



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE ESTUDIOS SUPERIORES IZTACALA

ANÁLISIS BIOINFORMÁTICO DEL EFECTO DE LOS  
ALELOS PATOGENICOS EN GENES DE  
REPARACIÓN DEL ADN ASOCIADOS A LA  
SUSCEPTIBILIDAD AL SÍNDROME DE CÁNCER DE  
MAMA Y OVARIO HEREDITARIO

TESIS  
QUE PARA OPTAR POR EL GRADO DE:  
BIÓLOGA

PRESENTA  
**NEFTE MENI YAÑEZ MENDOZA**

DIRECTOR DE TESIS  
Dr. FELIPE VACA PANIAGUA

MÉXICO, CIUDAD DE MÉXICO, 2022



Universidad Nacional  
Autónoma de México



**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.

Universidad Nacional Autónoma de México  
Facultad de Estudios Superiores Iztacala  
Unidad de Biomedicina  
Laboratorio Nacional en Salud:  
Diagnóstico Molecular y Efecto Ambiental en  
Enfermedades Crónico-Degenerativas

## AUTORES

### TESISTA:

Nefte Meni Yañez Mendoza

Facultad de Estudios Superiores Iztacala, UNAM

### DIRECTOR DE TESIS:

Dr. Felipe Vaca Paniagua

Facultad de Estudios Superiores Iztacala, UNAM

UBIMED

Instituto Nacional de Cancerología

## **Agradecimientos institucionales**

A la Universidad Nacional Autónoma de México y en especial a la Facultad de estudios superiores Iztacala.

A mi tutor Dr. Felipe Vaca Paniagua; a mi asesor Dr. Héctor Martínez Gregorio y a los miembros de mi comité sinodal: Dr. Miguel Santibáñez Andrade, Dr. Aldo Hugo De La Cruz Montoya, Dra. Clara Estela Díaz Velázquez y la Dra. Tania Garrido Garduño.

A los integrantes del Laboratorio de Epigenética y Medio Ambiente, y el Laboratorio de Carcinogénesis y Medio Ambiente del Instituto Nacional de Cancerología (INCan).

## **Agradecimientos personales**

El otro día mi abuelita Gloria me dijo que no dejara que el tiempo pasara en vano, me dejó pensando por un buen rato y en retrospectiva me di cuenta de la enorme fortuna que me ha brindado la vida.

Le doy gracias a mis amados padres, quienes me han dado todo el apoyo y amor de manera inmensurable, me han enseñado que es amor, nobleza y fuerza en su máxima expresión; a mi hermano por ser el mejor compañero de vida que uno podría haber deseado, gracias por ser siempre mi protector de primera línea.

A mis abuelos por darme tanto amor y otorgarme de su sabiduría; a Gloris por mostrarme que todo se logra con perseverancia y paciencia. A mi abuelo Cruz por enseñarme a disfrutar de lo esencial, aquello que es invisible para los ojos. A mis abuelos Crispina y Ramiro por ser un ejemplo de fortaleza. A Irais por ser una guía de continuo crecimiento como científica y como persona desde mi primer día en este mundo, y gracias por traer a esta vida a alguien con niveles de locura inimaginables que se hace llamar Luciana.

Gracias por a todos ustedes por siempre protegerme, amarme y acompañarme, gracias por darme una familia increíble.

Sin duda mis pilares son mi familia, pero ¿qué es uno sin los lazos que genera con su transitar?, en este rubro tampoco he dejado pasar el tiempo en vano. Gracias a mis amigos entrañables Jessica, Florence, Shaday, Claudia Bazán, Javier Aragón, Adri Chavira, Karla Lugo, Alan Ramírez, Benny, Luis Bustos, Alexis Mendoza, Eduardo Cerón, Jair G. Dimas, Romina Avendaño, Paulina, Emiliano Betancourt, Keneth Ventura, Cristian Araiza, Jonathan Garibay, Jorge Pineda, Rubén Moctezuma, Javier Lozolla, Miguel y Fernanda Largo. Gracias a aquellas personas que vivieron conmigo y se convirtieron en mi familia temporal; a mis compañeros del laboratorio 13, en especial a los doctores Clara y Aldo quienes siempre me apoyaron en decisiones complejas, a Hector por ser quien me ha guiado desde el inicio y finalmente gracias a los doctores más guapos del laboratorio 4, Miguel S. y Marel M. por mostrarme el lado divertido de la ciencia y por convertirse en más que mentores.

La cronología los coloca en distintas etapas de mi vida, algunos de ustedes han estado desde el inicio, otros han aparecido recientemente, pero sin duda cada uno de usted ha sembrado algo vital en mí, con cada uno de usted he tenido el privilegio de crear un lazo y compartir la filosofía de “vive al tope hasta morir”, los amo con todo el corazón.

¡Gracias totales!

# ÍNDICE

## 1. INTRODUCCIÓN

### 1.1 Cáncer

### 1.2 Epidemiología

### 1.3 Cáncer de mama

### 1.4 Síndrome de cáncer de mama y ovario hereditario

#### 1.4.1 Variantes patogénicas

### 1.5 Mecanismos de reparación

#### 1.5.1 Recombinación homóloga

#### 1.5.2 Mismatch repair

### 1.6 Firmas mutagénicas asociadas a cáncer de mama

## 2. JUSTIFICACIÓN

## 3. HIPÓTESIS

## 4. OBJETIVO

### 4.1 Objetivo general

### 4.2 Objetivo particular

## 5. MATERIALES Y MÉTODOS

## 6. RESULTADOS

## 7. DISCUSIÓN

## 8. CONCLUSIONES

## 9. REFERENCIAS

## 10. ANEXOS

## Lista de abreviaturas

CM	Cáncer de mama
<i>BRCA1</i>	Gen de susceptibilidad de cáncer de mama 1
<i>BRCA2</i>	Gen de susceptibilidad de cáncer de mama 2
BRCT	Dominio repetido carboxilo terminal
SCMOH	Síndrome de cáncer de mama y ovario hereditario
MR	Mecanismos de reparación
RH	Recombinación homóloga
IGV	Integrative Genomics Viewer
PoN	Panel de normales (por sus siglas en Inglés)
SNP	polimorfismos de nucleótido único (por sus siglas en Inglés)
BSQR	Base Quality Score Recalibration (por sus siglas en Inglés)
MQ	Mapping quality (por sus siglas en Inglés)
BQ	Base quality (por sus siglas en Inglés)
MMBQ	Media mínima de calidad de base
MRPS	Min read per strand (por sus siglas en Inglés)
fSB	f score beta (por sus siglas en Inglés)
UARC	Unique alt read count (por sus siglas en Inglés)
RE	Receptor de estrógeno
RP	Receptor de progesterona
HER2 Inglés)	Receptor 2 del factor de crecimiento epidérmico humano (por sus siglas en
TN	Triple negativo

CCHNP	Cáncer colorrectal hereditario no polipósico
AF	Anemia de Fanconi
MMR	Reparación por bases mal apareadas (por sus siglas en inglés)
SBS	Sustitución única de base (por sus siglas en inglés)
DBS	Sustitución doble de base (por sus siglas en inglés)
ID	Inserciones y deleciones (siglas en inglés)
LACAM hereditario	Consortio Latinoamericano para el estudio del cáncer de mama y ovario hereditario
HC	Hematopoyesis clonal
APOBEC polipéptido	Enzima editora de la apolipoproteína B mediante ARNm semejante al catalítico
Pol n	Polimerasa eta
NF1	Neurofibromatosis tipo I
APC	Adenomatous Polyposis Coli
WES	Secuenciación de exoma completo
WGS	Secuenciación de genoma completo
MLH1	Homólogo MutL 1
MSH2	Homólogo MutS 2
MSH6	Homólogo MutS 6
PMS2	Segregación posmeiótica 2
CDH1	E-cadherina
PALB2	Socio y localizador de BRCA2 " <i>Partner and localizer of BRCA2</i> "
BRIP1 interacting	Helicasa C-terminal de proteína que interactúa con BRCA1 " <i>BRCA1-protein C-terminal helicase 1</i> "
PTEN	Homólogo de fosfatasa y tensina " <i>Phosphatase and tensin homologue</i> "

*CDH3* P-cadherina 3

PARP poli ADP-ribosa polimerasa "*Poly(ADP-Ribose) Polym*

## INTRODUCCIÓN

### 1.1 Cáncer

El cáncer es un conjunto de enfermedades de origen multifactorial que se caracteriza por una proliferación exacerbada de células anormales con capacidad invasiva (OMS, 2022); estas células adquieren características que permiten el desarrollo tumoral, entre las que se encuentran la inestabilidad genómica y mutaciones, el mantenimiento de las señales proliferativas, la evasión de supresores de crecimiento, la evasión de la apoptosis, el potencial de replicación, la angiogénesis, la invasión y metástasis, así como la desregulación del metabolismo celular y la evasión de la respuesta inmunitaria. Recientemente, se han descrito cuatro características emergentes más, las cuales fungen favorecen el desarrollo de la patología: i) la reprogramación epigenética no mutacional, ii) la posibilitación de la plasticidad fenotípica, iii) la presencia de microbioma polimórficos y iv) las células senescentes (Hanahan D. & Weinberg, 2011; Hanahan, 2022). Esta caracterización permite tener un panorama global de esta patología a nivel molecular, destacando la sensibilidad de respuesta celular y su gran capacidad adaptativa, lo que produce un programa molecular entre la selección clonal de células cancerosas y la homeostasis del organismo.

Las alteraciones genéticas tumorales se desarrollan a través de diferentes procesos mutagénicos que actúan en conjunto; intrínsecos como el metabolismo celular, el estrés oxidativo, el microambiente, los defectos en la reparación y replicación del ADN, así como en la transcripción; o extrínsecos como el ambiente (la exposición a la luz ultravioleta, contaminación del aire, entre otros) y el estilo de vida (consumo de tabaco, consumo de alcohol, deficiencias alimenticias, entre otros) (Chatterjee N. & Walker, 2017). Estos cambios en el ADN son seleccionados, generando clonas con mutaciones conductoras, sucesivamente adquieren nuevas mutaciones que son seleccionadas dando origen a estirpes o subclonas celulares con mayores capacidades de sobrevivir y proliferar en ambientes desfavorables (Nowell P., 1976; Maley *et al.*, 2017). Uno de los mecanismos moleculares más importantes por su contribución a la evolución tumoral y el desarrollo de la heterogeneidad intratumoral es la inestabilidad genómica, la cual acelera la acumulación

de alteraciones genómicas (Sharma *et al.*, 2010; Hassanpour & Dehghani, 2017; Hanahan & Weinberg, 2011). La inestabilidad genómica es causada por daños en mecanismos de reparación del ADN, estrés durante la replicación y/o transcripción, defectos durante la segregación cromosómica mitótica y alteraciones en el mantenimiento de los telómeros.

Se conocen dos tipos de inestabilidad genómica; cromosómica y de microsatélite. El desarrollo de la inestabilidad genómica es progresivo, lo que permite la acumulación de alteraciones en genes encargados del mantenimiento de la integridad del ADN y que desempeñan un papel clave en ciertos procesos como la división celular, la proliferación, mecanismos de reparación, apoptosis y control del ciclo celular. Por consiguiente, una mutación en un gen conductor produce cambios dinámicos en el genoma; confiriendo ventajas adaptativas para proliferar y sobrevivir, tras la acumulación de dos a ocho mutaciones o una segunda mutación somática en un alelo con una variante germinal, permitiendo el desarrollo del fenotipo mutante y la progresión del cáncer (Negrini *et al.*, 2010; Vogelstein, 2013).

La desregulación de vías de señalización ocasionada por la inestabilidad genómica es una de las principales causas para desarrollar cáncer hereditario y una de las primeras marcas en adquirirse en el cáncer espontáneo, ya que permite una adaptación continua al medio (Negrini *et al.*, 2010; Vogelstein, 2013).

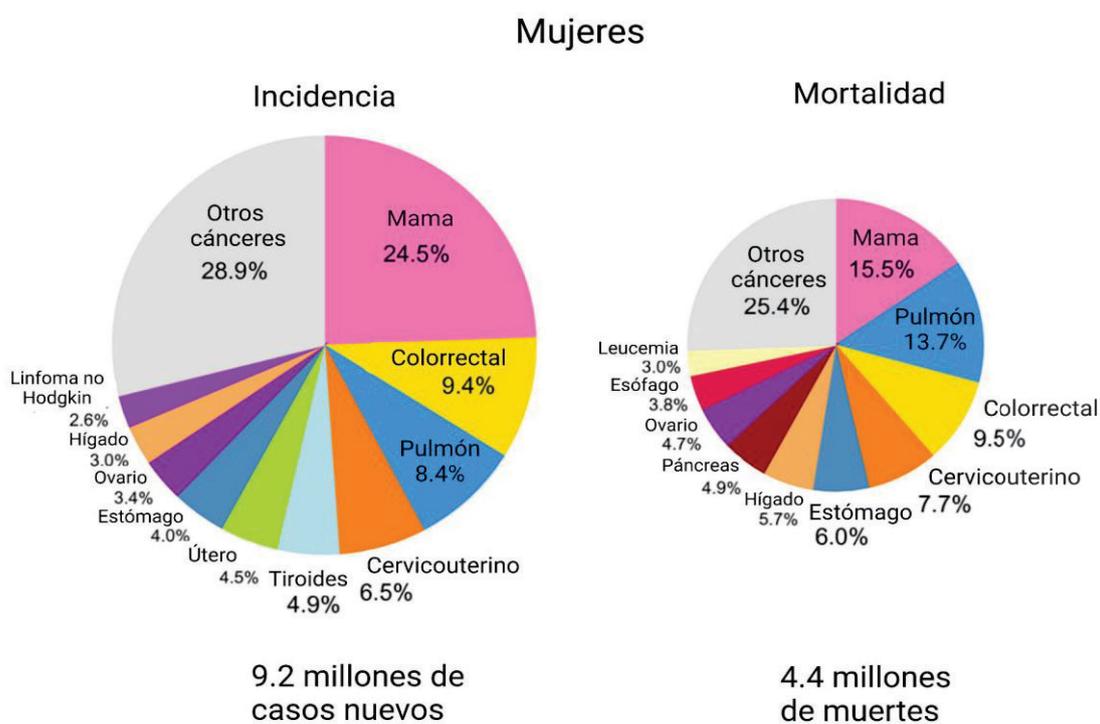
## **1.2 Epidemiología del cáncer**

El cáncer es una de las principales causas de muerte a nivel mundial (Bray F. *et al.* 2021). De acuerdo con la OMS, en 2020 se observó un incremento a nivel mundial en la tasa de incidencia y mortalidad de cáncer, con una estimación de más de 19.3 millones de casos nuevos y 10 millones de muertes, siendo el cáncer de mama (CM) la neoplasia maligna con la mayor tasa de incidencia y mortalidad en mujeres (Figura 1).

En América Latina, se estimó para ambos sexos una incidencia de 20.9% y una mortalidad de 14.2%, siendo el cáncer de mama el más incidente y el quinto en mortalidad,

no obstante, para mujeres fue la neoplasia con la mayor tasa de incidencia (24.5%) y mortalidad (15.5%) (Sung *et al.*, 2021; GCO,2020).

Este problema de salud pública se agrava en países en vías de desarrollo debido a que presentan un incremento en la tasa de incidencia del 88% y un aumento del 17% en la tasa de mortalidad. Lo anterior se ve reflejado en los reportes de América Latina ya que es la neoplasia maligna más incidente y con mayor número de muertes asociadas a cáncer (Sung *et al.*, 2021; GCO, 2020).



Tomado y modificado de GLOBOCAN 2020

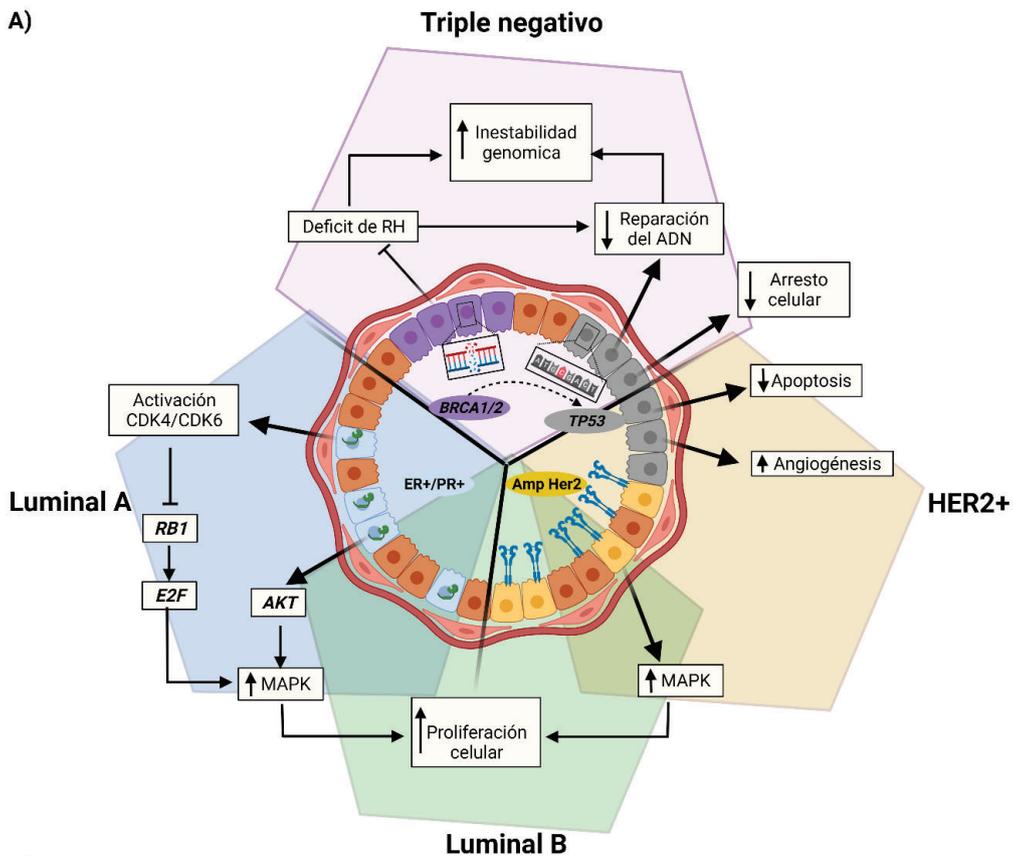
**Figura 1 Cáncer de mama, un reto para la salud pública.** En 2020 el cáncer de mama presentó la mayor incidencia y mortalidad en mujeres con un alza en países en vías de desarrollo.

### 1.3 Cáncer de mama

El CM es una neoplasia compleja y heterogénea a nivel molecular, histológico y clínico. Es causada por factores ambientales, genéticos y hormonales. Los principales factores de riesgo son la edad, contaminación, factores ambientales, etnicidad, menarquia temprana, menopausia tardía, nuliparidad, edad tardía de embarazo, periodo de lactancia, historial familiar; así como el estilo de vida, incluyendo el consumo de alcohol, la obesidad o el sedentarismo (McPherson, 2000; Harbeck *et al.*, 2019).

De acuerdo con sus características clínicas, el CM se clasifica en subtipos moleculares e histológicos, con base en los perfiles de expresión del receptor de estrógenos (RE), del receptor de progesterona (RP) y del receptor 2 del factor de crecimiento epidérmico humano (HER2). Existen en cuatro subtipos moleculares a nivel de expresión de ARNm, definidos por el sistema PAM50, los cuales presentan características propias del epitelio mamario involucrando dos principales vías de señalización, la primera implica genes asociados a la producción de estrógeno y progesterona, y la segunda a genes involucrados en proliferación y el ciclo celular (Figura 2) (Harbeck *et al.*, 2019; Perou *et al.*, 2000).

Dentro de los casos reportados de CM, el 95% de ellos son adenocarcinomas de lobulillos o conductos y pueden ser *in situ* o invasivos. Los carcinomas *in situ* proliferan de manera intraductal conteniéndose en el área del ducto o lobulillos, no obstante, poseen una tendencia a ser invasivos. Por el contrario, las células malignas de los carcinomas invasivos proliferan al estroma de la mama (Makki J., 2015). Ambas clasificaciones son clave para establecer un tratamiento óptimo y personalizado para pacientes en etapas tempranas o tardías. Un ejemplo es el uso de terapias blanco como los inhibidores de PARP, administrados en mujeres con SCMOH con variantes patogénicas en los genes *BRCA1* y *BRCA2* (*BRCA1/2*) (Daly *et al.*, 2021).



**B)**

	Triple negativo	HER2+	Luminal B	Luminal A
		No triple negativo		
Expresión de receptores	-RE, -RP, -HER2	HER2+	RE+/RP+	
% de cáncer de mama esporádico	15 - 20 %	10 - 15 %	25 - 30%	40 - 45%
Grado histológico	Alto grado		Bajo grado	
Genes alterados	<i>TP53</i> 75%	84%	32%	12%
	<i>PIK3CA</i> 7%	7%	32%	49%
	<i>PTEN</i> 8%		<i>MAPK3K1</i> 5%	<i>GATA3</i> 14%
	<i>BRCA1/2</i> 30/6%			
Terapia dirigida disponible	Sin terapia dirigida disponible	Anti-HER2 (trastuzumab)		
Pronóstico	Pobre		Bueno	
Sobrevida a los 5 años	30%	35 - 50%	50 - 70%	95%

**Figura 2 Clasificación del cáncer de mama e implicaciones de relevancia clínica.** A) Vías de señalización y genes participantes en la desregulación celular de acuerdo a los cuatro subtipos clínicos y moleculares. B) Descripción clínica y epidemiológica de los subtipos moleculares con base en la expresión de RE, RP y HER2 (basado en Perou *et al.*, 2020).

El CM al igual que otras neoplasias tiene dos orígenes: a) esporádico, el cual representa el 90-95% de los casos, se desarrolla por factores ambientales y hormonales; y b) hereditario, con una incidencia del 5-10%, relacionado con la historia familiar, etnicidad y variantes patogénicas en genes de alta penetrancia como *BRCA1/2*, entre otros (Mahdavi *et al.* 2018; Harbeck *et al.*, 2019).

#### **1.4 Síndrome de mama y ovario hereditario**

El nivel de riesgo del factor hereditario es el mayor contribuyente para desarrollar CM y destaca en pacientes con familiares de primero y segundo grado que presentan la patología. En estas familias se presenta el síndrome de cáncer de mama y ovario hereditario (SCMOH). Esta patología conlleva un aumento en el riesgo para desarrollar CM primario, mama contralateral, mama bilateral, ovárico, próstata, páncreas y melanoma (cutáneo y ocular); siendo mama y ovario los órganos blanco con mayor riesgo con ~80% y ~40%, respectivamente (van der Groep *et al.* 2006; Kobayashi *et al.*, 2013; Petrucelli *et al.*, 2022). Los individuos que presentan variantes patogénicas en genes asociados al SCMOH exhiben en la mayoría de los casos la pérdida de heterocigosidad expresando solo uno de los loci, esta pérdida de heterocigosidad es causada por la presencia de una variante germinal con una segunda mutación adquirida de manera somática, manteniendo una actividad alterada del gen afectado o bien una completa pérdida de la función (Knudson, 1971; Venkitaraman, 2001).

El SCMOH presenta características clínicas que lo catalogan como una patología compleja y agresiva puesto que i) se presenta a una edad temprana (< 45 años), ii) tiene tendencia a desarrollar tumores primarios sincrónicos con riesgo de metástasis a hígado, pulmón y próstata, y iii) alta probabilidad (80-90%) de ser subtipo triple negativo (TN), el más agresivo de todos (Daly *et al.*, 2021). Otra de las características distintivas es que a pesar de ser una enfermedad autosómica dominante no todos los portadores de variantes patogénicas desarrollan la enfermedad, lo cual se conoce como penetrancia incompleta. Además, se presenta una heterogeneidad de locus (múltiples genes afectados que generan el mismo fenotipo) e inestabilidad genómica que promueve el desarrollo, evolución y complejidad de la patología (Nielsen *et al.*, 2016; Duijf *et al.*, 2019).

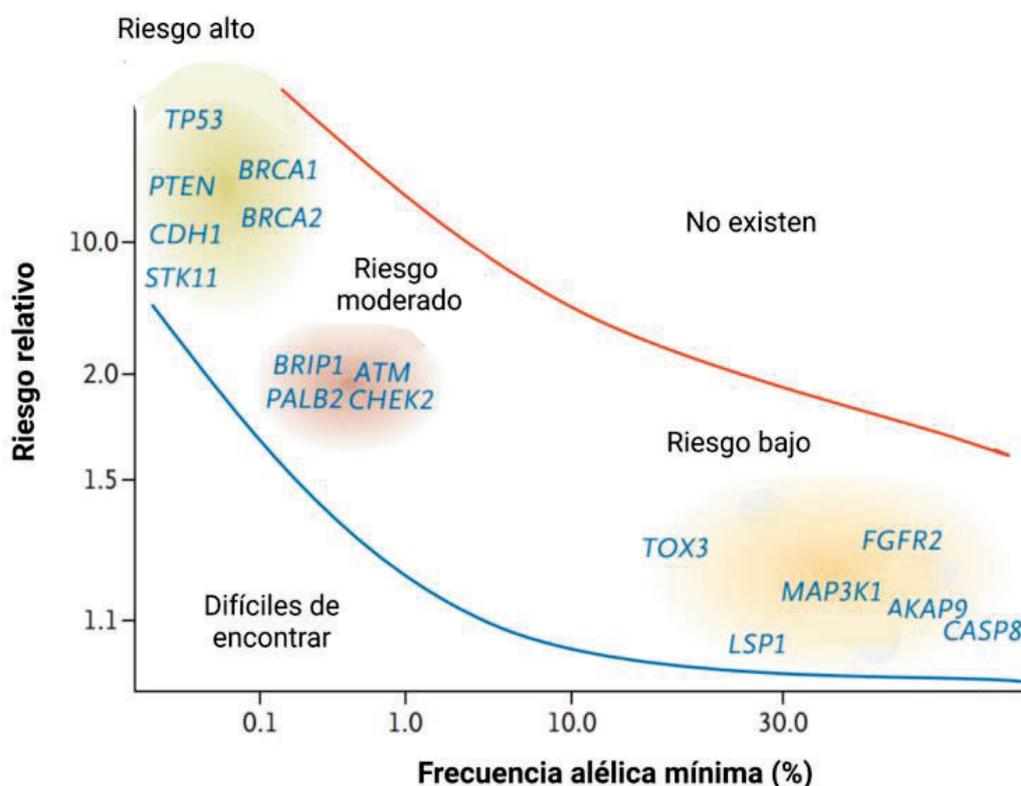
### 1.4.1 Variantes patogénicas asociadas al SCMOH

Las principales variantes patogénicas de alta penetrancia asociadas al SCMOH ocurren en *BRCA1* (66% de los casos) y *BRCA2* (34% de los casos) (Daly *et al.*, 2021). Sin embargo, se han reportado otros genes de relevancia clínica que incrementan el riesgo para el desarrollo del SHCMO como *CHEK2*, *ATM*, *PALB2*, *RAD51D*, *STK11*, *PTEN*, *CDH1* y *NF1* (Figura 3), muestra de la alta heterogeneidad a nivel de locus de esta enfermedad. Además, existe un riesgo aumentado cuando los pacientes presentan historial familiar. Los genes de alta penetrancia para CM son *BRCA1/2*, *PALB2*, *TP53*, *PTEN*, *STK11* y *CDH1*, y riesgo moderado en *ATM*, *CHEK2*, *PALB2*, *BRIP1* y *RAD51D* (Foulkes, 2008; Nielsen *et al.*, 2016; Hu C. *et al.*, 2021). Estas variantes producen proteínas no funcionales o con actividad deficiente a causa de mutaciones puntuales o inserciones y deleciones (indels) que generan la pérdida o ganancia de un codón de paro o bien, una sustitución de aminoácido, lo cual puede afectar a dominios clave como el dominio repetido carboxilo terminal (BCRT) en *BRCA1/2* o en estructuras terciarias de la proteína como los dedos de zinc característicos de *BRCA1*; lo que altera de manera global a las vías de señalización en la que se encuentren implicadas (Narod S. & Foulkes, 2004; Leung & Glover, 2011).

Además de la caracterización de variantes de riesgo para SCMOH, se han descrito otros genes de mediana o baja penetrancia cuyas variantes patogénicas podrían incrementar el riesgo de desarrollo y progresión de esta patología, incluyendo *MLH1*, *MSH2*, *PMS2* y *MSH6*, involucrados en el mecanismo de reparación por bases mal apareadas y reportados como de alta penetrancia en el síndrome de Lynch. Los genes *FANCD2*, *FANCA* y *FANCC*, asociados a la reparación de entrecruzamiento de cadenas y a la patología Anemia de Fanconi también presentan variantes patogénicas en pacientes con SCMOH, al igual que otros genes supresores de tumor como *SKT11* y *PTEN*, cuya inactivación favorece la acumulación de rupturas de doble cadena del ADN (Kobayashi *et al.*, 2013; Roberts *et al.*, 2018).

Otro factor de gran importancia es la correlación entre el subtipo molecular y una alta incidencia de mutaciones en algunos genes como el caso del CM ER- el cual presenta daños en *ATM*, *CDH1* y *CHEK2*; y comparte variantes patogénicas en *BARD1*, *RAD51C* y *RAD51D*

con el CM de subtipo TN. También se ha descrito que la amplificación de HER2 desregula proteínas encargadas de respuesta a daño de ADN como BRCA1, RPA1 y H2AX, que junto a mutaciones germinales en *BRCA1* promueve el desarrollo de CM subtipo basal, caracterizado por ser TN en 80-90% de los casos y con la expresión de citoquinas 5 y 14, vimetina, EGRF y CHD3. Por ello, los estudios genómicos sobre el perfil de expresión del CM han permitido caracterizar la actividad de cascadas de señalización que promueven a la iniciación y progresión tumoral del tumor (Duijf *et al.*, 2019, Dorling *et al.*, 2021; Hu C. *et al.*, 2021).



Tomado y modificado de Foulkes, 2008

**Figura 3 Arquitectura molecular del cáncer de mama.** Asociación entre la frecuencia alélica poblacional y el nivel de riesgo relativo de alteraciones en genes de baja, mediana y alta penetrancia susceptibilidad para desarrollar cáncer de mama.

## 1.5 Mecanismos de reparación

De manera fisiológica, existe una tasa de error  $10^{-3}$  a  $>10^{-6}$  en la incorporación de nucleótidos durante la replicación del ADN celular. Estos valores están en función del tipo de polimerasa y la capacidad de corregir los errores que se generan durante la síntesis del ADN, con la finalidad de mantener la integridad del genoma. Además de la actividad de exonucleasa 3'→5' que tienen ciertas polimerasas como primer respuesta, existen mecanismos adicionales de reparación del ADN que se encuentran activos de manera constitutiva en todas las células y se inician en diferentes fases del ciclo celular. Estos procesos aseguran la integridad y fidelidad del material genético entre los descendientes, y alcanzan una tasa de error inferior a  $10^{-9}$  (McCulloch & Kunkel, 2008; Chatterjee N. & Walker, 2017).

Además de las mutaciones asociadas a los errores durante la procesividad existen agentes genotóxicos de origen extrínseco e intrínseco que pueden dañar la estructura primaria del ADN causando rupturas de doble cadena, rupturas de cadena sencilla, generación de aductos, mal apareamiento de las bases, entrecruzamiento de cadenas, gaps de cadena única, desaminación de bases, entre otros; estos daños son censados y activan la respuesta a daño al ADN. Se han descrito cinco mecanismos para la reparación del ADN: reparación por escisión de bases (BER), reparación por escisión de nucleótidos (NER), reparación por bases mal apareadas (MMR), recombinación homóloga (RH) y unión por extremos no homólogos (NHEJ), que actúan en distintas fases del ciclo celular disminuyendo la tasa de mutagénesis. Cuando no hay una reparación del daño y este persiste en las células, estas se van a apoptosis; sin embargo, cuando los mecanismos fallan se genera una acumulación de variantes patogénicas que desencadenan en diferentes trastornos, como el cáncer y el envejecimiento (Chatterjee & Walker, 2017; Wolters S. & Schumacher B., 2013; Alexandrov *et al.*, 2013). En el caso del cáncer, las variantes patogénicas acumuladas son variables de acuerdo al tipo celular y los agentes causales (Alexandrov *et al.*, 2013).

### **1.5.1 Recombinación homóloga**

Uno de los mecanismos de reparación del ADN más importantes para el SCMOH es RH, debido a que alteraciones en sus componentes (*BRC1/2* y *PALB2*) se asocian a una mayor predisposición a la enfermedad. Este mecanismo se encarga de la reparación de las rupturas de doble cadena en el ADN. Estos daños son reparados durante la fase S y G2 del ciclo celular, con un aumento niveles en los niveles de expresión de las proteínas de este mecanismo durante la fase S (Venkitaraman, 2001; Chatterjee N. & Walker, 2017; Nielsen *et al.*, 2016)

Esta respuesta a la ruptura de doble cadena ocurre de manera espacio-estructural y requiere alertas de los puntos de control del ciclo celular que por consiguiente genera el reclutamiento secuencial y coordinado de los factores de la maquinaria de reparación (Figura 4). Esta respuesta consta de subvías que le permiten utilizar como molde la cadena de la eucromatina hermana, sintetizando con alta fidelidad.

Como se describe en la Figura 4; se puede comprender la importancia de los factores que componen el andamiaje encargado de la reparación, ya que además de reparar vía RH también se ven involucrados en reparar entrecruzamientos de hebras en coordinación con los genes de AF; mantener la estabilidad de la horquilla de replicación que por ende causará un impacto en la transcripción de ciertos genes. Cuando no se genera una correcta reparación de la ruptura de doble cadena o entrecruzamiento de cadenas, se genera una mayor susceptibilidad a agentes genotóxicos e inestabilidad cromosómica afectando la segregación durante la metafase (Venkitaraman, 2001; Chatterjee N. & Walker, 2017; Kobayashi *et al.*, 2013; Nielsen *et al.*, 2016).

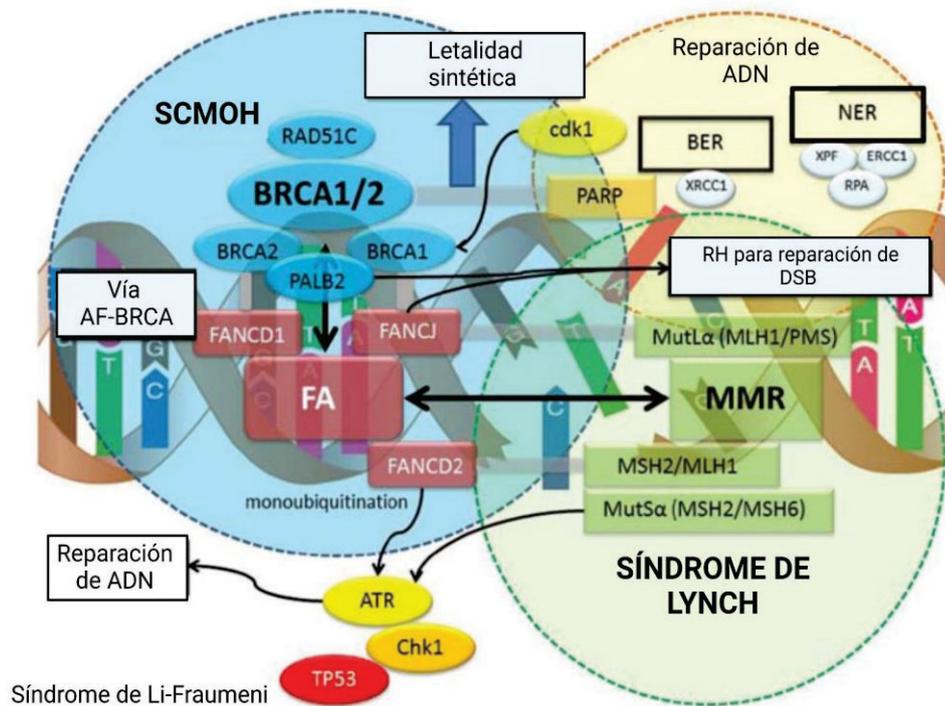
### **1.5.2 Reparación por bases mal apareadas**

La vía de MMR es altamente conservada en eucariontes y procariontes; su importancia radica en la eliminación de bases mal apareadas, inserciones y deleciones; generadas a partir de una mala incorporación de nucleótidos por la ADN polimerasa (Malik *et al.*, 2018;

Modrich & Lahue., 1996). Esta respuesta a daño actúa durante la fase S del ciclo celular y se caracteriza por la generación de heterodímeros para su correcto funcionamiento; se divide en cuatro fases: i) reconocimiento de la lesión, ii) inicio de la reparación, iii) eliminación de la lesión y iv) síntesis *de novo* del ADN como se muestra en la Figura 4 (Chatterjee N. & Walker, 2017; Pećina-Šlau *et al.*, 2020).

El fallo en esta vía de reparación genera sustitución de bases, deleciones, inserciones, incremento en la tasa de mutaciones espontáneas y alteraciones en la longitud de los microsatélites lo que genera una de inestabilidad microsatélite; característica clave en la etiología del cáncer. Aunado a esto, se describe una correlación entre secuencias repetitivas e interacciones con sitios remodeladores de histonas y cromatina lo que genera un impacto a nivel de expresión (Kobayashi *et al.*, 2013; Pećina-Šlau *et al.*, 2020).

Un ejemplo de susceptibilidad a cáncer es el SCMOH y cáncer colorrectal hereditario no polipósico, también conocido como síndrome de Lynch; el cual presenta variantes patogénicas germinales de alta penetrancia en *MLH1*, *MSH2*, *MSH6* y *PMS2* (Nielsen *et al.*, 2016; Chatterjee N. & Walker, 2017). En el caso de SCMOH, estos genes son de baja o penetrancia aún desconocida, sin embargo, la pérdida de estas proteínas promueve la quimioresistencia (Malik *et al.*, 2018).



Tomado de Kobayashi *et al.*, 2013

**Figura 4. Interrelación de mecanismos de reparación de ADN.** Convergencia de factores correspondientes a cuatro mecanismos de respuesta a daño de ADN y patologías asociadas a la falla de estos.

### 1.6 Firmas mutagénicas

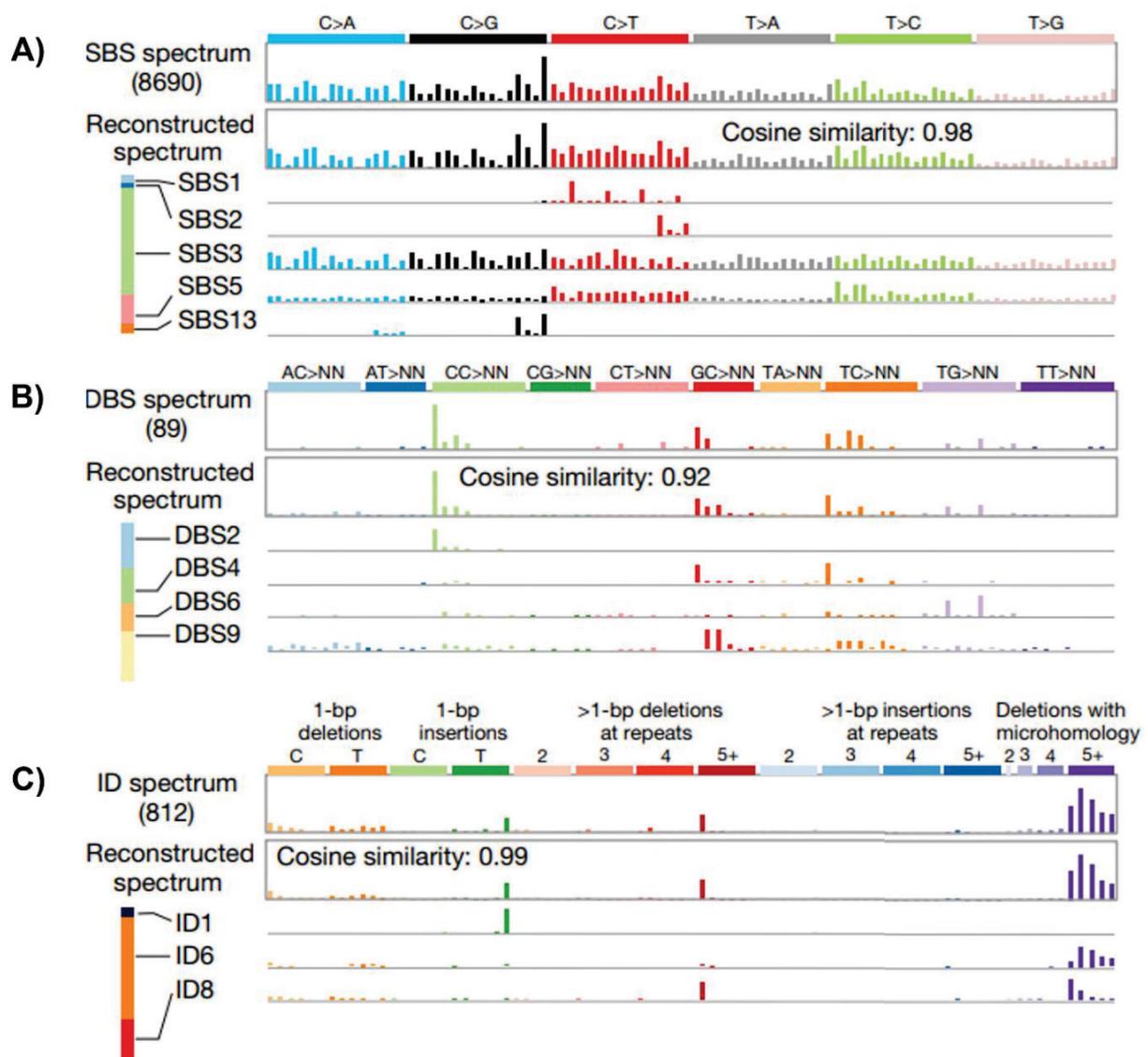
Existen múltiples factores que producen mutaciones a lo largo de la vida de un individuo, Los procesos mutacionales causados por factores endógenos como la edad, deficiencias en los MR del ADN, así como factores exógenos como el consumo de tabaco, exposición a luz UV, entre otros, dejan patrones específicos de mutación en el ADN denominados firmas mutacionales (Polak *et al.*, 2017; Nik-Zainal *et al.*, 2012; Alexandrov *et al.*, 2013).

Las firmas mutacionales se clasifican en cuatro grandes grupos; i) sustitución única de nucleótido con un contexto de 96 SBS, considerando seis tipos de posibles sustituciones; ii) sustitución doble de nucleótido con un total de 78 tipos de DBS; iii) variaciones de número

de copias con un esquema de 48 clases; iv) e inserciones y deleciones pequeñas con un total de 83 clases (Alexandrov *et al.*, 2013; Alexandrov *et al.*, 2020; COSMIC 2020).

Como se muestra en la Figura 5, el CM presenta firmas en el contexto de SBS, DBS e ID. Estos cambios distintivos son ocasionados por diferentes procesos mutagénicos como desaminación de 5 metilcitosina (SBS 1) deficiencias en MMR (SBS 6, 20 y 26), actividad de APOBEC (SBS 2 y 13), deficiencia en *BRCA1/2* (SBS 3) y como consecuencia de la firma 3 se genera un incremento en la incidencia de la firma 8. (Nik-Zainal & Morganella, 2017; Alexandrov *et al.*, 2020; COSMIC, 2020).

La presencia de distintas firmas mutacionales en CM es muestra de su etiología heterogénea, definida por la combinación de distintos procesos mutagénicos; a diferencia del cáncer de piel o pulmón, los cuales se caracterizan principalmente por la firma 7 (asociadas a la luz ultravioleta) y la firma 4 (fumar tabaco), respectivamente (Nik-Zainal *et al.*, 2016; Gerstung *et al.*, 2017).



**Figura 5. Firmas mutagénicas características de CM.** Descripción e identificación de firmas mutagénicas en tres contextos diferentes. A) Cambios de nucleótido único con presencia de firma mutagénica S1, S2, S3, S5 y S13. B) Contexto de doble nucleótido con las firmas S2, S4, S6 y S9 C) e indels con las firmas S1, S6 y S8 presentes en tejido tumoral mamario.

La importancia de la identificación de patrones moleculares radica en la mejora del diagnóstico, tratamiento y pronóstico del paciente, por lo que las mutaciones somáticas postcigóticas como las sustituciones de un solo nucleótido o indels presentes a lo largo del desarrollo son de gran relevancia, ya que participan en la carcinogénesis, como causa

directa o coadyuvante (Maley *et al* 2017). Actualmente falta describir y desarrollar sistemas para establecer distinciones clínicas relevantes entre cómo evolucionan los diferentes tipos de tumores o como afectan de manera global al individuo portador de variantes patogénicas de origen germinal. Sin embargo, la identificación de dichos cambios somáticos en el organismo, tanto en tejido normal y tumoral, representa un reto debido a la complejidad para determinar la diferencia entre firmas que se empalman durante el desarrollo de la neoplasia. Aunado a esto, la baja fracción alélica de mutaciones causadas por procesos mutacionales menos activos presenta un reto para la clasificación y detección de nuevos patrones mutacionales (Martincorena *et al.*, 2015; Alexandrov *et al.*, 2020).

No obstante se han realizado grandes esfuerzos con el fin de detectar la contribución de factores mutagénicos sobre tejidos tumorales y normales; como el caso de la contribución de clonas a partir de la exposición solar en piel sana; el análisis global de exomas de distintos tipos de cáncer a partir de bases de datos para discernir la carga mutacional a partir de variables germinales; o la plétora de mutaciones somáticas en tejido sano asociadas a la mutación germinal de *MUTYH* y su contribución a las firmas descritas por Alexandrov (Wei *et al.*, 2021; Vali-Pour *et al.*, 2022; Robinson *et al.*, 2022; Alexandrov *et al.*, 2020). Una perspectiva inexplorada del panorama molecular del CM es analizar el impacto que generan las variantes patogénicas germinales asociadas a MR en tejido sanguíneo sano. Debido a que su tasa proliferativa elevada ( $2 \times 10^{11}$  eritrocitos y plaquetas; y  $7 \times 10^{10}$  granulocitos), los procesos mutacionales en pacientes con variantes patogénicas germinales podrían producir la acumulación de mutaciones con características distintivas útiles para establecer un patrón y en consecuencia tener una mejor estrategia de detección y/o tratamiento (Slavin *et al.* 2017; Bellantuono I., 2004). Sin embargo, actualmente los enfoques que se tienen en cuanto al análisis de mutaciones en el tejido sanguíneo lo han contemplado la identificación de la fracción alélica mutacional cuando existe hematopoyesis clonal en personas sanas en comparación con pacientes que presentan patologías hemáticas; así como la evaluación del impacto de las mutaciones adquiridas en linfocitos T en leucemia aplásica (Genovese *et al.*, 2014; Lundgren *et al.*, 2021).

Por tales motivos, el siguiente trabajo propone evaluar e identificar el patrón mutacional asociado a defectos en mecanismos de reparación de ADN de origen germinal, presentes en el tejido sanguíneo de pacientes con SCMOH, con el fin de identificar un patrón mutacional característico de los portadores de variantes patogénicas hereditarias que permita mejorar la detección temprana y aumente nuestra comprensión de los eventos moleculares que influyen en la progresión del SCMOH.

## JUSTIFICACIÓN

El síndrome de cáncer de mama y ovario hereditario es una patología compleja con múltiples agravantes; se caracteriza por tener heterogeneidad de locus y variantes patogénicas implicadas en respuesta a daño de ADN como *BRCA1/2*, responsables del mecanismo de recombinación homóloga o en genes encargados de reparar bases mal apareadas. Por ende, la falta de eficacia en los mecanismos de reparación puede ser ocasionada por la falla de uno o varios componentes del complejo, lo que compromete la integridad del ADN modificando la tasa de mutaciones espontáneas durante la replicación celular. Dichos defectos en los mecanismos de reparación del ADN a lo largo de décadas podrían generar la acumulación de mutaciones con una fracción alélica muy baja en las células sanguíneas de las personas portadoras que desarrollan SCMOH. En estas pacientes, los genes involucrados en el mecanismo por recombinación homóloga *BRCA1/2* presentan las variantes patogénicas hereditarias con mayor incidencia y penetrancia. Los genes asociados MMR presentan una baja incidencia, pero generan una alta tasa de mutaciones somáticas. A la fecha, no existe una caracterización del impacto mutacional que tienen las variantes de origen germinal en la tasa de mutaciones somáticas asociadas a la falla de RH y MMR, lo cual permitiría tener un mayor entendimiento de los mecanismos mutacionales de esta patología heterogénea.

Por tal motivo, el siguiente trabajo se enfoca en describir los patrones de mutaciones en pacientes de Latinoamérica con SCMOH con deficiencias en los mecanismos de reparación del ADN de RH, MMR.

## **OBJETIVO GENERAL**

Identificar patrones de mutación causados por alelos patogénicos germinales en genes de reparación del ADN en pacientes con el síndrome de cáncer de mama y ovario hereditario.

## **OBJETIVOS PARTICULARES:**

- 1) Generar una base de datos de pacientes con SCMOH secuenciadas con variantes patogénicas en genes involucrados en los mecanismos de reparación del ADN.
- 2) Evaluar la tasa de mutaciones somáticas en un panel de 143 genes (618,732 pb) en las pacientes con deficiencias en los mecanismos de reparación del ADN.
- 3) Caracterizar los cambios de nucleótido único y el tipo de mutaciones en las muestras con deficiencias en los mecanismos de reparación del ADN.
- 4) Identificar las firmas mutagénicas generadas por las variantes patogénicas en genes que participan en los mecanismos de reparación del ADN.

## **HIPÓTESIS**

Las mujeres con SCMOH que presentan variantes patogénicas en genes que participan en los mecanismos de reparación del ADN por recombinación homóloga y bases mal apareadas tendrán una mayor tasa de mutaciones somáticas en células sanguíneas que controles sin mutaciones en estos genes.

## **PREGUNTA DE INVESTIGACIÓN**

¿Las variantes patogénicas germinales en genes de reparación de ADN favorecen la acumulación de mutaciones somáticas en las células sanguíneas?

## MATERIALES Y MÉTODOS

### Población de estudio

Se utilizaron 186 archivos fastq (R1 y R2) de secuenciación masiva de nueva generación (Illumina DNA-seq); los cuales se obtuvieron a partir del análisis de un panel de genes asociados a cáncer hereditario en muestras de gDNA extraído de sangre periférica de individuos con SCMOH e individuos normales. Estos datos derivan del proyecto realizado por el Consorcio Latinoamericano para el estudio del cáncer de mama y ovario hereditario (LACAM) (Oliver *et al.*, 2019) y del estudio realizado por Quezada y colaboradores en 2018.

Para este trabajo se incluyeron aquellas pacientes que cumplieron los siguientes criterios de selección: i) muestras secuenciadas con Hiseq 4000, ii) profundidad mínima de 500x, iii) mujeres con SCMOH con variantes patogénicas en genes involucrados en RH y MMR, iv) muestras con variantes con significado clínico incierto, y v) muestras de individuos control sin predisposición a SCMOH. Los datos clínicos se obtuvieron de los cuestionarios sobre hábitos y estilo de vida; a través de la base de datos del Consorcio LACAM en la plataforma RedCap (Quezada *et al.*, 2018; Oliver *et al.*, 2019).

Los datos analizados se generaron a partir de un panel de genes asociados a susceptibilidad a cáncer hereditario en distintos órganos blanco (Figura suplementaria 1). El primer panel utilizado fue NGHS-013X, "Human Cancer Predisposition Panel" generado por la empresa QIAGEN, el cual cuenta con una tecnología de enriquecimiento dirigido basada en PCR multiplex y un algoritmo de primers para maximizar la cobertura de 143 genes y minimizar la amplificación de regiones no específicas. De este panel se usaron un total de 93 muestras de las cuales 28 contenían variantes patogénicas en *BRCA1* o *BRCA2*, 5 en genes de MMR, 30 muestras de individuos con SCMOH sin variantes patogénicas en *BRCA1/2* (Figura 7), y 30 muestras de individuos control sanos sin SCMOH (panel de normales) pareados por quinquenio de edad y código postal con población de estudio.

El otro panel usado fue "Custom Hereditary Cancer Solution for the LACAM Consortium" generado por el Consorcio LACAM en la plataforma Sophia Genetics; el cual posee una

tecnología de enriquecimiento por captura de híbridos detectando la región exónica completa de 74 genes. Se analizaron un total de 20 muestras; 9 de ellas contienen variantes patogénicas en *BRCA1* o *BRCA2*, 2 en genes de MMR y 9 muestras de individuos con SCMOH sin variantes patogénicas en *BRCA1/2*.

### **Análisis bioinformático**

Los archivos fastq de secuenciación masiva para Qiagen y Sophia genetics fueron procesados en sus respectivas plataformas usando hg19 y human\_g1k\_v37 como referencia del genoma humano, respectivamente, para la generación de archivos de lectura bam.

Con base en los archivos fastq evaluados con el panel de 74 genes de Sophia genetics se generó un alineamiento local de las secuencias con referencia al genoma hg19 a partir del BWA, usando el algoritmo bwa-mem; tras el alineamiento se ordenó por coordenadas y se convirtió a un archivo binario (".bam"), al cual se le marcó y organizó las regiones duplicadas y se le generó un índice (".bai"); después de ello se generó una tabla de recalibración para el BSQR y finalmente se generó un archivo binario con las correcciones del error sistemático. Mediante el uso del algoritmo samtools se evaluó la profundidad y el número de lecturas de manera aleatoria.

### **Llamados de variantes germinales**

Se utilizaron los programas germinales de QIAGEN, Sophia genetics y Haplotypecaller para comparar el llamado de variantes a partir de algoritmos de detección de variantes patogénicas de origen germinal con el fin de seleccionar aquellas variantes con un porcentaje de mutación por región <20% en comparación con la referencia de genoma hg19 (fracciones alélicas <20%).

Tras el llamado inicial de cada uno de los programas se analizó el archivo vcf y se generó un anotado con ANNOVAR (wANNOVAR, s. f.) de las regiones exónicas; se filtraron los datos, incluyendo aquellos que presentaban fracción alélica <20%; profundidad mayor a 500x; frecuencias alélicas menor a 0.01 en las bases de datos 1000genomes, ESP6500si y

gnomAD; y variantes heterocigotas variantes de alta calidad de acuerdo con los criterios propios de cada algoritmo.

### **Panel basado en amplicones (QIAGEN)**

Se evaluaron 6 muestras pertenecientes al panel de 143 genes, tres de ellas con detección de variantes patogénicas en *BRCA1* y las otras 3 mujeres con SCMOH sin variantes patogénicas en genes relacionados a mecanismos de RH. Las muestras se interpretaron de manera directa con las especificaciones propias establecidas por QIAGEN para una detección de variantes con significado clínico previamente reportado en bases de datos.

### **Panel basado en captura de híbridos (SOPHIA Genetics)**

Se valoraron 18 muestras alineadas con el algoritmo bowtie y el genoma human\_g1k\_v37 del panel de 74 genes, nueve de ellas provenientes de individuos con variantes probablemente patogénicas. y nueve con variantes patogénicas en genes involucrados en RH (*BRCA1/2*). Estas muestras se interpretaron de manera directa en el software DDM Sophia genetics generando un archivo vcf por muestra. Para el filtrado del anotado se seleccionaron muestras arriba de 1000x.

### **Llamado de variantes germinales por reensamblaje de haplotipos (Haplotypecaller)**

Se evaluaron 18 muestras descritas anteriormente, correspondientes al panel de 74 genes alineadas con human\_g1k\_v37 y modificadas con la adición del encabezado. Las muestras se evaluaron con Haplotypecaller para generando un llamado de variantes germinales como se describe en GATK (Poplin *et al.*, 2017).

### **Algoritmos específicos para llamado de variantes somáticas**

Se utilizaron los programas Platypus, VarDict, Somaticseq, Needlestack y Muetct2 para comparar la detección de variantes somáticas en modalidades que no requieren un llamado pareado (tejido normal y tejido tumoral).

Todos los programas se evaluaron con muestras pertenecientes al panel de 143 genes de QIAGEN a excepción de Needlestack el cual utilizó 18 muestras del panel de 74 genes. Posterior al llamado de cada programa se hizo un anotado en ANNOVAR con criterios de selección específicos, como se describe en el flujo de trabajo final (Figura 6B "Filtrado").

### **Ensamblaje de novo de haplotipos para identificación de variantes (Platypus)**

Para evaluar el algoritmo especializado en datos de secuenciación con enfoque clínico (Platypus) se usaron las seis muestras provenientes de 143 genes de QIAGEN; a partir de los archivos vcf se realizó un anotado, el cual se filtró incluyendo todos los datos que pasaron los filtros de calidad, muestras de individuos heterocigotos con frecuencia alélica  $< 0.01$  reportadas en las tres bases de datos anteriormente descritas.

### **Identificación de variantes con aplicación en investigación del cáncer (VarDict)**

Utilizamos las 6 muestras pertenecientes al panel de 143 genes de QIAGEN, descritas anteriormente. El programa se ejecutó de acuerdo a las recomendaciones descritas en GitHub. El archivo vcf fue anotado con wANNOVAR y filtrado para eliminar las variantes con artefactos por secuenciación, incluyendo las muestras de heterocigotos con frecuencia alélica menor a 0.01 (Lai, 2014).

### **Clasificador de predicción de mutaciones a través de cinco *callers* (Somaticseq)**

El llamador somático fue probado con tres muestras provenientes del panel de 143 genes de QIAGEN. Este integrador de programas fue ejecutado con la paquetería de anaconda 3 y para la detección de variantes se utilizó Mutect, VarDict, Lofreq, Scalpel y Strelka. Para el ensamblaje e integración de los archivos vcf se le adiciona un archivo bed con las regiones presentes en el panel de genes para restringir la cantidad de regiones a las cuales se genera el llamado (Fang *et al.*, 2015).

### **Identificación ultra-sensible de variantes somáticas (Needlestack)**

El llamador de variantes somáticas y germinales fue ejecutado con las 18 muestras anteriormente descritas del panel de 74 genes las cuales fueron alineadas con el genoma hg19 y el algoritmo BWA (Delhomme et al., 2020). Las muestras procesadas se conjuntaron en un solo archivo el cual fue anotado y curado incluyendo las muestras que pasaron los filtros de calidad establecidos por el programa, variantes con frecuencia alélica  $<0.001$ , variantes sin strand bias, variantes heterocigotas y se excluyeron aquellas variantes con fracción alélicas mayor al 20%.

Se evaluó por separado la detección de variantes tras modificar el valor preestablecido por Needlestack de BQ (calidad de base) de 13 a 30 y MQ (calidad de mapeo) de 0 a 30.

### **Modelo bayesiano para identificación de variantes somáticas (Mutect2)**

Se utilizó el programa de llamado de variantes somático Mutect2, modalidad (*tumor-only*) con el genoma de referencia hg19 para las muestras provenientes de los dos paneles de genes (Quezada et al., 2018; Oliver et al., 2019). Este programa se evaluó con diversas modificaciones con el fin de establecer una mayor detección con un nivel óptimo de precisión.

Para la interpretación biológica de los datos se generaron anotados con ANNOVAR y se obtuvo el archivo delimitado por comas con los resultados de las regiones exónicas.

Para definir las variantes verdaderas de alta confianza se aplicaron los siguientes criterios de selección: 1) variantes que pasaron los filtros de calidad (PASS), 2) variantes con frecuencia alélica poblacional  $<0.01$  en las bases de datos 1000 genome, gnomAD y ESP6500, 3) variantes heterocigotas; para eliminar el sesgo de error y obtener una fracción alélica derivada de mutaciones somáticas.

## **Llamado basal**

Se generó el flujo de trabajo como se describe en las prácticas de GATK; sin modificación alguna de los parámetros establecidos, usando la versión 4.1.81. de GATK (Cibulskis *et al.*, 2013, Broad Institute, 2019, Broad Institute, 2021).

## **Delimitación de región de llamado de variantes somáticas**

Se generó un archivo de texto BED con las regiones genómicas que componen al panel de genes de 143 genes usando como referencia el genoma hg19.

Al llamado de variantes con Mutect 2 “*tumor-only*” se le adicionó el archivo BED para reducir el tiempo de ejecución e identificar las regiones blanco, evitando falsos positivos en regiones inespecíficas.

## **Optimización de parámetros predeterminados**

Las seis muestras del panel de 143 genes se llamaron con el archivo BED y se analizó por separado la detección de variantes tras modificar el valor preestablecido por Mutect2 de la MMBQ 20 (Media mínima de calidad de base) a 25 y 30; MRPS 0 (mínimo de lecturas por cadena) a 1, 2 y 3; FDR 0.05 (tasa máxima de falsos positivos) a 0.01; fSB 1 (f-score beta) de 1 a 0.9; y UARQ 0 (límite inferior en el número de lecturas únicas del alelo alternativo) a 1; como control se utilizó MMMQ 30 (media mínima de calidad de mapeo) .

Tras las modificaciones individuales se generó un llamado con el conjunto de modificaciones; se evaluando los cambios con los siguientes parámetros; MMBQ de 25 + MRPS de 2 + fSB; MMBQ de 30 + MRPS de 2 + fSB; MMBQ de 30 + MRPS de 3; y MMBQ de 30 + MRPS de 2.

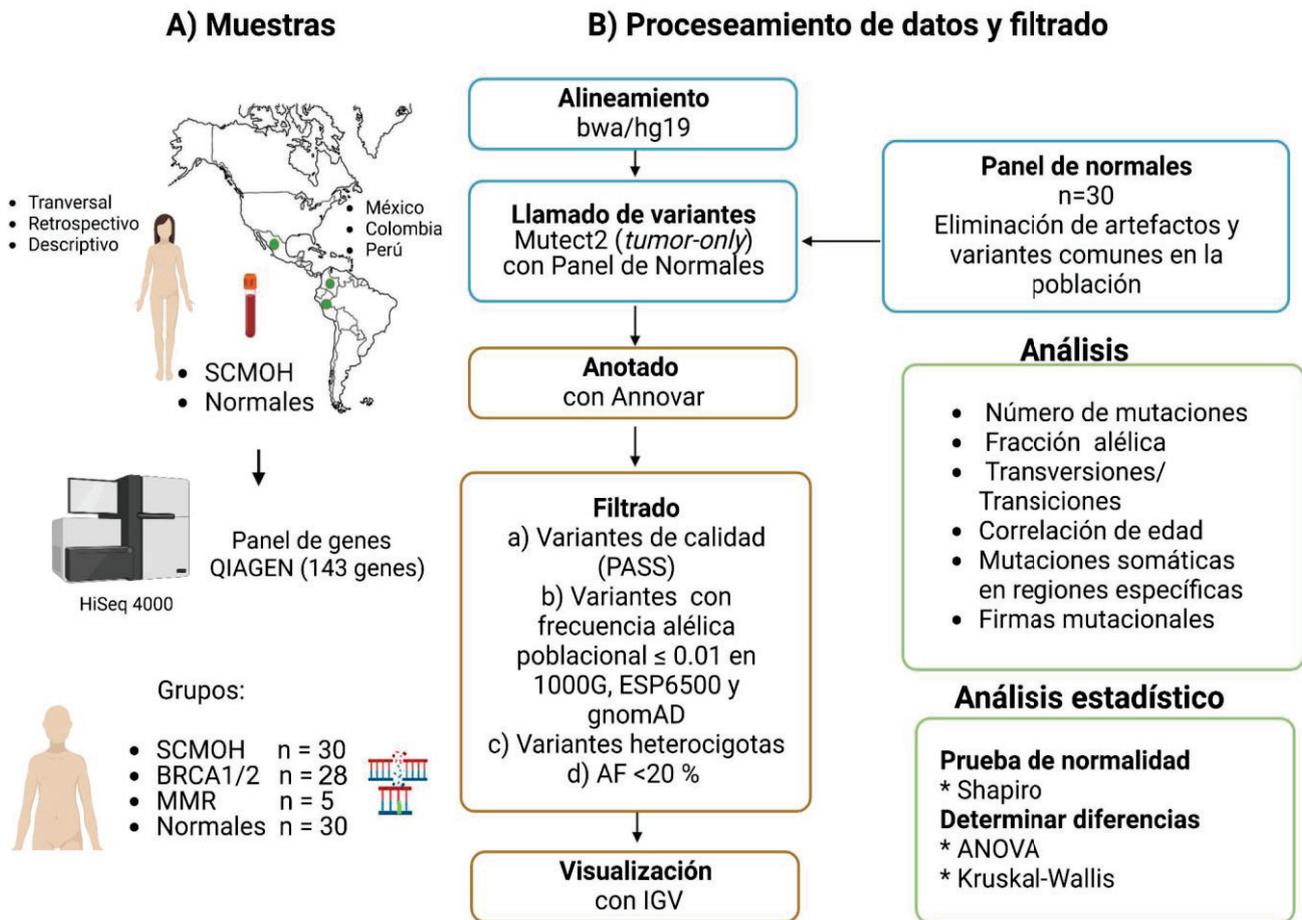
Los 54 archivos fueron anotados y curados bajo las especificaciones descritas (Figura 6).

## Llamado con el panel de normales

Se secuenciaron 30 muestras de individuos sanos (sin SCMOH), en el equipo Hiseq400 con una profundidad mínima de 1000x con el panel de 143 genes de QIAGEN. Se generó un conjunto de datos (panel de normales) para filtrar variantes a nivel de sitio, eliminando artefactos de secuenciación y variantes frecuentes en la población (Figura suplementaria 2).

Para obtener la mayor precisión y sensibilidad del llamado de variantes se ejecutó el programa con la versión de GATK 4.1.8.1, panel de normales, archivo BED, la habilitación de todas las regiones para la generación de haplotipos (parámetro - *force active TRUE*), se eliminó la restricción de 50 lecturas para la búsqueda de variantes (parámetro - *max read pear alignmend strand 0*) y se filtró con una calidad mínima de MQ 30, de acuerdo a la escala Phred.

Posteriormente se hizo un anotado de cada muestra y se filtró de acuerdo con los criterios establecidos en la Figura 6.



**Fig 6. Diseño del estudio y esquema de flujo de trabajo con Mutect2 (*tumor -only*).** A) Características de las muestras y grupos establecidos para el estudio. B) Procesamiento de datos y análisis posteriores.

### Visualización en Integrative Genomics Viewer (IGV)

La valoración de las variantes obtenidas en los programas de llamado de variantes somáticas y llamado de variantes germinales se realizó a través del visualizador Integrative Genomics Viewer. A partir del archivo bam y bai cargados a IGV se compaginó que la mutaciones somáticas estuvieran presentes en la cadena antisentido y sentido, descartando las variantes con sesgo en una de las cadenas (strand bias), aunado a esto en el caso de las muestras del panel de 143 genes se verificó que las mutaciones no estuvieran a los

extremos de los amplicones ya que el enriquecimiento de las muestras provenientes de este panel fue a partir de PCR, caso contrario a el panel de 74 genes el cual fue enriquecido por captura de híbridos.

### **Análisis estadístico**

Los datos fueron analizados mediante pruebas estadísticas con la paquetería “nortest”, “tidyr”, “DataCar” de R. Para determinar la distribución de los datos se usó la prueba de Shapiro, seguido de esto se realizó la prueba de Levene para identificar si los datos presentaban homocedasticidad. Con base en estos resultados se realizó la prueba ANOVA o Kruskal-Wallis para determinar diferencias significativas. Para identificar la diferencia entre grupos se utilizó la prueba de Tukey y la corrección de Holm.



**Tabla 1. Características clínicas y moleculares de las pacientes seleccionadas este estudio.**

SCMOH			BRCA			MMR		
ID	Edad	Sexo	ID	Edad	Sexo	ID	Edad	Sexo
IMS720105	33	Femenino	ISEM-07	51	Femenino	PJA_231	.	.
IMS720106	43	Femenino	ISEM-39	49	Femenino	PJA_206	.	.
IMS720109	42	Femenino	IMS720052	39	Femenino	PJA_095	.	.
IMS720110	45	Femenino	IMS720055	34	Femenino	Col_517	41	.
IMS720112	41	Femenino	IMS720076	48	Femenino	IMS720061	44	Femenino
IMS720113	40	Femenino	IMS720079	38	Femenino			
IMS720114	50	Femenino	IMS720086	49	.			
IMS720126	50	Femenino	IMS720090	.	Femenino			
IMS720127	43	Femenino	IMS720101	42	Femenino			
IMS720128	50	Femenino	IMS720104	40	Femenino			
IMS720129	42	Femenino	IMS720108	46	Femenino			
IMS720130	48	Femenino	IMS720115	49	Femenino			
IMS720131	43	Femenino	IMS720118	42	Femenino			
IMS720132	37	Femenino	IMS720143	44	Femenino			
IMS720133	50	Femenino	IMS720145	32	Femenino			
IMS720134	48	Femenino	IMS720151	56	Femenino			
IMSS72111	47	Femenino	IMS0400139	33	Femenino			
IMSS72120	37	Femenino	IMS0400145	40	Femenino			
IMSS72125	50	Femenino	IMS0400155	44	Femenino			
IMS0400137	42	Femenino	IMS0400179	37	Femenino			
IMSS7272	45	Femenino	MDH-4539	.	.			
IMS720152	50	Femenino	JRS-4763	.	.			
IMS720153	50	Femenino	Col_501	45	.			
IMS720154	52	Femenino	Col_539	.	.			
IMS720155	46	Femenino	Col_549	.	.			
IMS720156	50	Femenino	IMS720065	47	Femenino			
IMS720157	44	Femenino	ZHC_4630	.	.			
IMS720158	45	Femenino	ISM0026	50	Femenino			
IMS720159	50	Femenino						
IMS720160	50	Femenino						

### **Detección de variantes producto de artefactos de secuenciación**

Las variantes detectadas por los programas bioinformáticos (QIAGEN, SOPHIA genetics, Haplotypecaller, Platypus, VarDict, Somaticseq, Needlestack y Mutect 2 tumor-only sin panel de normales) mostraron una heterogeneidad en número de variantes totales y fracción alélica; lo que generó sobreestimaciones al correr los programas de modo predeterminado y subestimaciones al modificar el parámetro de calidad de bases, aumentando el valor mínimo de 30 con respecto a la escala Phred (escala logarítmica de error). También se observó una tasa de falsos positivos >20% y la presencia de errores recurrentes al detectar mutaciones en una sola hebra del ADN. La inspección en IGV reveló que las mutaciones de baja fracción alélica eran errores de secuenciación y se encontró un error compartido en la detección de falsos positivos en los extremos de los amplicones en las muestras correspondientes al panel de 143 genes, el cual utiliza PCR para enriquecer las secuencias durante el proceso de secuenciación (Figura suplementaria 3).

### **Precisión en el llamado de variantes somáticas con Mutect2 (*tumor-only*)**

Como se muestra en la Tabla 1, hubo una disminución en el número de variantes cuando se delimitó la región blanco y se realizó el ajuste del parámetro de calidad de mapeo al valor de MQ30 (correcto mapeo dentro de las coordenadas genómicas de valor 30 en escala Phred). A pesar de la implementación de estos parámetros de calidad del llamado aún se detectaron falsos positivos. El porcentaje de bases mutadas con respecto al total (fracción alélica) eran de origen germinal (~ >50%), no había complementariedad total en las cadenas y se detectaron falsos positivos en una sola de las cadenas del ADN y variantes en los extremos de los amplicones a pesar de delimitar las regiones exónica del panel de 143 genes.

**Tabla 2. Promedio de variantes somáticas detectadas por paciente y rango de fracción alélica detectada sin el filtro de frecuencia alélica <0.01%.**

	QIAGEN	AF QIAGEN	SOPHIA	AF SOPHIA
<i>Tumor-only</i>	241.6	<25%	863.5	<25% y 75%
<i>Tumor-only</i> BED	51	<30%	43	<1% y 25-50%

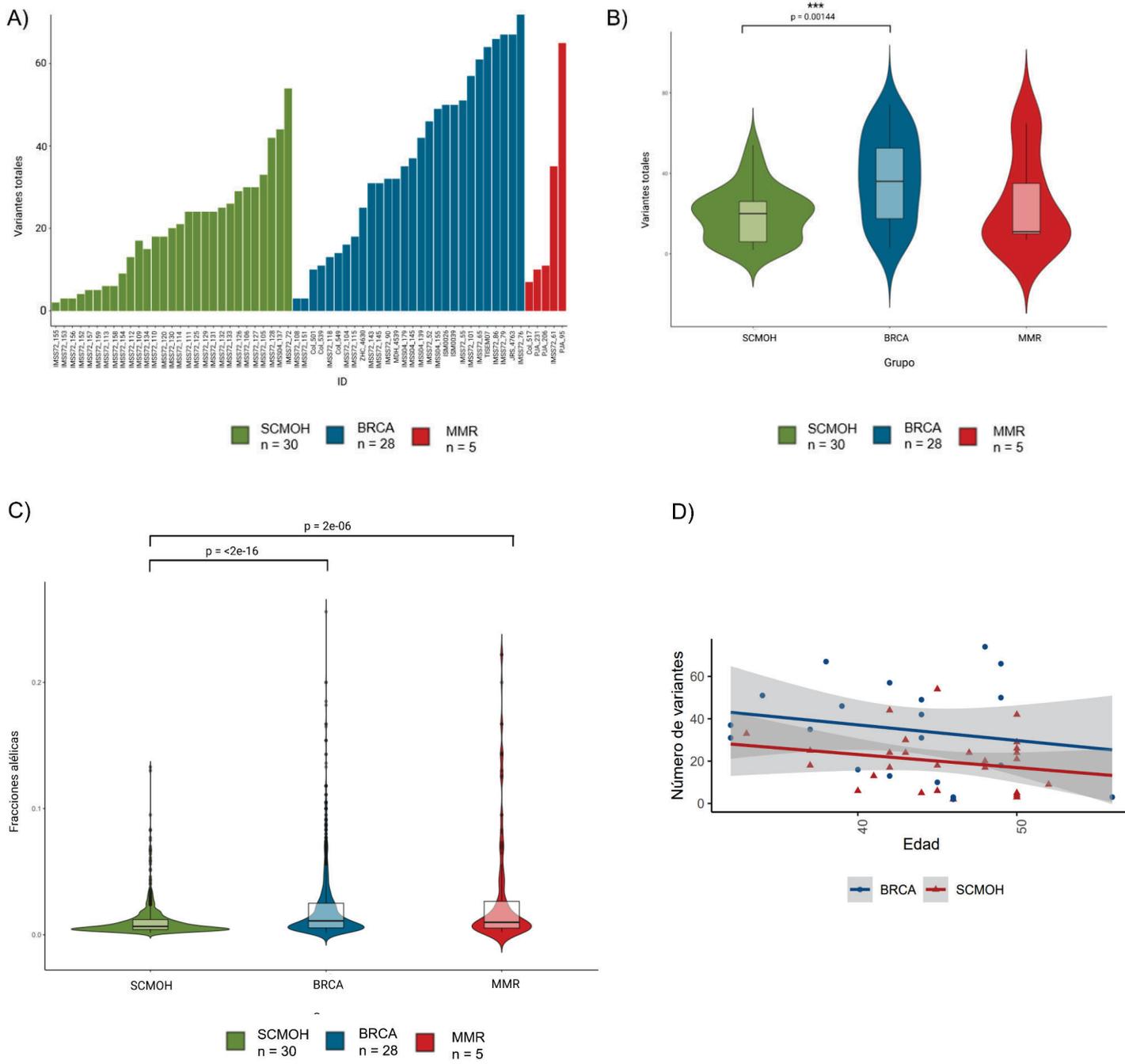
El uso de los parámetros de calidad de alineamiento, calidad de base, tasas de falsos positivos, f-score, lecturas mínimas en el alelo alternativo y la combinación de estas mostraron una tasa de error >20%. Ya que el algoritmo no podía por sí solo incrementar la precisión en la detección de variantes se generó un panel de normales para hacer el llamado más preciso.

### **Eliminación de artefactos de secuenciación**

El uso de GATK versión 4.1.81, los parámetros a) - *-max read pear alignmend strand 0* b) - *-force active TRUE*, panel de normales (muestras secuenciadas con panel de 143 genes) (Figura suplementaria 2) y un mínimo de calidad de mapeo y de base de 30, permitió hacer un llamado óptimo sin la detección de artefactos. Lo anterior, en conjunto con la eliminación de variantes frecuentes en la población (punto de corte=0.001), la eliminación de errores en extremos de los amplicones permitió la detección de datos de buena calidad. Las métricas de profundidad de las variantes obtenidas con Mutect2 se confirmaron en IGV y una similitud en el reporte de profundidades entre lo establecido por Mutect2 y IGV, corroborando. Con el resultado de los parámetros de optimización mencionados, se seleccionó este flujo de trabajo de procesamiento de datos para hacer los análisis de los tres grupos de estudio de las pacientes.

### **Mayor tasa de mutaciones somáticas en pacientes con defectos por mecanismos de recombinación homóloga**

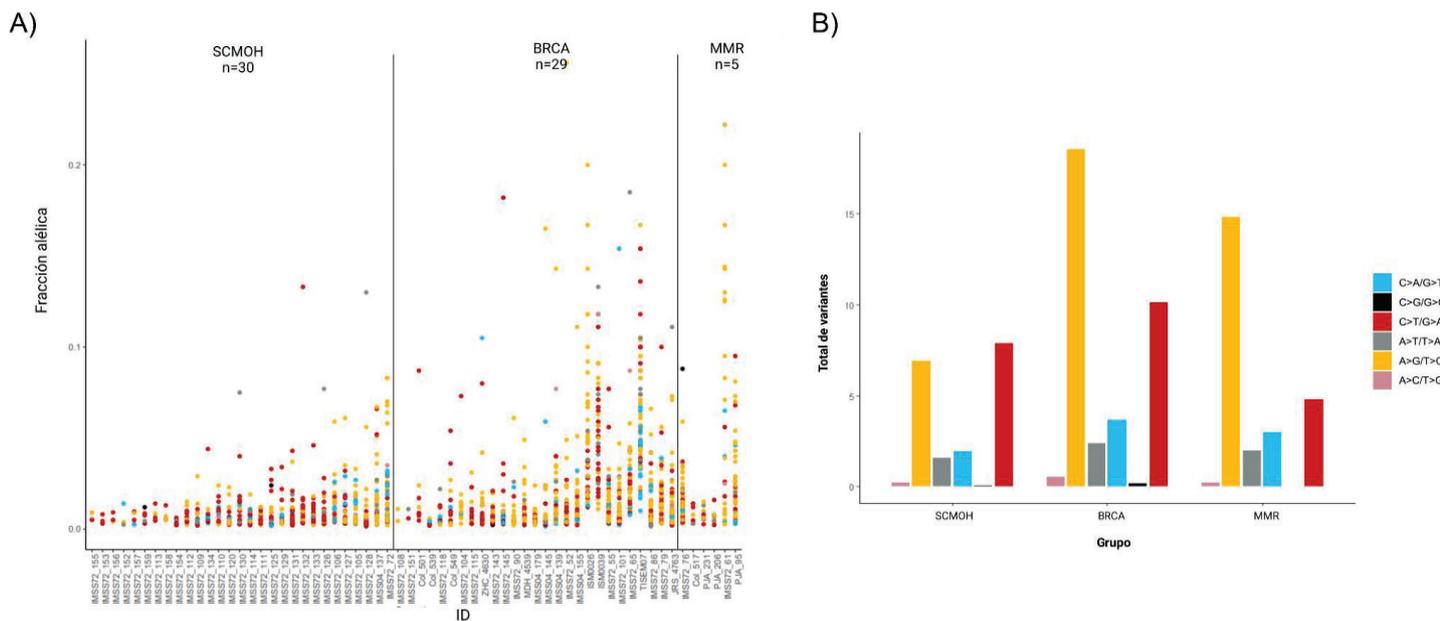
Para identificar el número de variantes somáticas presentes en los diferentes grupos con SCMOH, se realizó el análisis bioinformático de datos provenientes de la secuenciación masiva de bibliotecas enriquecidas, como ya se definió en la sección de “Materiales y métodos”. Se identificaron 1757 variantes totales (SCMOH 574, BRCA1/2 1055 y MMR 128; Figura 2A). La distribución de las mutaciones entre los grupos fue heterogénea con una media de 19.79, 37.67, 25.60 en SCMOH, BRCA1/2, MMR, respectivamente (Figura 8). Las pacientes con BRCA1/2 tuvieron la mayor tasa de mutaciones, con diferencia significativa ( $p=0.00144$ ) entre SCMOH y BRCA1/2. Por otra parte, se evaluó las fracciones alélicas entre los grupos, se identificó una media de 0.0111 en SCHMO, 0.02135 en BRCA1/2 y 0.0303 en MMR, con una diferencia significativa entre el grupo SCMOH y *BRCA1/2* ( $p=<2 \times 10^{-16}$ ); y SCMOH y MMR ( $p=2 \times 10^{-06}$ ) (Figura 8 C). Para determinar la correlación entre la edad y el número de variantes somáticas se realizó una regresión lineal. No se encontró correlación entre la edad y la acumulación de mutaciones somáticas (Figura 8D).



**Figura 8. Variantes somáticas de los 3 grupos de pacientes.** A) Número total de variantes por paciente y B) por grupo en las pacientes con SCMOH. Para establecer la significancia estadística se realizó la prueba de ANOVA y la prueba de Tukey con un  $p < 0.05$  \*. C) Distribución de la fracción alélica ( $AF < 0.25$ ) por grupo. Análisis estadístico realizado con Kruskal-Wallis seguido por la prueba de Holm con diferencia significativa de  $p < 0.05$  \*. D) Correlación de edad y mutaciones adquiridas en las pacientes con SCMOH.

## Predominio de transiciones en cambios de nucleótido único

Para identificar los tipos de mutaciones de cambio de nucleótido único se realizó una clasificación en dos grupos: i) transversiones (C>A/G>T, C>G/G>C, A>T/T>A y A>C/T>G) y ii) transiciones (C>T/G>A y A>G/T>C). De acuerdo con la distribución de la fracción alélica por paciente y su tipo de mutación se observó una mayor AF en la transición (A>G/T>C). Los grupos SCMOH y BRCA1/2 tuvieron una diferencia significativa entre los tipos de SNV ( $p < 0.05^*$ ) (Figura 9). Las pacientes pertenecientes a SCMOH tuvieron diferencias significativas entre todos los tipos de cambio de nucleótido único, excepto las transversiones C>A/G>T con A>T/T>A y C>G/G>C con A>C/T>G con  $p=0.664$  y  $0.381$  (Figura 8B). El grupo BRCA1/2 obtuvo diferencias significativas ( $p < 0.05^*$ ) en el tipo de cambio de nucleótido a excepción de A>T/T>A con respecto a C>A/G>T ( $p = 0.3184$ ) (Figura 9B). En contraste, el grupo MMR no presentó diferencias significativas (Figura suplementaria 4).

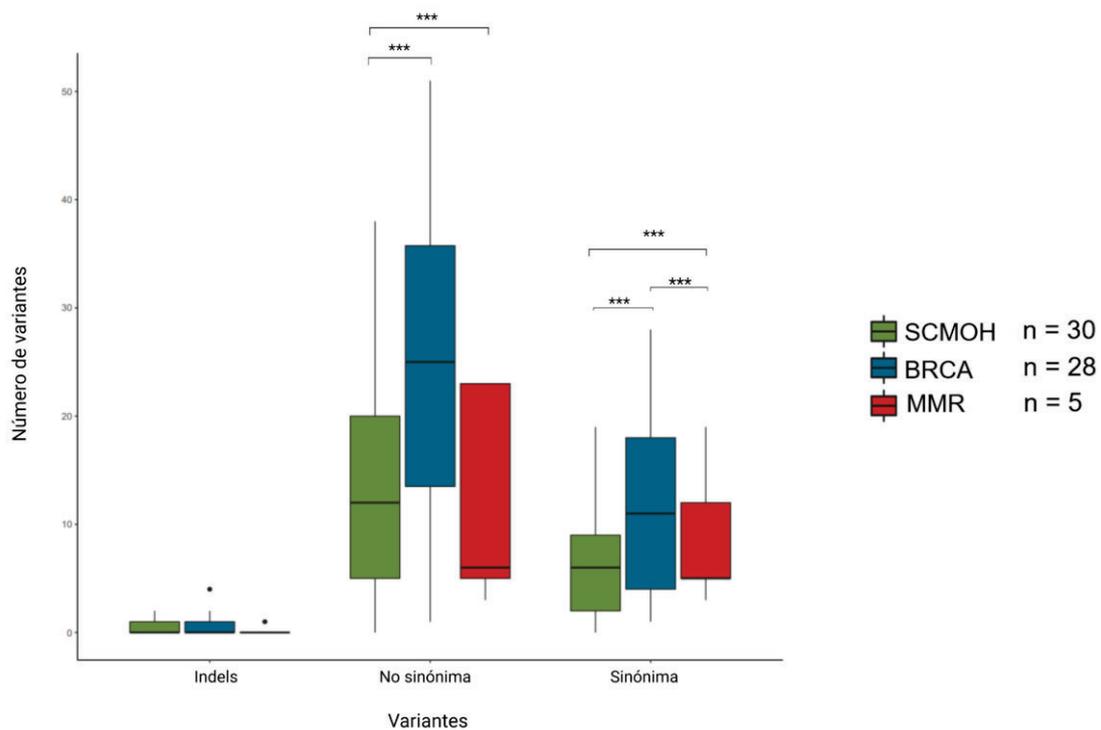


**Figura 9. Mayor tasa de transiciones.** A) Distribución de las variantes obtenidas de cada uno de los individuos con respecto a la fracción alélica y cambio de nucleótido único. B) Conteo de tipo de cambio único de nucleótido con respecto a los tres grupos.

### **El grupo de MMR presenta mayor número de mutaciones no sinónimas y sinónimas**

Para identificar el número y el tipo de mutaciones presentes en los tres grupos de pacientes, se clasificaron las variantes en tres grupos: a) indels, deleciones e inserciones pequeñas; b) mutaciones no sinónimas, variantes que codifican para un aminoácido diferente o adquieren una ganancia o pérdida de codones de paro; y c) mutaciones sinónimas, variantes que no cambian la estructura de la proteína (Figura 10). Los tres grupos presentaron una media de indels de 0.48, 0.67, 0.2 para SCMOH, BRCA1/2, MMR, respectivamente. No se observó ninguna diferencia significativa entre los grupos.

Las mutaciones no sinónimas se distribuyeron de manera heterogénea entre los grupos, con una media de 13.68 SCMOH, 24.85 BRCA1/2 y 12 MMR. A partir de los datos normalizados, se observó una diferencia significativa entre MMR, BRCA1/2 con respecto a SCMOH ( $p=0.0095$  y  $p=0.0031$ , respectivamente). Las mutaciones sinónimas presentaron una media de 6.03, 12.14 y 8.80 en SCMOH, BRCA1/2 y MMR, respectivamente. A partir de los datos normalizados se encontró diferencias significativas entre MMR, BRCA1/2 con respecto a SCMOH ( $p=0.0016$  y  $0.004$ , respectivamente) y entre BRCA1/2 y MMR ( $p=0.004$ ) (Figura 10).



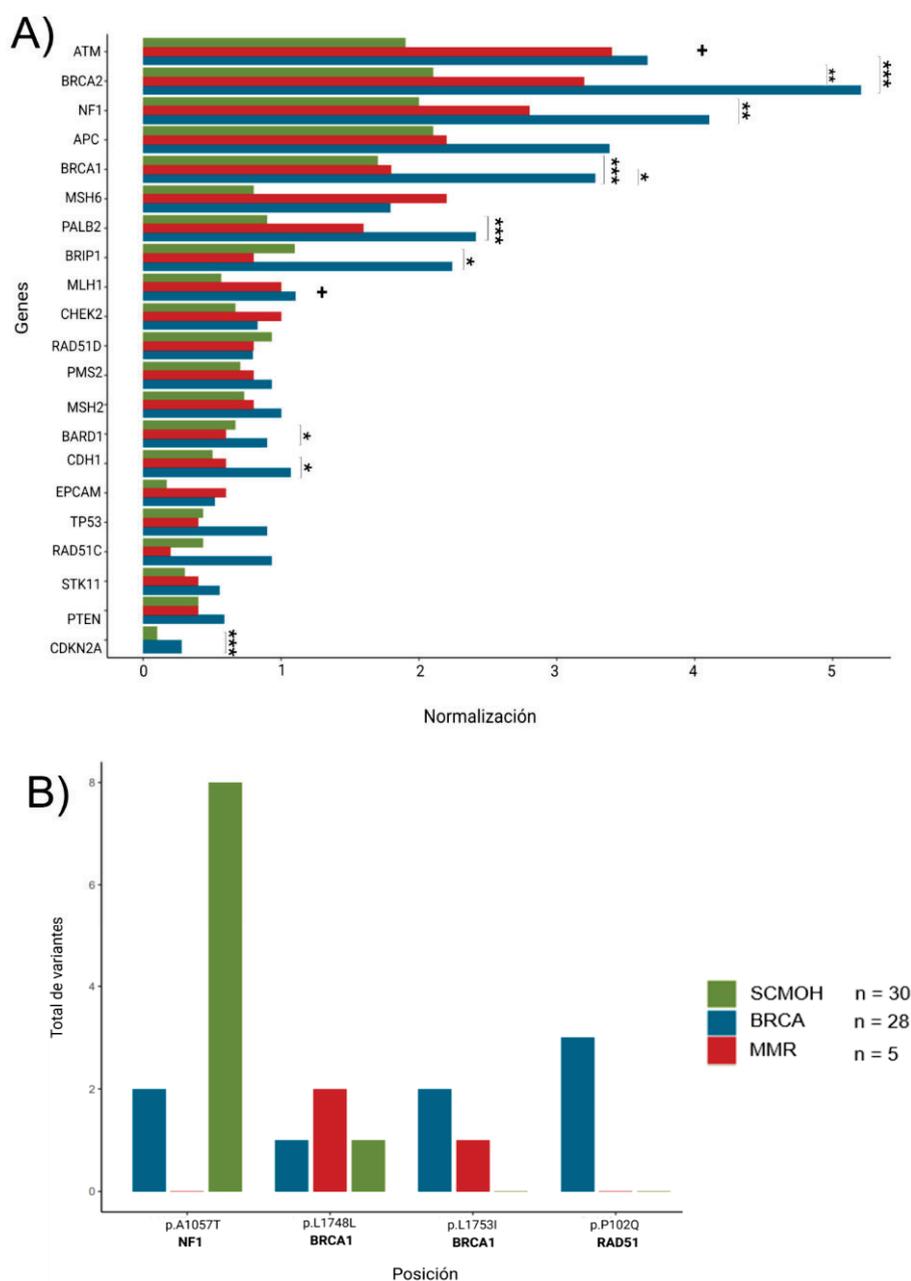
**Figura 10. Cambios de nucleótido único con impacto en el marco de lectura.** Los datos se normalizaron de acuerdo al número de muestras de cada grupo. Se realizó la prueba de ANOVA en el grupo de los indels ( $p < 0.05^*$ ) y Kruskal-Wallis para mutaciones no sinónimas y sinónimas ( $p < 0.05^*$ ), evaluando la diferencia entre grupos con prueba de Tukey y corrección de Holm, respectivamente.

### Mutaciones adquiridas en genes implicados en la respuesta al daño de ADN

Para comprender mejor el impacto de las variantes somáticas detectadas en vías de señalización relevantes para el proceso de carcinogénesis, se evaluaron las variantes en genes conductores del cáncer. También se evaluó la recurrencia de las variantes en las mismas posiciones genómicas.

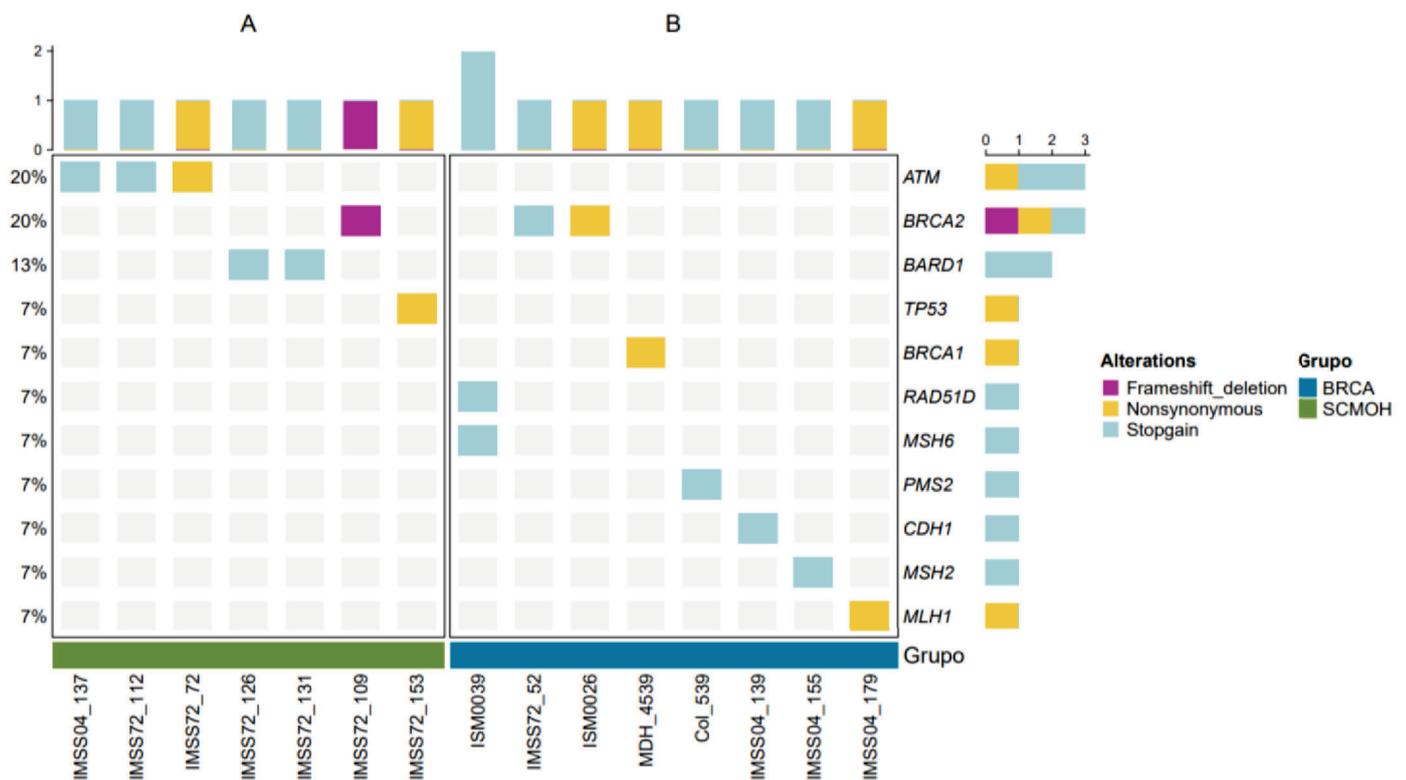
En los datos obtenidos con el panel de 143 genes, se detectaron variantes en 21 genes y sólo 10 ellos presentaron diferencias significativas (*BRCA2*, *NF1*, *ATM*, *BRCA1*, *BRIP1*, *PALB2*, *BARD1*, *MLH1*, *CDH1* y *CDK2A*) con  $p < 0.05$  (Figura 11A). En estos genes se analizó la posición genómica donde se encontraba cada una de las variantes somáticas; identificando 4 sitios recurrentes repetidos más de dos veces (Figura 11B). Con base a los datos de ClinVar, las variantes en la p.P102Q en *RAD51D* son de significado clínico incierto, el cambio

p.A1057T en *NF1* se describe como posible benigna. Las variantes p.L1753I y p.L1748L en *BRCA1* no tuvieron información en esta base de datos.



**Figura 11. Genes con mutaciones de nucleótido único.** A) Se detectaron mutaciones en 21 genes de susceptibilidad a cáncer hereditario, con diferencias significativas ( $p < 0.05^*$ ,  $p < 0.01^{**}$ ,  $p < 0.001^{***}$ ); *MLH1* y *ATM* mostraron diferencias significativas entre los tres grupos (+). B) Se identificaron 4 posiciones recurrentes en *NF1*, *BRCA1* y *RAD51D* ( $n > 2$ ).

Se identificaron variantes patogénicas en 15 pacientes, 7 en el grupo SCMOH y 8 en *BRCA1/2*. Las variantes patogénicas se hallaron en 11 genes (*ATM*, *BRCA2*, *BARD1*, *TP53*, *BRCA1*, *RAD51D*, *MSH6*, *PMS2*, *CDH1*, *MSH2* y *MLH1*) (Figura 12).



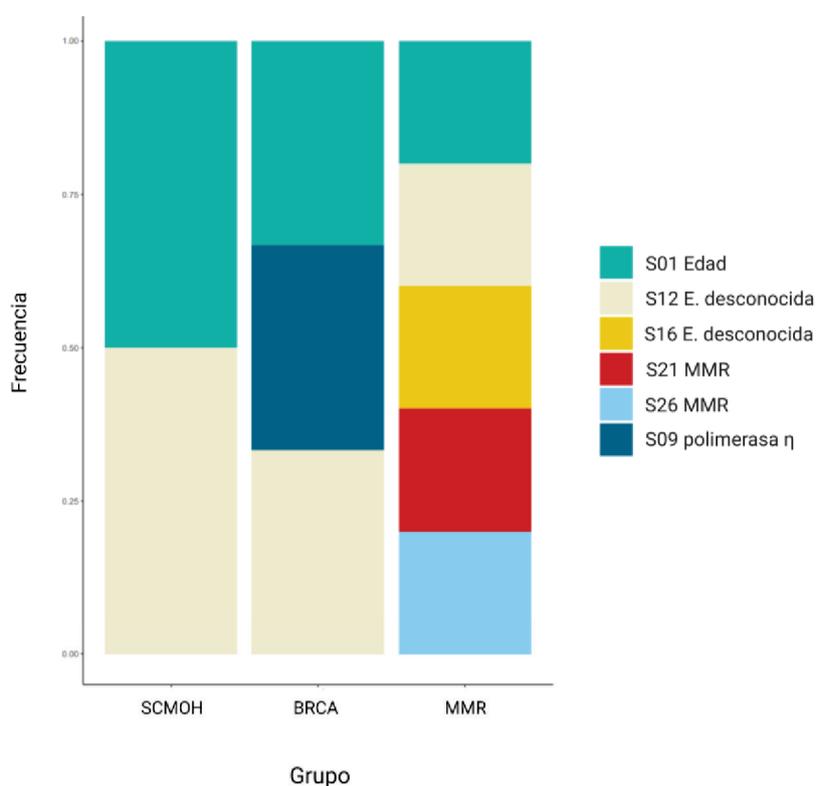
**Figura 12. Composición de variantes patogénicas no germinales.** Pacientes con variantes patogénicas somáticas en los grupos SCMOH y BRCA1/2.

### Identificación de firmas mutacionales propias de daño en MMR

Para identificar los procesos mutagénicos que actúan en las células sanguíneas de las pacientes del estudio se realizó un análisis de detección de patrones mutacionales conocidos como firmas mutacionales.

En los tres grupos se presentó la firma S01, asociada a las mutaciones acumuladas a lo largo de la edad y la firma S12 de etiología desconocida. También se detectó la firma S09

en el grupo de *BRCA1/2*. Esta firma está asociada a las mutaciones inducidas por la polimerasa  $\eta$  durante la replicación de linfocitos como respuesta inmune adaptativa en el proceso de hipermutación somática. En el grupo MMR encontramos firmas propias de la deficiencia de reparación con bases mal apareadas con la identificación de las firmas S21 y S26 (Figura 13).



**Figura 13. Frecuencia relativa de las firmas mutacionales de cada grupo.** Se observaron dos firmas mutagénicas compartidas en los tres grupos (S01 y S12). Adicionalmente, el grupo BRCA1/2 presentó la firma S09 y el grupo MMR dos firmas asociadas a la deficiencia en el mecanismo de MMR (S16 y S21).

## Discusión

El propósito de este estudio fue identificar patrones mutacionales causados por la deficiencia germinal en la función de genes que participan en la reparación del ADN por RH y MMR en pacientes con SCMOH. En esta enfermedad la pérdida monoalélica de la función de los genes *BRCA1/2* se asocia a un alto riesgo de desarrollar distintos tipos de cáncer. Por otro lado, los genes participantes en la reparación del ADN por el mecanismo de MMR tienen una menor prevalencia de pérdida de la función en estos pacientes, sin embargo, se asocian con altas tasas de mutación somática e inestabilidad microsatélite, como se ha observado en el cáncer colorrectal (Kobayashi H., *et al* 2013; Roberts *et al.*, 2018). Los efectos a nivel somático de la inactivación germinal de los genes *BRCA1*, *BRCA2*, *MLH1*, *MSH2*, *MSH6*, *EPCAM* y *PMS2* se han descrito con bastante detalle en muestras de tejido tumoral. No obstante, se ha estudiado muy poco el impacto de dichas alteraciones germinales en tejidos normales (Roberts *et al.*, 2018; Wei *et al.*, 2021; Robinson *et al.*, 2022).

Por ello, en este análisis retrospectivo se seleccionaron muestras de sangre periférica de 63 muestras de pacientes con SCMOH que presentan variantes germinales patogénicas en dichos genes, a partir de datos de 1000 pacientes secuenciadas por el consorcio LACAM (Oliver J., *et al* 2019). Los datos de secuenciación se obtuvieron empleando la tecnología más avanzada, con el uso del equipo HiSeq 4000 con profundidades promedio de 2000X, a partir de un panel de 143 genes con susceptibilidad a cáncer hereditario. Para identificar variantes somáticas que se acumulan de manera progresiva a lo largo de la vida de estas pacientes y comprender mejor su impacto en el contexto de pacientes con defectos germinales en genes de reparación del ADN, realizamos los siguientes análisis: i) la identificación de variantes somáticas totales con fracción alélica  $<0.20$ ; ii) la correlación de la edad y la tasa de mutaciones; iii) la composición de los cambios de nucleótido único; iv) el efecto biológico de cada mutación; v) la evaluación de genes de relevancia clínica que presentaron con mutaciones somáticas; vii) y la descripción del patrón de firmas mutacionales.

En la primera etapa de nuestro análisis realizamos una estandarización técnica exhaustiva de los procedimientos bioinformáticos de análisis de las variantes, con el objetivo de eliminar los errores sistemáticos de la secuenciación. De esta manera, buscamos contar con datos confiables que no tuvieran variantes falsas positivas ni causadas por las señales de fondo (“ruido de fondo”) de la tecnología de secuenciación.

Para la estandarización se utilizaron programas de identificación de mutaciones germinales y somáticas, ya que existen varios programas de alta sensibilidad en detección de mutaciones somáticas a partir de muestras de tejido normal y tejido adyacente al tumor (tejido normal) y pocos programas tienen algoritmos capaces de procesar muestras solo del tumor, por lo contrario, los programas diseñados para identificar variantes patogénicas germinales no requieren una muestra pareada y únicamente hacen uso del genoma de referencia. Debido al origen del tejido de la muestra (tejido sanguíneo), se optó por el uso de *callers* de variantes germinales, eliminando la detección germinal (fracción alélica >50%) y reteniendo los datos clasificados como de origen somático (fracción alélica <20%), sin embargo, las variables detectadas presentaban falsos positivos y una alta cantidad de artefactos, esto es concordante con el diseño de los algoritmos que precisan y especifican la detección basada en datos de frecuencia alélica poblacional y clasifican de manera general los datos de frecuencias alélica <0.01 (Cibulskis *et al.*, 2013; Koboldt, 2020). Para los programas especializados en la detección de variantes somáticas se requiere un llamado pareado, sin la muestra de tejido adyacente al tumoral no genera una detección; en cambio los programas que tiene un llamado con la modalidad de solo tumor detectan muchos artefactos de secuenciación como SNV de baja fracción alélica; además se reporta una alta tasa de falsos positivos al usar PCR multiplex (McCall *et al.*, 2014). Los falsos positivos generados son difíciles de filtrar con este tipo de programas debido al balance que tiene el algoritmo entre precisión y sensibilidad. Sin embargo, Mutect2 *tumor-only* permite la adición de múltiples parámetros y la construcción de un panel de normales que elimina artefactos de secuenciación y variantes frecuentes en la población, lo que incrementa la sensibilidad y precisión del llamado de variantes, dando una detección certera en comparación al resto de programas a pesar de tener una muestra de origen sanguíneo; razón por la que se utilizó este programa (Cibulskis *et al.*, 2013; Benjamin D. *et al.*, 2021).

Se observó una tasa mayor de mutaciones en el grupo de BRCA1/2, con una media de 37.67 y con diferencia significativa ( $p=0.00144$ ) con respecto al grupo SCMOH. Estos hallazgos son consistentes con las funciones reparadoras de los genes *BRCA1* y *BRCA2*. Ambos genes participan en la reparación de rupturas de doble cadena, de entrecruzamiento de cadenas y gaps de cadena sencilla que generan estrés durante la replicación celular, evitando el colapso de la horquilla de replicación (Venkitaraman A., 20001; Kobayashi *et al.*, 2013; Deans & West, 2011). Los genes *BRCA1* y *BRCA2* también participan en la reducción del estrés celular causado durante la transcripción, la exposición a radiaciones ionizantes, la quimioterapia con agentes intercalantes, la exposición a agentes alquilantes como sales de platino, entre otros. Por lo tanto, nuestros datos sugieren que el aumento en el número de mutaciones somáticas en las células sanguíneas observado en el grupo BRCA1/2 puede estar asociado a la deficiencia germinal monoalélica en esos genes. (Chatterjee N. & Walker, 2017; Krejci *et al.*, 2012; Deans & West, 2011).

Por el otro lado, la tasa de mutaciones que presentó en el grupo MMR fue superior al grupo SCMOH a pesar de ser elevada, con un promedio de 25.60, no mostró una tasa exacerbada de mutaciones como se observa en los tejidos tumorales. En los tumores con deficiencia de MMR debida a la pérdida de la función de *MLH1*, *MSH2*, *MSH6*, y *PMS2*, se presenta inestabilidad microsatélite, resultando en una alta tasa de mutaciones. Sin embargo, cabe resaltar que la inestabilidad microsatélite actúa en regiones repetitivas en el ADN, las cuales constituyen el ~3% del genoma. La mayoría de las regiones microsatélites o repeticiones en tandem se encuentran en zonas intrónicas y sitios de splicing, y en menor medida en regiones codificantes, como las que se evaluaron en este estudio. Esto podría explicar el por que no detectamos una alta tasa de mutación que supera los niveles de los otros grupos, ya que en este trabajo empleamos un panel de genes que incluye únicamente regiones codificantes y de splicing (Pećina-Šlaus *et al.*, 2020; Kobayashi H., *et al* 2013; Vogelstein *et al.*, 2013). Otro factor que podría influir en el número de mutaciones observadas en el grupo MMR es que sólo se contó con cinco pacientes de este grupo. Dos de las cinco pacientes (ID IMSS72\_61 e ID PJA\_95) presentaron un alto número de mutaciones (>30).

Para evaluar el efecto mutagénico acumulativo de las variantes germinales en genes de reparación del ADN analizamos la correlación entre la edad de las pacientes y el número de mutaciones. En los grupos SCMOH y BRCA1/2 no se encontró una correlación entre la tasa de mutaciones y la edad de los pacientes. Estos resultados difieren con los antecedentes descritos en la literatura, en los cuales se ha descrito en tejidos sin cáncer el efecto progresivo en la acumulación de mutaciones ocasionadas por procesos intrínsecos como la replicación celular, así como mutaciones ocasionadas por la exposición a mutágenos a lo largo de la vida (Martincorena I. & Campbell P. J., 2015; Wei *et al.*, 2021; Robinson *et al.*, 2022). Esta tasa también se incrementa en algunos tipos de cáncer presentes en tejidos que se renuevan continuamente como el tejido epitelial y sanguíneo (Vogelstein *et al.*, 2013; Yoshida *et al.*, 2020). Las edades de las pacientes del grupo SCMOH se encuentran en un intervalo de 33-52 años y las del grupo BRCA1/2 en 32-56 con una media de 45.4 y 43.5, respectivamente. Es posible que la acumulación progresiva de mutaciones sea detectable en intervalos de edad más grandes a los de las pacientes reclutadas para este estudio. Para mejorar este análisis sería necesario realizar un estudio de diseño prospectivo, dando seguimiento a cada paciente a lo largo de varios años o décadas, para evaluar si existe un aumento progresivo de las mutaciones somáticas en sus muestras sanguíneas.

En nuestros datos, las pacientes heterocigotas mostraron variantes somáticas con fracciones alélicas de 0.0111-0.0303. Con la eliminación de las variantes con tracción alélica  $\geq 0.2$  descartamos la inclusión de variantes de origen germinal que pudieran tener efecto confusor para nuestro análisis. Adicionalmente, eliminamos las variantes comunes de las poblaciones latinas que incluyen mexicanos, a partir de los datos de gnomAD, 1000genome y ESP, empleando un corte de frecuencia alélica de 0.001. De esta manera excluimos variantes genéticas y polimorfismos compartidos en la población de estudio. Con esta estrategia nos fue posible determinar las mutaciones somáticas adquiridas en las células sanguíneas de las pacientes (Figura 6C). Los valores de fracción alélica mutante que encontramos no coinciden en su totalidad con la detección de variantes en hematopoyesis clonal ya que se reporta una fracción alélica de 20-30% pero sí coinciden con las fracciones alélicas reportadas en estudios realizados a tejidos epiteliales sanos. En estos trabajos se reportan fracciones alélicas de 1-10%, a pesar de ser tejidos distintos (Genovese *et al.*, 2014; Wei L., *et al* 2021).

La presencia de variantes con baja fracción alélica es congruente con un proceso somático de acumulación aleatoria de mutaciones a lo largo del genoma. De acuerdo al apareamiento canónico de las bases se puede distinguir entre dos tipos de mutaciones, transiciones y transversiones. En nuestros análisis obtuvimos como resultado una alta tasa de transiciones ( $A>G/T>C$  y  $C>T/G>A$ ), destacando la transición de pirimidinas. Esta transición está asociada a procesos de mutación exógenos y endógenos como el consumo de alcohol, la acción directa de radiación ionizante y radicales libres producidos por el metabolismo de xenobióticos o ionización del agua (Chatterjee & Walker, 2017). Un fenómeno común ocasionado por radicales libres es la oxidación de guaninas (oxiguanina) lo que inhibe el apareamiento canónico entre G-C y apareo G-A; otro factor que ocasiona transiciones de pirimidinas es la exposición a rayos UVB. (McCulloch & Kunkel, 2008; Chatterjee & Walker, 2017, Alexandrov *et al.*, 2013).

El contexto específico de cada mutación determina los efectos en la función genética. Por ello, evaluamos el impacto biológico de las variantes somáticas detectadas en relación a los cambios en el marco de lectura, sustituciones no sinónimas o indels. Los resultados muestran una baja cantidad de indels, destacando que para el grupo MMR no se identificó ninguna. No encontramos diferencias significativas del tipo de variantes detectadas entre los grupos. Esta observación puede ser explicada por la actividad del mecanismo MMR presente en los grupos BRCA1/2 y SCMOH, ya que es muy probable que este mecanismo opere de manera correcta. Por el contrario, la falta de indels en el grupo MMR no coincide con lo descrito en la literatura, ya que las inserciones y deleciones producidas durante la replicación celular son reparadas por este mecanismo, por lo que su deficiencia debería producir un incremento de indels (Chatterjee & Walker, 2017; Modrich P. & Lahue R., 1996; Pećina-Šlaus *et al.*, 2020). En contraste, las mutaciones no sinónimas fueron las más abundantes en todos los grupos, con una media de 13.68, 24.85 y 12 en SCMOH, BRCA1/2 y MMR, respectivamente. Además presentan diferencias significativas entre los grupos SCMOH y MMR, así como entre los grupos SCMOH y BRCA1/2 (datos normalizados). El grupo BRCA1/2 tuvo la tasa de mutaciones más alta lo cual se asocia a su función biológica en mantenimiento de integridad del ADN, al reparar diferentes daños; en el grupo de MMR es congruente la alta tasa de mutaciones de nucleótido único por la falla en este mecanismo. Sin embargo, cabe resaltar que las mutaciones ocurren de manera aleatoria a lo largo de

todo el genoma, pero el conocer qué impacto genera es de vital importancia tanto a nivel estructural como nivel funcional; nuestros datos provenientes de panel de genes indican que existe una modificación estructural en las proteínas, este impacto se puede analizar en los genes que presentan dichas mutaciones, así como las vías implicadas (Chatterjee N. & Walker, 2017).

Para entender mejor el impacto biológico potencial de las variantes somáticas detectadas evaluamos la función molecular de los genes que presentaron variantes. Dentro del panel de 143 genes únicamente detectamos mutaciones en 21 de ellos (*ATM, APC, BRCA2, NF1, BRCA1, MSH6, PALB2, BRIP1, MLH1, CHEK2, RAD51D, PMS2, MSH2, BARD1, CDH1, EPCAM, TP53, RAD51C, STK11, PTEN, CDKN2A*), de los cuales los que presentaron mayor frecuencia de mutaciones son *BRCA2, NF1* y *ATM*. De manera global, todos genes anteriormente mencionados se encuentran implicados en mecanismos de respuesta a daño de ADN, además de causar un aumento en la susceptibilidad a cáncer de mama. Dentro de los genes con mutaciones somáticas destacan cuatro grupos, i) genes implicados en mecanismos de recombinación homóloga (*BRCA1, BRCA2, PALB2, BRIP1, RAD51D, BARD1C, CHEK2* y *BARD1*) y ii) genes encargados de reparación de bases mal apareadas (*MSH6, MSH2, MLH1* y *PMS2*); iii) genes supresores de tumor (*CDKN2A, ATM, TP53, STK11* y *PTEN*); y iv) genes involucrados en adhesión celular y proliferación (*APC, CDH1, EPCAM* y *NF1*). Las mutaciones somáticas presentes en el grupo (i) de estos genes concuerdan con la susceptibilidad reportada en cáncer de mama hereditario (Foulkes, W. D., 2008; Hu, C. et al., 2021). Asimismo, se ha descrito que la adquisición de mutaciones en los genes *STK11, PTEN, CDKN2A* promueven la acumulación de rupturas de doble cadena, así como la inestabilidad cromosómica; lo cual es consistente con la tasa mayor de mutaciones detectada en el grupo BRCA1/2 (Kobayashi *et al.*, 2013; Nielsen *et al.*, 2016; Slavin *et al.*, 2017). Las proteínas codificadas por los genes del grupo (ii) participan en la reparación de bases mal apareadas e identificación de indels en la cadena de ADN, así como de generar el andamiaje de proteínas que permite la degradación del ADN dañado e inicio de síntesis de novo del ADN (Negrini *et al.*, 2010; Pećina-Šlaus *et al.*, 2020). Si bien estos genes son responsables del MMR, también se ha descrito un solapamiento de la vía MMR con la de anemia de Fanconi (FA-BRCA), vía encargada de reparación entrecruzamiento de cadenas (Kim, H., & D'Andrea, 2012). En contraste, las mutaciones en estos genes son

menores en pacientes con cáncer de mama hereditario pero muy alta en síndrome de Lynch, el cual al igual que SCMOH es una enfermedad autosómica dominante que también presenta metástasis en ovario (Ferrer Márquez et al., 2011). El grupo de genes supresores de tumor permite las mutaciones conductoras las cuales favorecen de manera global al fenotipo mutador ya que se ve comprometido el mantenimiento de la integridad del genoma. Los genes *APC*, *CDH1*, *EPCAM* y *NF1* implicados en adhesión celular y proliferación se encuentran asociados a cáncer de colon y ovario; sin embargo, la interconexión de vías de señalización permiten tener múltiples respuestas como es el caso particular de *APC*, el cual es característico de síndrome de Lynch y se encuentran en un 70% mutado o hipermetilado en cáncer de mama esporádico; se ha descrito en modelos *in vitro* e *in vivo* que la pérdida de *APC* contribuye a la quimioresistencia a través de la disminución de la respuesta de daño a ADN (Kim, H., & D'Andrea, 2012; VanKlombenberg et al. 2015, Stefanski et al., 2017). Estos hallazgos son de gran relevancia, ya que se detectaron cambios de relevancia clínica en 11 genes con variantes patogénicas (Foulkes, W. D., 2008).

La perspectiva global del entendimiento de los cambios de nucleótido único, cambios en el marco de lectura en regiones exónicas y el impacto que tiene este sobre genes conductores, nos permite establecer un panorama general sobre las vías y procesos de señalización que se ven afectados o desregulados. El impacto que tiene la acumulación de mutaciones de nucleótido único muestra un perfil mutacional basado en la sustitución de una sola base que discrimina entre transiciones y transversiones; permitiendo establecer una posible etiología molecular. En este estudio se identificaron dos firmas compartidas entre los tres grupos S01 y S12 caracterizadas por la transición de pirimidinas, la primera es una firma asociada a la edad ya que se ve un incremento en la tasa de mutación con respecto a la edad. Estas mutaciones son producidas por la desaminación de la 5-metilcitosina, lo que produce un cambio de C > T. Esta firma se encuentra activa en todos tejidos sanos del cuerpo y en todos los tipos de cáncer (Nik-Zainal et al., 2012; Alexandrov et al., 2013); En contraste la S12 es de etiología desconocida, se caracteriza por un cambio de T>C y se presenta en su mayoría en el cáncer de hepatocelular. Estas dos firmas fueron las únicas identificadas para el grupo SCMOH, debido a que los genes con variantes patogénicas en este grupo no desarrollan una firma mutacional característica (Swanton et al., 2015; Alexandrov et al., 2015). En el grupo BRCA1/2 se detectó la firma S09 asociada a la deficiencia en la polimerasa  $\eta$ ,

polimerasa de baja fidelidad sin actividad exonucleasa, presente en mayor cantidad durante la fase G2 en comparación a la fase S de la síntesis celular (Narod S. A., & Foulkes., 2004; Kobayashi *et al.*, 2013; Duijf *et al.*, 2019). Posee un rol en síntesis de translesión (mecanismo de reparación del ADN posterior a la replicación) que asegura la duplicación del genoma y la supervivencia, en hipermutación somática durante la especialización de la respuesta inmune de los linfocitos B, y en BER para la síntesis de ADN tras la reparación de entrecruzamiento de cadenas (Chatterjee N. & Walker, 2017; Cruet-Hennequart *et al.*, 2009). Si bien la pol  $\eta$  no participa en RH, se encuentra en mayor cantidad en el momento del ciclo celular donde la recombinación homóloga actúa; aunado a esto la detección de esta firma concuerda con el origen de nuestras ya que se ve implicada en la hipermutación somática y se asocia a linfomas (Alexandrov *et al.*, 2013; COSMIC | SBS - Mutational Signatures, 2020).

Por último, el grupo MMR mostró tres firmas adicionales, S16, S21 y S26. La firma 16 es de etiología desconocida y se asocia a daños en las cadenas no transcritas del ADN y se ha reportado en el cáncer hepatocelular (Alexandrov *et al.*, 2013; Nik-Zainal *et al.* 2016 COSMIC | SBS - Mutational Signatures, 2020). No obstante, se caracteriza por una transición de pirimidinas T>C, el cambio más prevalente en nuestros resultados. Las firmas S21 y S26 están relacionadas a la deficiencia en el mecanismo de reparación de bases mal apareadas e inestabilidad microsatélite. Se caracterizan por la transición de pirimidinas y se ha reportado una asociación entre ambas firmas (Alexandrov *et al.*, 2013; Nik-Zainal *et al.* 2016). La distribución de la S21 y S26 en los tejidos concuerda con el tipo de cáncer que tiene nuestra población de estudio (SCMOH) ya que estas firmas se distribuyen en mama, colon, ovario, entre otros (COSMIC | SBS - Mutational Signatures, 2020).

Los hallazgos de este estudio nos permiten vislumbrar la complejidad e interconectividad de vías de señalización, observando el impacto a nivel clínico lo que repercute en la calidad de vida de los pacientes, teniendo un incremento en la tasa de mutaciones somáticas en mujeres con variantes patogénicas en *BRCA1/2* como respuesta a esta mutación germinal, aunado a esto, los resultados obtenidos del análisis de tejido normal nos permitieron identificar variantes patogénicas adquiridas y la identificación de firmas mutagénicas. Lo que posibilita un acercamiento a una herramienta de prevención, detección temprana y mejora de tratamiento para mujeres con síndrome de cáncer de mama y ovario hereditario.

## Conclusiones

- Las mujeres con SCMOH con variantes patogénicas en genes que participan en los mecanismos de recombinación homóloga y bases mal apareadas tienen una mayor tasa de mutaciones somáticas en células sanguíneas que controles sin mutaciones en estos genes, por lo que nuestros datos sugieren que variantes patogénicas germinales en genes responsables de reparación de ADN favorecen la acumulación de mutaciones somáticas en las células sanguíneas.
- Las pacientes que presentan variantes patogénicas germinales en *BRCA1/2* muestran una mayor tasa de mutaciones somáticas en las células sanguíneas.
- Las variantes germinales en MMR y RH favorecen la acumulación de transiciones (A>G/T>C)
- El patrón de firmas mutacionales sugiere una asociación con las variantes patogénicas germinales.

## **Perspectivas**

- Incrementar el número de muestras con mutaciones en genes responsables de MMR para tener mayor precisión estadística.
- Obtener el historial clínico completo de todos los participantes en el estudio para correlacionar el estadio clínico, terapia oncológica, subtipo molecular y subtipo histológico con la cantidad de mutaciones adquiridas.
- Para determinar con mayor precisión la composición de las firmas mutacionales y el impacto de las mutaciones somáticas utilizar secuenciación de exoma completo o genoma completo.

## LITERATURA CITADA

- Alexandrov L.B., Nik-Zainal S., Wedge D.C., Aparicio S. A., Behjati S., Biankin A. V., Bignell G. R., Bolli N., Borg A., Børresen-Dale A. L., Boyault S., Burkhardt B., Butler A. P., Caldas C., Davies H. R., Desmedt C., Eils R., Eyfjörd J. E., Foekens J. A., Greaves M., Stratton M. R. 2013. Firmas de procesos mutacionales en el cáncer humano. *Nat.* 500(7463), 415- 421.
- Alexandrov L. B., Nik-Zainal S., Wedge D. C., Campbell P. J., & Stratton, M. R. 2013. Deciphering signatures of mutational processes operative in human cancer. *Cell reports*, 3(1), 246–259. <https://doi.org/10.1016/j.celrep.2012.12.008>
- Alexandrov, L. B., Jones, P. H., Wedge, D. C., Sale, J. E., Campbell, P. J., Nik-Zainal, S. & Stratton, M. R. (2015). Clock-like mutational processes in human somatic cells. *Nature Genetics*, 47(12), 1402-1407. <https://doi.org/10.1038/ng.3441>
- Alexandrov, L. B., Kim, J., Haradhvala, N. J., Huang, M. N., Tian Ng, A. W., Wu, Y., Boot, A., Covington, K. R., Gordenin, D. A., Bergstrom, E. N., Islam S. M. A., Lopez-Bigas, N., Klimczak, L. J., McPherson, J. R., Morganella, S., Sabarinathan, R., Wheeler, D. A., Mustonen, V., Getz, G., . . . Stratton, M. R. (2020). The repertoire of mutational signatures in human cancer. *Nature*, 578(7793), 94-101. <https://doi.org/10.1038/s41586-020-1943-3>
- Benjamin D., Sato T., Lichtenstein L. & Shand M.(2021). Notes on Mutect 2. Broad Institute
- Bellantuono I. (2004). Haemopoietic stem cells. *The International Journal of Biochemistry & Cell Biology*, 36(4), 607–620. <https://doi.org/10.1016/j.biocel.2003.10.008>
- Broad Institute. (2019). GATK. GATK. Recuperado 3 de diciembre de 2022, de <https://gatk.broadinstitute.org/hc/en-us>
- Broad Institute, (2021). Mutect2 - somatic variant calling with/without matched normal sample. <https://gatk.broadinstitute.org/hc/en-us/community/posts/360057810051-Mutect2-somatic-variant-calling-with-without-matched-normal-sample>
- Bray F., Laversanne M., Weiderpass E. & Soerjomataram I. (2021). The ever-increasing importance of cancer as a leading cause of premature death worldwide. *Cancer*. In press.
- Chatterjee, N., & Walker, G. C. (2017). Mechanisms of DNA damage, repair, and mutagenesis. *Environmental and molecular mutagenesis*, 58(5), 235–263. <https://doi.org/10.1002/em.22087>
- Cibulskis K., Lawrence M., Carter S., Sivachenko A., Jaffe D., Sougnez C., Stacey G., Meyerson M., Lander E. S. & Getz G.(2013). Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* 31, 213–219

- COSMIC | SBS - Mutational Signatures. (2020, 7 abril). COSMIC. Recuperado 20 de octubre de 2022, de <https://cancer.sanger.ac.uk/signatures/sbs/>
- Cruet-Hennequart, S., Gallagher, K., Sokòl, A. M., Villalan, S., Prendergast, I. M. & Carty, M. P. (2009). DNA Polymerase  $\eta$ , a Key Protein in Translesion Synthesis in Human Cells. *Subcellular Biochemistry*, 189-209. [https://doi.org/10.1007/978-90-481-3471-7\\_10](https://doi.org/10.1007/978-90-481-3471-7_10)
- Deans, A. J., & West, S. C. (2011). DNA interstrand crosslink repair and cancer. *Nature reviews. Cancer*, 11(7), 467–480. <https://doi.org/10.1038/nrc3088>
- Delhomme, T. M., Avogbe, P. H., Gabriel, A. A. G., Alcalá, N., Leblay, N., Voegele, C., Vallée, M., Chopard, P., Chabrier, A., Abedi-Ardekani, B., Gaborieau, V., Holcatova, I., Janout, V., Foretová, L., Milosavljevic, S., Zaridze, D., Mukeriya, A., Brambilla, E., Brennan, P., . . . Foll, M. (2020). Needlestack: an ultra-sensitive variant caller for multi-sample next generation sequencing data. *NAR Genomics and Bioinformatics*, 2(2). <https://doi.org/10.1093/nargab/lqaa021>
- Dou Y., Gold H. D., Luquette J.L., Park J. P. 2018. Detecting somatic Mutations in normal cells. *CP 34*, 545-557
- Dorling, L., Carvalho, S., Allen, J., González-Neira, A., Luccarini, C., Wahlström, C., Pooley, K. A., Parsons, M. T., Fortunato, C., Wang, Q., Bolla, M. K., Dennis, J., Keeman, R., Alonso, M. R., Álvarez, N., Herraes, B., Fernandez, V., Núñez-Torres, R., Osorio, A., Easton, D. F. 2021. Breast Cancer Risk Genes — Association Analysis in More than 113,000 Women. *New England Journal of Medicine*, 384(5), 428–439. <https://doi.org/10.1056/nejmoa1913948>
- Daly MB, Pal T, Berry MP, Buys SS, Dickson P, Domchek SM, Elkhanany A, Friedman S, Goggins M, Hutton ML; CGC, Karlan BY, Khan S, Klein C, Kohlmann W; CGC, Kurian AW, Laronga C, Litton JK, Mak JS; LCGC, Menendez CS, Merajver SD, Norquist BS, Offit K, Pederson HJ, Reiser G; CGC, Senter-Jamieson L; CGC, Shannon KM, Shatsky R, Visvanathan K, Weitzel JN, Wick MJ, Wisinski KB, Yurgelun MB, Darlow SD, Dwyer MA. Genetic/Familial High-Risk Assessment: Breast, Ovarian, and Pancreatic, Version 2.2021.2021. *NCCN Clinical Practice Guidelines in Oncology. J Natl Compr Canc Netw*. 2021 Jan 6;19(1):77-102. doi: 10.6004/jnccn.2021.0001. PMID: 33406487.
- Duijf, P. H., Nanayakkara, D., Nones, K., Srihari, S., Kalimutho, M. & Khanna, K. K. (2019). Mechanisms of Genomic Instability in Breast Cancer. *Trends in Molecular Medicine*, 25(7), 595-611. <https://doi.org/10.1016/j.molmed.2019.04.004>
- Fang, L. T., Afshar, P. T., Chhibber, A., Mohiyuddin, M., Fan, Y., Mu, J. C., Gibeling, G., Barr, S., Asadi, N. B., Gerstein, M. B., Koboldt, D. C., Wang, W., Wong, W. H. & Lam, H. Y. (2015). An ensemble approach to accurately detect somatic mutations using SomaticSeq. *Genome Biology*, 16(1). <https://doi.org/10.1186/s13059-015-0758-2>

Ferrer Márquez, M., Reina Duarte, N., Maturana Ibáñez, V., Belda Lozano, R., Rubio Gil, F., Blesa Sierra, I. & Rico Morales, M. D. M. (2011, enero). Síndrome de Lynch: genética y cirugía. *Cirugía Española*, 89(1), 3-9. <https://doi.org/10.1016/j.ciresp.2010.06.010>

Foulkes, W. D. (2008). Inherited Susceptibility to Common Cancers. *New England Journal of Medicine*, 359(20), 2143-2153. <https://doi.org/10.1056/nejmra0802968>

Gerstung M., Jolly C., Leshchiner I., Dentro S. C., Gonzalez S., Rosebrock D., Mitchell T. J., Rubanova Y., Anur P., Yu K., Tarabichi.M., Deshwar A., Wintersinger J., Kleinheinz K., Vázquez-García I., Haase K., Jerman L., Sengupta S., Macintyre G., Malikic S., Donmez ., Livitz D. G., Cmero M., Demeulemeester J., Schumacher S., Fan Y., Yao X., Lee J., Schlesner M., Boutros P.C., Bowtell D.D., Zhu H., Getz G., Imielinski M., Beroukhi R., Sahinalp S. C., Ji Y., Peifer M., Markowitz F., Mustonen V., Yuan K., Wang W., Morris Q.D., PCAWG Evolution & Heterogeneity Working Group, Spellman P.T., Wedge D.C., Van Loo P. & PCAWG Consortium. 2017. The evolutionary history of 2,658 cancers. *Nat.* 578:122-128

14

Genovese G. Anna K. Kähler, Robert E. Handsaker, Johan Lindberg, Samuel A. Rose, Samuel F. Bakhom, Kimberly Chambert, Eran Mick, Benjamin M. Neale, Menachem Fromer, Shaun M. Purcell, Oscar Svantesson, Mikael Landén, Martin Höglund, Sören Lehmann, Stacey B. Gabriel, Jennifer L. Moran, Eric S. Lander, Patrick F. Sullivan, Pamela Sklar, Henrik Grönberg, Christina M. Hultman, Steven A. McCarroll. 2014. Clonal Hematopoiesis and Blood-Cancer Risk Inferred from Blood DNA Sequence. *N. E. J. of Med.* 371;26.

Global Cancer Observatory. 2018. Estimated number of new cases in 2020, worldwide, both sexes, all ages

Hanahan D. & Weinberg R. A. 2011. Hallmarks of Cancer: The Next Generation. *C.* 10.1016/j.cell.2011.02.013

Hanahan D. 2022. Hallmarks of Cancer: New Dimensions. *Cancer Discov* 1 January 2022; 12 (1): 31–46

Hassanpour H. & Dehghani S. M. 2017. Review of cancer from perspective of molecular. *J. C.R. P.* 10.1016/j.jcrpr.2017.07.001

Harbeck N., Penault-Llorca F., Cortes J., Gnant M., Houssami N., Poortmans P., Ruddy K., Tsang J. & Cardoso F. 2019. Breast cancer. *Nat.* 5: 66

Hu, C., Hart, S. N., Gnanaolivu, R., Huang, H., Lee, K. Y., Na, J., Gao, C., Lilyquist, J., Yadav, S., Boddicker, N. J., Samara, R., Klebba, J., Ambrosone, C. B., Anton-Culver, H., Auer, P., Bandera, E. V., Bernstein, L., Bertrand, K. A., Burnside, E. S., Couch, F. J. 2021. A Population-Based Study of Genes Previously Implicated in Breast Cancer. *New England Journal of Medicine*, 384(5), 440–451. <https://doi.org/10.1056/nejmoa2005936>

- Koboldt, D. C. (2020). Best practices for variant calling in clinical sequencing. *Genome Medicine*, 12(1). <https://doi.org/10.1186/s13073-020-00791-w>
- Kobayashi H., Ohno S., Sasaki Y. & Matsuura M. 2013. Hereditary breast and ovarian cancer susceptibility genes (Review). *Onc. Rep.* 30: 1019-1029
- Knudson, A.G., Jr. (1971). Mutation and cancer: Statistical study of retinoblastoma. *Proc. Natl. Acad. Sci. USA*, 68, 820–823.
- Krejci, L., Altmannova, V., Spirek, M. & Zhao, X. (2012). Homologous recombination and its regulation. *Nucleic Acids Research*, 40(13), 5795-5818. <https://doi.org/10.1093/nar/gks270>
- Kim, H., & D'Andrea, A. D. (2012). Regulation of DNA cross-link repair by the Fanconi anemia/BRCA pathway. *Genes & development*, 26(13), 1393–1408. <https://doi.org/10.1101/gad.195248.112>
- Lai, A. (2014). GitHub - AstraZeneca-NGS/VarDict: VarDict. GitHub. Recuperado 3 de diciembre de 2022, de <https://github.com/AstraZeneca-NGS/VarDict>
- Leung, C. C. Y. & Glover, J. M. (2011). BRCT domains. *Cell Cycle*, 10(15), 2461-2470. <https://doi.org/10.4161/cc.10.15.16312>
- Lundgren, S., Keränen, M., Kankainen, M., Huuhtanen, J., Walldin, G., Kerr, C. M., Clemente, M., Ebeling, F., Rajala, H., Brück, O., Lähdesmäki, H., Hannula, S., Hannunen, T., Ellonen, P., Young, N. S., Ogawa, S., Maciejewski, J. P., Hellström-Lindberg, E., & Mustjoki, S. 2021. Somatic mutations in lymphocytes in patients with immune-mediated aplastic anemia. *Leukemia*, 35(5), 1365–1379. <https://doi.org/10.1038/s41375-021-01231-3> Malik SS, Masood N, Asif M, Ahmed P, Shah ZU, Khan JS.2019. Expressional analysis of MLH1 and MSH2 in breast cancer. *Curr Probl Cancer*. 43 (2): 97-105
- Martincorena I.& Campbell P, J.2015. Somatic mutation in cancer and normal cells. *Science*. 349. 6255
- Mahdavi M., Nassiri M., Kooshyar M.M., M Vakili-Azghandi, Avan A., Sandry R., Pillai S., Yin Lam A.K., Gopalan V. 2018. Hereditary breast cancer; Genetic penetrance and current status with BRCA. *J.Cell Physiol.* 234,5741–5750
- Maley, C., Aktipis, A., Graham, T., Sottoriva A., Boddy A., Janiszewska M., Ariosto S., Gerlinger M., Yuan Y., Pienta K., Anderson K., Gatenby R., Swanton C., Posada D., Wu C., Schiffman J., Hwang E., Polyak K., Anderson A., Brown J., Greaves M. and Shibata D. 2017. Classifying the evolutionary and ecological features of neoplasms. *Nat Rev Cancer* 17, 605–619.
- Makki J. 2015. Diversity of Breast Carcinoma: Histological Subtypes and Clinical Relevance. *Clinical medicine insights. Pathology*, 8, 23–31. <https://doi.org/10.4137/CPath.S31563>

- McPherson, K., Steel, C. M., & Dixon, J. M. 2000. ABC of breast diseases. Breast cancer-epidemiology, risk factors, and genetics. *BMJ (Clinical research ed.)*, 321(7261), 624–628. <https://doi.org/10.1136/bmj.321.7261.624>
- McCulloch, S. D., & Kunkel, T. A. 2008. The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases. *Cell research*, 18(1), 148–161. <https://doi.org/10.1038/cr.2008.4>
- McCall, C. M., Mosier, S., Thiess, M., Debeljak, M., Pallavajjala, A., Beierl, K., Deak, K. L., Datto, M. B., Gocke, C. D., Lin, M. T. & Eshleman, J. R. (2014). False Positives in Multiplex PCR-Based Next-Generation Sequencing Have Unique Signatures. *The Journal of Molecular Diagnostics*, 16(5), 541-549. <https://doi.org/10.1016/j.jmoldx.2014.06.001>
- Modrich P. & Lahue R. 1996. Mismatch repair in replication fidelity, genetic recombination, and cancer biology. *Biochem.*, 65, 101–133
- Narod S. A., & Foulkes, W. D. 2004. BRCA1 and BRCA2: 1994 and beyond. *Nature Reviews Cancer*, 4(9), 665–676. <https://doi.org/10.1038/nrc1431>
- Negrini S., Gorgoulis V. & Halazonetis, T. 2010. Genomic instability — an evolving hallmark of cancer. *Nat Rev Mol Cell Biol* 11, 220–228
- Nik-Zainal S., Alexandrov L. B., Wedge D.C., Van Loo P., Greenman C., Raine K., Jones D., Hinton J., Marshall J., Stebbings L., Menzies A., Martin S., Leung K., Chen L., Leroy C., Ramakrishna M., Rance R., Lau K., Mudie L., Varela I., McBride D., Bignell G., Cooke S., Shlien A., Gamble J., Whitmore I., Maddison M., Tarpey P., Davies H., Papaemmanuil E., Stephens P., McLaren S., Butler A., Teague J., Jönsson G., Garber J., Silver D., Miron P., Fatima A., Boyault S., Langerød A., Tutt A., Martens J., Aparicio S., Borg K., Salomon A., Thomas G., Børresen-Dale A.L., Richardson A., Neuberger M., Futreal P., Campbell P., Stratton M. 2012. Mutational Processes Molding the Genomes of 21 Breast Cancers. *Elsev. Cell*. 149, 979–993.
- Nik-Zainal S, Davies H, Staaf J, Ramakrishna M, Glodzik D, Zou X, Martincorena I, Alexandrov LB, Martin S, Wedge DC, Van Loo P, Ju YS, Smid M, Brinkman AB, Morganella S, Aure MR, Lingjærde OC, Langerød A, Ringnér M, Ahn SM, Boyault S, Brock JE, Broeks A, Butler A, Desmedt C, Dirix L, Dronov S, Fatima A, Foekens JA, Gerstung M, Hooijer GK, Jang SJ, Jones DR, Kim HY, King TA, Krishnamurthy S, Lee HJ, Lee JY, Li Y, McLaren S, Menzies A, Mustonen V, O'Meara S, Pauporté I, Pivot X, Purdie CA, Raine K, Ramakrishnan K, Rodríguez-González FG, Romieu G, Sieuwerts AM, Simpson PT, Shepherd R, Stebbings L, Stefansson OA, Teague J, Tommasi S, Treilleux I, Van den Eynden GG, Vermeulen P, Vincent-Salomon A, Yates L, Caldas C, van't Veer L, Tutt A, Knappskog S, Tan BK, Jonkers J, Borg Å, Ueno NT, Sotiriou C, Viari A, Futreal PA, Campbell PJ, Span PN, Van Laere S, Lakhani SR, Eyfjord JE, Thompson AM, Birney E, Stunnenberg HG, van de Vijver MJ, Martens JW, Børresen-Dale AL, Richardson AL, Kong G, Thomas G, Stratton MR. 2016. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nat.* 2;534(7605):47-54.

- Nik-Zainal, S. & Morganella, S. (2017). Mutational Signatures in Breast Cancer: The Problem at the DNA Level. *Clinical Cancer Research*, 23(11), 2617-2629. <https://doi.org/10.1158/1078-0432.ccr-16-2810>
- Nielsen, F. C., van Overeem Hansen, T., & Sørensen, C. S. (2016). Hereditary breast and ovarian cancer: new genes in confined pathways. *Nature reviews. Cancer*, 16(9), 599–612. <https://doi.org/10.1038/nrc.2016.72>
- Nowell P. 1976. The Clonal Evolution of Tumor Cell Populations. *Science* 194, 4260, 23-28.
- Organización Mundial de la Salud.(2022)
- Oliver, J., Quezada Urban, R., Franco Cortés, C. A., Díaz Velásquez, C. E., Montealegre Paez, A. L., Pacheco-Orozco, R. A., Castro Rojas, C., García-Robles, R., López Rivera, J. J., Gaitán Chaparro, S., Gómez, A. M., Suarez Obando, F., Giraldo, G., Maya, M. I., Hurtado-Villa, P., Sanchez, A. I., Serrano, N., Orduz Galvis, A. I., Aruachan, S., Vaca Paniagua, F. (2019). Latin American Study of Hereditary Breast and Ovarian Cancer LACAM: A Genomic Epidemiology Approach. *Frontiers in Oncology*, 9. <https://doi.org/10.3389/fonc.2019.01429>
- Petrucci N, Daly MB, Pal T. BRCA1- and BRCA2-Associated Hereditary Breast and Ovarian Cancer. 1998 Sep 4 [Updated 2022 May 26]. In: Adam MP, Mirzaa GM, Pagon RA, et al., editors. *GeneReviews*. Seattle (WA): University of Washington, Seattle; 1993-2022. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK1247/>
- Petljak M. & Alexandrov L.B.2016. Understanding mutagenesis through delineation of mutational signatures in human cancer. *Oxford Un. Press*. 37(6):531-540
- Perou, C. M., Sùrlie T., Eisen M., Rijn M., Jeffreyk S., Rees C., Pollack J., Ross D., Johnsen H., Akslen L.,Flugel, Pergamenschikov A., Williams C., Zhu S., Lunning P., Bùrresen-Dale A., Brown P., Botstein D. 2000. Molecular portraits of human breast tumours. *Nature* 406, 747–752.
- Pećina-Šlaus, N., Kafka, A., Salamon, I., & Bukovac, A. 2020. Mismatch Repair Pathway, Genome Stability and Cancer. *Frontiers in Molecular Biosciences*, 7. <https://doi.org/10.3389/fmolb.2020.00122>
- Polak P., Kim J., Braunstein L.Z., Karlic R., Haradhavala N.J., Tiao G., Rosebrock D., Livitz D., Kübler K., Mouw K.W., Kamburov A., Maruvka J. E., Leshchiner I., Lander E.S., Golub T.R., Zick A., Orthwein A., Lawrence M.S., Batra R.N., Caldas C., Haber D.A., Laird P.W., Shen H., Ellisen L.W., D'Andrea A.D., Chanock S.J., Foulkes W. D., & Getz G.2017. A mutational signature reveals alterations underlying deficient homologous recombination repair in breast cancer. *Nat. Genetics*.49(10)

- Poplin R, Ruano-Rubio V, DePristo MA, Fennell TJ, Carneiro MO, Van der Auwera GA, Kling DE, Gauthier LD, Levy-Moonshine A, Roazen D, Shakir K, Thibault J, Chandran S, Whelan C, Lek M, Gabriel S, Daly MJ, Neale B, MacArthur DG, Banks E. (2017). Scaling accurate genetic variant discovery to tens of thousands of samples bioRxiv, 201178. DOI: 10.1101/201178
- Quezada Urban R, Díaz Velásquez CE, Gitler R, Rojo Castillo MP, Sirota Toporek M, Figueroa Morales A, Moreno García O, García Esquivel L, Torres Mejía G, Dean M, Delgado Enciso I, Ochoa Díaz López H, Rodríguez León F, Jan V, Garzón Barrientos VH, Ruiz Flores P, Espino Silva PK, Haro Santa Cruz J, Martínez Gregorio H, Rojas Jiménez EA, Romero Cruz LE, Méndez Catalá CF, Álvarez Gómez RM, Fragoso Ontiveros V, Herrera LA, Romieu I, Terrazas LI, Chirino YI, Frecha C, Oliver J, Perdomo S, Vaca Paniagua F. 2018. Comprehensive Analysis of Germline Variants in Mexican Patients with Hereditary Breast and Ovarian Cancer Susceptibility. *Cancers (Basel)*. 27;10(10):36
- Riaz N., Blecua P., Lim R.S. , Shen R. ,Higginson D.S., Weinhold N., Norton L., Weigelt B., Powell S.N. y Reis-Filho J.S. .2017. Pan-cancer analysis of bi-allelic alterations in homologous recombination DNA repair genes. *Nat Commun* 8, 857
- Rial Boubeta A. & Valera Mallou J. 2008. Estadística práctica para la investigación en ciencias de la salud.España. Netbiblo
- Roberts, M. E., Jackson, S. A., Susswein, L. R., Zeinomar, N., Ma, X., Marshall, M. L., Stettner, A. R., Milewski, B., Xu, Z., Solomon, B. D., Terry, M. B., Hruska, K. S., Klein, R. T., & Chung, W. K. 2018. MSH6 and PMS2 germ-line pathogenic variants implicated in Lynch syndrome are associated with breast cancer. *Genetics in Medicine*, 20(10), 1167–1174. <https://doi.org/10.1038/gim.2017.254>
- Robinson, P. S., Thomas, L. E., Abascal, F., Jung, H., Harvey, L., West, H. D., Olafsson, S., Lee, B., Coorens, T., Lee-Six, H., Butlin, L., Lander, N., Truscott, R., Sanders, M. A., Lensing, S. V., Buczacki, S., Ten Hoopen, R., Coleman, N., Brunton-Sim, R., Rushbrook, S. Stratton, M. R. 2022. Inherited MUTYH mutations cause elevated somatic mutation rates and distinctive mutational signatures in normal human cells. *Nature communications*, 13(1), 3949. <https://doi.org/10.1038/s41467-022-31341-0>
- Stefanski, C. D., Keffler, K., McClintock, S., Milac, L., & Prospero, J. R. (2019). APC loss affects DNA damage repair causing doxorubicin resistance in breast cancer cells. *Neoplasia (New York, N.Y.)*, 21(12), 1143–1150. <https://doi.org/10.1016/j.neo.2019.09.002>
- Sharma S., Kelly T. K., & Jones P. A.2010.Epigenetics in cancer. *Carcinogenesis*,31(1), 27–36.
- Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*, 71(3), 209–249. <https://doi.org/10.3322/caac.21660>

- Slavin, T. P., Maxwell, K. N., Lilyquist, J., Vijai, J., Neuhausen, S. L., Hart, S. N., Ravichandran, V., Thomas, T., Maria, A., Villano, D., Schrader, K. A., Moore, R., Hu, C., Wubbenhorst, B., Wenz, B. M., D'Andrea, K., Robson, M. E., Peterlongo, P., Bonanni, B., Ford, J. M. & Couch, F. J. 2017. The contribution of pathogenic variants in breast cancer susceptibility genes to familial breast cancer risk. *NPJ breast cancer*, 3, 22. <https://doi.org/10.1038/s41523-017-0024-8>
- Swanton, C., McGranahan, N., Starrett, G. J., & Harris, R. S. (2015). APOBEC Enzymes: Mutagenic Fuel for Cancer Evolution and Heterogeneity. *Cancer discovery*, 5(7), 704–712. <https://doi.org/10.1158/2159-8290.CD-15-0344>
- van der Groep P., Bouter, A., van der Zanden R., Siccama I., Menko FH, Gille JJ, van Kalken C., van der Wall, E., Verheijen, RH y van Diest, PJ. 2006. Distinción entre cáncer de mama hereditario y esporádico en base a datos clínico-patológicos. *Revista de pat. Clín.* 59 (6), 611-617.
- VanKlompberg, M. K., Bedalov, C. O., Soto, K. F. & Prosperi, J. R. (2015). APC selectively mediates response to chemotherapeutic agents in breast cancer. *BMC Cancer*, 15(1). <https://doi.org/10.1186/s12885-015-1456-x>
- Vali-Pour, M., Lehner, B., & Supek, F. 2022. The impact of rare germline variants on human somatic mutation processes. *Nature communications*, 13(1), 3724. <https://doi.org/10.1038/s41467-022-31483-1>
- Venkitaraman, A. R. 2001. Functions of BRCA1 and BRCA2 in the biological response to DNA damage. *Journal of Cell Science*, 114(20), 3591–3598. <https://doi.org/10.1242/jcs.114.20.3591>
- Vogelstein, B., Papadopoulos, N., Velculescu, V. E., Zhou, S., Diaz, L. A. & Kinzler, K. W. (2013). Cancer Genome Landscapes. *Science*, 339(6127), 1546-1558. <https://doi.org/10.1126/science.1235122>
- wANNOVAR. (s. f.). Recuperado 23 de octubre de 2022, de <https://wannovar.wglab.org/>
- Wolters, S., & Schumacher, B. 2013. Genome maintenance and transcription integrity in aging and disease. *Frontiers in genetics*, 4, 19. <https://doi.org/10.3389/fgene.2013.00019>
- Wei, L., Christensen, S. R., Fitzgerald, M. E., Graham, J., Hutson, N. D., Zhang, C., Huang, Z., Hu, Q., Zhan, F., Xie, J., Zhang, J., Liu, S., Remenyik, E., Gellen, E., Colegio, O. R., Bax, M., Xu, J., Lin, H., Huss, W. J., Foster, B. A., Paragh, G. 2021. Ultradeep sequencing differentiates patterns of skin clonal mutations associated with sun-exposure status and skin cancer burden. *Science advances*, 7(1), eabd7703. <https://doi.org/10.1126/sciadv.abd7703>
- Xie M., Charles Lu, Jiayin Wang, Michael D McLellan, Kimberly J Johnson, Michael C Wendl, Joshua F McMichael, Heather K Schmidt, Venkata Yellapantula, Christopher A Miller, Bradley A Ozenberger, John S Welch, Daniel C Link, Matthew J Walter, Elaine R Mardis, John F Dipersio, Feng Chen, Richard K Wilson, Timothy J Ley, Li Ding. 2014. Age-

related mutations associated with clonal hematopoietic expansion and malignancies. *Nat. med.* 20: 1472-1478

Yoshida, K., Gowers, K. H. C., Lee-Six, H., Chandrasekharan, D. P., Coorens, T., Maughan, E. F., Beal, K., Menzies, A., Millar, F. R., Anderson, E., Clarke, S. E., Pennycuik, A., Thakrar, R. M., Butler, C. R., Kakiuchi, N., Hirano, T., Hynds, R. E., Stratton, M. R., Martincorena, I., . . . Campbell, P. J. (2020). Tobacco smoking and somatic mutations in human bronchial epithelium. *Nature*, 578(7794), 266-272. <https://doi.org/10.1038/s41586-020-1961-1>

ANEXO

PANELES MULTIGÉNICOS: 143 genes y 74 genes

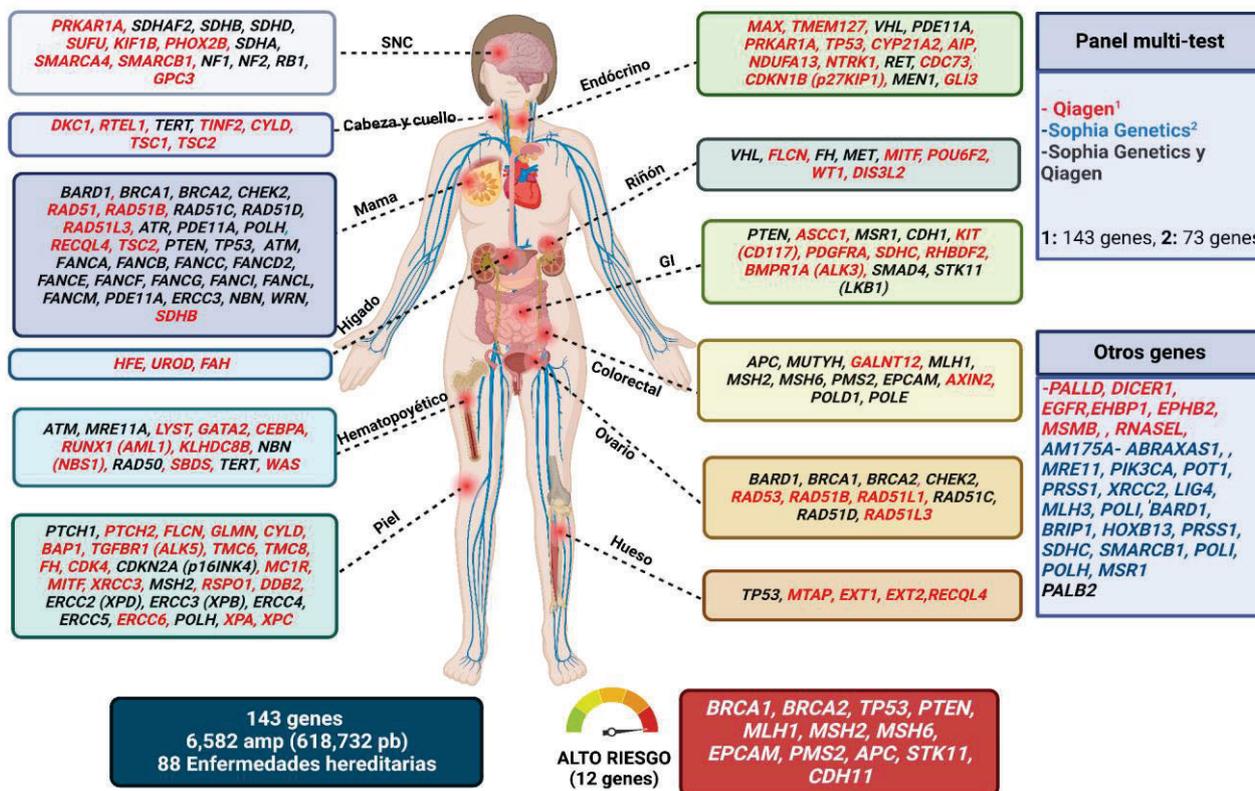
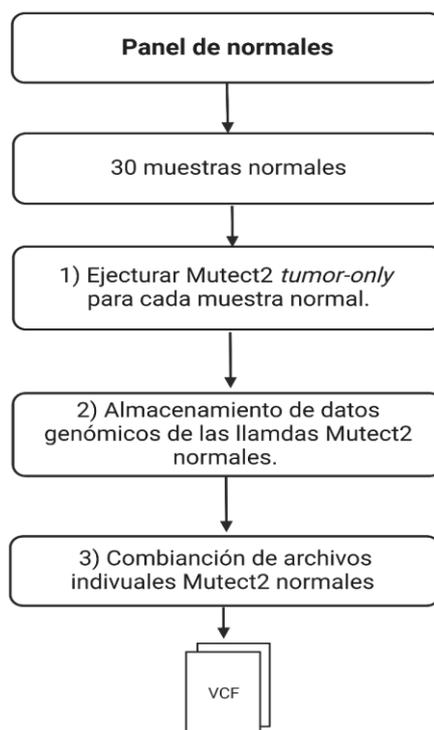
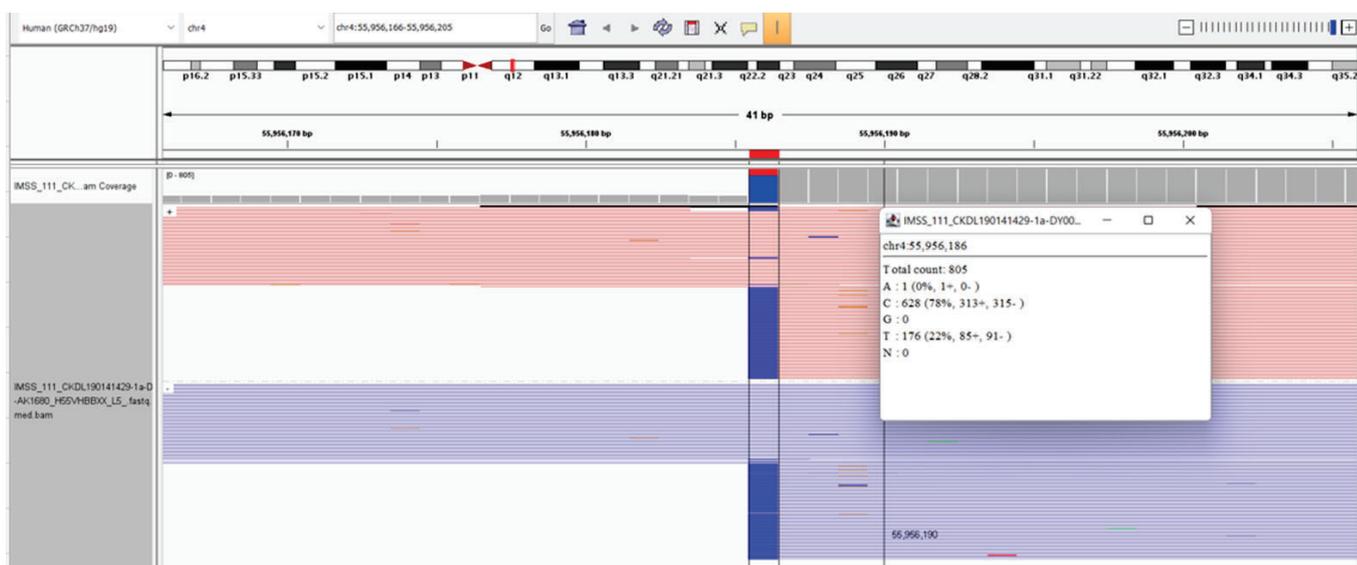


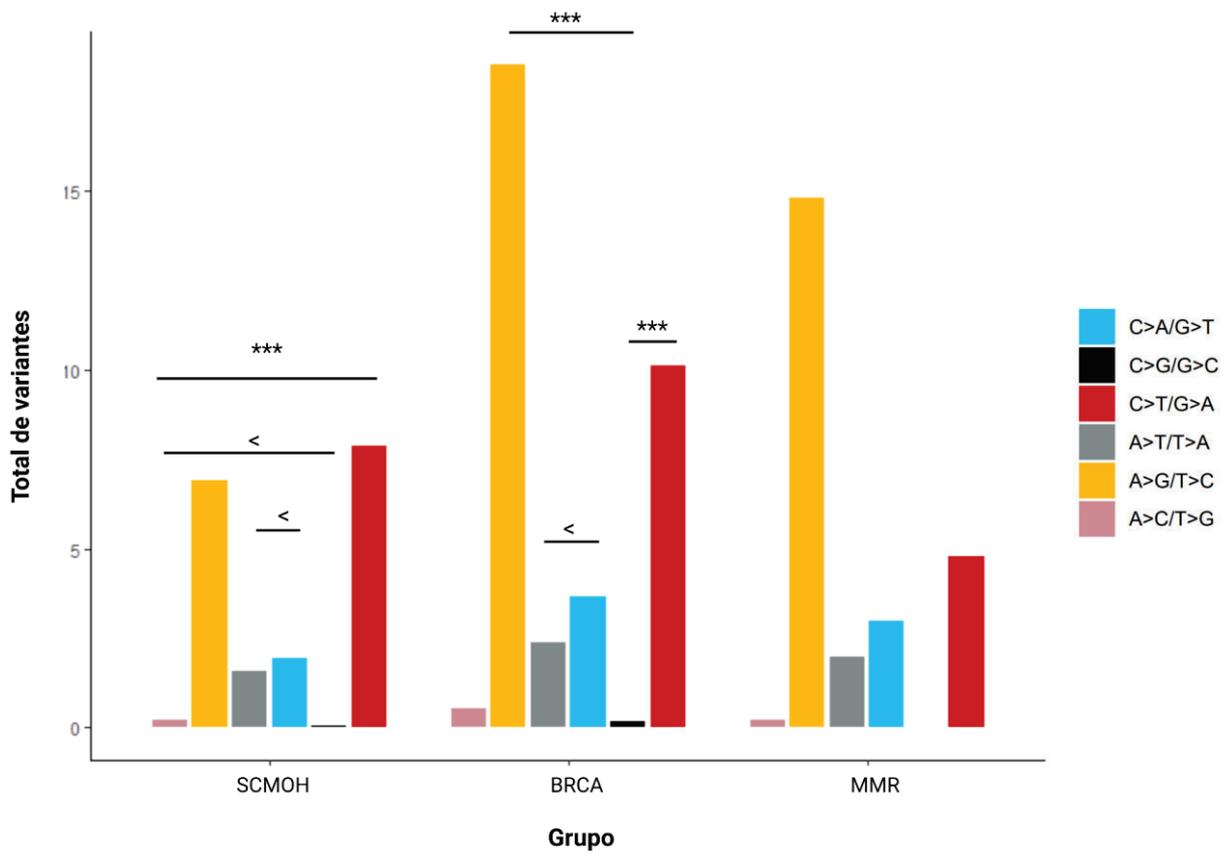
Figura Suplementaria 1. Panel de 143 genes y 74 genes susceptibles a cáncer hereditario. Relación de los genes con la distribución corporal.



**Figura Suplementaria 2. Flujo de trabajo para generar el Panel de Normales.** Procesamiento de muestras normales para generar una matriz de datos genómicos.



**Figura Suplementaria 3. Artefactos en extremos de amplicones.** Sin la delimitación de regiones Mutect 2 busca variantes en todas las regiones que estén presentes en el archivo.



**Figura suplementaria 4. Conteo de tipo de cambio único de nucleótido con respecto a los tres grupos de estudio.** La diferencia significativa ( $p < 0.05$ ) se marca con  $<0.05^*$ ,  $<0.01^{**}$  y  $<0.001^{***}$  los grupos que no muestran diferencia significativa entre ellos se señalan con “<”.