDirection="over")
resultBPmCC<-hyperGTest(paramsmCC)
GOfilenamemCC <- file.path(resultsDir, paste("GOResultsmamaCC.html",
sep=""))
htmlReport(resultBPmCC, file=GOfilenamemCC,
summary.args=list("htmlLinks"=TRUE))
resultBPmCC

#próstata
#BIOLOGICAL PROCESS
require(mouse4302.db)
whichGenesp <- topTabp["adj.P.Val"]<0.05
topTabIDp <- row.names(topTabp)[whichGenesp]
genes.totalidp<-row.names(eset_rmaprostate)
geneIdsp<-
unique(getEG(as.character(topTabIDp[whichGenesp]),"mouse4302.db"))
geneIdsp<-geneIdsp[!is.na(geneIdsp)]
entrezUniversep<-unique(getEG(as.character(genes.totalidp),
"mouse4302.db"))
entrezUniversep<-entrezUniversep[!is.na(entrezUniversep)]

```

```

paramsp<-new("GOHyperGParams", geneIds=geneIdsp,
universeGeneIds=entrezUniversep, annotation=c("mouse4302.db"),
ontology="BP", pvalueCutoff=0.05, conditional=FALSE,
testDirection="over")
resultBPp<-hyperGTest(paramsp)
GOfilenamep <- file.path(resultsDir, paste("GOResultsprostateBP.html",
sep=""))
htmlReport(resultBPp, file=GOfilenamep,
summary.args=list("htmlLinks"=TRUE))
resultBPp

#MOLECULAR FUNCTION
paramspMF<-new("GOHyperGParams", geneIds=geneIdsp,
universeGeneIds=entrezUniversep, annotation=c("mouse4302.db"),
ontology="MF", pvalueCutoff=0.05, conditional=FALSE,
testDirection="over")
resultBPpMF<-hyperGTest(paramspMF)
GOfilenamepMF <- file.path(resultsDir,
paste("GOResultsprostateMF.html", sep=""))
htmlReport(resultBPpMF, file=GOfilenamepMF,
summary.args=list("htmlLinks"=TRUE))
resultBPpMF

#CELLULAR COMPONENT
paramspCC<-new("GOHyperGParams", geneIds=geneIdsp,
universeGeneIds=entrezUniversep, annotation=c("mouse4302.db"),
ontology="CC", pvalueCutoff=0.05, conditional=FALSE,
testDirection="over")
resultBPpCC<-hyperGTest(paramspCC)
GOfilenamepCC <- file.path(resultsDir,
paste("GOResultsprostateCC.html", sep=""))
htmlReport(resultBPpCC, file=GOfilenamepCC,
summary.args=list("htmlLinks"=TRUE))
resultBPpCC

#non-small cell lung
#BIOLOGICAL PROCESS
whichGenes1 <- topTabl["adj.P.Val"]<0.05
topTabID1 <- row.names(topTabl)[whichGenes1]
genes.totalid1<-row.names(eset_rmalung)
geneIdsl<-
unique(getEG(as.character(topTabID1[whichGenes1]),"hgu133plus2.db"))
geneIdsl<-geneIdsl[!is.na(geneIdsl)]
entrezUniverse1<-unique(getEG(as.character(genes.totalid1),
"hgu133plus2.db"))
entrezUniverse1<-entrezUniverse1[!is.na(entrezUniverse1)]
paramsl<-new("GOHyperGParams", geneIds=geneIdsl,
universeGeneIds=entrezUniverse1, annotation=c("hgu133plus2.db"),
ontology="BP", pvalueCutoff=0.05, conditional=FALSE,
testDirection="over")
resultBP1<-hyperGTest(paramsl)
GOfilename1 <- file.path(resultsDir, paste("GOResultslungBP.html",
sep=""))
htmlReport(resultBP1, file=GOfilename1,

```

```

summary.args=list("htmlLinks"=TRUE))
resultBP1

#MOLECULAR FUCTION
params1MF<-new("GOHyperGParams", geneIds=geneIds1,
universeGeneIds=entrezUniverse1, annotation=c("hgu133plus2.db"),
ontology="MF", pvalueCutoff=0.05, conditional=FALSE,
testDirection="over")
resultBP1MF<-hyperGTest(params1MF)
GOfilename1MF <- file.path(resultsDir, paste("GOResultslungMF.html",
sep=""))
htmlReport(resultBP1MF, file=GOfilename1MF,
summary.args=list("htmlLinks"=TRUE))
resultBP1MF

#CELLULAR COMPONENT
params1CC<-new("GOHyperGParams", geneIds=geneIds1,
universeGeneIds=entrezUniverse1, annotation=c("hgu133plus2.db"),
ontology="CC", pvalueCutoff=0.05, conditional=FALSE,
testDirection="over")
resultBP1CC<-hyperGTest(params1CC)
GOfilename1CC <- file.path(resultsDir, paste("GOResultslungCC.html",
sep=""))
htmlReport(resultBP1CC, file=GOfilename1CC,
summary.args=list("htmlLinks"=TRUE))
resultBP1CC

```

9.2. KEGG pathways

```

#coLon
KEGGparamsc = new("KEGGHyperGParams",
  geneIds=geneIdsc, universeGeneIds=entrezUniversec,
  annotation="hgu133a.db",
  pvalueCutoff=0.05, testDirection="over")
KEGGhyperc = hyperGTest(KEGGparamsc)
KEGGhyperc

KEGGfilenameec <-file.path(resultsDir, paste("KEGGResultscolon.html",
sep=""))
htmlReport(KEGGhyperc, file = KEGGfilenameec,
summary.args=list("htmlLinks"=TRUE))

#epidermoid
KEGGparamse = new("KEGGHyperGParams",
  geneIds=geneIdse, universeGeneIds=entrezUniversee,
  annotation="hgu133plus2.db", pvalueCutoff=0.05, testDirection="over")
KEGGhypere = hyperGTest(KEGGparamse)
KEGGhypere

KEGGfilenameee <-file.path(resultsDir,
paste("KEGGResultsepidermoid.html", sep=""))
htmlReport(KEGGhypere, file = KEGGfilenameee,
summary.args=list("htmlLinks"=TRUE))

#mama
KEGGparamsm = new("KEGGHyperGParams",

```

```

geneIds=geneIdsm, universeGeneIds=entrezUniverse,
annotation="hgu133plus2.db", pvalueCutoff=0.05, testDirection="over")
KEGGhyper = hyperGTest(KEGGparams)
KEGGhyper

KEGGfilenamem <-file.path(resultsDir, paste("KEGGResultsmama.html",
sep=""))
htmlReport(KEGGhyper, file = KEGGfilenamem,
summary.args=list("htmlLinks"=TRUE))

#prostate
KEGGparamsp = new("KEGGHyperGParams", geneIds=geneIdsp,
universeGeneIds=entrezUniverse, annotation="mouse4302.db",
pvalueCutoff=0.05, testDirection="over")
KEGGhyperp = hyperGTest(KEGGparamsp)
KEGGhyperp

KEGGfilenameep <-file.path(resultsDir,
paste("KEGGResultsprostate.html", sep=""))
htmlReport(KEGGhyperp, file = KEGGfilenameep,
summary.args=list("htmlLinks"=TRUE))

#Lung
KEGGparamsl = new("KEGGHyperGParams", geneIds=geneIdsl,
universeGeneIds=entrezUniverse, annotation="hgu133plus2.db",
pvalueCutoff=0.05, testDirection="over")
KEGGhyperl = hyperGTest(KEGGparamsl)
KEGGhyperl

KEGGfilenamel <-file.path(resultsDir, paste("KEGGResultslung.html",
sep=""))
htmlReport(KEGGhyperl, file = KEGGfilenamel,
summary.args=list("htmlLinks"=TRUE))

```

10. Búsqueda de genes en las 3 vías seleccionadas

```
library(globaltest)
```

```

#colon
#Pathways in cancer KEGG pathways
groupcolon<-c(0,0,1,1,1)
rescolon<-gtKEGG(groupcolon, eset_rmacolon, id=c("05200"),
multitest="BH")
ftcolon <- features(rescolon, alias=hgu133aSYMBOL, plot=FALSE)
head(ftcolon,10)

#colon
#small cell lung cancer KEGG pathway
rescolon2<-gtKEGG(groupcolon,eset_rmacolon, id=c("05222"),
multitest="BH")
ftcolon2 <- features(rescolon2, alias=hgu133aSYMBOL, plot=FALSE)
head(ftcolon2,10)

#colon
#p53 signaling KEGG pathway
rescolon3<-gtKEGG(groupcolon,eset_rmacolon, id=c("04115"),

```

```

multttest="BH")
ftcolon3 <- features(rescolon3, alias=hgu133aSYMBOL, plot=FALSE)
head(ftcolon3,20)

#mama
#Pathways in cancer KEGG pathways
groupmama<-c(0,0,1,1)
resmama<-gtKEGG(groupmama, eset_rmamama, id=c("05200"), multttest="BH")
ftmama <- features(resmama, alias=hgu133plus2SYMBOL, plot=FALSE)
head(ftmama,120)

#mama
#small cell Lung cancer KEGG pathway
resmama2<-gtKEGG(groupmama,eset_rmamama, id=c("05222"), multttest="BH")
ftmama2 <- features(resmama2, alias=hgu133plus2SYMBOL, plot=FALSE)
head(ftmama2,40)

#mama
#p53 signaling KEGG pathway
resmama3<-gtKEGG(groupmama,eset_rmamama, id=c("04115"), multttest="BH")
ftmama3 <- features(resmama3, alias=hgu133plus2SYMBOL, plot=FALSE)
head(ftmama3,50)

#epidermoid
#Pathways in cancer KEGG pathways
groupepidermoid<-c(0,0,1,1)
resepidermoid<-gtKEGG(groupepidermoid, eset_rmaepidermoid,
id=c("05200"), multttest="BH")
ftepidermoid <- features(resepidermoid, alias=hgu133plus2SYMBOL,
plot=FALSE)
head(ftepidermoid,180)

#epidermoid
#small cell Lung cancer KEGG pathway
resepidermoid2<-gtKEGG(groupepidermoid,eset_rmaepidermoid,
id=c("05222"), multttest="BH")
ftepidermoid2 <- features(resepidermoid2, alias=hgu133plus2SYMBOL,
plot=FALSE)
head(ftepidermoid2,80)

#epidermoid
#p53 signaling KEGG pathway
resepidermoid3<-gtKEGG(groupepidermoid,eset_rmaepidermoid,
id=c("04115"), multttest="BH")
ftepidermoid3 <- features(resepidermoid3, alias=hgu133plus2SYMBOL,
plot=FALSE)
head(ftepidermoid3,200)

#Lung
#Pathways in cancer KEGG pathways
grouplung<-c(0,0,1,1)
reslung<-gtKEGG(grouplung, eset_rmalung, id=c("05200"), multttest="BH")
ftlung <- features(reslung, alias=hgu133plus2SYMBOL, plot=FALSE)
head(ftlung,190)

```

```

#Lung
#small cell Lung cancer KEGG pathway
reslung2<-gtKEGG(grouplung,eset_rmalung, id=c("05222"), multttest="BH")
ftlung2 <- features(reslung2, alias=hgu133plus2SYMBOL, plot=FALSE)
head(ftlung2,80)

#Lung
#p53 signaling KEGG pathway
reslung3<-gtKEGG(grouplung,eset_rmalung, id=c("04115"), multttest="BH")
ftlung3 <- features(reslung3, alias=hgu133plus2SYMBOL, plot=FALSE)
head(ftlung3)

#prostate
#Pathways in cancer KEGG pathways
groupprostate<-c(0,0,0,1,1,1)
resprostate<-gtKEGG(groupprostate, eset_rmaprostate, id=c("05200"),
multttest="BH")
ftprostate <- features(resprostate, alias=mouse4302SYMBOL, plot=FALSE)
head(ftprostate)

#prostate
#small cell Lung cancer KEGG pathway
resprostate2<-gtKEGG(groupprostate,eset_rmaprostate, id=c("05222"),
multttest="BH")
ftprostate2 <- features(resprostate2, alias=mouse4302SYMBOL,
plot=FALSE)
head(ftprostate2)

#prostate
#p53 signaling KEGG pathway
resprostate3<-gtKEGG(groupprostate,eset_rmaprostate, id=c("04115"),
multttest="BH")
ftprostate3 <- features(resprostate3, alias=mouse4302SYMBOL,
plot=FALSE)
head(ftprostate3)

```