



Universidad Nacional Autónoma de México

Centro de Física Aplicada y Tecnología Avanzada

Identificación de variantes genéticas candidatas a desarrollar melanoma familiar y su exploración funcional

T E S I S

Que para obtener el título de
Licenciada en Tecnología

P R E S E N T A

Rebeca Olvera León

Directora de tesis
Dra. Carla Daniela Robles Espinoza

Codirector de tesis
Dr. Christian Molina Aguilar



Juriquilla, Querétaro, 2020



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

La presente tesis está dedicada con todo el cariño y amor a mis padres, sin quienes nada de esto sería posible. Gracias infinitas por su apoyo, entrega y dedicación a lo largo de toda mi vida. Por su acompañamiento en mi carrera universitaria, por siempre creer en mí, por confiar en mis decisiones, por entender mis desvelos, por soportar mis frustraciones, por ayudarme en todo lo que pudieron. Por siempre cuidar que no me faltara nada, por todas las horas de sueño sacrificadas, por todos los consejos, por toda la paciencia, por todo el amor.

Gracias, Rosa Elena y Rodolfo.
Esto es de ustedes.

El presente trabajo fue realizado en el Grupo de Genética del Cáncer y Bioinformática, del Laboratorio Internacional de Investigación sobre el Genoma Humano, en el campus Juriquilla de la Universidad Nacional Autónoma de México. Con financiamiento del Consejo Nacional de Ciencia y Tecnología a través de un donativo FOSISS (A3-S-31603), de la Fundación Wellcome (Ref. 204562/Z/16/Z), del Medical Research Council (Ref. MR/S01473X/1), de la Academy of Medical Sciences (Ref. NAF\R2\180782) y del Programa de Apoyo a Proyectos de Investigación Tecnológica (Refs. IA200318 e IA202020).

Agradecimientos

A la Universidad Nacional Autónoma de México, especialmente al Centro de Física Aplicada y Tecnología Avanzada y a todos mis profesores, por la formación académica y profesional durante mi etapa universitaria.

Un especial agradecimiento a mi tutora, la Dra. Carla Daniela Robles Espinoza, por creer en mí desde el primer momento. Gracias por siempre brindarme su apoyo, tiempo y disposición. Gracias por la confianza, formación y aprendizaje recibidos durante todo este tiempo.

A mi tutor, el Dr. Christian Molina Aguilar, por todo su tiempo y paciencia brindado en el aprendizaje de las técnicas de laboratorio. Gracias por la libertad y confianza al momento de realizar experimentos, por la formación, apoyo y disposición en todo momento.

Al Grupo de Genética del Cáncer y Bioinformática del Laboratorio Internacional de Investigación sobre el Genoma Humano (LIIGH), especialmente a la LBT Estefanía Vázquez, por su apoyo en la realización de experimentos y al Dr. Raúl Ossio Vela, por su aportación en el análisis bioinformático.

Al personal académico y administrativo del LIIGH, especialmente a Abigayl Hernández Jiménez, QIA Alejandra Castillo Carbajal y QIA Carina Uribe Díaz. Al Ing. Jair García e Ing. Luis Aguilar, también miembros del Laboratorio Nacional de Visualización Científica Avanzada, por su apoyo con los servidores de la UNAM en todo momento.

Al grupo de trabajo del Dr. David Adams del Wellcome Sanger Institute, por la donación de plásmidos empleados en los experimentos y por el apoyo de su personal en la realización de técnicas, así como por la contribución de fondos para nuestro proyecto (Ref. 204562/Z/16/Z). A la Prof. Julia Newton-Bishop y al Prof. Tim Bishop de la Universidad de Leeds, por su apoyo en el estudio de casos de melanoma. Al Medical Research Council (Ref. MR/S01473X/1), The Academy of Medical Sciences (Ref. NAF\R2\180782) y al Consejo Nacional de Ciencia y Tecnología (CONACYT) (Ref. A3-S-31603) por los fondos provistos para la realización de este proyecto.

Al Programa de Apoyo a Proyectos de Investigación Tecnológica (PAPIIT) de la UNAM (Refs. IA200318 e IA202020), por la asignación de becas de licenciatura.

A mi comité evaluador: Dra. Luz María López Marín (CFATA), Dra. Rocío Brenda Anguiano Serrano (INB), Dr. Diego Ortega Del Vecchyo (LIIGH), y Dra. Blanca Edith Millán Chiu (CFATA). Gracias por los comentarios y aportaciones hechas a este trabajo.

A mi familia, amigos, compañeros y todos aquellos que me acompañaron durante estos años. Gracias por su apoyo, motivación y cariño. Gracias por todo lo compartido. Ustedes hicieron el camino más ligero.

Contenido

Resumen.....	1
Estructura de la tesis.....	2
Capítulo 1: Análisis bioinformático de secuencias obtenidas de familias con melanoma para la identificación de variantes genéticas de alto riesgo.....	3
1.1. Introducción.....	3
1.1.1. Cáncer	3
1.1.2. NGS en el diagnóstico y tratamiento del cáncer.....	4
1.1.3. Melanoma	7
1.1.3.1. Factores genéticos en el desarrollo de melanoma.....	8
1.1.4. Variantes genéticas	10
1.1.4.1. Formato de Identificación de Variantes	10
1.2. Justificación	12
1.3. Antecedentes.....	12
1.4. Objetivos	12
1.5. Hipótesis.....	13
1.6. Metodología.....	13
1.6.1 Datos de secuenciación	13
1.6.2 Filtrado de variantes	14
1.7. Resultados y discusión.....	16
1.7.1. <i>ESR1</i> Y537C.....	17
1.8. Conclusión.....	18
Capítulo 2: Evaluación funcional de la mutación <i>ESR1</i> Y537C en la línea celular NIH 3T3.....	19
2.1. Introducción.....	19
2.1.1. Estrógenos y receptores de estrógenos.....	19
2.1.1.1. REs en cáncer de mama	21
2.1.1.1.1 Mutaciones en el aminoácido Y537	22
2.1.1.2. Melanoma y estrógenos.....	22
2.2. Justificación	23
2.3. Objetivo	24
2.4. Hipótesis.....	24
2.5. Metodología.....	24
2.5.1 Reactivos.....	24

2.5.2 Digestiones y ligaciones	26
2.5.2.1 Digestión del plásmido MA-RQ	26
2.5.2.2 Digestión del plásmido BABE-puro	26
2.5.2.3 Ligación del plásmido BABE-puro	26
2.5.3 Transformación de bacterias XL-Blue	27
2.5.3.1 Purificación y cuantificación del plásmido	28
2.5.4 Cultivo y transfección de células NIH 3T3	29
2.5.4.1 Transfección de células	29
2.5.5 Ensayo de formación de focos	30
2.5.5.1 Tinción de células	31
2.6. Resultados y discusión	31
2.6.1 Trabajo de laboratorio	31
2.6.2 Ensayo de formación de focos	32
2.7. Conclusiones	35
2.8. Perspectivas	35
Anexo I: Experimentos adicionales	36
AI.1 Réplica de ensayo de formación de focos	36
AI.2 Ensayo de raspado celular	38
AI.3 Ensayo de proliferación celular	41
Anexo II: Material complementario	43
Referencias	47

Resumen

El melanoma cutáneo es el cáncer de piel más mortal. Es un tumor maligno originado en los melanocitos cuyo factor de riesgo ambiental más importante es la exposición a la radiación ultravioleta. Sin embargo, se estima que el 10% de casos de melanoma son familiares, por lo que la causa es muy probablemente genética. Algunas de las variantes de línea germinal que son causales en este tipo de cáncer se han visto en los genes *CDKN2A*, *CDK4*, *BAP1*, *POT1* y *TERT*, conformando alrededor del 50% de casos familiares; no obstante, se desconoce la causa genética del porcentaje restante, ya que consiste en familias que no presentan variantes en estos genes.

En el presente trabajo se realizó la identificación de variantes genéticas candidatas a desarrollar melanoma a partir del análisis bioinformático de genomas y exomas de 314 miembros de 135 familias europeas, australianas y estadounidenses con este padecimiento, las cuales no poseen variantes en genes de predisposición conocidos. Se realizó un filtrado de acuerdo con criterios de presencia en catálogos de variación genética humana y posibles consecuencias en la proteína, con lo que se obtuvo una lista de variantes candidatas, las cuales pueden ser evaluadas posteriormente por métodos experimentales.

Una de las variantes obtenidas fue Y537C, correspondiente al gen *ESR1* (el cual codifica para el receptor de estrógenos alfa), y resulta sumamente interesante debido a que no se ha reportado ni en melanoma ni en línea germinal. Aunado a ello, el gen *ESR1* es un foco mutacional en cáncer de mama, ya que mutaciones en el aminoácido 537 han sido relacionadas con resistencia endócrina a tratamiento en pacientes con cáncer de mama metastásico. Es por ello que esta mutación es una buena candidata en la búsqueda de variantes genéticas que pueden contribuir al desarrollo de melanoma.

Como parte de este proyecto, se evaluó el efecto de *ESR1* Y537C en células NIH 3T3 (fibroblastos de ratón) mediante ensayos de proliferación. Se utilizó esta línea celular como parte del establecimiento de controles en el laboratorio. Los resultados obtenidos sugieren que esta mutación induce un efecto oncogénico en NIH 3T3; sin embargo, se requiere la obtención de más réplicas para su validación, así como de su evaluación en células melanocíticas y de melanoma para conocer su papel en el desarrollo de este cáncer.

Estructura de la tesis

La presente tesis está dividida en dos capítulos: el primero corresponde a un análisis bioinformático y el segundo a trabajo de laboratorio. El capítulo 1 se titula *Análisis bioinformático de secuencias obtenidas de familias con melanoma para la identificación de variantes genéticas de alto riesgo* y consiste en el análisis de datos de secuenciación obtenidos de familias con melanoma cuya causa de padecimiento de este cáncer es desconocida. La finalidad de esta parte del proyecto fue encontrar nuevas variantes genéticas que pudieran estar involucradas en el desarrollo de melanoma.

El capítulo 2 se titula *Evaluación funcional de la mutación ESR1 Y537C en la línea celular NIH 3T3* y consiste en la evaluación *in vitro* de la variante candidata encontrada en el análisis bioinformático realizado.

Asimismo, al final de ambos capítulos se incluyen dos anexos. El anexo 1 consta de experimentos adicionales que fueron realizados como parte de la estandarización de técnicas y protocolos, los cuales no fueron presentados como parte de los resultados de este trabajo debido a que los controles positivos y negativos fallaron. Y, finalmente, el anexo 2 contiene material complementario.

1

Análisis bioinformático de secuencias obtenidas de familias con melanoma para la identificación de variantes genéticas de alto riesgo

1.1. Introducción

1.1.1. Cáncer

El cáncer es un conjunto de enfermedades caracterizadas por la proliferación descontrolada de células debido a la acumulación de mutaciones, por lo cual es considerada una enfermedad del genoma¹. En condiciones fisiológicas, la división celular es el mecanismo encargado de reemplazar células que han envejecido o que han sufrido daños y han entrado en apoptosis, necrosis o necroptosis. Cuando los genes que regulan este proceso sufren alteraciones, las células pueden comenzar a dividirse incesantemente, lo que puede conducir a la formación de un tumor y, eventualmente, a la capacidad de éste para invadir otras partes del cuerpo a través de la sangre o los vasos linfáticos, produciendo metástasis [Figura 1.1]^{2,3}. Estas alteraciones pueden ser causadas por agentes exógenos como la exposición a carcinógenos químicos, físicos o biológicos (como el tabaco, la radiación UV e infecciones por virus, respectivamente), o por factores endógenos, como mutaciones en la maquinaria de reparación de ADN celular, las cuales pueden ser adquiridas con el tiempo (somáticas) o pueden estar presentes desde el nacimiento (de línea germinal)^{2,5}.

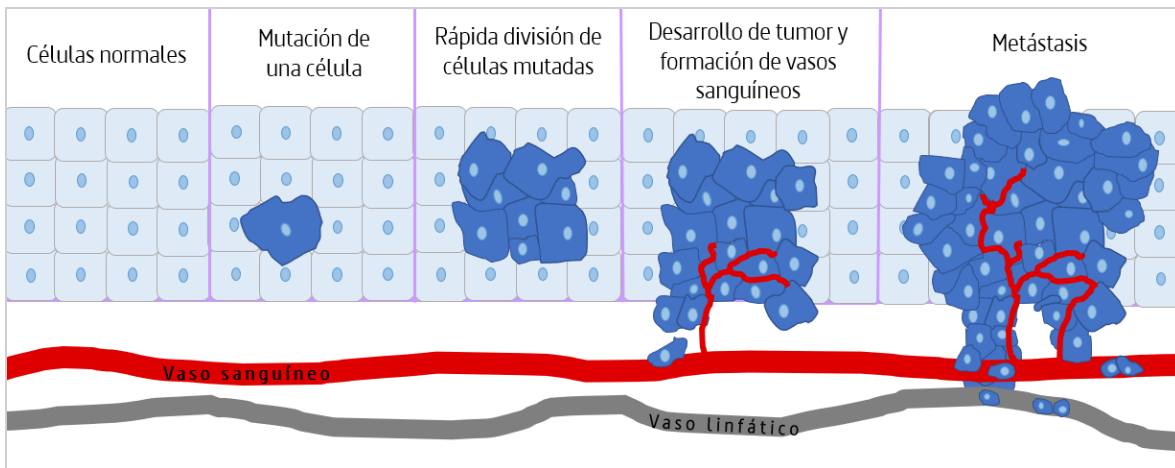


Figura 1.1 | Representación esquemática del desarrollo del cáncer. (Adaptada de The Cancer Institute NSW).

De acuerdo con la Organización Mundial de la Salud (OMS), el cáncer es la segunda causa principal de muerte en todo el mundo, justo debajo de las enfermedades cardiovasculares, siendo el cáncer de pulmón, mama y colorrectal los tipos más comunes. En 2018 se registraron 18.1 millones de nuevos casos de cáncer, así como 9.6 millones de muertes debidas a esta enfermedad. Asimismo, se estima que 1 de cada 5 hombres y 1 de cada 6 mujeres desarrollarán cáncer a lo largo de su vida⁶.

Debido a lo anterior, actualmente las líneas de investigación sobre el cáncer se enfocan en explicar los mecanismos biológicos detrás de este padecimiento, en la identificación de potenciales objetivos terapéuticos, así como en la detección temprana. En relación a esta última, existen varias técnicas que en su conjunto permiten el diagnóstico del cáncer, tales como exámenes físicos, pruebas de imagen, exámenes endoscópicos, biopsias y pruebas genéticas^{7,8}. Las pruebas genéticas son utilizadas para buscar mutaciones que podrían aumentar el riesgo a padecer un tipo específico de cáncer y, aunque no es necesariamente hereditario, las personas con antecedentes familiares de esta enfermedad son más propensas a desarrollarla, por lo que se le recomienda someterse a este tipo de pruebas^{9,10}.

1.1.2. NGS en el diagnóstico y tratamiento del cáncer

La secuenciación de nueva generación (NGS, por sus siglas en inglés) es una tecnología que consiste en la secuenciación masiva y paralela del genoma completo o de regiones específicas, de una manera rápida y con pequeñas cantidades de muestra, lo que es de gran ayuda en la prevención, diagnóstico, tratamiento y seguimiento de una enfermedad^{11,12}.

Entre la secuenciación del ADN, hay tres tipos principales utilizados en el cáncer: secuenciación del genoma completo (SGC), secuenciación del exoma completo (SEC), y secuenciación dirigida. Como su nombre lo indica, SGC se utiliza para secuenciar todo el genoma y, por lo tanto, permite conocer las alteraciones genómicas, el descubrimiento de mutaciones conductoras y firmas mutacionales en los genomas del cáncer¹³. Estas últimas corresponden a combinaciones únicas de mutaciones asociadas a determinados factores, las cuales han sido clasificadas por posible agente etiológico. Ejemplos de ellas son la catalogada firma mutacional 4, la cual ha sido observada en cánceres asociados con el consumo de tabaco (cáncer de pulmón, hígado, cabeza y cuello) o la catalogada como firma mutacional 7, la cual está asociada a mutaciones inducidas por radiación UV y ha sido observada en melanoma y carcinoma escamoso de cabeza y cuello¹⁴ [Figura 1.2].

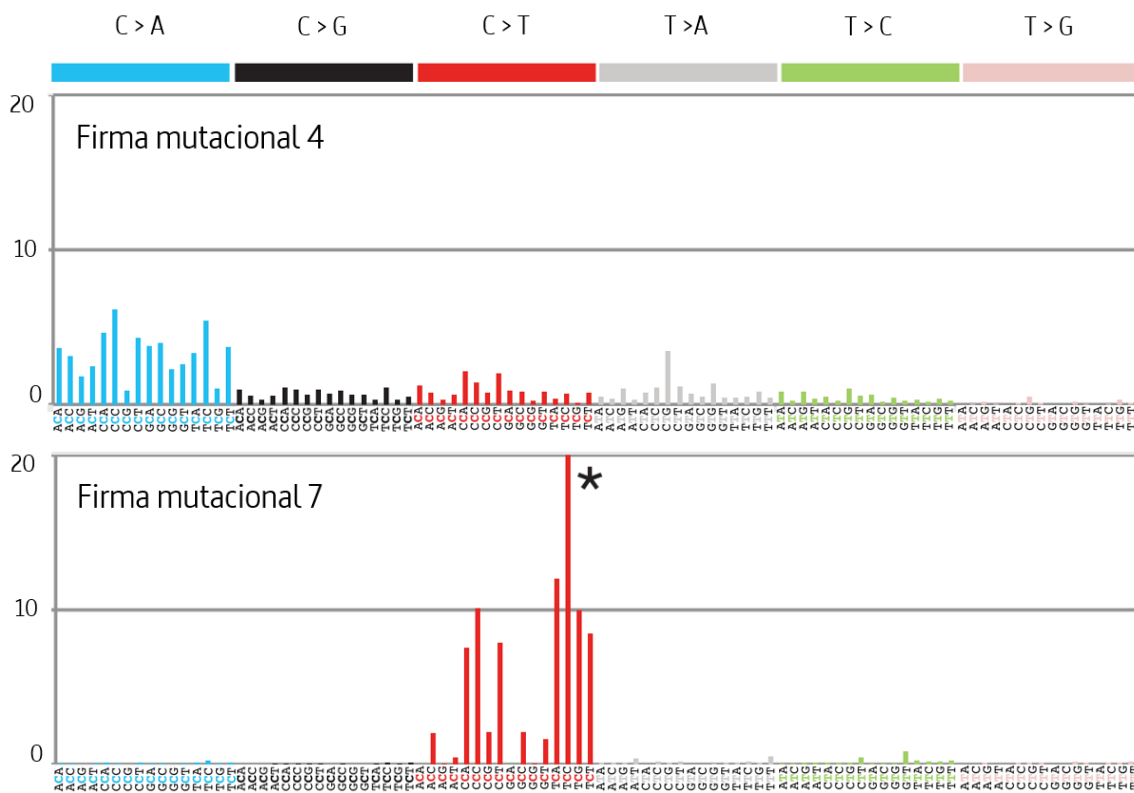


Figura 1.2 | Firmas mutacionales 4 y 7. En las gráficas se muestran firmas debidas a mutaciones por sustitución de base: las seis clases de sustitución se muestran en diferentes colores y se señalan en la parte superior; el contexto genómico en el que ocurre la mutación se presenta en el eje horizontal (compuesto por 96 combinaciones de tres pares de bases) y en el eje vertical se muestra el porcentaje de mutaciones atribuidas a cada tipo de mutación. El asterisco en la firma 7 indica un tipo de mutación superior al 20% del total. (Tomada y modificada de Alexandrov *et al.* 2013¹⁴).

Dado que aproximadamente solo el 2.5% del genoma humano está constituido por regiones codificantes (exones), SEC permite encontrar variantes raras o comunes asociadas con un trastorno o fenotipo sólo en las regiones génicas. Finalmente, la secuenciación dirigida es la secuenciación de genes de interés para una enfermedad específica y, por lo tanto, permite concentrar todos los recursos en las áreas de interés¹¹.

Las secuencias de ADN empleadas para NGS se pueden obtener de células tumorales, saliva o muestras de sangre. Después del proceso de secuenciación, se realiza un análisis bioinformático a través de algoritmos con lo cual se identifican alteraciones y, en consecuencia, se obtiene una lista de variantes que facilita la interpretación de datos genómicos¹⁵. Las variantes que se encuentran en los datos de NGS se separan en dos tipos: las que son predichas patógenas (y que probablemente modifican la secuencia de la proteína, quizás alterando su función, lo que podría elevar el riesgo a padecer cáncer) y las que están predichas como benignas¹⁶. Eventualmente, el trabajo de laboratorio puede brindar soporte experimental a la predicción de la función biológica de las variantes encontradas.

NGS también puede ser útil en el diseño de terapias dirigidas, de tal manera que, a partir de la detección de alteraciones genéticas en muestras de tumores, se podrían tomar decisiones con respecto al tratamiento. Esto es importante debido a que aquellos pacientes que no se someten a una prueba genética no pueden tratarse con un medicamento dirigido y, con frecuencia, eso puede resultar dañino. Por lo tanto, la medicina personalizada (también conocida como medicina de precisión) es uno de los objetivos principales en la investigación del cáncer^{12,17,18}. Además, es posible conocer, mediante la detección de mutaciones posteriores al tratamiento, la razón por la cual un paciente deja de responder a una terapia específica¹⁵. Ejemplos de NGS aplicado al tratamiento del cáncer pueden ser cuando una mutación en el gen *EGFR* está presente en cáncer de pulmón, de tal manera que el paciente es candidato para terapia con inhibidores de la tirosina quinasa¹⁹, o en el caso de melanoma metastásico en pacientes con mutaciones *BRAF V600E* o *V600K*, los cuales pueden ser tratados con una combinación de dabrafenib y trametinib²⁰.

1.1.3. Melanoma

El melanoma cutáneo es el cáncer de piel más agresivo y es altamente resistente a quimioterapia²¹. Si bien no es el cáncer de piel más común, ya que representa menos del 5% de casos²², sí resulta ser el más peligroso debido a que, si no se detecta y se trata a tiempo, tiene una gran capacidad de propagarse a otras partes del cuerpo, como pulmón, hígado, cerebro y huesos^{23,24}. Este cáncer se presenta cuando los melanocitos comienzan a proliferar descontroladamente. Los melanocitos se encuentran en la capa externa de la piel, la epidermis, y su función principal es la producción de melanina, el pigmento que le da color a la piel y al cabello. La melanina, además, contribuye en la protección de la piel por daños causados debido a la radiación UV²⁵, ya que es capaz de absorber esta radiación, evitando así que llegue al núcleo y produzca mutaciones en el ADN^{26,27}. Hasta ahora, la radiación UV ha sido considerada como el factor de riesgo ambiental más importante para el desarrollo de melanoma^{28,29}.

El melanoma se presenta en la piel como una formación anormal de tejido (tumor) que puede estar pigmentado [Fig. 1.3]. Generalmente, las células de melanoma continúan produciendo melanina, por lo cual los tumores son cafés o negros; sin embargo, algunas células dejan de producir este pigmento, por lo que también pueden ser rosados o incluso blancos^{23,30,31}. En los hombres, este cáncer usualmente aparece en el tronco, la cabeza y el cuello, mientras que en las mujeres es más común que se manifieste en las piernas^{23,31}.

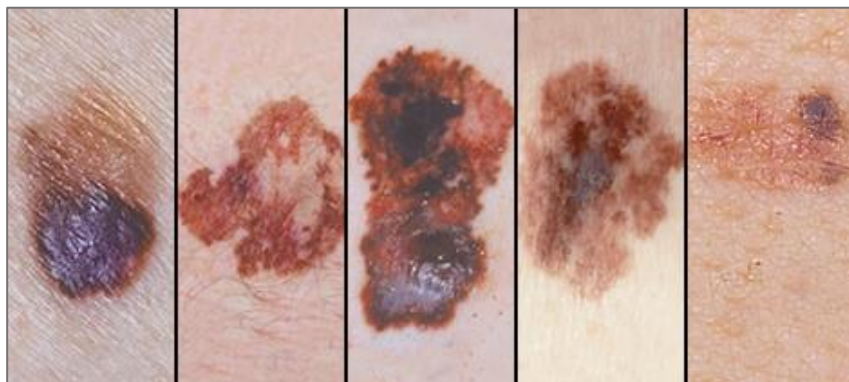


Figura 1.3 | Formas en las que el melanoma se presenta en la piel.
Derechos de la imagen: Mayo Foundation For Medical Education And Research.
Tomada de <https://n9.cl/diq3d>

La incidencia de melanoma se ha visto incrementada de una manera acelerada a nivel mundial en los últimos 50 años, principalmente en poblaciones de piel clara, probablemente debido a que

presentan una menor producción de melanina, lo que implica una disminución en la fotoprotección³². Regiones como Nueva Zelanda, Australia, América del Norte y Europa del Norte, presentan mayores tasas de incidencia de este cáncer [Fig. 1.4], lo que puede estar relacionado con un aumento en el tiempo de exposición a la luz solar o a luz UV artificial, así como a factores genéticos³³.

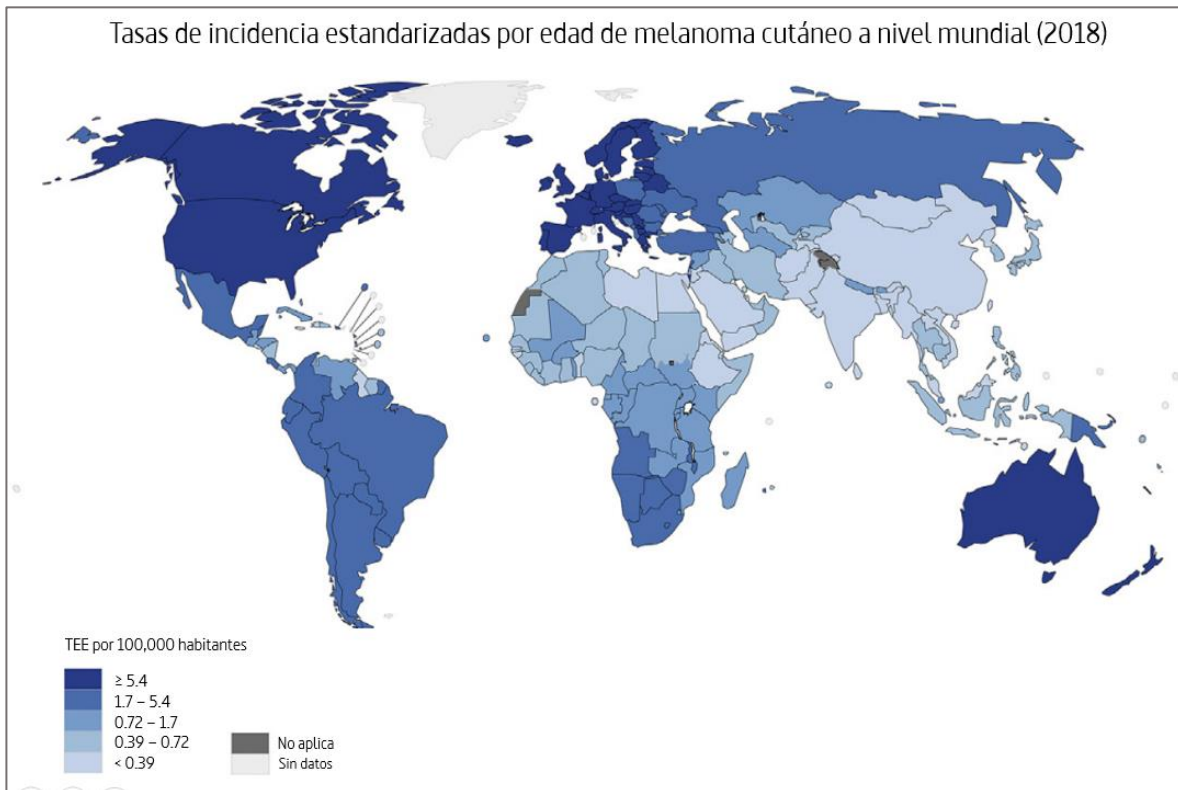


Figura 1.4 | Tasas estandarizadas por edad (TEE) de la incidencia mundial de melanoma en 2018. La incidencia de melanoma es mayor en poblaciones de piel clara, especialmente en regiones como Nueva Zelanda, Australia, Europa del Norte y América del Norte. (Tomada y modificada de Surg. Clin. North Am., Feb 2020, Carr, S., Smith, C. & Wernberg, J., Epidemiology and Risk Factors of Melanoma, Copyright (2020), con permiso de Elsevier³³).

1.1.3.1. Factores genéticos en el desarrollo de melanoma

La causa más frecuente de cáncer son las mutaciones somáticas, las cuales se presentan en cualquier célula del cuerpo (excepto los óvulos y espermatozoides). Sin embargo, las mutaciones en línea germinal (presentes en células sexuales) también juegan un papel importante en el riesgo a padecer cáncer, además de que estas se transmiten a las siguientes generaciones^{34,35}.

Aproximadamente el 10% de casos de melanoma son familiares³⁶, es decir, hay presencia de este cáncer en uno o más familiares de primer o segundo grado³⁷. Hasta el día de hoy, se han identificado varios genes con mutaciones en línea germinal que aumentan la susceptibilidad a desarrollar melanoma, los cuales a su vez pueden dividirse en penetrancia media y alta (siendo la penetrancia definida como la probabilidad de que un individuo portador de una mutación asociada a una enfermedad presente el padecimiento)^{35,38}; algunos de estos genes son *CDKN2A*, *CDK4*, *BAP1*, *TERT*, *POT1*, *MC1R* y *MITF*³⁹, siendo *CDKN2A* el gen más común, ya que está mutado en aproximadamente un 20-40% de casos de melanoma familiar [Tabla 1.1]³⁸⁻⁴⁰. En conjunto, estos genes están involucrados en aproximadamente el 50% de los casos de melanoma familiar; sin embargo, se desconoce la base genética del resto de familias que no presentan variantes en estos genes³⁹ [Fig. 1.5].

Tabla 1.1 | Descripción general de los genes de penetrancia alta y media en la susceptibilidad al melanoma.

Penetrancia	Gen	Proteína codificada	Función
Alta	<i>CDKN2A</i>	p16	Regulador del ciclo celular
		p14	
	<i>CDK4</i>	CDK4	Regulador del ciclo celular
	<i>BAP1</i>	BAP1	Diferenciación de melanocitos y respuesta al daño del ADN
	<i>TERT</i>	Subunidad catalítica de la enzima telomerasa	Elongación de los telómeros
	<i>POT1</i>	POT1	Mantenimiento de telómeros
Media	<i>MC1R</i>	MC1R	Síntesis de melanina y proliferación de melanocitos
	<i>MITF</i>	MITF	Desarrollo y diferenciación de melanocitos

(Tomada y modificada de Rossi *et al.*, 2019³⁸).

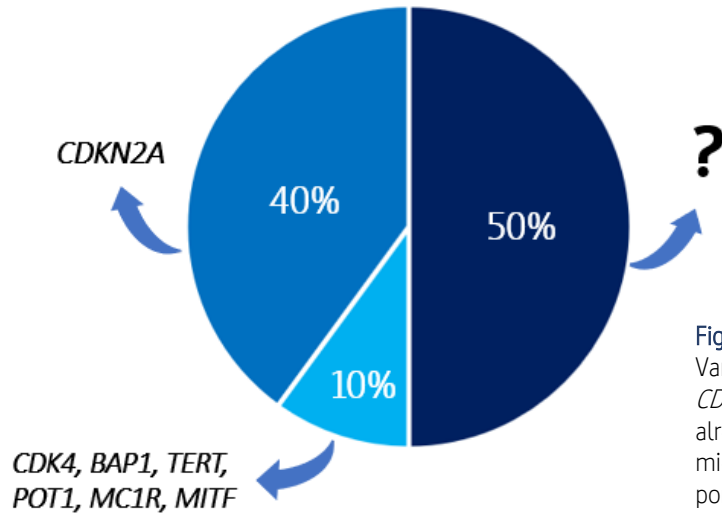


Figura 1.5 | Causas genéticas de melanoma. Variantes de línea germinal en los genes *CDKN2A*, *CDK4*, *BAP1*, *TERT*, *POT1*, *MC1R* y *MITF* conforman alrededor del 50% de casos de melanoma familiar, mientras que se desconoce la causa genética del porcentaje de casos restante³⁸.

1.1.4. Variantes genéticas

Una gran proporción de todo el ADN se comparte entre todos los humanos; sin embargo, el genoma de cualquier individuo posee millones de variaciones genéticas que hacen que cada persona sea única. Una variante genética es cualquier diferencia existente entre dos secuencias de ADN, se le llama así generalmente cuando se compara contra el genoma de referencia. Algunas de estas variantes han sido relacionadas con enfermedades, mientras que se desconocen los efectos de la mayoría. Ejemplos de ellas son los polimorfismos de un solo nucleótido (SNP), los cuales ocurren cuando un solo nucleótido cambia con respecto al genoma de referencia; los indels, que son aquellas inserciones o deleciones de nucleótidos a la secuencia de ADN; los polimorfismos multinucleótidos, los cuales alteran múltiples nucleótidos en sucesión; o las variantes estructurales, que ocurren cuando hay duplicación, eliminación o reorganización de secciones de un cromosoma^{35,41}.

1.1.4.1. Formato de Identificación de Variantes

Una de las aplicaciones de NGS es la posibilidad de descubrir variantes genéticas en un grupo de muestras secuenciadas. Esta tecnología ha brindado una enorme cantidad de datos genómicos que deben ser organizados en formatos para facilitar su manipulación e interpretación; uno de ellos es el formato de identificación de variantes (VCF, por sus siglas en inglés) el cual es utilizado para

almacenar anotaciones y ubicaciones de las variantes presentes en un grupo de muestras, incluidos los SNP, indels y variantes estructurales^{42,43}.

Un ejemplo de los campos presentes en un archivo VCF se muestra en la Figura 1.6. De manera general, este archivo se divide en dos secciones: encabezado y datos. El encabezado, a su vez, está conformado por líneas de metainformación (las cuales comienzan con los caracteres "##") y una línea de definición de campos (que comienza con un solo carácter "#") separada por tabulaciones. Las líneas de metainformación proporcionan una descripción de las etiquetas utilizadas en la sección de datos, así como también pueden contener información sobre la fecha de creación del archivo, la versión de la secuencia de referencia, el software utilizado o cualquier otra información relevante. La línea de definición de campos consta de varias columnas: CHROM, el cromosoma; POS, la posición donde se encuentra la variante; ID, el identificador de la variante; REF, el alelo de referencia; ALT, una lista separada por comas de alelos alternativos (es decir, los alelos encontrados en las muestras secuenciadas); QUAL, puntaje de calidad de la variante; FILTER, indica si la muestra pasó filtros de calidad aplicados anteriormente; tanto INFO (lista separada por punto y coma o con una barra vertical que contiene información adicional de las variantes) como FORMAT (lista de campos separados por dos puntos que contiene información estándar) contienen datos que estarán contenidos en cada columna de SAMPLE (las cuales corresponden a las muestras secuenciadas)⁴².

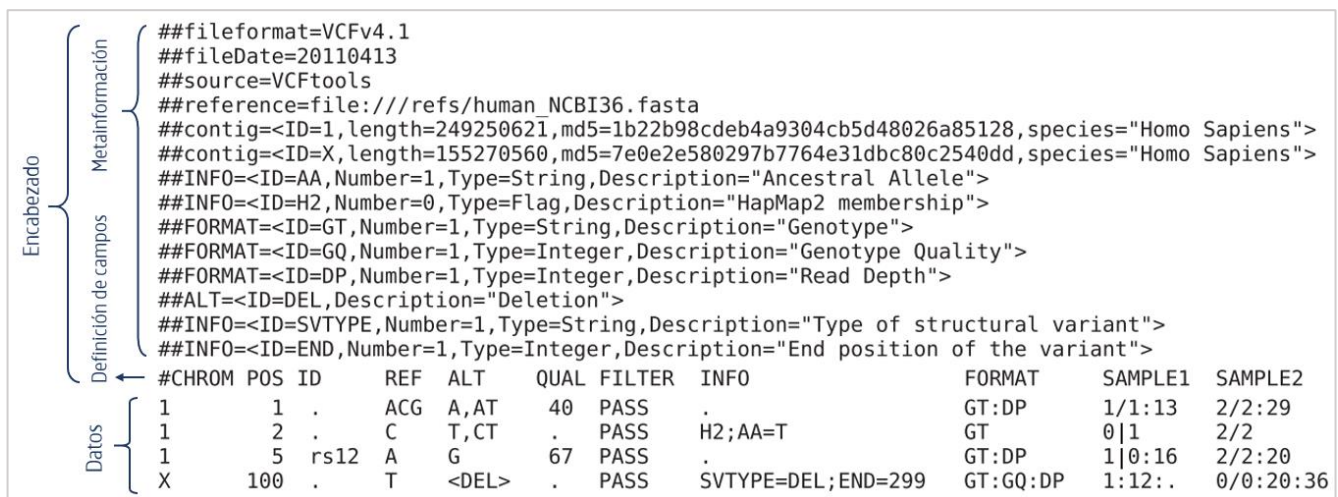


Figura 1.6 | Ejemplo de un archivo VCF. (Tomada y modificada de Danecek *et al.*, 2011⁴²).

1.2. Justificación

Actualmente, se desconoce la causa genética de aproximadamente el 50% de casos de melanoma familiar, los cuales consisten en familias que no presentan variantes en los genes de predisposición conocidos. Conocer el espectro de genes y variantes genéticas que aumentan el riesgo a padecer melanoma podría ayudarnos a aportar información sobre la biología de este cáncer, así como generar conocimiento para proveer de consejo genético a familias afectadas, esto es, informar a los pacientes de la probabilidad de desarrollar o transmitir el padecimiento, así como formas de prevención¹⁰.

1.3. Antecedentes

Los últimos datos obtenidos por nuestro grupo de trabajo, el Grupo de Genética del Cáncer y Bioinformática (UNAM, México) en colaboración con The Wellcome Sanger Institute y The University of Leeds (ambos en Reino Unido), incluyen la secuenciación de ADN por genoma y exoma completo de 314 miembros de 135 familias europeas, estadounidenses y australianas propensas a desarrollar melanoma que no presentaron variantes genéticas para los genes ya conocidos. A partir de estos mismos datos analizados anteriormente, pero con criterios distintos, se ha reportado la predisposición a melanoma de variantes en los genes *POT1*⁴⁴, *TERT*⁴⁵, *ACD* y *TERF2IP*⁴⁶.

1.4. Objetivos

- Identificar nuevas variantes genéticas candidatas a aumentar el riesgo a desarrollar melanoma mediante el análisis bioinformático de 314 secuencias provenientes de familias con este cáncer.
- Elegir aquella variante “poco frecuente” con mayor evidencia de asociación con el desarrollo de melanoma para su posterior validación experimental.

1.5. Hipótesis

Existen variantes genéticas de línea germinal poco frecuentes que aumentan el riesgo a desarrollar melanoma, las cuales serán evidenciadas mediante técnicas bioinformáticas.

1.6. Metodología

1.6.1 Datos de secuenciación

A partir de muestras de sangre, se secuenció el genoma o exoma completo de 314 individuos de 135 familias con melanoma por medio de la tecnología Illumina Hiseq 2000. Esto fue realizado en las instalaciones de The Wellcome Sanger Institute, Cambridge, Reino Unido. En la [Tabla 1.2](#) se muestra la procedencia y el número de muestras obtenidas de la totalidad de familias. Todos los casos se encontraron previamente negativos para variantes patogénicas en *CDKN2A* y *CDK4*.

Con los datos obtenidos, se generó un archivo VCF con las variantes presentes en las familias con melanoma, el cual fue anotado previamente con las consecuencias de cada mutación en proteínas por medio de la herramienta Variant Effect Predictor (VEP) de Ensembl⁴⁷. A partir de este punto, se comenzó con el trabajo descrito en este capítulo.

Tabla 1.2 | Procedencia de los genomas y exomas secuenciados de 135 familias con melanoma.

	Locación	Número de familias	Número de muestras	
Genomas	Dataset 1.0	32	123	
	Leiden	6	28	
Exomas	Leeds	66	77	
	Leiden	4	11	
	Barcelona	10	20	
	Pensilvania	1	3	
	Brisbane	4	15	
	Estocolmo	2	6	
	Sídney	7	28	
	Londres (KCL)	3	3	
	Total		135	314

1.6.2 Filtrado de variantes

Partiendo del archivo VCF anotado, se realizaron una serie de filtrados empleando programas hechos con el lenguaje de programación Perl, los cuales se aplicaron específicamente a la columna INFO del archivo. En la Figura 1.7 se muestran los datos que se pueden encontrar en dicha columna; resaltados en amarillo están aquellos campos que fueron extraídos del archivo: *Consequence* corresponde al efecto que puede tener cada alelo de la variante; *SYMBOL* corresponde al nombre común del gen; *Protein_position*, como su nombre lo indica, es la posición donde se encuentra en la proteína el aminoácido que sufrió el cambio; *Amino_acids* corresponde al aminoácido de referencia y al aminoácido alternativo observado (es decir, indica el cambio de aminoácido que produce esa variante); *Existing_variations* indica si esa variante existe en la base de datos dbSNP⁴⁸ (base de datos de SNP's) y/o COSMIC⁴⁹ (base de datos que posee una colección de mutaciones somáticas en cáncer), mostrando su ID en caso de ser así; y finalmente *SWISSPROT* muestra un ID de la variante en caso de encontrarse en UniProtKB/Swiss-Prot⁵⁰, la cual es una sección anotada manualmente de la base de datos de proteínas UniProtKB.

```
##INFO=<ID=CSQ,Number=.,Type=String,Description="Consequence annotations from Ensembl VEP. Format: Allele|Consequence|IMPACT|SYMBOL|Gene|Feature_type|Feature|BIOTYPE|EXON|INTRON|HGVS|HGVS|cdna_position|CDS_position|Protein_position|Amino_acids|Codons|Existing_variation|DISTANCE|STRAND|FLAGS|VARIANT_CLASS|SYMBOL_SOURCE|HGNC_ID|CANONICAL|TSL|APPRIS|CCDS|ENSP|SWISSPROT|TREMBL|UNIPARC|GENE_PHENO|SIFT|PolyPhen|DOMAINS|GMAF|AFR_MAF|AMR_MAF|EAS_MAF|EUR_MAF|SAS_MAF|AA_MAF|EA_MAF|ExAC_MAF|ExAC_Adj_MAF|ExAC_AFR_MAF|ExAC_AMR_MAF|ExAC_EAS_MAF|ExAC_FIN_MAF|ExAC_NFE_MAF|ExAC_OTH_MAF|ExAC_SAS_MAF|CLIN_SIG|SOMATIC|PHENO|PUBMED|MOTIF_NAME|MOTIF_POS|HIGH_INF_POS|MOTIF_SCORE_CHANGE">
```

Figura 1.7 | Columna del archivo VCF anotado correspondiente a INFO. En amarillo se encuentran resaltados los campos que se extrajeron para realizar el análisis.

En la Figura 1.8 se muestra un esquema con los pasos seguidos para la obtención de la lista de variantes candidatas. El primer filtro consistió en la extracción de variantes que tuvieran la anotación “missense variant”, la cual cambia una o más bases, dando como resultado una secuencia de aminoácidos diferente, pero sin cambiar la longitud de la proteína. Se eligió esta consecuencia ya que otras consecuencias más drásticas (como adquisición de codones de paro, variantes que causan cambios en el marco de lectura y aquellas que afectan el proceso de empalme) ya han sido analizadas en este conjunto de datos anteriormente⁴⁴⁻⁴⁶.

En el segundo filtro se seleccionaron aquellas variantes presentes en la base de datos COSMIC. Como se mencionó anteriormente, esta base de datos contiene información sobre una gran cantidad de mutaciones somáticas presentes en diferentes tipos de cáncer; así, extraer variantes presentes en COSMIC permite considerar aquellas que cuentan con evidencia publicada en oncología o bien aquellas encontradas a partir de secuenciación de muestras de tejido canceroso en varios tipos de cáncer.

Para maximizar la posibilidad de que las variantes encontradas fueran funcionales, el tercer filtro consistió en buscar variantes que tuvieran un impacto en modificaciones postraduccionales, específicamente aquellas que interrumpían sitios de fosforilación, ya que es una manera de encontrar cuáles podrían afectar en mayor grado la función de la proteína. Para ello, se buscó la presencia de las variantes en la base de datos de modificaciones postraduccionales de proteínas dbPTM⁵¹, en casos de fosforilación. dbPTM incluye los ID de todos los sitios de modificaciones postraduccionales validados experimentalmente.

Por último, para destacar aquellas variantes con evidencia publicada, se realizó una búsqueda sistemática de las variantes restantes en la base de datos PubMed, de NCBI (<https://www.ncbi.nlm.nih.gov/pubmed>), con la finalidad de obtener el número de artículos científicos existentes tras buscar el nombre del gen, el aminoácido afectado y la palabra “cancer” en el título o resumen de cada artículo. Las variantes con al menos un artículo científico fueron analizadas manualmente y se obtuvo una lista final.

(Programas utilizados se presentan en Anexo II).



Figura 1.8 | Esquema de los filtros aplicados al archivo VCF anotado con las consecuencias de las variantes.

1.7. Resultados y discusión

El VCF original con la información de la secuenciación de todos los individuos contenía 120,196 variantes. Con el primer filtro, el cual descartó todas aquellas variantes que no tuvieran como consecuencia “missense_variant”, el número se redujo a 107,803. El siguiente filtro, que buscaba mantener solamente las variantes presentes en la base de datos COSMIC, dejó solamente 10,930 variantes. Con la aplicación del tercer filtro, el cual extrajo sólo aquellas variantes que se encuentran anotadas en la base de datos dbPTM, restaron 130 variantes. Finalmente, después de realizar la búsqueda en NCBI, el total de variantes que tenían al menos un artículo científico con el nombre del gen, número de aminoácido y la palabra “cancer” en el título o resumen, fue de 13 [Tabla 1.3].

Tabla 1.3 | Número de variantes resultantes después de aplicar cada filtro.

Variantes iniciales	Filtro	Variantes eliminadas	Variantes restantes
120,196	Missense variant	12,393	107,803
107,803	Presencia en COSMIC	96,873	10,930
10,930	Presencia en dbPTM	10,800	130
130	Presencia en artículo	117	13

La lista final obtenida se muestra en la Tabla 1.4; se presenta una columna con el nombre del gen, el número de aminoácido afectado, el cambio de aminoácido que ocasiona la variante detectada, su ID en la base de datos COSMIC, su ID en la base de datos dbPTM (que utiliza los IDs de UniProtKB/Swiss-Prot) y el número de publicaciones encontradas en PubMed que contuvieran las palabras clave.

Se analizó manualmente cada una de las publicaciones encontradas para comprobar que las palabras clave realmente se refirieran al nombre de un gen y a un aminoácido. Se encontró que la mayoría de estas publicaciones sí hablaban del gen en cuestión, pero no del aminoácido, esto debido a que el número de aminoácido buscado generalmente se refería a un número cualquiera (cantidades o concentración de algún reactivo, número de casos, números que conformaban el doi del artículo, número de volumen o issue de la revista, por mencionar algunos). Por ejemplo, en el caso del gen *REST*, el cual tuvo el mayor número de publicaciones, ninguno de los resultados correspondió al nombre del gen, sino a la palabra en inglés “rest”.

Luego de la búsqueda manual, sólo una variante fue encontrada en la mayoría de las publicaciones arrojadas: *ESR1* Y537C (subrayado en amarillo en la tabla). De las cinco publicaciones encontradas para esta variante, en cuatro de ellas se mencionaba al gen *ESR1* (el cual codifica para el receptor de estrógenos alfa [RE α]), el aminoácido 537 y la palabra “cancer”⁵²⁻⁵⁵.

Tabla 1.4 | Lista final de variantes obtenidas.

Gen	AA	Cambio	ID COSMIC	ID dbPTM	Publicaciones
<i>REST</i>	750	S/T	COSM3302227	Q13127	20
<i>ESR1</i>	537	Y/C	COSM1074637	P03372	5
<i>ENDOG</i>	12	S/L	COSM4163392	Q14249	4
<i>KRT18</i>	4	T/I	COSM1128552	P05783	3
<i>BCR</i>	634	T/M	COSM268752	P11274	3
<i>CALCA</i>	41	T/M	COSM1675853	P01258	2
<i>KRT18</i>	34	S/R	COSM1666686	P05783	2
<i>S100A7A</i>	34	T/M	COSM3385198	Q86SG5	1
<i>AGXT</i>	9	T/N	COSM1406872	P21549	1
<i>WRN</i>	1292	S/Y	COSM3648356	Q14191	1
<i>MRPS2</i>	30	T/N	COSM3214668	Q9Y399	1
<i>KRT18</i>	100	S/R	COSM4590815	P05783	1
<i>KRT16</i>	215	T/S	COSM4066411	P08779	1

En amarillo se encuentra la única variante de la cual cuatro de las cinco publicaciones arrojadas sí correspondieron con el aminoácido en cuestión.

1.7.1. *ESR1* Y537C

La variante Y537C, la cual induce un cambio de tirosina (Y) por cisteína (C) en el aminoácido 537 del RE α , se encontró en una familia inglesa (procedente de la ciudad de Leeds). De esta familia, sólo se cuenta con datos de secuenciación de tres miembros (dos individuos de la segunda generación y uno de la tercera), de los cuales dos presentan la variante (madre e hijo). En la familia también hay dos casos de cáncer de mama (CM) en la primera y tercera generación, respectivamente.

Esta variante resulta muy interesante ya que no solo satisface todos los filtros aplicados (causar un cambio que afecta la fosforilación de una proteína, el cual ya se ha reportado somáticamente en la literatura), sino que tampoco existe en bases de datos de línea germinal como gnomAD⁵⁶, lo cual indica que es extremadamente rara en la población.

Además, el gen *ESR1* es un foco mutacional en CM y se cree que mutaciones en el residuo 537 proveen resistencia endócrina a tratamiento en pacientes con metástasis, así como también pueden fungir como posibles objetivos terapéuticos^{52,53,57-59}. Aunado a ello, también se ha mutado a Y537 y se ha encontrado que la presencia de ciertos aminoácidos en esta posición puede mantener activo al RE α aunque no haya presencia de ligando^{54,55,60}.

Debido a lo anterior, y a que nunca se ha descrito ni en melanoma ni en línea germinal, *ESR1* Y537C representa un buen candidato en la búsqueda de variantes genéticas que pueden conferir un mayor riesgo a padecer melanoma.

1.8. Conclusión

El análisis bioinformático de genomas y exomas completos de individuos puede ser aplicado para identificar variantes candidatas a aumentar el riesgo a padecer enfermedades. De acuerdo con los datos mostrados y al análisis empleado, se propone que la variante *ESR1* Y537C es una buena candidata para fungir como factor de riesgo en el desarrollo de melanoma cutáneo. Asimismo, estos hallazgos abren investigaciones y proponen una caracterización experimental en líneas celulares para complementar la información obtenida. Parte de esta investigación se describe en el siguiente capítulo.

2

Evaluación funcional de la mutación *ESR1* Y537C en la línea celular NIH 3T3

2.1. Introducción

2.1.1. Estrógenos y receptores de estrógenos

Las hormonas esteroideas sexuales se sintetizan a partir del colesterol y se clasifican en andrógenos, progestágenos y estrógenos⁶¹. Hay tres tipos de estrógenos: estriol, estrona y estradiol, siendo este último responsable del desarrollo y funcionamiento del sistema reproductor femenino⁶². Sin embargo, los estrógenos no son exclusivamente de mujeres, ya que también existen tejidos diana en los hombres, como la próstata⁶³.

Los efectos de los estrógenos están mediados por proteínas receptoras intracelulares llamadas receptores de estrógenos (RE), que actúan como factores transcripcionales⁶⁴. Hay dos tipos de RE, el receptor de estrógeno alfa (RE α) compuesto por 595 aminoácidos y el receptor de estrógeno beta (RE β) compuesto por 530 aminoácidos, ambos codificados por diferentes genes, *ESR1* y *ESR2*, respectivamente^{62,65}. Los RE están compuestos por cinco dominios funcionales: A/B, C, D y E [Fig. 2.1]. A/B es el dominio de amino terminal y, contiene, a su vez, el dominio de transactivación AF-1, el cual recluta coactivadores y correpresores; C es el dominio de unión al ADN y contiene aminoácidos específicos capaces de unirse a elementos estrógeno-responsivos (EER) en

potenciadores o promotores; el dominio de unión al ligando (E) es reconocido por la hormona 17 β -estradiol (E2), la cual una vez unida al receptor, cambia su conformación y activa su transcripción; el dominio de bisagra flexible (D) une a D y E, además de ser importante para los cambios conformacionales en el receptor; y finalmente, no se conoce en su totalidad la función del dominio F^{59,62,65}.

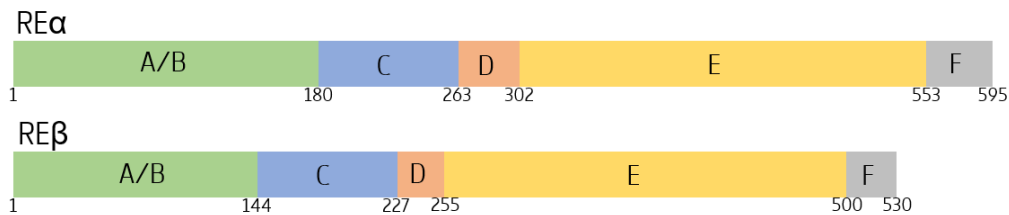


Figura 2.1 | Representación esquemática de los dominios funcionales de los RE α y RE β .

Se muestra el número de aminoácidos de cada receptor. (Adaptada de Yaşar *et al.* 2016⁶²).

La unión del ligando E2 al dominio E puede ocurrir en el citoplasma o en el núcleo [Fig. 2.2], ya que los estrógenos son lipófilos y pueden dispersarse fácilmente a través de la membrana y, en consecuencia, el RE forma homo (RE $\alpha\alpha$, RE $\beta\beta$) o heterodímeros (RE $\alpha\beta$) que se unen directamente a EER, los cuales a su vez activan una cascada de señalización que culmina con la transcripción de genes como *FFP* (cuya expresión está en el útero y promueve la proliferación celular), *EBAG9* (presente en CM e inhibidor de la proliferación celular), *NR2D* (receptor de glutamato, expresado en distintas partes del encéfalo), entre otros^{59,64,66,67}. El homodímero RE $\alpha\alpha$ tiene un efecto proliferativo en el crecimiento de las células mamarias, mientras que el homodímero RE $\beta\beta$ tiene un efecto antiproliferativo; las funciones de los heterodímeros aún no han sido demostradas en su mayoría, pero se sabe que se unen a EER con una afinidad relacionada a la de los homodímeros, así como también se sabe que los heterodímeros son más estables^{68,69}.

En cuanto a la distribución, RE α se expresa en órganos como ovario, útero, vagina, glándula mamaria e hipotálamo; de igual manera, se ha detectado RE β en ovario, próstata, testículo, cerebro, colon y piel^{67,70}.

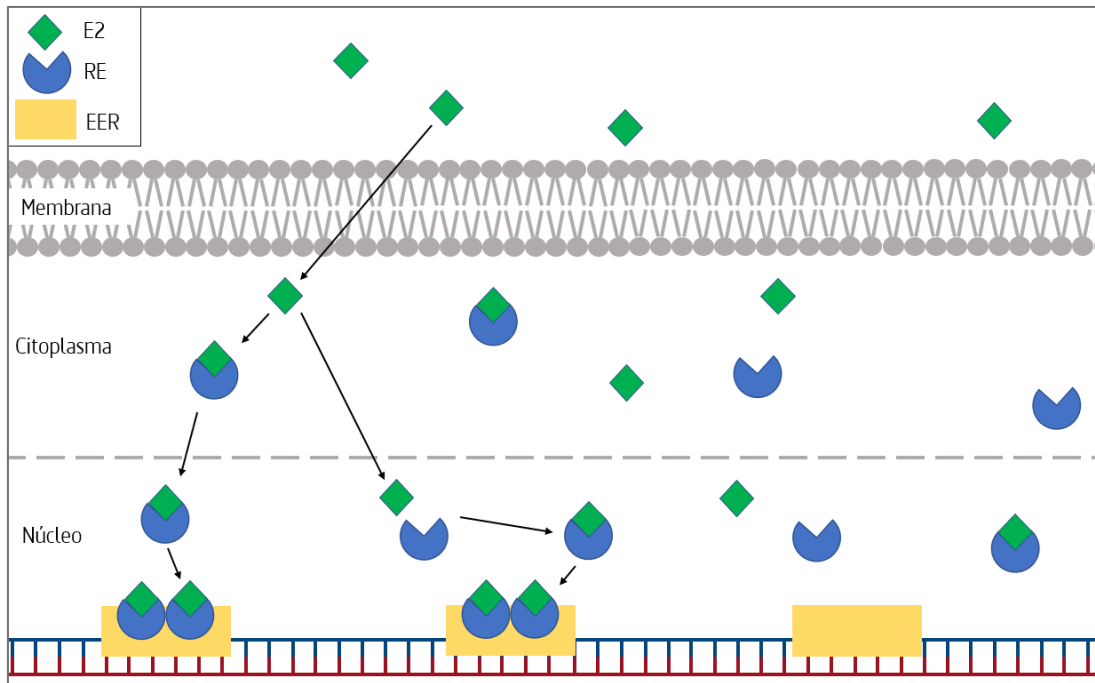


Figura 2.2 | Esquema de la unión de estrógenos (E2) al receptor de estrógenos (RE) y su translocación al núcleo, en donde se una a promotores específicos conocidos como elementos estrógenos responsables (EER). (Adaptada de Tecalco & Ramírez, 2017⁵⁹).

2.1.1.1. REs en cáncer de mama

La actividad de los RE está regulada por diferentes factores involucrados en su estabilización o degradación, como lo son la unión de ligandos, la fosforilación o la ubiquitinación; sin embargo, una señalización anormal en las vías involucradas en estos procesos puede conducir al desarrollo de padecimientos como el cáncer, enfermedades metabólicas o cardiovasculares, entre otras^{59,65,71}. Además, los estrógenos pueden estimular una proliferación celular irregular (debida a un desequilibrio entre la expresión del RE α y el RE β) en órganos como las glándulas mamarias, el útero y los ovarios, lo que eventualmente puede promover la formación de tumores⁷².

El RE α controla la proliferación y diferenciación celular en el tejido mamario y se expresa en aproximadamente el 10% de su epitelio. Además, es el principal impulsor en el desarrollo y progresión del CM, donde su expresión aumenta en alrededor de un 70%. Pacientes con CM que presentan una sobreexpresión de este receptor, son denominados RE α positivos (o RE α ⁺), por lo cual se someten a terapias anti-hormonales con la finalidad de inhibirlo; ejemplos de ellas son el tamoxifeno (compite con el E2 para unirse al receptor) y fulvestrán (se une al receptor haciendo que

se degrade). Sin embargo, aproximadamente 30% de pacientes con tratamiento de tamoxifeno adquieren resistencia a la terapia^{59,65}.

Por otro lado, el RE β se expresa en aproximadamente el 80% del tejido mamario, siendo su expresión disminuida (o perdida) conforme el CM progresa. Se sabe que la re-expresión de este receptor en CM inhibe la proliferación celular, promueve la apoptosis y propicia una eficacia en quimioterapia. Además, la pérdida de este receptor está asociada con resistencia a tratamiento⁶⁵.

Mutaciones en el gen *ESR1* (que codifica para el RE α) han sido identificadas en CM como un mecanismo de resistencia endocrina a tratamiento, además de que pueden fungir como un posible objetivo terapéutico^{52,57}. La mayoría de mutaciones identificadas en CM avanzado se producen en el dominio de unión al ligando del RE α , con una mayor frecuencia en los aminoácidos L536, D538 e Y537⁵³.

2.1.1.1.1 Mutaciones en el aminoácido Y537

La tirosina 537 (Y537) se encuentra en el dominio de unión al ligando del RE α , siendo una región importante en la regulación de su actividad transcripcional. Sin embargo, la presencia de tirosina en este sitio no es necesaria para la unión del E2, ya que se ha demostrado que la sustitución de aminoácidos como alanina (A), ácido glutámico (E), lisina (K) o serina (S), en este residuo pueden inducir un estado activo del RE α , aunque no haya presencia de ligando⁶⁰.

E2 induce fosforilación en los residuos S118, S294, S341 e Y537, la cual es necesaria para la degradación del receptor; sin embargo, su degradación es evitada por mutaciones en Y537, como Y537C, Y537N, e Y537S, las cuales han sido identificadas en muestras de CM metastásico y han sido relacionadas con resistencia a tratamiento⁵⁹. Asimismo, se ha demostrado que las mutaciones Y537C, Y537N y D538G inducen una actividad independiente del E2, siendo relativamente resistente al tamoxifeno y fulvestrán⁵⁵.

2.1.1.2. Melanoma y estrógenos

La relación entre diferencias de sexo y la incidencia y mortalidad del melanoma ha sido poco estudiada, pero algunas investigaciones han generado gran interés en este tema. La incidencia de melanoma en hombres generalmente se presenta en el tronco, la cabeza y el cuello, mientras que en las mujeres este cáncer aparece principalmente en las extremidades inferiores^{31,73}. Los hombres

tienen una mayor incidencia y niveles de mortalidad que las mujeres: a nivel mundial, en 2018, la incidencia de melanoma en las mujeres fue de 137,025 casos con 25,881 muertes, mientras que en los hombres se reportaron 150,698 casos y 34,831 muertes⁷⁴. Se ha informado que el hecho de que las mujeres tengan niveles más altos de supervivencia y menor riesgo de metástasis que los hombres, puede estar relacionado con la distribución del melanoma, la influencia de los estrógenos en las mujeres y la esperanza de vida⁷⁵.

La piel es el órgano principal no reproductivo sobre el cual actúan los estrógenos; de hecho, se ha sugerido que la presencia de melanoma progresivo podría estar influenciada por los estrógenos^{76,77}. La expresión del RE β (receptor que tiene un efecto antiproliferativo en células de tejido mamario) en la piel está relacionada con el sexo y la edad; se ha observado que los hombres muestran niveles más bajos de RE β que las mujeres, en quienes la expresión de este receptor disminuye con la edad y más rápidamente después de la menopausia (donde ocurre una disminución en los niveles de estrógeno)^{70,76,78}. Además, en un estudio que examinó las asociaciones entre los factores reproductivos, el uso exógeno de estrógenos y la incidencia de melanoma en mujeres posmenopáusicas no hispanas, se demostró que existe un mayor riesgo de padecer melanoma en mujeres que presentaron una menarquía temprana y una menopausia tardía (ya que implicaría una exposición prolongada a estrógeno endógeno)⁷⁹. Asimismo, análisis inmunohistoquímicos del nivel de RE β en tejidos de melanoma han demostrado que la expresión de este receptor disminuye en melanomas más gruesos y más invasivos^{80,81}.

2.2. Justificación

El análisis bioinformático de los genomas y exomas de 135 familias con melanoma presentados en el capítulo anterior, sugieren que la mutación Y537C del gen *ESR1*, correspondiente al RE α , es un factor de predisposición para el desarrollo de melanoma. Esta mutación es notable porque está ausente de todos los catálogos de variación genética humana publicados hasta el momento y porque es un foco mutacional en cáncer de mama. Dada la presencia de la mutación Y537C en familias con melanoma y el hecho de que el potencial oncogénico de esta mutación no ha sido comprobado en este tipo de cáncer, se propone investigar su papel en células melanocíticas y de melanoma. Sin embargo, es necesario establecer la técnica y los controles necesarios en el

laboratorio, por lo que primero se evaluará en la línea celular NIH 3T3 (correspondiente a fibroblastos de ratón).

2.3. Objetivo

Evaluar el potencial oncogénico de la mutación *ESR1* Y537C en la línea celular NIH 3T3 mediante el ensayo de formación de focos (Focus Formation Assay).

2.4. Hipótesis

La mutación *ESR1* Y537C inducirá un efecto oncogénico en la línea celular NIH 3T3.

2.5. Metodología

2.5.1 Reactivos

Plásmido BABE-puro

El plásmido BABE-puro (Addgene®, plasmid #1768) o pBABE-puro, es un vector retroviral para clonar y expresar genes de interés y es utilizado en células de mamíferos; tiene un tamaño de 5062 pb, resistencia bacteriana a Ampicilina y selección de células de mamífero con resistencia a la Puromicina⁸³.

Plásmido MA-RQ

El plásmido MA-RQ (pMA-RQ) tiene un tamaño de 4141 pb. Cuenta con gen de resistencia bacteriana a Ampicilina, pero no posee gen de selección de células de mamífero.

(Mapa de los plásmidos en Anexo II).

Bacterias XL1-Blue

La cepa XL-Blue es una cepa hospedadora eficaz para clonación con plásmidos, con resistencia intrínseca a la Tetraciclina, velocidad de duplicación alta y gran eficiencia en la transformación.

Línea celular NIH 3T3

La línea celular NIH 3T3 (ATCC®, CRL-1658) corresponde a fibroblastos de ratón y es muy útil en estudios de transfección de ADN⁸⁴. La morfología de las células es adherente y además presentan una fuerte inhibición por contacto, así como también se considera una línea fácil de cultivar. La introducción de un gen sobreexpresado en las células NIH 3T3 puede promover la transformación morfológica y la pérdida de la inhibición por contacto, lo que puede ayudar a determinar el potencial oncogénico de ese gen de interés⁸⁵.

Gen *K-RAS* G12V

Gen que codifica para la proteína KRAS, la cual participa en las vías de señalización celular que controlan el crecimiento, la maduración y la muerte de las células. La mutación G12V tiene un gran potencial oncogénico. Es utilizado como control positivo en ensayos de proliferación celular. Tiene un peso de 750 pb.

Gen *ESR1*

Gen que codifica para el RE α . Tiene un peso de 1800 pb.

Gen *ESR1* Y537C

Mutación del gen *ESR1* con una sustitución de G→A en la posición 1610. Induce un cambio de aminoácido Y537C (tirosina por cisteína) en el RE α . Tiene un peso de 1800 pb.

2.5.2 Digestiones y ligaciones

2.5.2.1 Digestión del plásmido MA-RQ

Se mandó sintetizar el gen *ESR1* y *ESR1* Y537C con el cambio idéntico al portado por la familia con la mutación a la compañía Invitrogen® (No. Ref. 2273748 y 2273749, respectivamente) la cual retornó las secuencias en el plásmido denominado MA-RQ. El plásmido MA-RQ no posee gen con resistencia para selección de células de mamíferos. Este plásmido contenía los genes *ESR1* y *ESR1* Y537C (cada uno por separado) por lo cual se realizó su digestión mediante enzimas de restricción BamHI y EcoRI [Fig. 2.3a]. La reacción consistió de 6 µl de agua libre de nucleasas, 4 µl de buffer Tango, 8 µl de pMA-RQ, 6 µl de enzima Bam HI y 6 µl de enzima EcoRI; seguidamente, fue puesta en incubación con agitación durante 30 minutos a 37°C. Posteriormente, se corrió un gel de integridad el cual fue preparado con 0.4 g de agarosa, 40 µl de buffer TAE y 4 µl de SYBR Safe DN, durante 90 minutos a 90 V y 50 mA. De este gel se extrajo el ADN correspondiente a los genes *ESR1* y *ESR1* Y537C, respectivamente, siguiendo el protocolo establecido por el Kit QIAquick Gel Extraction de QIAGEN® (No. Ref. 28704) [Fig. 2.3b]. Finalmente, se midió la concentración del ADN purificado utilizando un equipo Qubit® 4 (Thermo Fisher Scientific®).

2.5.2.2 Digestión del plásmido BABE-puro

El plásmido BABE-puro (obsequio del Laboratorio del Dr. Adams, Wellcome Sanger Institute, Inglaterra), el cual cuenta con gen para selección de células de mamífero con resistencia a Puromicina, también fue digerido para posteriormente insertarle los genes de interés [Fig. 2.3c]; asimismo, este plásmido contenía el gen *K-RAS* con la mutación G12V, el cual tiene un gran potencial oncogénico y actúa a su vez como nuestro control positivo. El procedimiento fue el mismo que el realizado con el plásmido MA-RQ. De igual manera, se corrió un gel de integridad, pero en esta ocasión se extrajo el ADN correspondiente al pBABE-puro.

2.5.2.3 Ligación del plásmido BABE-puro

Para el caso de la ligación con *ESR1* y *ESR1* Y537C, el proceso consistió en una reacción con 9 µl de pBABE-puro vacío [Fig. 2.3d], 20 µl de gen, 2 µl de 10x T4 DNA Ligase Buffer y 1 µl de T4 DNA Ligase. Por otro lado, (y como parte de un control negativo) se ligaron los extremos del pBABE-puro y se le denominó CASSETTE; la reacción consistió de 20 µl de pBABE-puro vacío, 2 µl de 10x T4 DNA Ligase Buffer y 1 µl de T4 DNA Ligase. Las reacciones de ligación se colocaron en incubación

con agitación durante 10 minutos a 22°C. Posteriormente, se corrió un gel de integridad y se extrajo y purificó el ADN correspondiente a la ligación pBABA+ *ESR1*, pBABA+ *ESR1* Y537C y CASSETTE [Fig. 2.3e]. Finalmente, se midió la concentración del ADN purificado utilizando un equipo Qubit® 4.

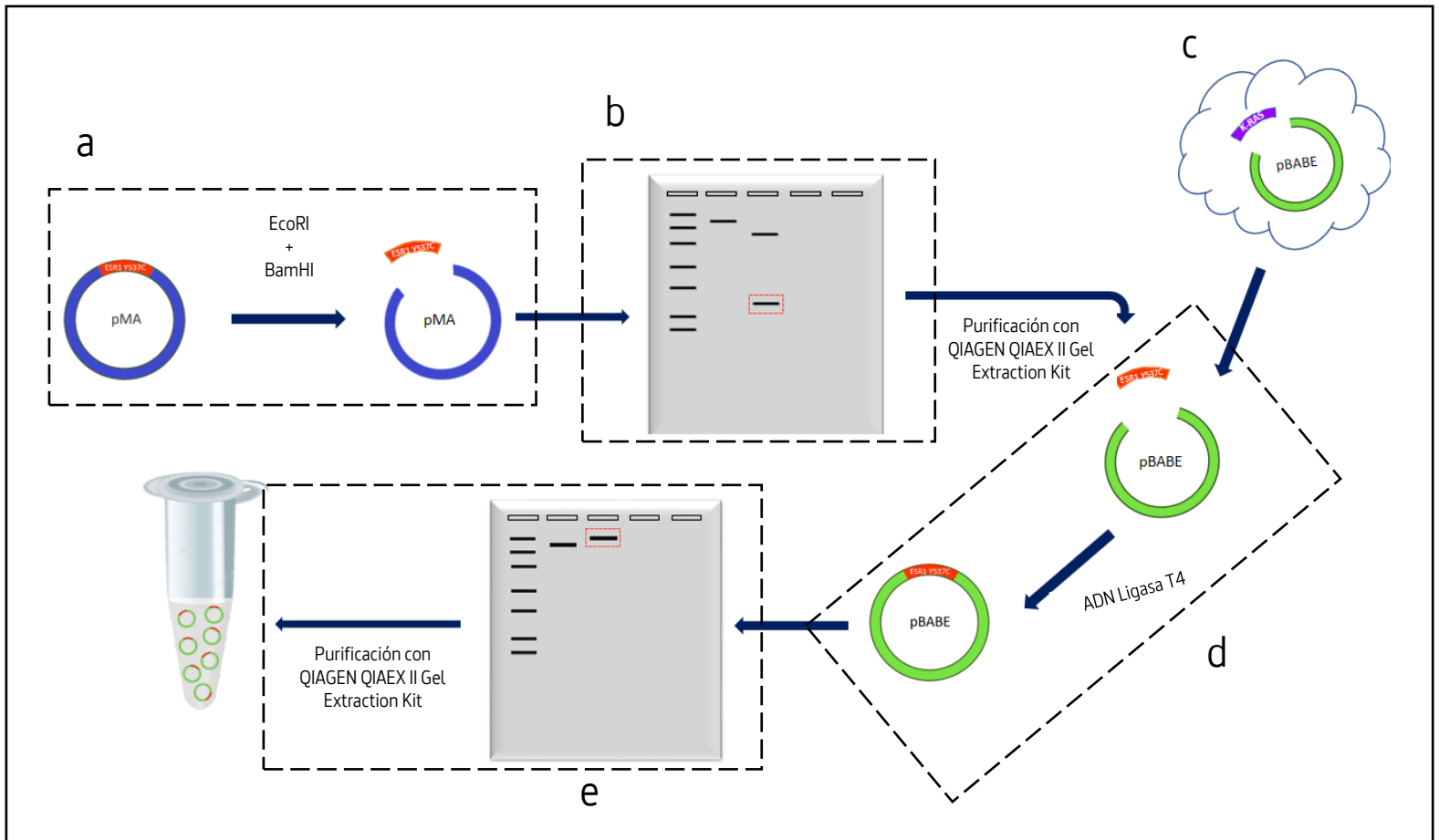


Figura 2.3 | Esquema de la digestión y ligación de plásmidos. El plásmido con el inserto de interés es digerido con enzimas de restricción EcoRI y BamHI (a); posteriormente, se corre un gel de integridad y se extrae el ADN correspondiente al inserto (b). Por otro lado, y siguiendo el mismo proceso, el plásmido BABA-puro también es digerido (c). Finalmente, el inserto y el plásmido BABA digerido se ligan con ADN ligasa T4 (d) y se corre nuevamente un gel de integridad para purificar los plásmidos ligados (e).

2.5.3 Transformación de bacterias XL-Blue

Para la amplificación de las ligaciones obtenidas se realizó la transformación de bacterias XL-Blue mediante choque térmico (42°C) con 30 ng de cada tipo plásmido ligado [Fig. 2.4a]; se incubaron con 400 ml de medio LB a 37°C durante 30 minutos (en agitación), se centrifugó a 13,000 rpm por

dos minutos y se desechó el sobrenadante. Con un asa de siembra se retiró el pellet y se esparció en una caja de Petri que contenía Ampicilina (se utilizó una caja de Petri para cada una de las ligaciones) y se incubó a 37°C durante 24 horas [Fig. 2.4b]. Pasado el tiempo de incubación, se resembraron 5 colonias de bacterias de la caja de Petri en medio LB con Ampicilina a una concentración de 100 µg/ml y se dejó a 37°C en agitación durante 12 horas [Fig. 2.4c].

2.5.3.1 Purificación y cuantificación del plásmido

Para la purificación de cada uno de los plásmidos obtenidos, se siguió el protocolo establecido por el Kit de purificación de ADN plasmídico de QIAGEN® (No. Ref. 12143) [Fig. 2.4d]. Finalmente, se midió la concentración de los plásmidos purificados en un equipo Qubit® 4.

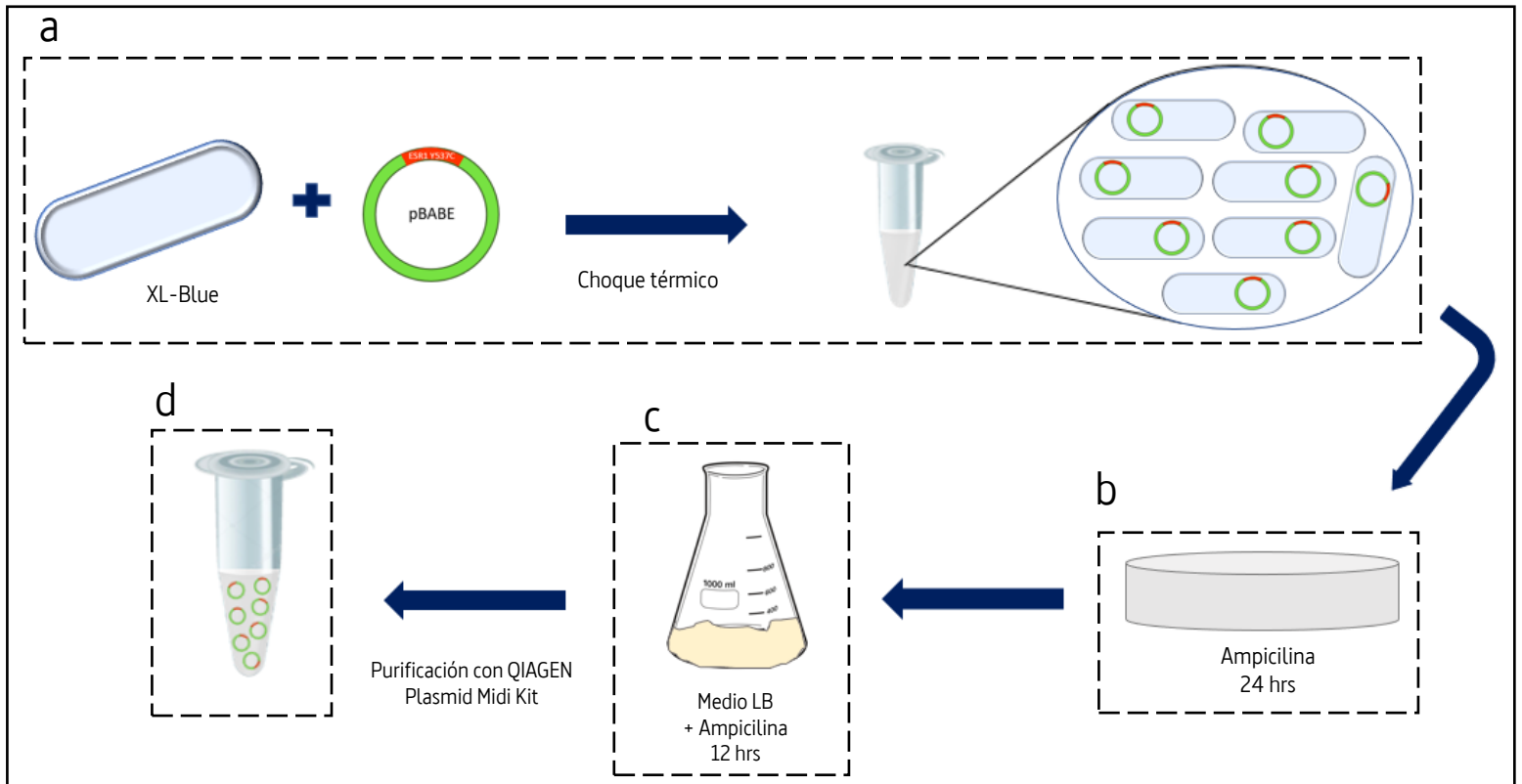


Figura 2.4 | Esquema de la amplificación de plásmidos. Las bacterias XL-Blue fueron transformadas con cada uno de los plásmidos ligados por choque térmico (a), posteriormente fueron sembradas en medio con ampicilina para su selección (b), para luego ser resembradas en medio LB con ampicilina para su amplificación (c) y, finalmente, los plásmidos fueron purificados (d).

2.5.4 Cultivo y transfección de células NIH 3T3

Las células NIH 3T3 se descongelaron a baño María, se transfirieron a un tubo Falcon de 15 ml y se les agregaron 10 ml de medio DMEM + suero BCS al 10% (al que se le llamará a partir de ahora *medio DB*). Posteriormente, se centrifugaron a 400 g durante 5 minutos y se desechó el sobrenadante. Las células se resuspendieron en 10 ml de medio DB y se transfirieron a una caja de Petri. Se incubaron a 37°C y 5% de CO₂ hasta alcanzar el 60% de confluencia para realizar pases a cajas nuevas y mantener la línea, además de criopreservar alícuotas con las células del pase 1.

2.5.4.1 Transfección de células

Pasado el tiempo de incubación, se sembraron un millón de células en una placa de seis pozos [Fig. 2.5] (se realizó por duplicado, para un total de 12 pozos) y se les agregaron 2 ml de medio DB. Las condiciones fueron las siguientes: a) pozo 1A sólo contiene células y medio, funge como control negativo ya que las células no fueron transfectadas; b) al pozo 1B se le añadió 1.5 µl de Lipofectamina 3000 (No. de catálogo L3000008, Invitrogen®) y c) a partir del pozo 2A se añadió una concentración de 800 ng de cada plásmido (dos pozos por tipo de plásmido), así como 1.5 µl de Lipofectamina 3000. Posteriormente, se seleccionaron con Puromicina a 2 µg/ml por 2 días de tratamiento. Así, los grupos experimentales son cinco: NIH 3T3 WILD TYPE (WT), NIH 3T3 + CASSETTE, NIH 3T3 + *K-RAS* G12V, NIH 3T3 + *ESR1* y NIH 3T3 + *ESR1* Y537C.

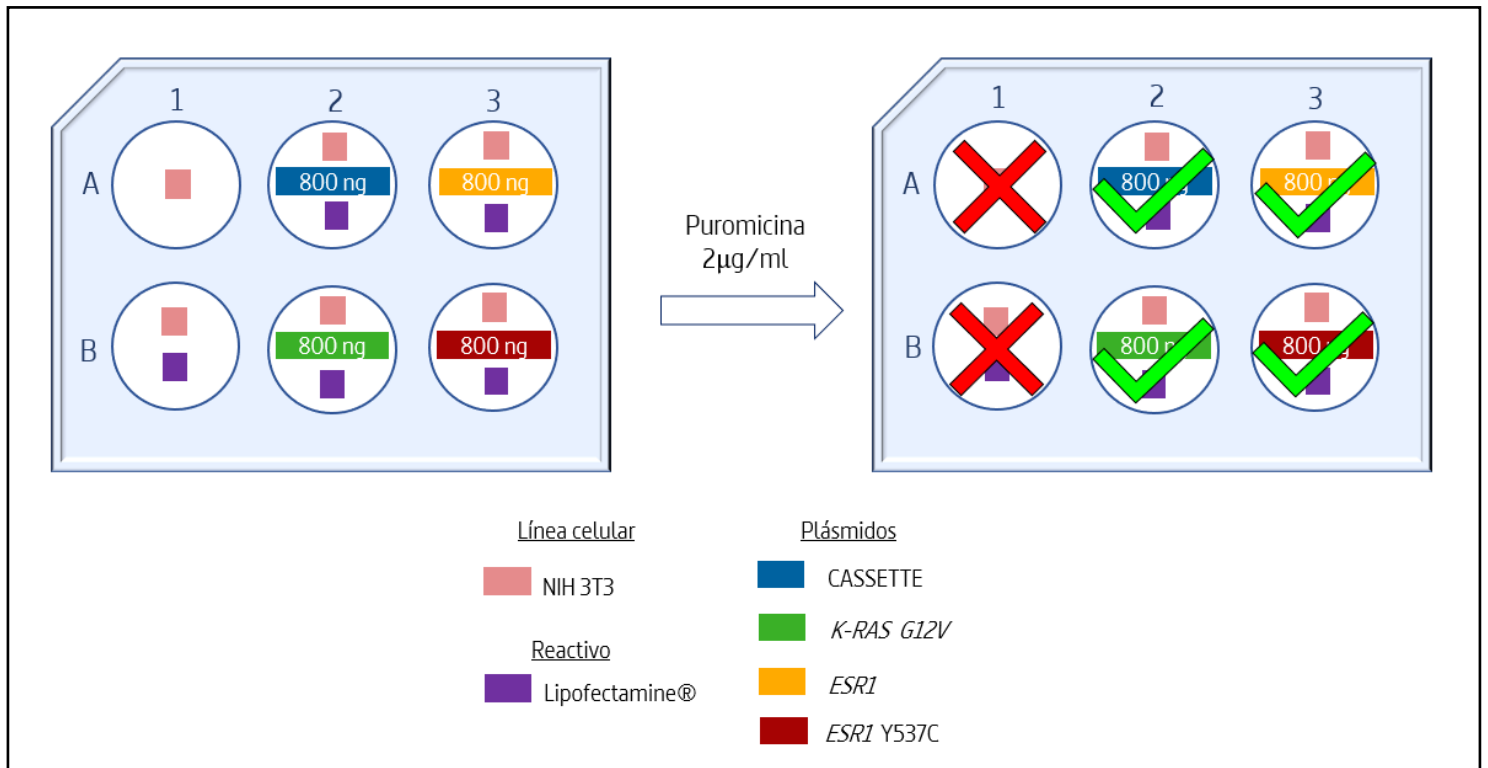


Figura 2.5 | Esquema de la transfección de células. En una placa de 6 pozos se sembraron células con diferentes condiciones. En el pozo 1A sólo se sembraron células; en el pozo 1B se sembraron células con lipofectamina; a partir del pozo 2A se sembraron células con 800 ng de los distintos plásmidos. Posteriormente, se seleccionaron con puromicina a $2\mu\text{g}/\text{ml}$ durante 48 horas, de tal manera que sólo aquellas células que fueron transfectadas exitosamente, sobrevivieron.

2.5.5 Ensayo de formación de focos

El ensayo de formación de focos (Focus Formation Assay) es una de las técnicas más utilizadas para evaluar el potencial oncogénico de un gen de manera *in vitro*. Consiste en la transfección de la línea celular NIH 3T3 con el gen de interés, de tal manera que si este es un oncogén, dará como resultado que las células pierdan la inhibición por contacto y crezcan en múltiples capas formando "foci" (focos)⁸⁵.

Para realizar este ensayo se sembraron 2 millones de células, por duplicado, de cada grupo experimental en cajas de Petri con 7 ml de medio DB durante 21 días, reemplazándose el medio según fuera necesario, de acuerdo al protocolo de Álvarez et al. (2014)⁸⁵. Durante este tiempo se esperó la formación de *foci*.

2.5.5.1 Tinción de células

Pasados los 21 días, se aspiró el medio DB de las cajas de Petri y se realizaron dos lavados con PBS frío. Las células se fijaron con 3 ml de metanol frío manteniendo la caja de Petri en hielo durante 10 minutos. Pasado ese tiempo, se retiró el cultivo del hielo, se aspiró el metanol y se agregaron 3 ml de cristal violeta al 0.5% en 25% de metanol. Se incubó a temperatura ambiente durante 5 minutos. Posteriormente, se aspiró la solución de cristal violeta y se enjuagó con PBS hasta que no desprendiera color. Se dejó secar toda la noche sin cubrirla.

2.6. Resultados y discusión

2.6.1 Trabajo de laboratorio

Al momento de comenzar con este proyecto, el Laboratorio Internacional de Investigación sobre el Genoma Humano acababa de ser establecido, por lo que aproximadamente el 80% de lo presentado en este capítulo corresponde a la estandarización de técnicas y protocolos. Las técnicas estandarizadas se listan a continuación:

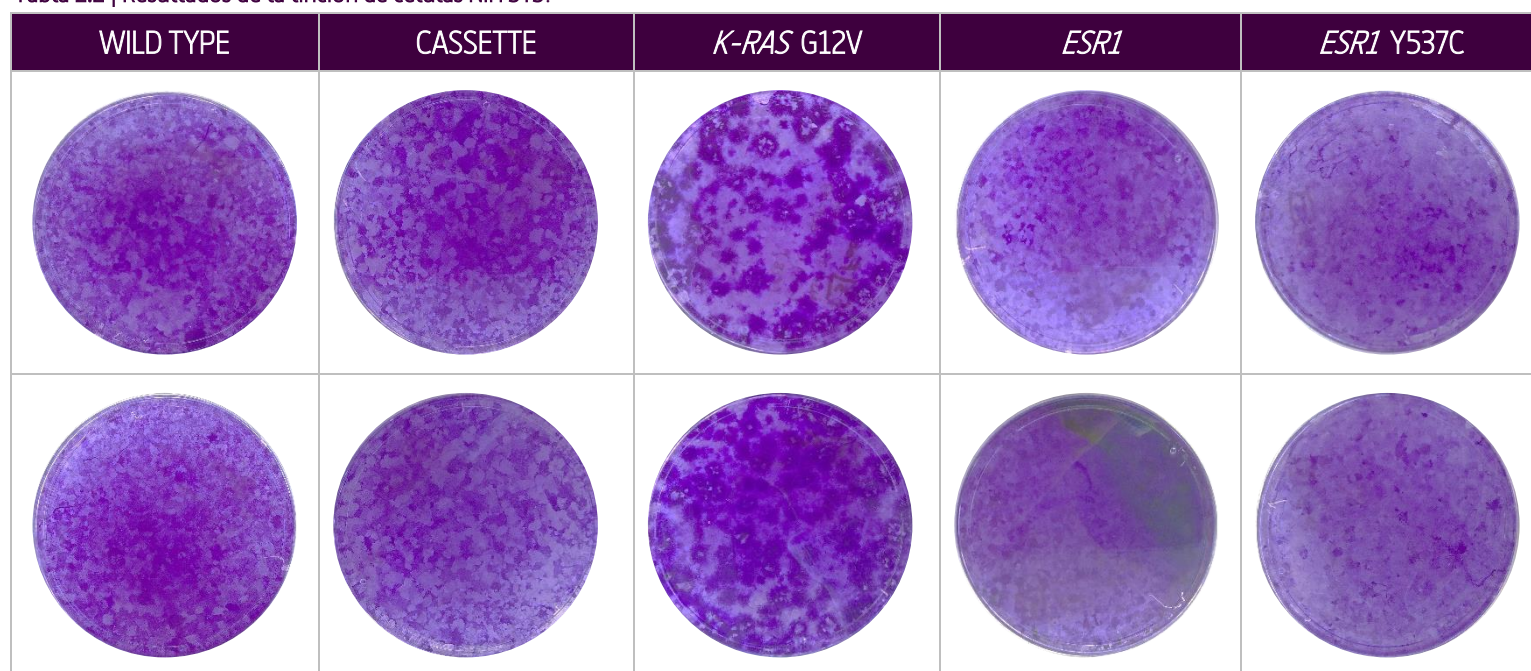
1. Geles de agarosa
2. Digestión y ligación del plásmido BABE-puro y MA-RQ
3. Transformación bacteriana con la cepa XL-Blue
4. Cultivo de células NIH 3T3
5. Transfección de células NIH 3T3 con plásmido BABE-puro
6. Ensayo de formación de focos

Debido a lo anterior, los resultados presentados a continuación son preliminares, ya que se requieren de más réplicas para poder ser validados. Por la misma razón, los resultados obtenidos no presentan análisis estadísticos.

2.6.2 Ensayo de formación de focos

En la [Tabla 2.2](#) se presentan los resultados de la tinción de las células NIH 3T3 del ensayo de formación de focos (al que se le llamará a partir de ahora *ensayo oncogénico*). Las cajas teñidas muestran a simple vista que el único grupo que formó *foci* es el *K-RAS G12V*. Por otro lado, las cajas de WILD TYPE y CASSETTE presentan el comportamiento más cercano entre ellas, mientras que pareciera ser que *ESR1* y *ESR1 Y537C* se comportan de manera similar entre sí; en estas últimas se observan más espacios entre conglomeraciones que en las dos primeras.

Tabla 2.2 | Resultados de la tinción de células NIH 3T3.



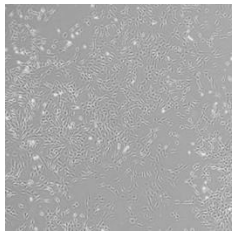
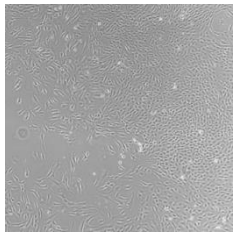
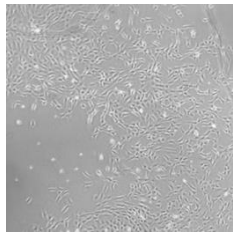
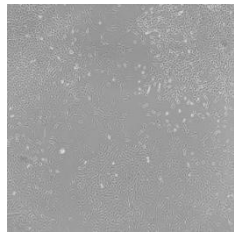
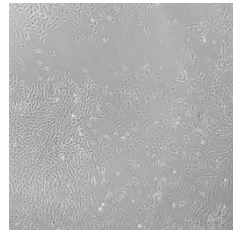
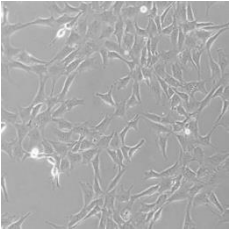
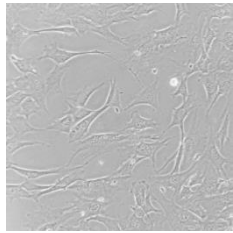
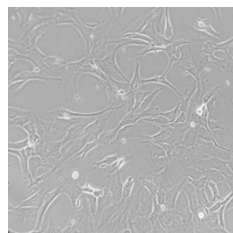
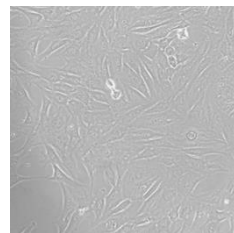
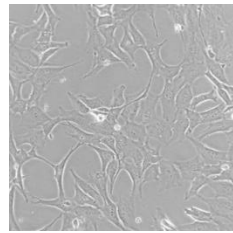
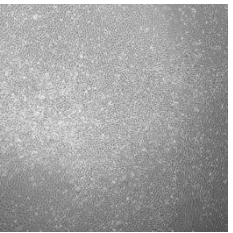
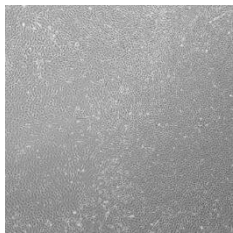
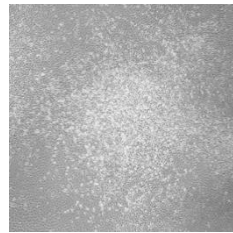

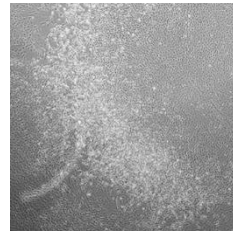
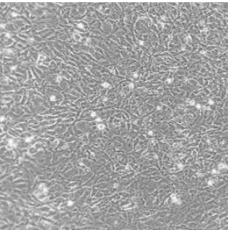
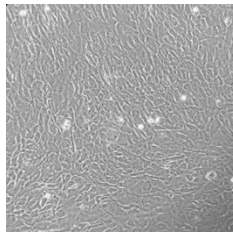
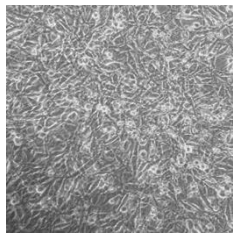
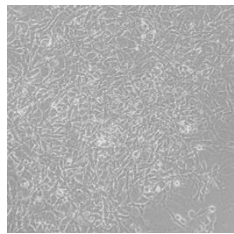
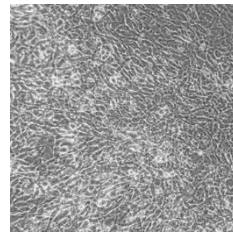
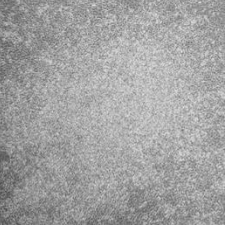

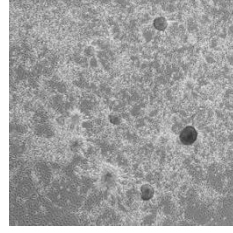
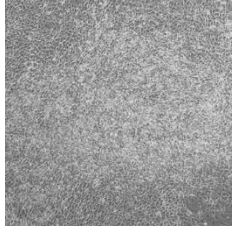
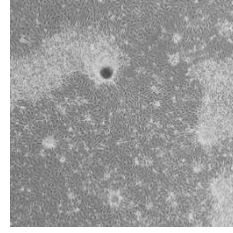
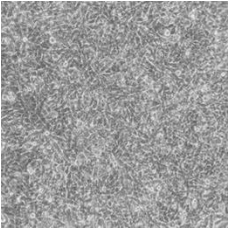
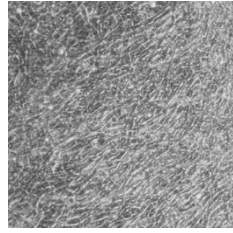
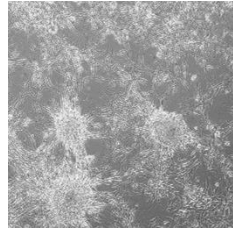
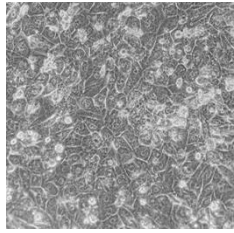
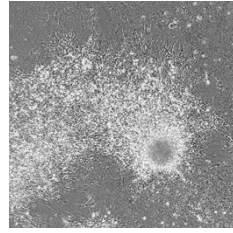
En cada columna se observan los grupos experimentales (por duplicado). Las áreas con mayor coloración corresponden a conglomeraciones de células, mientras que la presencia de *foci* solo es notable a simple vista en el grupo *K-RAS G12V* (correspondiente a los círculos más oscuros).

En la [Tabla 2.3](#) se presentan los resultados obtenidos del ensayo oncogénico a escala microscópica. Durante la primera semana de experimento no se observan cambios evidentes en la confluencia entre grupos. En la segunda semana las cajas se encuentran al 100% de su capacidad y se puede observar que las células comienzan a agruparse; específicamente, el grupo *K-RAS G12V* y el grupo *ESR1 Y537C* presentan una mayor confluencia celular con respecto al resto de grupos, lo cual indica que se comportan de una manera similar. Para la tercera semana, en estos dos últimos grupos se observa la presencia de *foci* (correspondientes a los círculos más oscuros en las imágenes). De

manera general y de acuerdo con lo esperado, el grupo WILD TYPE y CASSETTE presentan similitudes en su comportamiento; además, pareciera ser que el grupo *ESR1* se comporta similar a estos controles negativos.

La formación de *foci* en el grupo *K-RAS* G12V demuestra el funcionamiento del control positivo. Asimismo, el hecho de que microscópicamente se haya observado formación de *foci* en el grupo *ESR1* Y537C, sugiere que esta mutación puede tener un posible efecto oncogénico en las células NIH 3T3, ya que la pérdida de inhibición por contacto y el crecimiento en multicapas, son efecto de un oncogén. Además, en los resultados de la tinción el único grupo al que se le puede observar la formación de *foci* a simple vista es el *K-RAS* G12V, lo que indica que, en este resultado, sólo de manera microscópica se puede observar la diferencia entre este grupo y *ESR1* Y537C.

Tabla 2.3 | Resultados del ensayo oncogénico en células NIH 3T3.

Semana		WILD TYPE	CASSETTE	<i>K-RASG12V</i>	<i>ESR1</i>	<i>ESR1Y537C</i>
1	4x					
	20x					
2	4x					
	20x					
3	4x					
	20x					

En cada columna se presentan los grupos experimentales y en cada fila los resultados de cada semana a un aumento 4x y 20x, respectivamente.

2.7. Conclusiones

Los resultados obtenidos en este trabajo confirman que *K-RAS* G12V es un oncogén funcional en células NIH 3T3. Además, debido a que los grupos experimentales *K-RAS* G12V y *ESR1* Y537C son los únicos que presentan formación de *foci* en el ensayo oncogénico, se hipotetiza que la mutación Y537C tiene un efecto oncogénico en células NIH 3T3. Sin embargo, es necesaria la realización de réplicas que permitan validar lo presentado en este trabajo.

De manera general, los resultados presentados en este estudio son preliminares, debido a que aproximadamente el 80% del trabajo realizado fue hecho en la estandarización de técnicas y protocolos. No obstante, durante este proceso fue posible la comprensión de los protocolos, además del aprendizaje de los distintos factores a considerar tanto en el cultivo y transfección celular, como en el ensayo oncogénico.

2.8. Perspectivas

Se propone que, una vez validados los resultados, estas técnicas sean implementadas en células melanocíticas y de melanoma, para de esa manera poder conocer el papel que desempeña la mutación *ESR1* Y537C en el desarrollo y progresión de melanoma.

Asimismo, se propone la implementación de experimentos complementarios, como lo serían ensayos de proliferación y ensayos de raspado celular (los cuales se describen en el Anexo I), así como la introducción del factor E2, esto para identificar si la sobreestimulación de esta hormona puede influir en un aumento en la actividad del $RE\alpha$, lo que eventualmente podría ser de gran ayuda como posible objetivo terapéutico.

Finalmente, se propone la implementación de una cámara digital en el microscopio empleado en la observación de las células, esto únicamente con el fin de mejorar la resolución de las fotografías tomadas.

Anexo I

Experimentos adicionales

Debido a que la gran parte del trabajo realizado en el área experimental corresponde a estandarización de técnicas y protocolos, en esta sección se describen otros experimentos realizados para la evaluación de la mutación *ESR1* Y537C, entre los cuales se encuentra una réplica del ensayo de formación de focos, así como el inicio de la implementación de ensayos de raspado y proliferación celular. Sin embargo, estos experimentos no forman parte de los resultados debido a que los controles positivos y negativos no funcionaron correctamente, por lo cual todos estos procesos son parte de la estandarización.

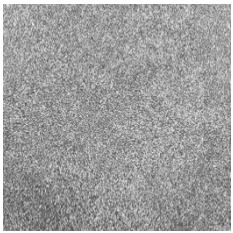
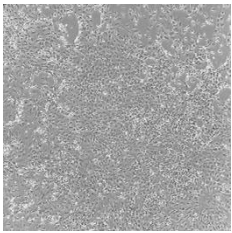
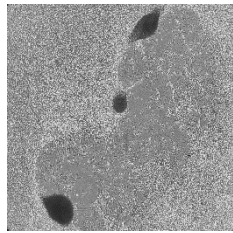
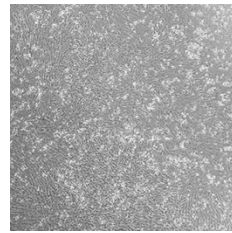
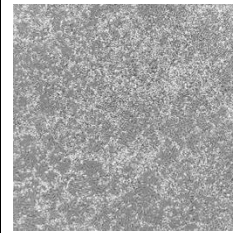
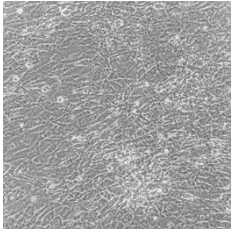
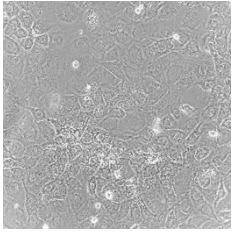
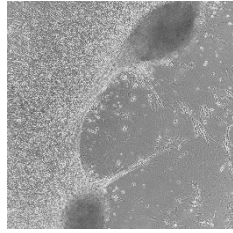
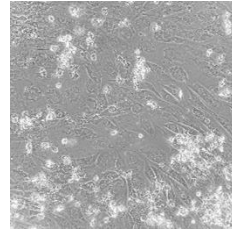
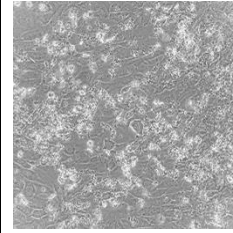
A1.1 Réplica de ensayo de formación de focos

En la Tabla A1.1 se muestran los resultados de una réplica del ensayo oncogénico. Se presentan las imágenes correspondientes a la tercera semana. Es esta réplica las células no lograron adherirse por completo a la caja, razón por la cual se observan más espacios entre ellas en todos los grupos experimentales, así como un cambio en su morfología. Las células NIH 3T3 son adherentes, por lo que la falta de esta capacidad puede ser debida a distintos factores, como, por ejemplo, contaminación en el medio, una recombinación en el plásmido empleado para su transfección o que se estén diferenciando.

En la Tabla A1.2 se pueden observar las cajas después de la tinción. El grupo CASSETTE debía comportarse similar al WT, pero presentó mayor desprendimiento de células. Una vez más, fue posible observar a simple vista la formación de *foci* en el grupo *K-RAS*G12V. En el caso tanto de *ESR1* como de *ESR1* Y537C, la confluencia celular fue tanta, que no se logró apreciar la formación de *foci*, debido al desprendimiento de las células. Asimismo, para este ensayo, se realizó una mejora en la técnica para tomar las imágenes de las cajas teñidas.

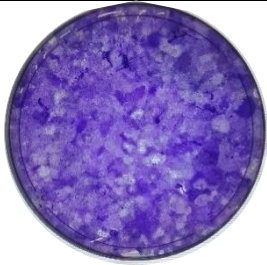
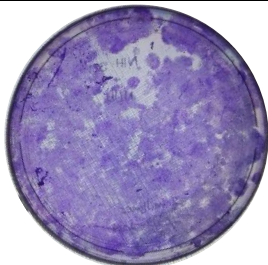
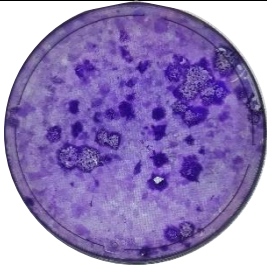
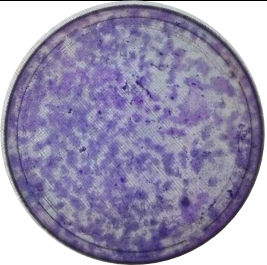
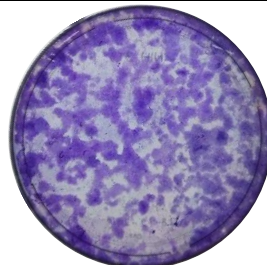


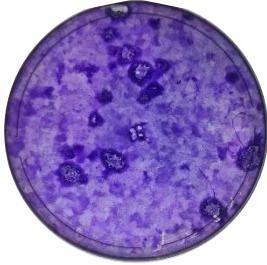
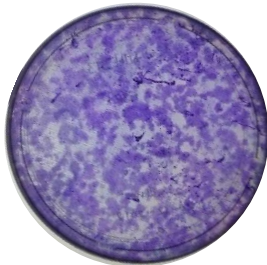
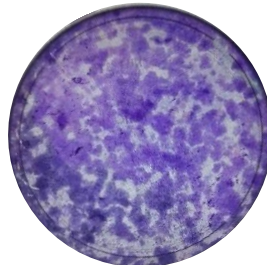
Con la realización de este experimento, se puede afirmar nuevamente que *K-RAS* G12V es un oncogen funcional en la línea celular NIH 3T3, mientras que no es posible determinar esto para *ESR1* Y537C, debido al desprendimiento celular.

Tabla A1.1 | Resultados de una réplica del ensayo oncogénico en células NIH 3T3.

Semana		WILD TYPE	CASSETTE	<i>K-RAS</i> G12V	<i>ESR1</i>	<i>ESR1</i> Y537C
3	4X					
	20X					

En cada columna se presentan los grupos experimentales y se muestran los resultados de la tercera semana a 4x y 20x.

Tabla A1.2 | Resultados de la tinción de células NIH 3T3.

WILD TYPE	CASSETTE	<i>K-RAS</i> G12V	<i>ESR1</i>	<i>ESR1</i> Y537C
				
				

En cada columna se observan los grupos experimentales (por duplicado). La presencia de *foci* solo es notable a simple vista en el grupo *K-RAS* G12V (correspondiente a los círculos más oscuros) y se pueden observar zonas vacías, lo que corresponde al desprendimiento de células.

A1.2 Ensayo de raspado celular

Los ensayos de raspado (Wound Scratch Assay) son utilizados para evaluar la migración y proliferación celular de células transfectadas, de tal manera que la velocidad de migración celular puede estar relacionada con la metástasis. El ensayo consiste en sembrar las células transfectadas y esperar a que formen una monocapa; posterior a ello, se raspa la caja con una punta de pipeta estéril y se analiza la distancia que recorren las células para cerrar la 'herida' realizada y el tiempo que les toma hacerlo⁸⁶.

Para realizar este ensayo se sembraron, por duplicado, 50 millones de células de cada tipo de cepa en una caja de Petri con 5 mL de medio DB. Las células se incubaron hasta que se adhirieron a la caja y formaron una monocapa. Previamente, se dibujó al reverso de las cajas de Petri una cuadrícula con nueve cuadrantes (tres filas). Posteriormente, se removió el medio y se realizó un raspado al centro de cada fila con una punta estéril de 200 μ l; las cajas fueron lavadas con PBS y se les agregaron 5 mL de medio DB. La herida realizada fue monitoreada a las 0 h (el día que se realizó el raspado), 24 y 48 h a través de la toma de fotos de cada uno de los cuadrantes [Fig. A1.1].

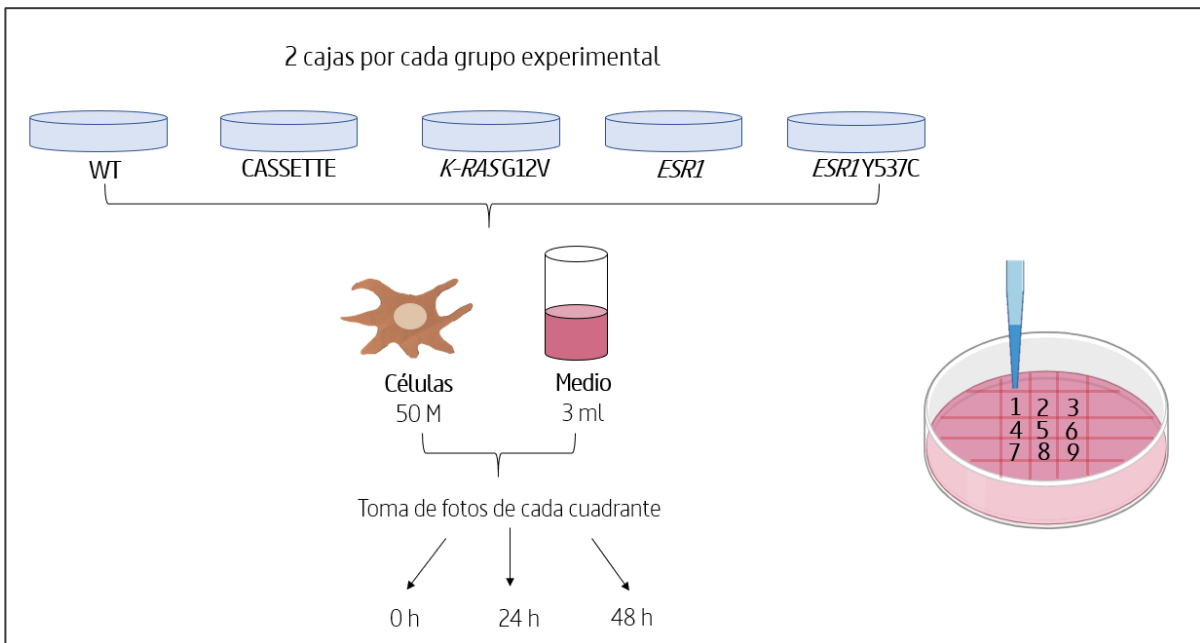


Figura A1.1 | Representación esquemática del ensayo de scratch. A la izquierda se presenta la disposición de los grupos experimentales. A la derecha se muestra la disposición de los cuadrantes; la herida es realizada en el centro de cada fila con una punta estéril (en azul).

Cuantificación de la distancia recorrida por las células

La longitud del raspado para los tres tiempos se analizó digitalmente con el software ImageJ [Fig. A1.2]. A la imagen de cada cuadrante se le agrega una cuadrícula; posteriormente se coloca una línea que va de extremo a extremo de la herida, de tal manera que el software indica en una tabla la longitud de dicha línea. Estas mediciones se promedian y se normalizan, ya que la longitud inicial de cada grupo experimental es distinta.

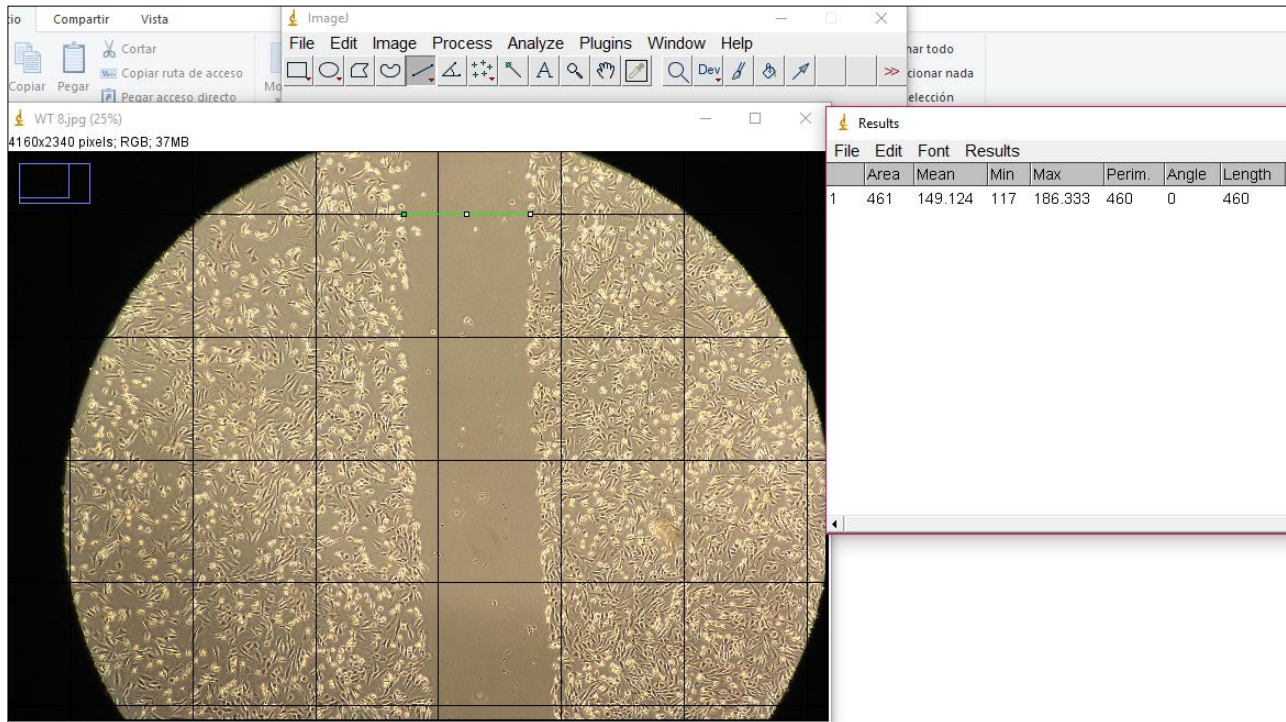


Figura A1.2 | Análisis de las distancias recorridas por las células utilizando ImageJ. La línea verde mide la distancia en píxeles (el resultado se puede observar en la tabla a la derecha).

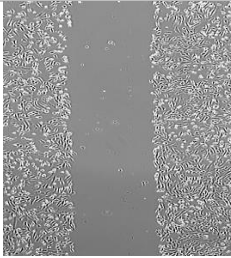
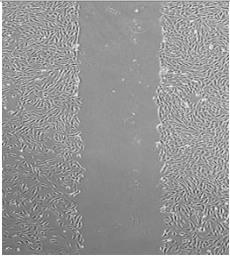
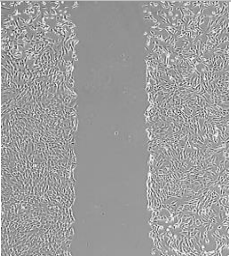
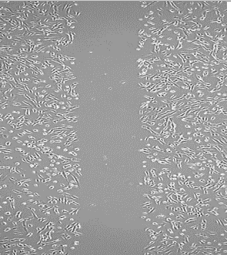
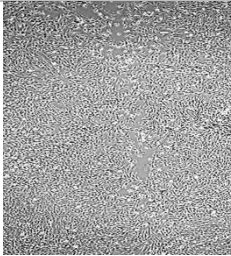
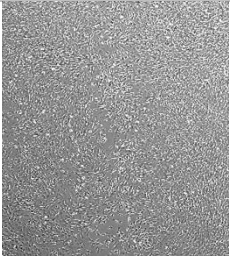
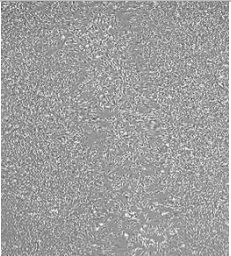
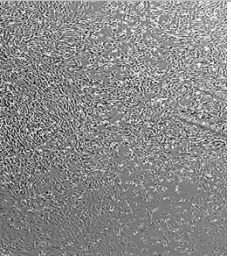
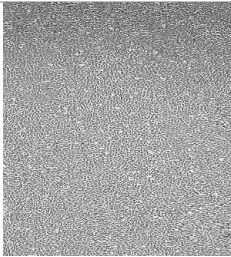
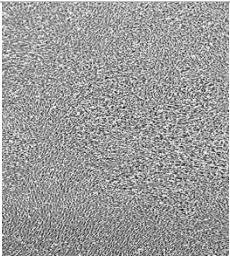
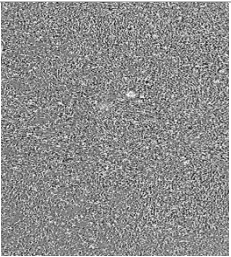
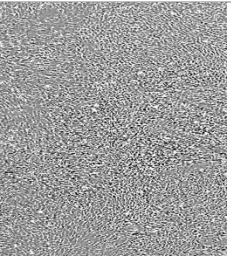
Resultados

En la Tabla A1.3 se presentan las imágenes de un solo cuadrante de cada grupo experimental; se puede observar que, a las 24 h de haber hecho el raspado, las células casi han cerrado por completo la herida, de tal manera que a las 48 h ya no es posible apreciarla. En la Figura A1.3 se muestran los valores normalizados de las distancias recorridas por las células, lo cuales fueron obtenidos a partir de análisis realizados a las imágenes tomadas en cada cuadrante utilizando ImageJ. Debido a que el tamaño de la herida a las 0h en cada grupo experimental es distinto, se normalizó a 100; así, todos los grupos comienzan con el mismo valor, de tal manera que en la gráfica se presenta el

porcentaje de la longitud del raspado a cada tiempo. Se puede observar que todos los grupos recorrieron distancias distintas, siendo CASSETTE el más rápido; sin embargo, esto indica que no se está comportando como el grupo WT, lo que representa que el control negativo no está funcionando correctamente. En este ensayo, el grupo *ESR1* recorrió una mayor distancia que el gen portador de la mutación. No se presentan resultados para *K-RASG12V* debido a contaminación en el cultivo celular.

Como perspectiva está reducir los tiempos en la captura de fotos, ya que la velocidad con que las células cierran la herida es muy alta y, por lo tanto, no es posible apreciarla en los tiempos establecidos en este ensayo.

Tabla A1.3 | Resultados del ensayo de raspado celular.

	WILD TYPE	CASSETTE	<i>ESR1</i>	<i>ESR1</i> Y537C
0 h				
24 h				
48 h				

En cada columna se presentan los grupos experimentales y en cada fila los resultados al cabo de 0 h (es decir, al momento de realizar la herida) a las 24 y 48 h.

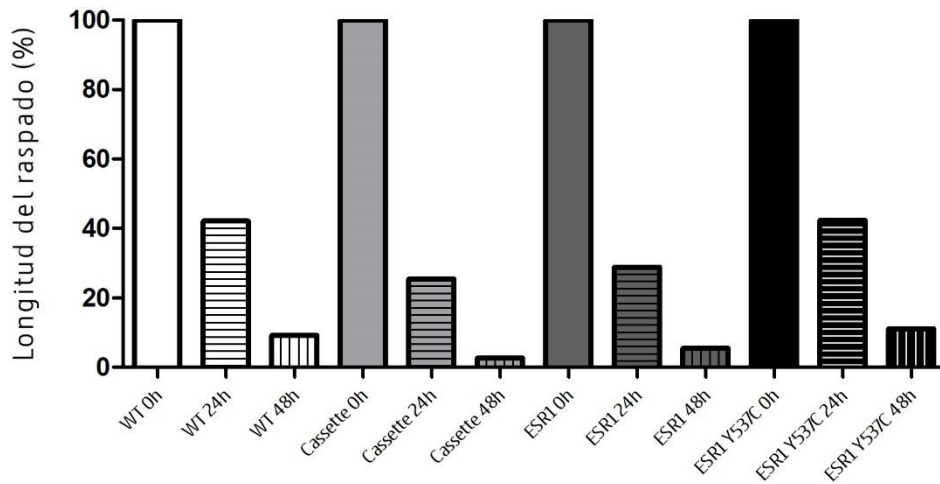


Figura A1.3 | Resultados del ensayo de raspado celular. Se presenta el porcentaje de la longitud del raspado a las 0, 24 y 48. Las distancias fueron obtenidas de las fotos tomadas en cada cuadrante y fueron normalizadas para su posterior comparación.

A1.3 Ensayo de proliferación celular

Se comenzó también con la implementación de ensayos de proliferación celular en presencia y ausencia de E2. Para este ensayo se sembraron, por duplicado, 50 millones de células de cada tipo de cepa en una caja de Petri con 5 mL de medio DB. Por otro lado, se realizó el mismo procedimiento, pero además se les agregó a las cajas estrógeno (E2) a una concentración de 10 nM/mL. En total se tuvieron 20 cajas, de las cuales 10 fueron cuantificadas 24 h después de haberlas sembrado y las 10 restantes fueron cuantificadas 48 h después de la siembra [Fig. A1.4]. La cuantificación celular se realizó utilizando cámara de Neubauer.

Resultados

En la Figura A1.5 se muestran los resultados del ensayo de proliferación. Se observa una disminución radical en el número de células. El grupo *K-RAS* G12V muestra una disminución en su proliferación, mientras que WT y CASSETTE se comportan de manera contraria; incluso, CASSETTE+E2 24 h muestra un mayor número de células que cualquier otro grupo. *ESR1* presenta mayor proliferación las 48 h en ausencia de E2, mientras que parece ser que *ESR1* Y537C en presencia de E2 disminuye. El grupo *ESR1* Y537C 48 h no formó parte del ensayo debido a contaminación encontrada en la caja. Todo lo anterior indica que ningún control funcionó

correctamente, además de que las células murieron en gran cantidad, ya que en todos los grupos la cantidad de células disminuyó.

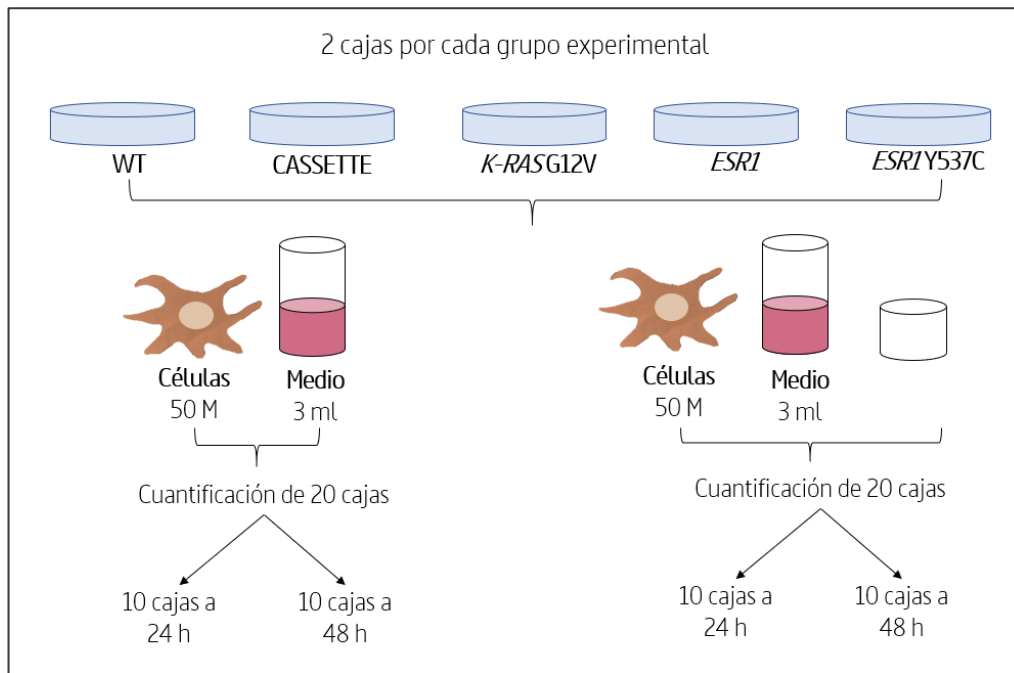


Figura A1.4 | Representación esquemática del ensayo de proliferación celular.

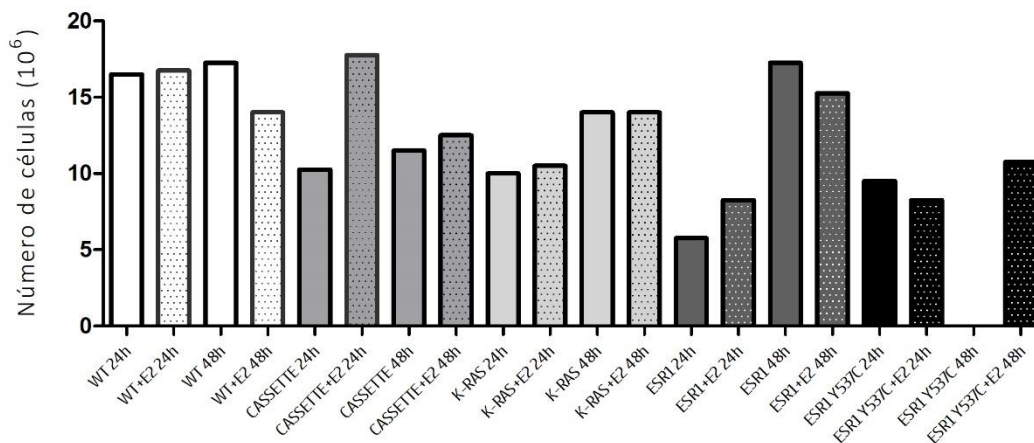


Figura A1.5 | Ensayo de proliferación. Cada barra representa el promedio de las cuantificaciones de cada caja.

Anexo II

Material complementario

```
#!/usr/local/bin/perl -w

use strict;

my $file = "whole_genome_vep_severe.filtered.vcf"; #archivo con variantes
my @lines = ();
my @miss = ();
my @campos = ();
my @info = ();
my @csq = ();
my @a = ();
my @b = ();
my @c = ();

#Se guarda el archivo VCF en un arreglo
open (FILE, "<$file")
    or die("I can't open $file: $!\n");
@lines = <FILE>;
close(FILE);

#Se seleccionan aquellas variantes que contengan la palabra 'missense'
for my $i (0..$#lines){
    chomp ($lines[$i]);
    if ($lines[$i] =~ m/missense\\w\\w/){
        push (@miss, $lines[$i]);}
}

#Se separan las columnas del archivo VCF, quedándonos con INFO
for my $j(0..$#miss){
    @a = split(/\\t/, $miss[$j]);
    push (@campos,$a[7]);
    @a = ();}
for my $k (0..$#campos){
    my $csq;
    if( $campos[$k] =~ m/CSQ\\=(\\^[^;]+) / ){
        $csq = $1; }
    push (@info, $csq);
    @b = ();}

#Se seleccionan las columnas Gen, AA, Cambio, ID COSMIC e ID UniProt
for my $m (0..$#info){
    @c = split(/\\|/, $info[$m]);
    my @selected = ( $c[3], $c[14], $c[15], $c[17], $c[29] );
    for my $att ( 0..$#selected ){
        if( !defined( $selected[$att] ) ){
            warn(
                "Valor no inicializado $att en $c[3], $c[14], $c[15], $c[17], $c[29]\\n" );
            $selected[$att] = "";
        }
    }
    push (@csq, join( " ", @selected ) );
    @c = ();}
```

Figura A2.1 | Código utilizado para extraer variantes 'missense'. En este código también se seleccionaron los campos correspondientes al nombre del gen, la posición del aminoácido que sufrió el cambio, el cambio producido, ID de COSMIC y el ID de UniProtKB/Swiss-Prot.

```

#!/usr/bin/perl -w

use strict;

my $file = "vcf_procesado.COSMIC.vcf"; #Archivo VCF con los filtros anteriores
my $base = "Phosphorylation.txt"; #Base de datos de fosforilación de dbPTM
my $output = "variantes_idphospho.txt";
my @cos = (), my @mic = (), my @pro = (),my @ID = (),my @variants = ();

#Se guarda el archivo VCF en un arreglo
open (FILE, "<$file")
    or die ("no: $!\n");
my @lineas = <FILE>;
chomp( @lineas );
close (FILE);

#Busca el ID de UniProt en la base de datos de fosforilación; si el ID coincide,
#se guardan los campo en el archivo de texto
open( OUTPUT, ">$output" )
    or die( "no pude crear output: $!\n" );
for my $i (0..$#lineas){
    my @fields = split( / /, $lineas[$i] );
    if( defined( $fields[4] ) ){
        my $protein = `grep $fields[4] $base`;
        if( $protein ne '' ){
            my @protein_lines = split( /\n/, $protein );
            foreach my $protein_line ( @protein_lines ){
                my @protein_fields = split( /\t/, $protein_line);
                if( $protein_fields[2] == $fields[1] ){
                    print( $lineas[$i] . "\n" );
                    print ( OUTPUT ( $lineas[$i] . "\n" ));
                }
            }
        }
        else{
            print( "No encuentre la proteina $fields[4]\n" );
        }
    }
    else{
        print( "Protein not defined for line $lineas[$i]\n" );
    }
}
close( OUTPUT );

```

Figura A2.2 | Código utilizado para extraer variantes que interrumpieran sitios de fosforilación. Se hizo una búsqueda en la base de fosforilación de dbPTM, de tal manera que al final se obtuvo una lista de las variantes que cumplieran con los filtros establecidos.

```

#!/usr/bin/perl -w

use strict;
use warnings;

my $utils = 'http://eutils.ncbi.nlm.nih.gov/entrez/eutils/esearch.fcgi?db=pubmed';
my $utils2 = 'term=__GENE__[Title/Abstract]+AND+__KEYWORD__[Title/Abstract]+AND+__AA__[Title/Abstract]';

my $lista_final = "variantes_idphospho.txt";
my $output = "lista_publicaciones.txt";
my @keywords = ( 'cancer' );
my @campos = ();
my $scores;

open( INPUT, $lista_final ) or die( "Unable to open input file $lista_final: $!\n" );
while( <INPUT> ){
    chomp( $_ );
    push( @campos, $_ );
}
close( INPUT );

open( OUTPUT, ">$output" ) or die( "Unable to open/create output file $output: $!\n" );
print OUTPUT ( "Gene_name\tNum_of_pubmed_records(gene+cancer+amino acid)\n" );
foreach my $campo ( @campos ){
    my @fields = split( / /, $campo );
    my $searched_name = $campo;
    foreach my $keyword ( @keywords ){
        my $search = $utils2;
        $search =~ s/\_\_\_GENE\_\_\_/$fields[0]/;
        $search =~ s/\_\_\_KEYWORD\_\_\_/$keyword/;
        $search =~ s/\_\_\_AA\_\_\_/$fields[1]/;
        my $web = $utils . "\\&" . $search;
        my $html = qx{wget --quiet --output-document=- $web};
        $html =~ m/\<Count\>(\d+)\</Count\>/;
        my $count = $1;
        $scores{ $campo } = $count;
        sleep( 2 );
        print( $searched_name . "\t" . $count . "\n" );
    }
}

foreach my $campo ( sort keys %scores ){
    print OUTPUT ( $campo . ": " . $scores{ $campo } . "\n" );
}
close( OUTPUT );

```

Figura A2.3 | Código utilizado para obtener el número de publicaciones de cada una de las variantes de la lista final.

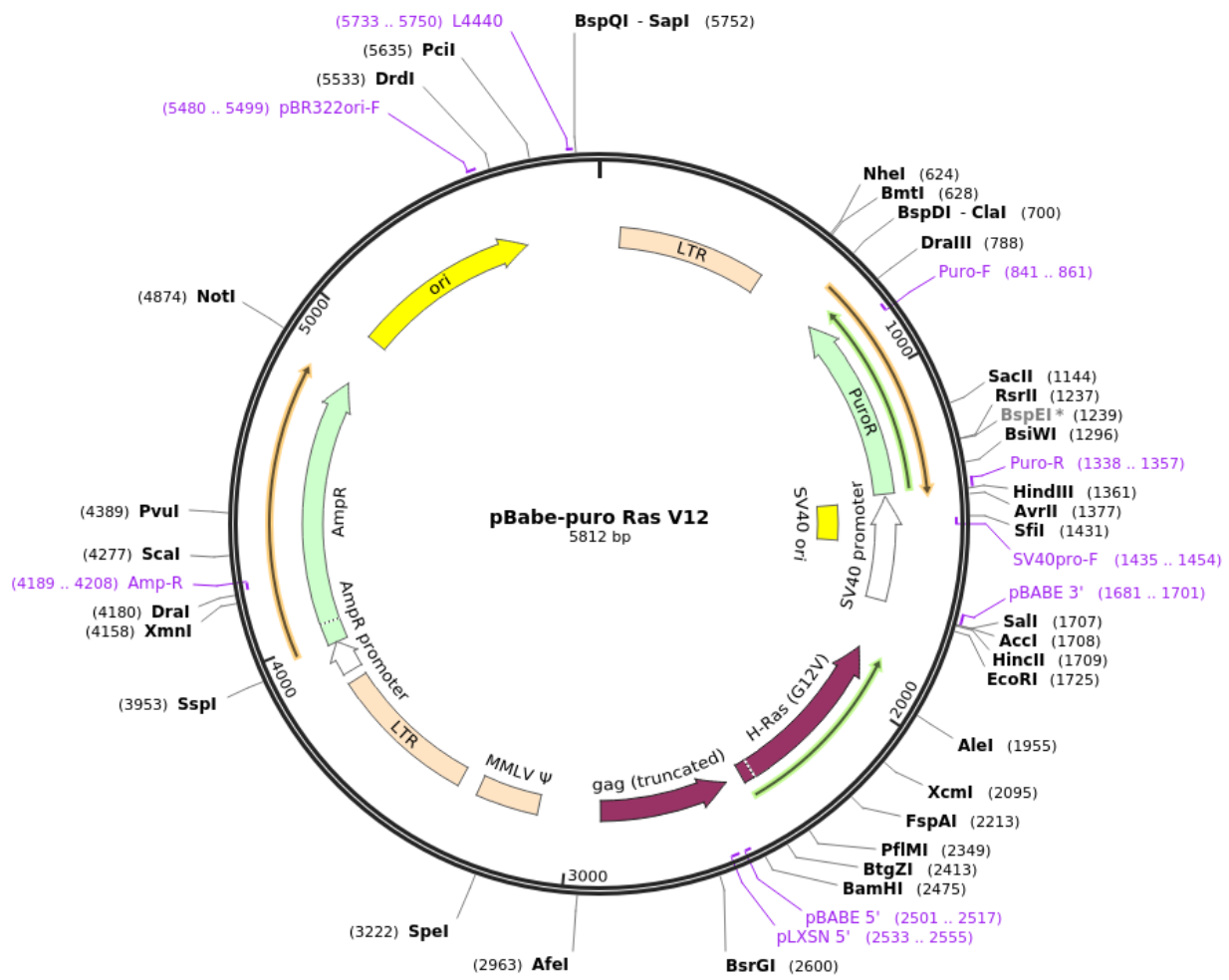
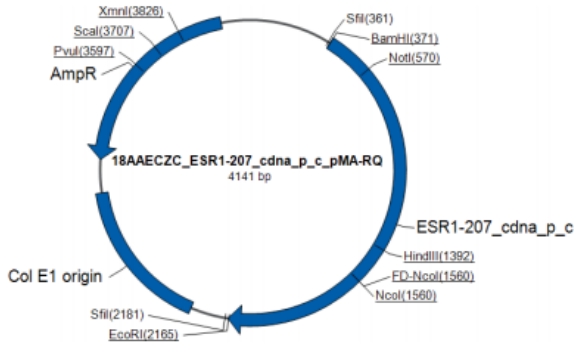


Figura A.4 | Mapa del plásmido BABE-puro con el inserto *H-RAS* G12V. Se muestra las enzimas que flanquean al inserto: Bam HI y EcoRI.

Plasmid DNA Description:

The synthetic gene *ESR1-207_cdna_p_c* was assembled from synthetic oligonucleotides and/or PCR products. The fragment was inserted into pMA-RQ. The plasmid DNA was purified from transformed bacteria and concentration determined by UV spectroscopy. The final construct was verified by sequencing. The sequence congruence within the insertion sites was 100%. 5 µg of the plasmid preparation were lyophilized for shipping.

Plasmid Map:



Quality Assurance Documentation: 18AAECZC

Ref. No.: 2273748

Designation: E.coli K12 DH10B™ T1R

Gene Name: ESR1-207_cdna_p_c

Gene Size: 1800 bp

Vector backbone: pMA-RQ

Cloning sites: SfiI / SfiI

Quantity: ~5 µg Plasmid DNA

Note: Please dissolve lyophilized DNA in 50 µl distilled water or 10 mM Tris-HCl (pH 8.0). We recommend sequence verification after each transformation step.

Date: 17 February 2018

Carmen Billerbeck

Quality control

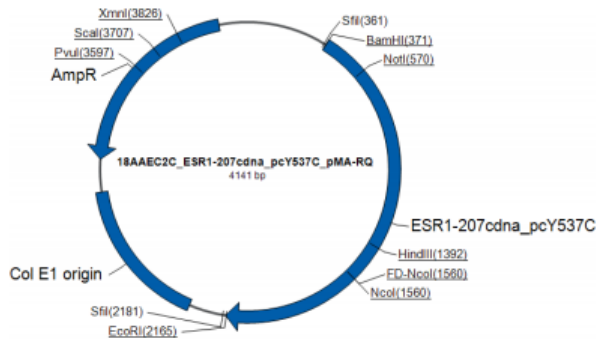
www.thermofisher.com

geneartsupport@thermofisher.com

Plasmid DNA Description:

The synthetic gene *ESR1-207cdna_pcY537C* was assembled from synthetic oligonucleotides and/or PCR products. The fragment was inserted into pMA-RQ. The plasmid DNA was purified from transformed bacteria and concentration determined by UV spectroscopy. The final construct was verified by sequencing. The sequence congruence within the insertion sites was 100%. 5 µg of the plasmid preparation were lyophilized for shipping.

Plasmid Map:



Quality Assurance Documentation: 18AAEC2C

Ref. No.: 2273749

Designation: E.coli K12 DH10B™ T1R

Gene Name: ESR1-207cdna_pcY537C

Gene Size: 1800 bp

Vector backbone: pMA-RQ

Cloning sites: SfiI / SfiI

Quantity: ~5 µg Plasmid DNA

Note: Please dissolve lyophilized DNA in 50 µl distilled water or 10 mM Tris-HCl (pH 8.0). We recommend sequence verification after each transformation step.

Date: 16 February 2018

Sarah Weingart

Quality control

www.thermofisher.com

geneartsupport@thermofisher.com

Figura A2.5 | Mapa del plásmido MA-RQ con el inserto *ESR1* y *ESR1 Y537C*, respectivamente. Se muestra especificaciones de cada plásmido y se puede observar que los insertos contienen las mismas enzimas de restricción.

Referencias

1. MacConaill, L. E. & Garraway, L. A. Clinical implications of the cancer genome. *Journal of Clinical Oncology* 28, 5219–5228 (2010).
2. Cooper, G. M. The development and causes of cancer. in *The Cell: A Molecular Approach* (Sunderland (MA): Sinauer Associates, 2000).
3. Mathur, G., Nain, S. & Sharma, P. K. Cancer : An Overview. *Acad. J. Cancer Res.* 8, 1–9 (2015).
4. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: The next generation. *Cell* 144, 646–674 (2011).
5. Tyagi, N., Sharma, G. N., Shrivastava, B., Chaydhary, N. & Sahu, N. Cancer: An Overview. *Int. J. Res. Dev. Pharm. Life Sci.* 6, 2740–2747 (2017).
6. World Health Organization. *Latest global cancer data: Cancer burden rises to 18.1 million new cases and 9.6 million cancer deaths in 2018.* (2018).
7. National Cancer Institute. How Cancer Is Diagnosed. (2019). Available at: <https://www.cancer.gov/about-cancer/diagnosis-staging/diagnosis>.
8. Mayo Clinic. Cancer. (2018). Available at: <https://www.mayoclinic.org/diseases-conditions/cancer/diagnosis-treatment/drc-20370594>.
9. National Cancer Institute. Genetic Testing for Inherited Cancer Susceptibility Syndromes. (2019). Available at: <https://www.cancer.gov/about-cancer/causes-prevention/genetics/genetic-testing-fact-sheet>.
10. Yang, M. & Kim, J. W. Principles of genetic counseling in the era of next-generation sequencing. *Annals of Laboratory Medicine* 38, 291–295 (2018).
11. Serrati, S. *et al.* Next-generation sequencing: advances and applications in cancer diagnosis. *Onco Targets Ther.* 2016, 7355–7365 (2016).
12. Zhang, X. *et al.* Application of next-generation sequencing technology to precision medicine in cancer: joint consensus of the Tumor Biomarker Committee of the Chinese Society of Clinical Oncology. *Cancer Biol. Med.* 16, 189–204 (2019).
13. Nakagawa, H. & Fujita, M. Whole genome sequencing analysis for cancer genomics and precision medicine. *Cancer Sci.* 109, 513–522 (2018).
14. Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* 500, 415–421 (2013).
15. Chatterjee, T. Next Generation Sequencing In Clinical Molecular Diagnosis of Cancer: A Journey towards Personalized Diagnosis and Therapeutics. *J. Med. Oncol. Ther.* 1, 76–80 (2016).
16. Kamps, R. *et al.* Next-Generation Sequencing in Oncology: Genetic Diagnosis, Risk Prediction and Cancer Classification. *Int. J. Mol. Sci.* 18, 1–57 (2017).
17. Pant, S., Ramarao, M., Green, G. A. & Montalto, M. C. Next-Generation Sequencing–Based Companion Diagnostics: From Biomarker Discovery to Clinical Implementation. in *Companion and Complementary Diagnostics* 135–163 (Elsevier Inc., 2019).
18. Gagan, J. & Van Allen, E. M. Next-generation sequencing to guide cancer therapy. *Genome Med.* 7, 1–10 (2015).
19. Thompson, J. C. *et al.* Detection of therapeutically targetable driver and resistance mutations in lung cancer patients by next-generation sequencing of cell-free circulating tumor DNA. *Clin. Cancer Res.* 22, 5772–5782 (2016).
20. Robert, C. *et al.* Improved overall survival in melanoma with combined dabrafenib and trametinib. *N. Engl. J. Med.* 372, 1–11 (2015).

21. Shannan, B., Perego, M., Somasundaram, R. & Herlyn, M. Heterogeneity in melanoma. in *Cancer Treatment and Research* 167, 1–15 (Kluwer Academic Publishers, 2016).
22. Turner, N., Ware, O. & Bosenberg, M. Genetics of metastasis: melanoma and other cancers. *Clin. Exp. Metastasis* 35, 379–391 (2018).
23. American Cancer Society. *About Melanoma Skin Cancer*. (2019).
24. Damsky, W. E., Theodosakis, N. & Bosenberg, M. Melanoma metastasis: new concepts and evolving paradigms. *Oncogene* 33, 2413–2422 (2014).
25. Lin, J. Y. & Fisher, D. E. Melanocyte biology and skin pigmentation. *Nature* 445, 843–850 (2007).
26. Losquadro, W. D. Anatomy of the Skin and the Pathogenesis of Nonmelanoma Skin Cancer. *Facial Plast. Surg. Clin. North Am.* 25, 283–289 (2017).
27. Merghoub, T., Polsky, D. & Houghton, A. N. Molecular Biology of Melanoma. in *The Molecular Basis of Cancer* 463–470 (2008).
28. Law, M. H., MacGregor, S. & Hayward, N. K. Melanoma genetics: Recent findings take us beyond well-traveled pathways. *J. Invest. Dermatol.* 132, 1763–1774 (2012).
29. Veierød, M. B., Adami, H.-O., Lund, E., Armstrong, B. K. & Weiderpass, E. Sun and solarium exposure and melanoma risk: Effects of age, pigmentary characteristics, and nevi. *Cancer Epidemiol. Biomarkers Prev.* 19, 111–120 (2010).
30. De La Fuente, A. & Ocampo, J. Melanoma cutáneo. *Gac. Med. Mex.* 146, 126–135 (2010).
31. Goldstein, B. G. & Goldstein, A. O. The Diagnosis and Management of Malignant Melanoma. *Am. Fam. Physician* 63, 33–66 (2001).
32. Matthews, N. H., Li, W.-Q., Qureshi, A. A., Weinstock, M. A. & Cho, E. Epidemiology of melanoma. in *Cutaneous Melanoma: Etiology and Therapy* (Codon Publications, 2017).
33. Carr, S., Smith, C. & Wernberg, J. Epidemiology and Risk Factors of Melanoma. *Surg. Clin. North Am.* 100, 1–12 (2020).
34. Geeleher, P. & Huang, R. S. Exploring the link between the germline and somatic genome in cancer. *Cancer Discov.* 7, 354–355 (2017).
35. Griffiths, A. J. F., Wessler, S. R., Carroll, S. B. & Doebley, J. *Introduction to Genetic Analysis*. (W.H. Freeman & Company, 2015).
36. Hayward, N. K. Genetics of melanoma predisposition. *Oncogene* 22, 3053–3062 (2003).
37. Gandini, S. *et al.* Meta-analysis of risk factors for cutaneous melanoma: III. Family history, actinic damage and phenotypic factors. *Eur. J. Cancer* 41, 2040–2059 (2005).
38. Rossi, M. *et al.* Familial melanoma: diagnostic and management implications. *Dermatol. Pract. Concept.* 9, 10–16 (2019).
39. Read, J., Wadt, K. A. W. & Hayward, N. K. Melanoma genetics. *J. Med. Genet.* 53, 1–14 (2015).
40. Goldstein, A. M. *et al.* Features associated with germline CDKN2A mutations: A GenoMEL study of melanoma-prone families from three continents. *J. Med. Genet.* 44, 99–106 (2007).
41. Brown, T. A. *Genomes 3*. (Garland Sciences, 2007).
42. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158 (2011).
43. Tollefson, G. A. *et al.* VIVA (Visualization of VAriants): A VCF File Visualization Tool. *Sci. Rep.* 9, 1–7 (2019).
44. Robles-Espinoza, C. D. *et al.* POT1 loss-of-function variants predispose to familial melanoma. *Nat. Genet.* 46, 478–481 (2014).

45. Harland, M. *et al.* Germline TERT promoter mutations are rare in familial melanoma. *Fam. Cancer* 15, 139–144 (2016).
46. Aoude, L. G. *et al.* Nonsense Mutations in the Shelterin Complex Genes ACD and TERF2IP in Familial Melanoma. *J. Natl. Cancer Inst.* 107, 1–9 (2015).
47. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol.* 17, 1–14 (2016).
48. Sherry, S. T. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* 29, 308–311 (2001).
49. Tate, J. G. *et al.* COSMIC: The Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res.* 47, D941–D947 (2019).
50. Boutet, E., Lieberherr, D., Tognolli, M., Schneider, M. & Bairoch, A. UniProtKB/Swiss-Prot. in *Plant Bioinformatics* 406, 89–112 (Humana Press, 2007).
51. Huang, K. Y. *et al.* dbPTM in 2019: exploring disease association and cross-talk of post-translational modifications. *Nucleic Acids Res.* 47, D298–D308 (2019).
52. T, R. *et al.* Association of ESR1 Mutations and Visceral Metastasis in Patients with Estrogen Receptor-Positive Advanced Breast Cancer from Brazil. *J. Oncol.* 2019, (2019).
53. Harrod, A. *et al.* Genomic modelling of the ESR1 Y537S mutation for evaluating function and new therapeutic approaches for metastatic breast cancer. *Oncogene* 36, 2286–2296 (2017).
54. Kumar, M. *et al.* 18F-fluoroestradiol PET imaging of activating estrogen receptor- α mutations in breast cancer. *J. Nucl. Med.* 60, 1247–1252 (2019).
55. Jeselsohn, R. *et al.* Emergence of constitutively active estrogen receptor- α mutations in pretreated advanced estrogen receptor-positive breast cancer. *Clin. Cancer Res.* 20, 1757–1767 (2014).
56. The Genome Aggregation Database (gnomAD). Available at: <https://www.nature.com/immersive/d42859-020-00002-x/index.html>. (Accessed: 1st August 2020)
57. Jeselsohn, R., Buchwalter, G., De Angelis, C., Brown, M. & Schiff, R. ESR1 mutations—a mechanism for acquired endocrine resistance in breast cancer. *Nature Reviews Clinical Oncology* 12, 573–583 (2015).
58. Martin, L.-A. *et al.* Discovery of naturally occurring ESR1 mutations in breast cancer cell lines modelling endocrine resistance. *Nat. Commun.* 8, 1–15 (2017).
59. Tecalco-Cruz, A. C. & Ramírez-Jarquín, J. O. Mechanisms that Increase Stability of Estrogen Receptor Alpha in Breast Cancer. *Clin. Breast Cancer* 17, 1–10 (2017).
60. Weis, K. E., Ekena, K., Thomas, J. A., Lazennec, G. & Katzenellenbogen, B. S. Constitutively active human estrogen receptors containing amino acid substitutions for tyrosine 537 in the receptor protein. *Mol. Endocrinol.* 10, 1388–1398 (1996).
61. Norman, A. W. Steroid Hormones. in *Hormones* 3166–3174 (2003). doi:10.1016/B978-0-08-091906-5.00002-1
62. Yaşar, P., Ayaz, G., User, S. D., Güpür, G. & Muyan, M. Molecular mechanism of estrogen–estrogen receptor signaling. *Reproductive Medicine and Biology* 16, 4–20 (2016).
63. Lemmen, J. G. *et al.* Expression of estrogen receptor alpha and beta during mouse embryogenesis. *Mech. Dev.* 81, 163–167 (1999).
64. Norman, A. W., Mizwicki, M. T. & Norman, D. P. G. Steroid-hormone rapid actions, membrane receptors and a conformational ensemble model. *Nat. Rev. Drug Discov.* 3, 27–41 (2004).
65. Jia, M., Dahlman-Wright, K. & Gustafsson, J. Å. Estrogen receptor alpha and beta in health and disease. *Best Pract. Res. Clin. Endocrinol. Metab.* 29, 557–568 (2015).
66. Vrtačnik, P., Ostanek, B., Mencej-Bedrač, S. & Marc, J. The many faces of estrogen signaling. *Biochem. Medica* 24, 329–342 (2014).

67. Muramatsu, M. & Inoue, S. Estrogen receptors: How do they control reproductive and nonreproductive functions? *Biochem. Biophys. Res. Commun.* 270, 1–10 (2000).
68. Coriano, C. G. *et al.* A Computational-Based Approach to Identify Estrogen Receptor α/β Heterodimer Selective Ligands. *Mol. Pharmacol.* 93, 197–207 (2018).
69. Chakraborty, S., Willett, H. & Biswas, P. K. Insight into estrogen receptor beta-beta and alpha-beta homo- and heterodimerization: A combined molecular dynamics and sequence analysis study. *Biophys. Chem.* 170, 42–50 (2012).
70. de Giorgi, V. *et al.* Influence of sex hormones on melanoma. *J. Clin. Oncol.* 29, e94–e95 (2011).
71. Pearce, S. T. & Jordan, V. C. The biological role of estrogen receptors α and β in cancer. *Critical Reviews in Oncology/Hematology* 50, 3–22 (2004).
72. Hah, N. & Kraus, W. L. Hormone-regulated transcriptomes: Lessons learned from estrogen signaling pathways in breast cancer cells. *Mol. Cell. Endocrinol.* 382, 652–664 (2014).
73. Roh, M. R., Eliades, P., Gupta, S., Grant-Kels, J. M. & Tsao, H. Cutaneous melanoma in women. *Int. J. Women's Dermatology* 1, 21–25 (2015).
74. Bray, F. *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA. Cancer J. Clin.* 68, 394–424 (2018).
75. Joosse, A. *et al.* Gender differences in melanoma survival: Female patients have a decreased risk of metastasis. *J. Invest. Dermatol.* 131, 719–726 (2011).
76. Hall, G. & Phillips, T. J. Estrogen and skin : The effects of estrogen , menopause , and hormone replacement therapy on the skin. *Am. Acad. Dermatology* 53, 555–568 (2005).
77. Miller, J. G. & Mac Neil, S. Gender and cutaneous melanoma. *Br. J. Dermatol.* 136, 657–665 (1997).
78. Stevenson, S. & Thornton, J. Effect of estrogens on skin aging and the potential role of SERMs. *Clin. Interv. Aging* 2, 283–297 (2007).
79. Donley, G. M. *et al.* Reproductive factors, exogenous hormone use and incidence of melanoma among women in the United States. *Br. J. Cancer* 1–7 (2019).
80. de Giorgi, V. *et al.* Estrogen Receptor Expression in Cutaneous Melanoma. *Arch. Dermatol.* 145, 30–36 (2009).
81. Schmidt, A. *et al.* Oestrogen receptor-beta expression in melanocytic lesions. *Exp. Dermatol.* 15, 971–980 (2006).
82. Glatthaar, H., Katto, J., Vogt, T. & Mahlknecht, U. Estrogen receptor alpha (ESR1) single-nucleotide polymorphisms (SNPs) affect malignant melanoma susceptibility and disease course. *Genet. Epigenetics* 1, 1–6 (2016).
83. Morgenstern, J. P. & Land, H. Advanced mammalian gene transfer: High titre retroviral vectors with multiple drug selection markers and a complementary helper-free packaging cell line. *Nucleic Acids Res.* 18, 3587–3596 (1990).
84. Copeland, N. G. & Cooper, G. M. Transfection by exogenous and endogenous murine retrovirus DNAs. *Cell* 16, 347–356 (1979).
85. Alvarez, A., Barisone, G. A. & Diaz, E. Focus Formation: A Cell-based Assay to Determine the Oncogenic Potential of a Gene. *J. Vis. Exp.* 1–6 (2014).
86. Cormier, N., Yeo, A., Fiorentino, E. & Paxson, J. Optimization of the wound scratch assay to detect changes in murine mesenchymal stromal cell migration after damage by soluble cigarette smoke extract. *J. Vis. Exp.* 2015, e53414 (2015).