

BIOLOGÍA

TESIS

*IDENTIFICACIÓN DE REARREGLOS ESTRUCTURALES EN GENES
DE SUSCEPTIBILIDAD A CÁNCER DE MAMA Y OVARIO EN 500
PACIENTES LATINOAMERICANAS CON SHCMO*

QUE PARA OBTENER EL TÍTULO DE:

BIÓLOGA

PRESENTA:

JIMENA PAREDES DE LA VEGA

TUTORA: DRA. CLARA ESTELA DÍAZ VELÁSQUEZ

COMITÉ TUTOR:

DR. FELIPE VACA PANIAGUA

DR. ALDO HUGO DE LA CRUZ MONTOYA

DR. LUIS ENRIQUE ARIAS ROMERO

DR. LUIS IGNACIO TERRAZAS VALDÉZ





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.

DEDICATORIA:

*Dedico este trabajo a todos
aquellos que creyeron en mí. Gracias.*

*A mi tía Laura, quien lamentablemente falleció por
cáncer de mama en 2022.*

A mi mamá, Blanca, gracias por todo.

AGRADECIMIENTOS INSTITUCIONALES:

A la Universidad Nacional Autónoma de México por todas las oportunidades brindadas, por abrirme las puertas de la máxima casa de estudios para que pudiera desarrollarme como estudiante, como científica y como persona. Agradezco que en este tiempo he podido desenvolverme haciendo lo que me apasiona y me siento muy orgullosa de ser una puma de la UNAM. Así mismo, quiero agradecer a la Dirección General de Cooperación e Internacionalización (DGEI) de la UNAM, por haberme otorgado una beca de movilidad estudiantil en el extranjero a la universidad McGill en Canadá y a la Universidad de Málaga en España. Esas experiencias me cambiaron la vida y me abrieron las puertas del mundo.

El presente proyecto se realizó gracias a los fondos y el apoyo proporcionados por: El Consejo Nacional de Ciencia y Tecnología (CONACyT) Fondo Sectorial 272573, Fondo SEP CONACyT 285879, CONACyT Laboratorios Nacionales 2021 proyecto 315817, UNAM PAPIIT IN219217, UNAM PAPIIT IN225920 y recursos institucionales del Instituto Nacional de Cancerología. Agradezco a las instituciones de salud participantes tanto de México, como de Guatemala, Colombia y Perú por facilitarnos las muestras y demás datos e información cruciales para el desarrollo de este proyecto. Agradezco también a los integrantes del consorcio LACAM que hicieron posible este proyecto.

Este trabajo se desarrolló en el Laboratorio 13 de la Unidad de Biomedicina de la FES Iztacala de la UNAM en conjunto con el Laboratorio Nacional en Salud: Diagnóstico molecular y efecto ambiental en enfermedades Crónico-Degenerativas, también en la FES Iztacala, UNAM.

Agradezco especialmente al Dr. Felipe Vaca Paniagua por darme la oportunidad de participar en los proyectos de investigación del Laboratorio 13 de la Unidad de Biomedicina de la FES Iztacala desde mi primer semestre. También le agradezco por creer en el talento y la capacidad de los jóvenes y siempre apoyarnos a lograr nuestros objetivos. Estoy muy orgullosa de lo que he logrado y del equipo que hemos conformado. Agradezco que siempre me enseñara a mantener las expectativas y estándares de mí misma muy en alto y que ningún reto es demasiado para mí sin importar mi edad o experiencia. Quisiera que sepa que aprendí mucho con usted y con el equipo, mucho más de lo que se podría imaginar. A la Dra. Clara E. Díaz Velásquez, por siempre creer en mí y en mi capacidad. Le agradezco toda su paciencia y el tiempo que dedicó a enseñarme nuevas técnicas y métodos. De usted aprendí a trabajar con orden y rigor en un laboratorio tan formal como el Laboratorio Nacional en Salud.

Al Dr. Oliver, María y Juan Lu quiénes me ayudaron y enseñaron mucho durante mi estancia en la Universidad de Málaga. Fue una de las experiencias más enriquecedoras de mi formación académica y no habría sido posible sin su ayuda.

A todas las mujeres de América Latina que accedieron a ser partícipes del estudio, sin ustedes, nada de esto habría sido posible. De corazón espero que el cáncer no sea más que un obstáculo en sus vidas y a aquellas que ya no están, que en paz descansen. Su contribución a la ciencia permanecerá por siempre.

¡GRACIAS

AGRADECIMIENTOS PERSONALES:

En primer lugar, agradezco a mi madre, Blanca, por siempre apoyarme a cumplir mis metas, por nunca dudar de mi capacidad y siempre estar ahí en todos mis logros y fracasos. Todo lo que hago y he hecho en mi vida es porque quiero que te sientas orgullosa de mí, ese es mi mayor logro y el único. Tu siempre quisiste que volara con mis propias alas y siempre me educaste para proteger mi libertad y ser fuerte. Espero con todo mi corazón, que todo tu esfuerzo y todo lo que has dado por mí, haya valido la pena y estés, de hecho, orgullosa de mí. Te amo.

A mis abuelos, por siempre preocuparse por mí y darme su amor incondicional. A mis tíos Rodolfo, Luz María y Gaby, por abrirme las puertas de sus casas cuando lo necesité y apoyarme siempre. A mi familia en general, les agradezco su apoyo, su amor y su alegría contagiosa, nunca me han dejado sola y sé que siempre puedo contar con ustedes.

A mis amigos, aquellos que conozco desde el jardín de niños, y a aquellos que conocí en el camino, gracias por las aventuras, los momentos alegres y las memorias. A Valeria, César y Vania, pues nos conocimos desde el día 1 y compartimos logros, fracasos, aventuras buenos y malos momentos, gracias por darme su amistad y apoyo.

Al equipo del laboratorio 13 en general, les agradezco sus consejos, su amistad, y todos los buenos momentos, de todos aprendí algo y siempre los llevaré en mi memoria. A Fernando, porque empezamos juntos y contra todo pronóstico, estuvimos a la altura de los más experimentados, te agradezco que hayamos conformado este equipo que, aunque es mayormente disfuncional, siempre logra cada objetivo. A Rosie y Ernesto, los mejores mentores, pacientes y siempre dispuestos a ayudar a los menos experimentados, son un ejemplo a seguir.

A los amigos que conocí en el extranjero, gracias por estar conmigo en una de las mejores experiencias de mi vida y hacer de mi tiempo fuera de mi país, algo inolvidable. Gracias a ustedes mi perspectiva del mundo y la vida es distinta, y aunque yo misma he cambiado, también creo que soy mejor.

¡GRACIAS!

ÍNDICE

ABREVIATURAS	5
ÍNDICE DE TABLAS Y FIGURAS	6
RESUMEN	7
SUMMARY	8
INTRODUCCIÓN:	9
EL CÁNCER	9
FACTORES DE RIESGO	9
<i>Intrínsecos</i>	<i>9</i>
<i>No intrínsecos</i>	<i>10</i>
Endógenos:	10
Exógenos:	10
EL CÁNCER EN EL MUNDO	10
EL CÁNCER EN AMÉRICA LATINA	11
EL CÁNCER DE MAMA	12
FACTORES DE RIESGO DE CÁNCER DE MAMA	13
EL SÍNDROME HEREDITARIO DE CÁNCER DE MAMA Y OVARIO (SHCMO)	15
DIAGNÓSTICO DE SHCMO	15
EL MANEJO CLÍNICO DEL SHCMO	16
LAS MUTACIONES PUNTUALES EN EL SHCMO	16
GENES DE SUSCEPTIBILIDAD A CÁNCER DE MAMA Y OVARIO	17
LOS REARREGLOS ESTRUCTURALES EN EL SHCMO	18
ANTECEDENTES:	19
JUSTIFICACIÓN	22
HIPÓTESIS	24
OBJETIVOS	25
MATERIALES Y MÉTODOS	26
<i>MUESTRAS</i>	<i>26</i>
<i>EXPERIMENTACIÓN</i>	<i>26</i>
Análisis bioinformático:	28
<i>PANELCN.MOPS</i>	<i>30</i>
<i>CONSIDERACIONES DE CONTROL DE CALIDAD DE LOS DATOS</i>	<i>31</i>
<i>FILTRADO</i>	<i>31</i>
RESULTADOS	32
<i>CONTROLES</i>	<i>32</i>
<i>ANÁLISIS GLOBAL</i>	<i>32</i>
<i>GUATEMALA</i>	<i>32</i>
<i>MÉXICO</i>	<i>33</i>
<i>PERÚ</i>	<i>33</i>
<i>COLOMBIA</i>	<i>33</i>
<i>GENERACIÓN DE UN CÓDIGO ÚNICO PARA EL ANÁLISIS AUTOMATIZADO</i>	<i>33</i>
CONCLUSIÓN:	38
PERSPECTIVAS:	39
BIBLIOGRAFÍA:	40

ABREVIATURAS

AL	América Latina
aCGH-Microarrays	Hibridación Genómica Comparativa-Microarreglos (<i>Comparative Genomic Hybridization- Microarrays</i>)
DNA	Ácido Desoxi Ribonucleico (<i>Desoxi Ribonucleic Acid</i>)
<i>BRCA1</i> Ex9-12del	Deleción del exón 9 al 12 del gen <i>BRCA1</i>
HRR	(Homologue Recombination Repair)
LR	Grandes Rearreglos (<i>Large Rearrangements</i>)
MLPA	(<i>Multiplex Ligation-dependant Probe Amplification</i>)
NCCN	<i>National Comprehensive Cancer Network</i>
NGS	Secuenciación de próxima generación (<i>Next Generation Sequencing</i>)
PCR	Reacción en Cadena de la Polimerasa (<i>Polymerase Chain Reaction</i>)
SHCMO	Síndrome Hereditario de Cáncer de Mama y Ovario (<i>Hereditary Breast and Ovarian Cancer Syndrome</i>)
SIR	Radio de Incidencia Estandarizada (<i>Standarized Incidence Ratio</i>)
SNV	Variante de un solo Nucleótido (<i>Single Nucleotide Variant</i>)

*En algunas abreviaturas se utilizan las siglas en inglés por convención internacional.

ÍNDICE DE TABLAS Y FIGURAS

Figura 1. Gráficas de pastel que muestran las estadísticas de incidencia y mortalidad de los tipos de cáncer para ambos sexos (femenino y masculino) y todas las edades en el mundo en el año 2018 (Bray et al., 2018).

Figura 2. Gráficos de pastel que muestran la incidencia y mortalidad de cáncer de mama en América Latina, en el año 2018 (Bray et al., 2018). A) Incidencia de cáncer en ambos sexos, todas las edades. B) Mortalidad por cáncer para ambos sexos, todas las edades. C) Incidencia de cáncer en mujeres de todas las edades. D) Mortalidad por cáncer en mujeres de todas las edades.

Figura 3. Diagrama de círculos concéntricos que en el primer anillo muestra los genes seleccionados agrupados de acuerdo a las vías metabólicas en las que participan. En el segundo anillo se muestran las vías metabólicas normales y en el último anillo se presentan los Hallmarks del cáncer de los que son partícipes cuando están mutados.

Figura 4. Diagrama que muestra cómo se agruparon las muestras y controles para el análisis bioinformático.

Figura 5. Diagrama que ilustra el diseño experimental para la detección de la Ex9-12 del (tomado y modificado de Quezada-Urban et al., 2018). En el diagrama se señalan los exones 8 a 13 (no a escala real) del gen BRCA1 y la posición de los 3 primers diseñados. Se muestra que al producirse una banda de 400 pb en una electroforesis, se considera un alelo silvestre. En cambio, si se produce una banda de 400 pb y otra banda de 297 pb, se reconoce que la muestra presenta la delección de los exónes 9-12.

Figura 6. Gráfico de tipo *Oncoprint* que muestra la distribución de las variantes en el número de copias en las muestras. En el eje de las X están las muestras y en el eje Y están los genes ordenados de forma descendente de acuerdo con el número de variantes que se detectaron en cada gen. En azul se indican las pérdidas y en rojo las amplificaciones. Los puntos de colores añadidos sobre cada variante indican los programas con los que se detectó dicha variante. En el eje X en la parte superior de la gráfica se muestra el número de variantes por muestra en el mismo eje X en la zona inferior se muestra se indica la plataforma de secuenciación con la que se secuenciaron las muestras. Además, se indican los controles positivos con un símbolo + en la parte inferior.

Tabla 1. Genes de alta y moderada penetrancia con el respectivo riesgo estimado en cáncer de mama y ovario, así como en otras enfermedades (Nielsen, Hansen & Sørensen, 2012; NCCN, 2020).

Tabla 2. Número de muestras por país

RESUMEN

El cáncer de mama es la neoplasia maligna de mayor incidencia y mortalidad en América Latina. En 2018 se reportaron cerca de 200,000 nuevos casos, con una mortalidad superior a 50 mil personas. El síndrome hereditario de cáncer de mama y ovario (SHCMO) representa entre el 5 y 10% de los casos de cáncer de mama totales. Se conoce que, en estos pacientes, aproximadamente el 50% de las alteraciones ocurren en los genes *BRCA1* y *BRCA2* y la otra mitad en cerca de 20 genes diferentes.

En este trabajo analizamos, mediante herramientas bioinformáticas, rearrreglos estructurales en 14 genes de susceptibilidad a cáncer de mama y ovario en 500 pacientes reclutados en cuatro países de América Latina (Colombia, Guatemala, México y Perú) que cumplen con los criterios del SHCMO, las cuales fueron secuenciadas con los equipos MiSeq y HiSeq de Illumina. Como control positivo se emplearon muestras de pacientes con previa detección experimental de la mutación fundadora de la población mexicana, la delección de los exones 9-12 de *BRCA1* (*BRCA1* ex9-12del).

Se detectaron 36 variaciones en el número de copias en un total de 34 muestras, que representan el 6.8% del total (considerando ambos grupos muestrales: MiSeq y HiSeq). Asimismo, las variantes se detectaron en 12 de los genes de interés, siendo los más frecuentes *BRCA2* (1.4%, 7/500) y *BRCA1* (6 muestras) sin tomar en cuenta las muestras control y en general las muestras con una delección fundadora validada. Con respecto a la proporción entre amplificaciones y delecciones, se encontraron 17 delecciones y 17 amplificaciones.

No se detectó la presencia de la mutación fundadora *BRCA1 Ex9-12Del* en ninguna muestra que no fuera mexicana y tampoco se encontró esta variante en ninguna muestra que no se haya detectado antes mediante PCR, es decir, no se detectaron falsos positivos de esta variante.

Los resultados de este trabajo permitirán implementar la detección de rearrreglos estructurales en la línea germinal de pacientes con SHCMO a partir de datos de secuenciación masiva, con previa validación experimental. Lo anterior representa un avance importante en la detección de estas variantes de difícil identificación para mejorar el diagnóstico molecular y la descripción epidemiológica de esta enfermedad.

SUMMARY

Breast cancer is the malignant neoplasm with the highest incidence and mortality in Latin America. In 2018, nearly 200,000 new cases were reported, with a mortality of more than 50,000 people. Hereditary breast and ovarian cancer syndrome (HBOC) accounts for between 5% and 10% of all breast cancer cases. In these patients, approximately 50% of the alterations occur in the *BRCA1* and *BRCA2* genes, and the other half is distributed in about 20 different genes.

In this work we used bioinformatic tools to analyze large rearrangements in 14 breast and ovarian cancer susceptibility genes, in 500 patients recruited in four Latin American countries (Colombia, Guatemala, Mexico and Peru) who meet the SHCMO criteria. The germline was sequenced with the Illumina MiSeq and HiSeq machines. Samples from patients with previous experimental detection of the founding mutation of the Mexican population, the deletion of exons 9-12 of *BRCA1* (*BRCA1 ex9-12del*), were used as positive control.

The overview of the results shows that 36 copy number variations were detected in a total of 34 samples, which represent 6.8% of the total (considering both sample groups: Illumina MiSeq and Illumina HiSeq). Likewise, variants were detected in 12 of the genes of interest, the most frequent being *BRCA2* (7 samples) and *BRCA1* (6 samples) without considering the control samples and, in general, the samples with a validated founder mutation. Additionally, the ratio between amplifications and deletions, 17 deletions and 17 amplifications were found.

The presence of the founder mutation *BRCA1 Ex9-12Del* was not detected in any non-Mexican sample and this variant was not found in any sample that had not been previously detected by PCR, that is, no false positives of this variant were detected.

The results of this work will make it possible to implement the detection of structural rearrangements in the germ line of patients with HBOC from mass sequencing data, with prior experimental validation. This represents an important advance in the detection of these difficult-to-identify variants to improve the molecular diagnosis and epidemiological description of this disease.

INTRODUCCIÓN:

El cáncer

El cáncer se define como un conjunto de enfermedades de origen genético y epigenético en el que se presenta un crecimiento y proliferación anormal y descontrolado de células, las cuales pueden formar tumores malignos, migrar e invadir a otros tejidos, en un proceso denominado metástasis.

A diferencia de las células normales, las células cancerígenas poseen una serie de características específicas, las cuales se conocen como los “*Hallmarks*” del cáncer. Todas ellas son características biológicas anormales que adquieren las células cancerígenas durante el desarrollo de los tumores humanos como son: la evasión de reguladores de crecimiento, la capacidad de invasión y metástasis, entre otras. Dichas características son distintivas y complementarias y permiten el crecimiento de los tumores y la diseminación metastásica de los mismos (Hanahan & Weinberg, 2011).

A su vez, los tumores cancerígenos se definen como un tejido complejo compuesto por muchos tipos celulares distintos que interactúan de forma heterotípica entre sí, más que una simple masa de células anormales.

Factores de riesgo

Al cáncer se le considera un conjunto de enfermedades de origen multifactorial, que pueden ser causadas por muchos factores, tanto intrínsecos (factores no modificables) como no intrínsecos (ajenos al cuerpo). A continuación, se describen los factores del riesgo que pueden asociados al desarrollo del cáncer (Wu *et al.*, 2018).

Intrínsecos

Los factores de riesgo intrínsecos se refieren a aquellos factores no modificables que ocurren en el cuerpo, como los errores aleatorios en la replicación del DNA. El riesgo intrínseco surge de la tasa de mutación basal que opera en todas las células del cuerpo humano sin considerar ningún factor no intrínseco que pudiese causar dichas alteraciones en el genoma (Wu *et al.*, 2018).

En este sentido, la fijación de mutaciones adquiridas aleatoriamente en un tejido en particular depende de que la célula afectada sobreviva y se divida. Las mutaciones aleatoriamente adquiridas pueden ser de dos tipos: conductoras, que son necesarias para el desarrollo del cáncer; o pasajeras que, a pesar de ser alteraciones del genoma, no afectan el proceso del desarrollo de una neoplasia (Wu, *et al.*, 2018).

Hasta la fecha, se sugiere que los factores de riesgo intrínsecos debido a errores aleatorios en la replicación del DNA (que es común para todos los seres humano), individualmente representaría más del 60% de las mutaciones de genes conductores del cáncer (Tomassetti & Vogelstein, 2017).

No intrínsecos

Los factores de riesgo no intrínsecos son aquellos que se consideran modificables o parcialmente modificables y se clasifican en dos grupos:

Endógenos:

Se consideran factores parcialmente modificables, son parte del funcionamiento fisiológico-anatómico del paciente y tienen un componente genético. Algunos ejemplos de factores de riesgo endógenos son: las hormonas, la maquinaria de reparación del DNA y el envejecimiento.

Exógenos:

Por otra parte, los factores de riesgo exógenos son ajenos al cuerpo y son modificables, es decir, pueden evitarse o se puede controlar la exposición a los mismos. Algunos ejemplos de factores de riesgo exógenos son: la radiación ultravioleta, los virus causantes de cáncer (como el Virus de Papiloma Humano), la exposición a químicos carcinogénicos, entre otros (Wu *et al.*, 2018).

El cáncer en el mundo

El cáncer se reconoce como uno de los problemas de salud pública más graves a nivel mundial.

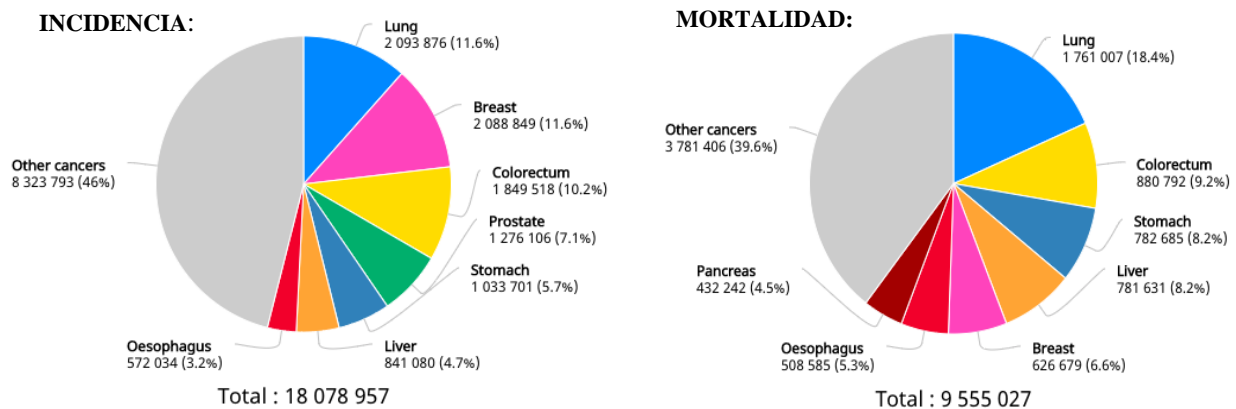


Figura 1: Gráficas de pastel que muestran las estadísticas de incidencia y mortalidad de los tipos de cáncer para ambos sexos (femenino y masculino) y todas las edades en el mundo en el año 2018 (Bray *et al.*, 2018).

En el año 2018 se registraron más de 18 millones de casos de cáncer a nivel mundial y cerca de 9 millones y medio de muertes atribuidas a algún tipo de cáncer. En cuanto a incidencia, los tipos de cáncer de pulmón, mama y colorrectal ocupan los primeros lugares; mientras que, en mortalidad, el cáncer de pulmón es la principal causa de muerte por cáncer, seguido de cáncer colorrectal y de estómago (Bray, *et al.*, 2018).

Particularmente cuando se refiere a los tipos de cáncer que más afectan al sexo femenino; a nivel mundial el cáncer de mama ocupa el primer puesto tanto en incidencia como en mortalidad. Tan sólo en el año 2018, el 24% de los casos de cáncer en mujeres, fueron cáncer de mama y el 15% de las muertes de mujeres por cáncer, también se atribuyeron a esta neoplasia. (Bray, *et al.*, 2018).

El cáncer en América Latina

Particularmente, en la región de América Latina y el Caribe en el año 2018, las estadísticas muestran que en cuanto a la incidencia de cáncer considerando ambos sexos y todas las edades, el cáncer de mama ocupa el primer lugar afectando a 2,288,283 personas, seguido por el cáncer de pulmón y colorrectal. En cuanto a la mortalidad, el primer lugar lo ocupa el cáncer de pulmón, que, en 2018, cobró la vida de 1,842,391 personas en América Latina. Sin embargo, cuando únicamente se consideran las estadísticas de incidencia y mortalidad de cáncer en mujeres, es claro que el primer lugar lo ocupa el cáncer de mama, que afectó a 2,288,283 mujeres y cobró la vida de 679,237 de ellas en 2018 (Bray, *et al.*, 2018).

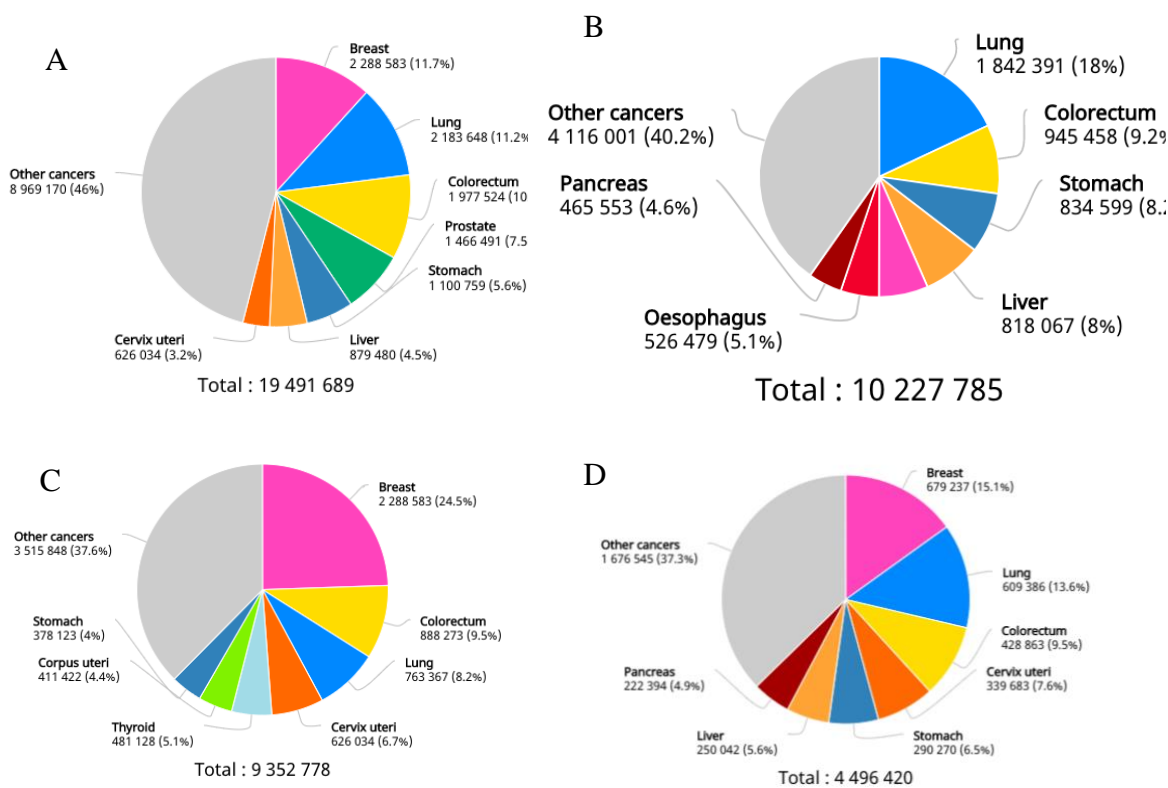


Figura 2: Gráficos de pastel que muestran la incidencia y mortalidad de cáncer de mama en América Latina, en el año 2018 (Bray *et al.*, 2018). A) Incidencia de cáncer en ambos sexos, todas las edades. B) Mortalidad por cáncer para ambos sexos, todas las edades. C) Incidencia de cáncer en mujeres de todas las edades. D) Mortalidad por cáncer en mujeres de todas las edades.

De acuerdo a estas estadísticas, podemos concluir que el cáncer de mama es un grave problema de salud pública en la región de América Latina.

El cáncer de mama

El cáncer de mama es un tipo de cáncer que surge del tejido mamario. Los tumores de mama son visibles en radiografías o como bultos de tejido en el seno. La mayoría de los casos de cáncer de mama suceden en mujeres, pero también se da en hombres (Varonesi *et al.*, 2017).

Esta neoplasia usualmente inicia en los ductos lactíferos que conducen la leche materna al pezón, y se conoce como *cáncer ductal*. Sin embargo, en algunos casos, se presenta cáncer de tipo *lobular*, el cual se origina en las glándulas que producen la leche materna. En casos raros, se pueden presentar *tumores filodes* o *angiosarcomas* en el seno. Además, el cáncer puede originarse de otros tejidos mamarios como *sarcomas* y *limfomas*, aunque estos no se consideran del todo como cáncer de mama (Veronesi *et al.*, 2017).

En muchos casos, el cáncer de mama puede ser detectado en estadios iniciales de la enfermedad y antes de que el cáncer invada otros tejidos aledaños o distantes. Detectar el cáncer de mama en una etapa temprana aumenta considerablemente las probabilidades de supervivencia del paciente.

El cáncer, en general, tiene como característica fundamental la capacidad de invasión de tejidos aledaños y distantes. El cáncer de mama en particular se vuelve invasivo cuando las células cancerosas entran al sistema circulatorio o sistema linfático y son transportadas a otras partes del cuerpo. Inicialmente, el cáncer de origen mamario invade el tejido de los nodos linfáticos cercanos a la axila o alrededor de la clavícula. En cuanto a los tejidos distantes, comúnmente las células cancerosas de mama invaden el tejido óseo, el hígado, entre otros (Varonesi *et al.*, 2017).

Clasificación molecular de tumores mamarios:

En el año 2000, Perou y colaboradores propusieron que el cáncer de mama podría clasificarse en subtipos moleculares según los patrones de expresión génica y los denominaron como: Tipo luminal, tipo Erb-B2+ y el tipo basal. A continuación se describen los subtipos a detalle (Perou CM, Sørlie T, Eisen MB, et al., 2000; Johnson, K. S., Conant, E. F., & Soo, M. S., 2021).

Los tumores de tipo **luminal** son el tipo más común de cáncer de mama y comprenden entre el 60 % y el 70 % de todos los tumores. Se caracterizan por una alta expresión genética del receptor de estrógeno (ER) entre otros genes expresados por las células epiteliales que recubren la luz de la unidad lobulillar del conducto terminal donde surgen la mayoría de los cánceres de mama. A menudo carecen de una expresión significativa del oncogén *HER2*. Así mismo, los tumores de tipo luminal se clasifican en dos tipos distintos conocidos como LuminalA y LuminalB. Si bien la expresión genética de los receptores hormonales (HR) es una característica compartida de los tumores luminales A y B; los tumores **luminales B** se diferencian del subtipo luminal A por tener una mayor expresión de genes proliferativos y/o del ciclo celular y una menor expresión de PR. Así mismo los tumores luminales B, a diferencia del luminal A, se asocian con una alta frecuencia de mutaciones de p53. Desde el punto de vista morfológico, los tumores **luminales A** son carcinomas bien diferenciados, mientras que los tumores luminales B están menos diferenciados y suelen ser de mayor grado (Johnson, K. S., Conant, E. F., & Soo, M. S., 2021).

En el caso de los tumores enriquecidos con *HER2* (**tipo Erb-B2+**), representan del 12 % al 20 % de todos los cánceres de mama. Se caracterizan por la sobreexpresión del oncogén *Erb-*

B2 y muestran bajos niveles de expresión de receptores de estrógeno (Johnson, K. S., Conant, E. F., & Soo, M. S., 2021)..

Finalmente, los tumores de tipo **basal** comprenden aproximadamente el 15 % de todos los cánceres de mama y expresan muchos de los mismos genes que se encuentran en las células mioepiteliales basales de la unidad lobular del ducto terminal y a menudo no logran expresar ER, entre muchos otros genes asociados con la expresión de ER (Johnson, K. S., Conant, E. F., & Soo, M. S., 2021)..

Además, se han informado diferencias críticas entre los subtipos en cuanto a incidencia, respuesta al tratamiento, progresión de la enfermedad, supervivencia y características de imagen. Conocidos como los "subtipos intrínsecos de cáncer de mama" (luminal A, luminal B, enriquecido con HER2 y de tipo basal), estos grupos moleculares han sido objeto de una extensa investigación durante los últimos 20 años (Johnson, K. S., Conant, E. F., & Soo, M. S., 2021).

A lo largo de los años, a partir de la primera clasificación molecular de los subtipos de cáncer de mama, se han propuesto nuevos subtipos incluyendo el *subtipo de claudina baja*, *subtipo apocrino molecular* y un *subtipo luminal HER2*. Sin embargo, estos subtipos aún están siendo investigados y fueron recientemente propuestos, demostrando la compleja heterogeneidad molecular del cáncer de mama (Johnson, K. S., Conant, E. F., & Soo, M. S., 2021; Tang P. & Tse GM., 2016).

Factores de riesgo de cáncer de mama

A continuación, se describen los principales factores de riesgo de padecer cáncer de mama.

Edad:

El riesgo de padecer cáncer de mama incrementa durante la edad reproductiva y decrece a la edad aproximada de 50 años que es la edad promedio en la que se presenta la menopausia en la mujer. La incidencia acumulativa de cáncer de mama en mujeres de Europa y América del Norte es de alrededor del 2.7% a la edad de 55 años, del 5% a la edad de 65 años y del 7.7% a los 75 años (Key, Verkasalo, & Banks, 2001).

La menarca y el ciclo menstrual:

En cuanto a la menstruación, se sabe que, entre más tiempo tarde una mujer en tener su primera menstruación, el riesgo de padecer cáncer de mama disminuye. Por 1 año de retraso en presentarse la primera menstruación, el riesgo disminuye en un 5%. Otros aspectos del ciclo menstrual como la duración o la regularidad no han sido vinculados al riesgo de padecer cáncer de mama (Key, Verkasalo, & Banks, 2001).

Paridad

En este sentido, las mujeres que han tenido al menos un embarazo a término tienen un 25% menos de riesgo de padecer cáncer de mama comparado con mujeres que no han tenido hijos. Además, la edad a la que se tiene el primer embarazo influye en el riesgo. Por ejemplo, mujeres que tienen un hijo antes de los 20, tienen 30% menos riesgo de padecer cáncer de mama que las mujeres cuyo primer hijo nació a sus 35 años (Key, Verkasalo, & Banks, 2001).

En cuanto la producción de leche materna y amamantar a los recién nacidos, el vínculo con el riesgo de cáncer de mama es controversial. Sin embargo, se sabe que las mujeres que

amamantan a su bebé por más de 25 meses tienen un 33% menos de riesgo de desarrollar una neoplasia mamaria en comparación a mujeres que no amamantaron (Layde *et al.*, 1989).

Menopausia

En cuanto a la relación de la menopausia con el riesgo de padecer cáncer de mama, se conoce que las mujeres que dejan de menstruar a una edad temprana tienen un 3% menos de riesgo de desarrollar cáncer de seno que aquellas mujeres que experimentan la menopausia a una edad más avanzada.

Hormonas

Los factores hormonales juegan un papel clave en el desarrollo del cáncer de mama. Un ejemplo claro es que el estradiol estimula la mitosis de las células epiteliales mamarias, y este efecto mitogénico puede aumentarse con la producción de otra hormona como es, la progesterona. Como se describió anteriormente, la menarquía temprana y la menopausia tardía (procesos biológicos mediados por hormonas) aumenta el riesgo de desarrollar cáncer en el seno (Veronesi *et al.*, 2017).

En mujeres posmenopáusicas, altas concentraciones del estradiol endógeno se asocian positivamente con el riesgo. Además, la obesidad en mujeres posmenopáusicas aumenta riesgo de cáncer de mama, en gran parte porque causa un aumento en concentraciones de estradiol y estrona en el cuerpo de la mujer.

Por otra parte, los anticonceptivos orales suprimen la producción de estradiol y progesterona, pero las hormonas endógenas son reemplazadas efectivamente por las hormonas exógenas, aunque no se tiene evidencia de que los anticonceptivos orales aumenten el riesgo de padecer esta neoplasia.

El embarazo tiene un doble efecto sobre el riesgo: a corto plazo, el riesgo aumenta seguido de una reducción a largo plazo del riesgo. Los mecanismos más probables de estos efectos son que, las altas concentraciones de hormonas durante el embarazo estimulan el crecimiento de pequeños tumores preclínicos, aumentando así el riesgo de padecer cáncer de mama (Key, Verkasalo, & Banks, 2001).

Factores genéticos y herencia

Con respecto al riesgo por cáncer familiar, varios estudios coinciden en que el riesgo de padecer cáncer de mama para los familiares directos (madre, hijas, hermanas) de un paciente diagnosticado con la enfermedad, es del doble con respecto a una persona sin historia familiar de cáncer de mama.

Con respecto a los factores genéticos hereditarios, hasta la fecha se conocen algunos genes de alta penetrancia que incrementan considerablemente el riesgo de que una persona padezca cáncer de mama en el futuro. Se conoce, por ejemplo, que variantes patogénicas germinales en *BRCA1* y *BRCA2* incrementan el riesgo en un 50% a 85%. Por otra parte, mutaciones germinales en el gen *PTEN* son responsables de desarrollar la enfermedad de Cowden, en la cual, el cáncer de seno es una característica común.

El factor hereditario es de gran importancia debido a que el cáncer no sólo afecta a un individuo, sino que también afecta a la descendencia del mismo. Uno de los síndromes más conocidos en cuanto al estudio del cáncer de mama de tipo hereditario es el Síndrome

Hereditario de Cáncer de Mama y Ovario (*SHCMO*) (Key, Verkasalo, & Banks, 2001; Veronesi *et al.*, 2017).

El Síndrome Hereditario de Cáncer de Mama y Ovario (SHCMO)

La mayoría (60-80%) de los casos de cáncer de mama son a causa de la combinación de factores ambientales y de estilo de vida, más que por factores genéticos heredados. La mayoría de las mujeres que padecen esta enfermedad, no tienen historia familiar de cáncer y la mayoría de las mujeres con familiares afectados, nunca desarrollan cáncer a lo largo de su vida (Layde *et al.*, 1989).

A diferencia del cáncer de mama de origen esporádico (es decir, que se origina por factores ambientales o no hereditarios), el cáncer de mama de tipo hereditario afecta al individuo que lo padece y a su futura descendencia portadora de las mutaciones patogénicas que pueden resultar en el desarrollo de cáncer de mama u ovario o incluso, ambos a edad temprana (ASCO, n.d.).

El Síndrome Hereditario de Cáncer de Mama y Ovario (SHCMO) se define como un desorden hereditario por el que el riesgo de padecer cáncer de mama u ovario (particularmente antes de los 50 años) es mayor, con respecto a personas que no presentan el síndrome. La herencia de este genotipo es de tipo mendeliana, autosómica dominante. Sin embargo, del total de casos de cáncer de mama, entre el 5 y el 10% son a causa de mutaciones germinales hereditarias (Kida, Murai & Yamauchi, 2017).

Diagnóstico de SHCMO

De acuerdo con la *Sociedad Americana de Oncología Clínica* (ASCO, por sus siglas en el inglés), para que un paciente sea diagnosticado con SHCMO, debe cumplir con uno o más de los criterios que a continuación se presentan:

- Una o más mujeres emparentadas son diagnosticadas con cáncer de mama antes de los 45 años.
- Una mujer es diagnosticada con cáncer de mama, antes de los 50 años, con historia familiar adicional de cáncer de otro tipo, como próstata o páncreas.
- Se han presentado varios casos de cáncer de mama u ovario en múltiples generaciones del mismo lado de la familia.
- Una mujer es diagnosticada con cáncer de mama por segunda vez en el mismo seno o en el otro.
- Una mujer es diagnosticada con cáncer de mama y ovario a la vez.
- Un familiar de sexo masculino es diagnosticado con cáncer de mama
- Tener ancestría judía Askenazi¹

¹ Se sabe que los miembros de la comunidad judía Ashkenazi tienen una probabilidad más alta de desarrollar cáncer de mama debido a que se trata de una comunidad cerrada.

El manejo clínico del SHCMO

El manejo clínico del cáncer de mama y ovario varía de acuerdo con el subtipo de tumor y el estadio en el que se encuentre. El procedimiento clínico propuesto por *Nielsen y colaboradores en 2016* comienza con la toma de una biopsia de los tumores de mama o de ovario que se envían para una clasificación basada en la expresión génica, que incluye perfiles de reparación por recombinación homóloga (HRR) o para clasificación basada en firmas mutacionales.

Posteriormente, se consideran otros criterios de diagnóstico. Si el paciente es mujer menor de 40 años, hombre o si el tumor de mama pertenece al tipo basal o luminal B, o los tumores de mama y ovario son HRR deficientes, se envían las muestras de sangre completa para hacer estudios genéticos de secuenciación masiva de próxima generación (NGS) y determinar si se trata de un caso de cáncer de mama y ovario hereditario (SHCMO). De encontrarse mutaciones patógenas en la línea germinal del paciente, las mutaciones se tratan y se asesora al paciente de acuerdo con las recomendaciones actuales para casos de SHCMO.

En el caso de identificación de una variante de significado desconocido (VUS), se puede realizar un análisis funcional para clasificar dicha variante. En el caso de que el VUS parezca ser benigno o el tumor de mama pertenece al grupo LumA, LumB, apocrino molecular o subtipos normales, el paciente puede someterse a terapia convencional y no requiere asesoramiento genético, pues el cáncer se considera de origen esporádico.

Las mutaciones puntuales en el SHCMO

Las mutaciones se definen como cambios en la secuencia de nucleótidos en el genoma de un ser vivo o virus que puede o no ser patogénica y que son heredables. Las mutaciones patogénicas son aquellas que dañan o imposibilitan el funcionamiento del gen mutado (Alberts *et al.*, 2002). Las mutaciones también son llamadas variantes genéticas. En esta tesis empleamos ambas denominaciones.

En el contexto de la biología del cáncer hereditario, las mutaciones se dan en células germinales (espermatozoides y/u óvulos), lo que resulta en un cigoto y un ser humano con dichas mutaciones en las células del cuerpo.

No todas las mutaciones tienen el mismo impacto en cuanto al riesgo de padecer cierto tipo de cáncer. En el caso particular del SHCMO, de acuerdo a la guía de la *National Comprehensive Cancer Network (NCCN)*, para el estudio de este síndrome los genes se dividen en dos grupos: alta penetrancia y penetrancia moderada (NCCN, 2019). La penetrancia se define como la proporción de individuos portadores de una mutación patogénica de importancia clínica que presentan los síntomas característicos del fenotipo esperado, entre los individuos que también presentan la mutación, pero no presentan síntomas.

Genes de susceptibilidad a cáncer de mama y ovario

Tabla 1: Genes de alta y moderada penetrancia con el respectivo riesgo estimado en cáncer de mama y ovario, así como en otras enfermedades (Nielsen, Hansen & Sørensen, 2012; NCCN, 2020)

Penetrancia	Gen	Riesgo estimado de cáncer de mama	Riesgo estimado de cáncer de ovario	Otras enfermedades relacionadas
Alta	<i>BRCA1</i>	57-65% a los 70 años	39-44% a los 70 años	Páncreas y próstata
	<i>BRCA2</i>	45-55% a los 70 años	11-18% a los 70 años	Páncreas, próstata y melanoma
	<i>TP53</i>	25% a los 70 años	No incrementa el riesgo	Páncreas y Síndrome de Li-Faumeni
	<i>CDHI</i>	42% a los 80 años	No incrementa el riesgo	Cáncer gástrico difuso
	<i>PTEN</i>	85% a los 70 años	No incrementa el riesgo	Síndrome de Cowden
	<i>STK11</i>	32% a los 60 años	13% a los 60 años	Cáncer colorrectal y de páncreas
Moderada	<i>CHEK2</i>	37% a los 70 años	No incrementa el riesgo	Cáncer de colon
	<i>PALB2</i>	35% a los 70 años	Desconocido	Cáncer de páncreas
	<i>MLH1</i>	19% a los 70 años	20% a los 70 años	Cáncer de colon, útero, páncreas y otros
	<i>MSH2</i>	11% a los 70 años	24% a los 70 años	Cáncer de colon, útero, páncreas y otros
	<i>PMS2</i>	SIR: 3.8	SIR:12.0	Cáncer de colon, útero, y otros
	<i>MSH6</i>	Desconocido	Desconocido	Cáncer de colon, útero, páncreas y otros
	<i>EPCAM</i>	Desconocido	Desconocido	Cáncer de colon, útero, páncreas y otros
	<i>CDKN2A</i>	No incrementa el riesgo	No incrementa el riesgo	Cáncer de páncreas

SIR: Standardized Incidence Ratio

La mayoría de los estudios de cáncer hereditario, se enfocan en las mutaciones puntuales en los genes, sin embargo, existen otros tipos de variantes genéticas que pueden resultar en el mismo fenotipo del cáncer, como son, los rearrreglos estructurales. Además, aunque la mayoría de los estudios acerca de cáncer de mama y ovario de tipo hereditario se enfocan en mujeres, los hombres portadores de mutaciones germinales en los genes *BRCA1* y *BRCA2* tienen un riesgo significativo de padecer cáncer de próstata, mama o páncreas (Liede, Karlan & Narod, 2004).

Los rearrreglos estructurales en el SHCMO

Los rearrreglos estructurales se definen como un tipo de variante genética que involucra cambios en porciones grandes del genoma. Las deleciones, duplicaciones, triplicaciones, inserciones, entre otras, son ejemplos de rearrreglos estructurales (LR). Este tipo de variantes se consideran patogénicas cuando afectan regiones críticas de algún gen y reducen o anulan por completo, el funcionamiento de los genes relacionados con síndromes hereditarios de cáncer (como es el caso del SHCMO) (Ford et al., 1998; Mancini-DiNardo *et al.*, 2019).

Al igual que las mutaciones puntuales, los LR también se han relacionado con síndromes hereditarios de cáncer. Por ejemplo, las variantes estructurales en *BRCA1* y *BRCA2* representan un riesgo de entre el 43% y 87% de padecer cáncer de mama en la vida (Stuppia *et al.*, 2012).

A pesar de que se reconoce la importancia de los LR en el cáncer hereditario, comúnmente son variantes que están poco representadas en resultados de secuenciación (Mancini-DiNardo *et al.*, 2019).

Las tecnologías que más comúnmente se emplean para analizar LR son el método de PCR multiplex dependiente de ligación (MLPA, por sus siglas en inglés) y el de hibridación genómica comparativa basada en microarreglos (CHG-microarrays, por su nombre en inglés). La técnica de MLPA tiene la importante limitante de que sólo se puede usar para detectar ciertos rearrreglos en determinados genes por el número y posición de las sondas. Por otra parte, los microarreglos, son eficaces para detectar la mayoría de los LR, sin embargo, no es posible detectar con certeza, las inserciones de retro elementos como las inserciones *Alu* (Qian *et al.*, 2017; Mancini-DiNardo *et al.*, 2019). Esta limitante es particularmente importante porque *BRCA1* y *BRCA2* tienen un alto contenido de elementos *Alu* en las regiones intrónicas, los cuales son la base molecular de la elevada tasa de inserción de dichos elementos en ambos genes (Ewald *et al.*, 2009).

Las nuevas tecnologías de secuenciación masiva de próxima generación (NGS, por sus siglas en inglés *Next Generation Sequencing*) permiten analizar múltiples secuencias y varios genes a la vez con una buena eficacia en la detección de variantes de un solo nucleótido y de LR (Mancini-DiNardo *et al.*, 2019). Esta tecnología representa una alternativa potencial para la detección eficiente de los LR.

ANTECEDENTES:

Los primeros trabajos que describen rearrreglos estructurales de línea germinal en genes de susceptibilidad a cáncer de mama y ovario se centran en los genes *BRCA1* y *BRCA2*. En el trabajo de Mazoyer en 2005 se describen por primera vez, varios rearrreglos estructurales con puntos de corte definidos en el gen *BRCA1*; estas variantes se distribuyen a lo largo del gen y consisten principalmente de inserciones, deleciones, triplicaciones o eventos que combinan más de un tipo de variante.

El gen *BRCA1* tiene una estructura genética particular con una densidad alta de repeticiones *Alu* intrónicas y una repetición del promotor que contiene un pseudogen de *BRCA1* lo cual propicia los eventos de recombinación homóloga desigual (Ewald *et al.*, 2009).

La frecuencia de rearrreglos estructurales en los genes *BRCA1* y *BRCA2* así como el impacto clínico de los mismos en los en el cáncer de mama se conoce en varias poblaciones del mundo como es Estados Unidos (aproximadamente 12%), Singapur (3%), Finlandia (0%), entre otros (Ewald *et al.*, 2009). Sin embargo, no fue hasta el año 2007 que la población de América Latina se incluyó en los estudios.

Weitzel y colaboradores, en 2007 publicaron uno de los trabajos más importantes para la población latina en cuanto a rearrreglos estructurales en los genes *BRCA1/2* en cáncer de mama y ovario. En su trabajo, estudiaron el DNA de 106 individuos hispanos sin una mutación exónica en los genes *BRCA*. En los resultados de su estudio, reportaron una deleción recurrente del exon 9 al 12 (*ex9-12del*) en el gen *BRCA1* en 5 (3.8%) familias no emparentadas; 4 de dichas familias eran mexicanas y una tenía ancestría de grupos nativos americanos y afroamericanos. La evidencia indica que se trata de una mutación fundadora de origen amerindio y mestizo.

Posteriormente, otro estudio realizado en la universidad de Pensilvania en Estados Unidos en el que estudiaron a 251 pacientes y los clasificaron en dos grupos no pertenecientes a la comunidad judía Askenazi y los judíos Askenazi. En este trabajo detectaron que en el grupo de no judíos Askenazi, el 26% presentaba algún rearrreglos estructural en los genes *BRCA1/2*; sin embargo, la frecuencia en el grupo de pacientes Askenazis fue del 18% (Palma *et al.*, 2008).

En el año 2010, en el estudio de Del Valle y colaboradores, usaron la tecnología de MLPA para analizar rearrreglos estructurales en los genes *BRCA1/2*, en este caso en 310 pacientes españoles diagnosticados con SHCMO. En este trabajo caracterizaron 6 rearrreglos distintos en el gen *BRCA1* y dos deleciones en *BRCA2* en la población española que, por razones históricas está estrechamente relacionada con la población de América Latina.

Posteriormente, en el año 2012, Judkins y colaboradores publicaron un trabajo en el que evaluaron pacientes clasificados en dos grupos: de alto riesgo (25,535 pacientes) y electivo (22,921). Además de la N del estudio, otra fortaleza del trabajo es que el grupo de pacientes evaluados provenía de diferentes grupos étnicos y enfrentaban un grado de riesgo distinto. Los rearrreglos estructurales se determinaron mediante secuenciación Sanger y MLPA. Los resultados del estudio arrojaron que La prevalencia general de la mutación *BRCA1 / 2* entre los pacientes de alto riesgo fue del 23.8% versus el 8.2% del grupo electivo. El perfil de mutación para los pacientes de alto riesgo fue 90.1% de secuenciación de mutaciones versus 9.9% de LR, y para los pacientes electivos, 94,1% de secuenciación versus 5.9% de LR. Esta diferencia puede reflejar el sesgo en los pacientes de alto riesgo de portar mutaciones en

BRCA1, que tiene una mayor penetrancia y frecuencia de LR en comparación con *BRCA2*. Particularmente los resultados de pacientes de etnicidad latina (1560 pacientes) 104 fueron positivos para un LR en *BRCA1/2*. Además, se reportaron rearrreglos estructurales recurrentes específicos para la población de ancestría latina; por ejemplo, la delección del gen *BRCA1* completo se encontró en 13 muestras de este grupo, también detectaron la delección fundadora de los exones 9 a 12 de *BRCA1* en 44 muestras, entre otros. En conclusión, este trabajo es probablemente el estudio más grande y detallado que se conoce en cuanto a los rearrreglos estructurales de los genes *BRCA1/2* en pacientes con SHCMO de la población de América Latina.

En 2014, un trabajo que evaluó los rearrreglos estructurales en poblaciones de la península ibérica (España y Portugal) mediante la tecnología de MLPA. En este estudio se detectaron nuevos rearrreglos en *BRCA1* en el 2.3% de los pacientes y en *BRCA2* en el 8.3%. Además, detectaron 2 rearrreglos previamente descritos, como la delección del gen *BRCA1* completo y la delección del exón 1 al 13 en *BRCA1*. El estudio de poblaciones de Portugal y España es relevante para comprender la composición genética de las poblaciones de América Latina por razones demográficas e históricas (Fachal *et al.*, 2014).

Hasta este punto, los únicos genes estudiados por rearrreglos estructurales relacionados al SHCMO son *BRCA1* y *BRCA2*. En 2016, una revisión publicada por Nielsen y colaboradores muestra la contribución genética de mutaciones puntuales en genes tanto *BRCA* como otros en el desarrollo de cáncer de mama y ovario en pacientes de alto riesgo. La relevancia de este trabajo es que muestra la importancia de mutaciones germinales en genes no *BRCA* en el desarrollo de cáncer de mama y ovario como: *MLH1*, *CHEK2*, *ATM*, *TP53*, entre otros. Aunque en este trabajo no se menciona la importancia de rearrreglos estructurales. Estos datos se describen más detalladamente en la tabla 1.

En 2017, se publicó un estudio de variación en el número de copias (CNV) y su importancia en cuanto al riesgo y pronóstico de cáncer de mama. En este estudio se analizaron 422 pacientes canadienses diagnosticadas con cáncer de mama invasivo. El punto clave de este trabajo, es que analizaron las muestras para identificar CNVs mediante un perfil de CNV de genoma completo y se tomaron en cuenta varios genes además de *BRCA1/2*, como el caso de delecciones en el gen *APOBEC3B* (Kumaran *et al.*, 2017).

Uno de los trabajos más estrechamente relacionados con esta tesis es el artículo de Quezada-Urban y colaboradores en 2018. En este estudio se analizaron las variantes de un solo nucleótido germinales en 300 pacientes mexicanos diagnosticadas con SHCMO mediante secuenciación masiva de un panel de 143 genes de susceptibilidad. En este estudio, detectaron variantes en 23 genes de susceptibilidad en el 22% de los pacientes, incluyendo la delección de los exones 9 a 12 en *BRCA1*. En este estudio, aunque sólo consideraron variaciones de un solo nucleótido y no rearrreglos estructurales; sí evaluaron la población mexicana directamente, y pacientes diagnosticados con SHCMO además de incluir un panel de 143 genes de susceptibilidad que incluye *BRCA1/2* y otros genes.

Uno de los antecedentes más recientes, fue el artículo de Mancini-DiNardo y colaboradores en 2019 en Estados Unidos, en el que evaluaron rearrreglos estructurales en 376,527 pacientes de cáncer hereditario con un panel de pan-cancer que incluye genes como *CDH1*, *CDKN2A*, *CHEK2*, *EPCAM*, entre otros, además de *BRCA1/2*. En este trabajo describen 3461 rearrreglos estructurales patogénicos. Determinaron que los LR varían de acuerdo al gen. Los LR fueron

más prevalentes en *STK11* (60.7% de todas las variantes patogénicas identificadas en el gen), *MSH2* (27.9%), *PMS2* (25.6%), *BMPRIA* (26.9%), *RAD51C* (21.1%) y *CDKN2A p14ARF* (16.7%). De los 2336 individuos que se encontraron portadores de LR patogénicos, 164 (7,0%) tenían al menos otra variante de la línea germinal. En este estudio sí consideraron los rearrreglos estructurales de línea germinal en genes *BRCA1/2* y otros genes de susceptibilidad, aunque fue en población de Estados Unidos no necesariamente latina, aunque por cuestiones demográficas la población de EU es importante para la población mexicana.

Otro antecedente relevante es el estudio de nuestro grupo de trabajo (Oliver *et al.*, 2019). En este trabajo estudiaron 222 pacientes diagnosticados con SHCMO de 5 países de América Latina: Argentina, México, Perú, Colombia y Guatemala. Se secuenció un panel de 143 genes con tecnología Illumina de secuenciación masiva. Detectaron que la prevalencia general de variantes de un solo nucleótido fue del 17%. Se detectaron variantes en *BRCA1* (20%), *BRCA2* (29%) y en otros genes de susceptibilidad como: *MSH2*, *MSH6*, *PALB2*, entre otros. Aunque este trabajo no consideró el análisis de rearrreglos estructurales, sí incluye a diferentes poblaciones de América Latina, entre ellas, las poblaciones objetivo de esta tesis: México, Guatemala, Perú y Colombia.

Finalmente, el antecedente directo más reciente de este trabajo fue publicado por Ossa-Gomez y colaboradores en el año 2022. En este estudio investigaron la prevalencia de las variantes patogénicas de línea germinal en pacientes de América Latina y Estados Unidos con sospecha de Cáncer de mama y/u ovario hereditario. Se identificaron variantes patogénicas en genes que aumentan el riesgo de cáncer de mama y de ovario en el 18,7 % de las personas evaluadas en México, América Central y el Caribe y en el 13,8 % de las personas evaluadas en América del Sur. Las variantes germinales patogénicas en *BRCA1* y *BRCA2* fueron los más comunes. Las tasas de variantes de significado incierto fueron similares en todas las regiones, y cuando hubo evidencia adicional disponible, > 90 % de las variantes de significado incierto se reclasificaron como benignas o probablemente benignas. Este estudio no está enfocado en variantes patogénicas del número de copias, sin embargo sí incluye diversas poblaciones de América Latina.

Adicionalmente, los genes de susceptibilidad seleccionados para este trabajo y parte de la información recopilada se tomaron de las guías de la *National Comprehensive Cancer Network, Genetic/Familial High-risk assessment: breast, ovarian and pancreatic (V2.21)* (NCCN, 2020).

JUSTIFICACIÓN

El cáncer de mama es un problema de salud pública a nivel mundial, en América Latina afecta a cerca de doscientos mil mujeres cada año. En el año 2018, más de seiscientas mil mujeres latinas murieron a causa de esta neoplasia.

El cáncer de mama y ovario de tipo hereditario representa alrededor del 10% de los casos del número total de casos de cáncer de mama y ovario en general. El síndrome hereditario de cáncer de mama y ovario afecta generalmente a mujeres muy jóvenes, en edades productivas y aumenta el riesgo de que su descendencia también desarrolle cáncer de mama, ovario (o ambos) u otro tipo de cáncer en el futuro, por lo que no sólo afecta al paciente en lo individual. Sin embargo y a pesar de que se reconoce la importancia de los rearrreglos estructurales en el desarrollo del cáncer de origen hereditario, la mayoría de los estudios que se conocen enfocados en las poblaciones de América Latina se limitan al estudio de mutaciones de un solo nucleótido. Aquellos trabajos que sí consideran los rearrreglos estructurales en sus resultados, están limitados en cuanto al número de genes que pueden estudiar.

Además, se conoce que en lo que respecta al SHCMO, los cambios en el número de copias tales como la *ex9-12del* en México, pueden ser específicas para determinadas poblaciones y producen el mismo fenotipo de cáncer. Tener información con respecto a las variantes únicas de una población, así como aquellas variantes que son compartidas entre distintas poblaciones puede ser útil en el ámbito de los estudios genéticos clínicos para los profesionales de la salud y los pacientes. Además, estos datos pueden proporcionar información relevante para la consejería genética que se brinda a los pacientes que han sido diagnosticados con cáncer de mama y ovario de origen hereditario, al igual que sus familias.

Particularmente, en referencia a los análisis estadísticos e informáticos enfocados al estudio de rearrreglos estructurales en general enfrentan varios problemas, por ejemplo:

- El balance inadecuado entre la sensibilidad y especificidad del análisis produce un número muy alto de falsos positivos.
- Las complicaciones en el control de calidad tanto de las variaciones en la calidad del propio DNA, preparación de librerías y la secuenciación en sí misma, además de otros factores como el contenido GC, puede llevar a diferentes resultados en cuanto al conteo de lecturas lo que hace muy difícil mantener un estándar de calidad aceptable sin comprometer la eficiencia del programa y evitar la pérdida de información.
- Sensibilidad baja para detectar los CNV de un tamaño determinado. Algunos programas pueden tener alta sensibilidad en detectar CNVs pequeños que parte de una región, pero deficientes en detectar CNV muy grandes que abarcan todo un gen o más de un gen. Sin embargo, en el ámbito clínico, es importante detectar ambos.
- Riesgo de hallazgos incidentales: dado que muchos paneles de genes están diseñados para cubrir varios tipos de enfermedades diferentes a la vez, los hallazgos incidentales se vuelven problemáticos en la interpretación y el informe clínico.

Por lo anterior, consideramos necesario el uso y estandarización de más de un programa especializado en el análisis de CNV que permita analizar muestras de línea germinal a partir de datos de secuenciación de exoma por panel. Esto con la finalidad de contrastar los resultados de los programas e identificar las coincidencias para reportar regiones candidato más certeras y poder aprovechar de manera más eficiente los métodos experimentales de validación.

Finalmente, según la revisión de a literatura, este trabajo es el primero que incluye el análisis de un grupo muestral tan grande de diferentes poblaciones latinas, para analizar rearrreglos estructurales en 14 genes de susceptibilidad a cáncer de mama y ovario.

HIPÓTESIS

Haciendo uso de herramientas bioinformáticas es posible detectar rearrreglos estructurales en 14 genes de susceptibilidad a cáncer de mama y ovario tales como los genes *BRCA1* y *BRCA2* en cerca del 7% de los casos; así como rearrreglos estructurales en otros genes, en frecuencias aún desconocidas para poblaciones de América Latina.

OBJETIVOS

Objetivo general

Establecer un flujo de trabajo bioinformático para detectar rearrreglos estructurales en 14 genes de predisposición al síndrome hereditario de cáncer de mama y ovario en un total de 500 muestras de línea germinal de pacientes mujeres de 4 países de América Latina: México, Guatemala, Perú y Colombia.

Objetivos particulares

1. Identificar rearrreglos estructurales en 500 muestras de 4 países de América Latina mediante herramientas bioinformáticas.
2. Diseñar y escribir un código único que permita analizar las muestras de forma rápida y eficiente con los 3 programas seleccionados cuyos códigos están escritos en python y R.
3. Generar una lista de CNV que hayan sido reportadas por al menos dos de los programas utilizados.
4. Reportar la frecuencia de rearrreglos estructurales detectados en 14 genes de susceptibilidad para el desarrollo del SHCMO.

MATERIALES Y MÉTODOS

Muestras

Se analizaron 500 muestras de pacientes diagnosticadas con SHCMO de 4 países de América Latina: México, Guatemala, Perú y Colombia. En la tabla 2 se muestran a detalle los datos sobre los grupos muestrales.

Tabla 2: Número de muestras por país

País	N
México	394 (78.8%)
Guatemala	16 (3.2%)
Perú	40 (8%)
Colombia	50 (10%)
TOTAL	500

Experimentación

Extracción de DNA

Se recolectaron 4 mL de sangre de las 500 pacientes seleccionadas y se almacenaron localmente a -80°C . Posteriormente, se extrajo el DNA de la sangre periférica con el kit *DNeasy Blood & Tissue* (Qiagen) siguiendo las instrucciones del fabricante. La concentración de DNA se cuantificó con el kit de ensayo *Qubit dsDNA HS Assay Kit* (Invitrogen), y se verificó la integridad y pureza del material mediante electroforesis en gel de agarosa y espectrofotometría, respectivamente.

Secuenciación masiva en paralelo

Se utilizó DNA de sangre periférica para la preparación de la biblioteca con el *kit GeneRead Cancer Predisposition V2* (Qiagen), que esta dirigido a 143 genes asociados con 88 enfermedades oncológicas heredadas según los datos del College of American Pathologists (CAP), las pautas de la NCCN, los ensayos clínicos en etapa tardía, The Cancer Genome Atlas (TCGA) y la base de conocimientos Ingenuity®. Posteriormente, la amplificación se dividió en reacciones de PCR de 4 grupos con un total de 6582 amplicones. La secuenciación de pares se realizó con la plataforma del sistema MiSeq y HiSeq. En el caso de las muestras que fueron secuenciadas por MiSeq, brevemente, se amplificaron $40 \pm 2,5$ ng de DNA con el kit *GeneRead DNAseq Gene Panel* (Qiagen) y se purificaron con perlas magnéticas *Agencourt AMPure XP* (Beckman Coulter). Los fragmentos amplificados se repararon en los extremos, se acoplaron con dA y se ligó el adaptador *GeneRead Adapter 1 Set plex* (Qiagen) utilizando el *GeneRead DNA Library 1 Core Kit*. A continuación, se seleccionó el tamaño de los segmentos amplificados (200-300 pb) utilizando perlas magnéticas *Agencourt AMPure XP* (Beckman Coulter). Se incorporaron códigos de barras de *New*

England Biolabs mediante amplificación por PCR en 10 ciclos de PCR y los productos se purificaron. Las bibliotecas se diluyeron a 4,0 nM y se combinaron en lotes de 60-80 muestras. La calidad de la biblioteca se evaluó mediante la cuantificación del DNA con *Qubit* después de la selección de tamaño y mediante el perfil de *Bioanalyzer* (Agilent). Las bibliotecas de códigos de barras agrupadas se diluyeron a 15 pM y se secuenciaron con un kit de reactivos *MiSeq V2 2 × 150 ciclos* (Illumina) para alcanzar una cobertura media teórica de 100× para cada muestra. En el caso de las muestras secuenciadas mediante la plataforma *Hiseq*, la preparación de bibliotecas se realizó en nuestro laboratorio siguiendo el mismo protocolo de *Miseq*, y posteriormente, estas se enviaron a los laboratorios de Novogene para secuenciación mediante la plataforma *Hiseq*.

Genes de alto riesgo

De acuerdo con la guía de cáncer hereditario de mama y ovario de la NCCN, se determinó analizar los siguientes genes: *BRCA1*, *BRCA2*, *TP53*, *PTEN*, *STK11*, *CDH1*, *MLH1*, *MSH2*, *MSH6*, *PMS2*, *EPCAM*, *CDKN2A*, *PALB2* y *APC* por su alto/moderado riesgo clínico y biológico en el desarrollo de esta enfermedad, así como por las opciones de tratamiento o intervención médica que se relacionan a los mismos.

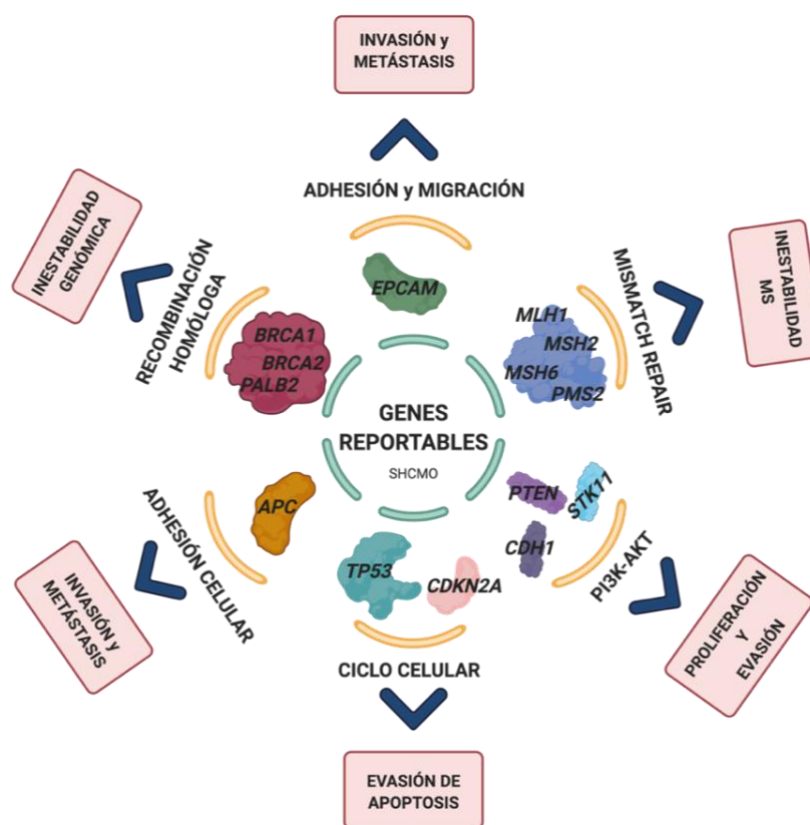


Figura 3. Diagrama de círculos concéntricos que en el primer anillo muestra los genes seleccionados agrupados de acuerdo a las vías metabólicas en las que participan. En el segundo anillo se muestran las vías metabólicas normales y en el último anillo se presentan los Hallmarks del cáncer de los que son partícipes cuando están mutados.

Análisis bioinformático:

Preprocesamiento

Los archivos Fastq.z obtenidos de la secuenciación se alinearon mediante el algoritmo BWA-mem y posteriormente se procesaron mediante la herramienta GATK4 hasta obtener los archivos Bam, los cuales contienen la información binaria comprimida del alineamiento de las secuencias (Van der Auwera, 2013). Como parte del diseño experimental, se clasificaron las muestras de acuerdo con la tecnología de secuenciación en la que fueron procesadas, es decir, dos grupos: *Miseq* y *Hiseq*.

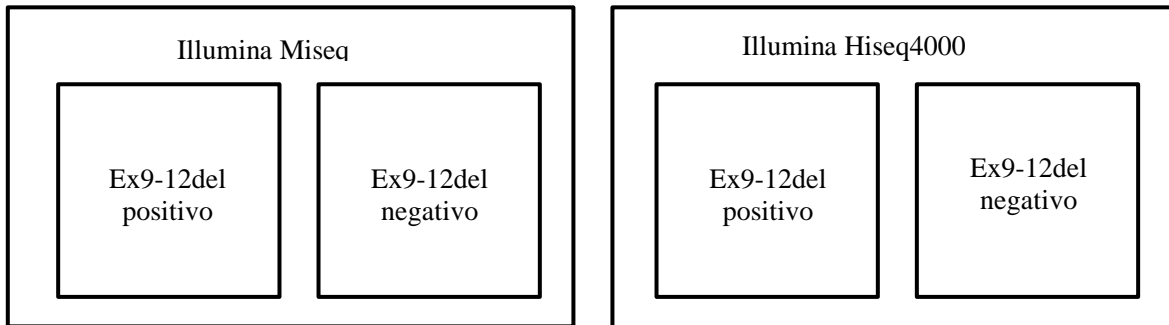


Figura 4: Diagrama que muestra cómo se agruparon las muestras y controles para el análisis bioinformático.

Controles

Como controles negativos (muestras normales), se utilizaron cinco muestras negativas a la Ex 9-12del previamente validadas por la técnica de reacción en cadena de la polimerasa (PCR). Por otro lado, como controles positivos se tomaron cinco muestras de diferentes corridas de secuenciación en las que se comprobó de manera experimental (PCR) la presencia de la Ex 9-12del.

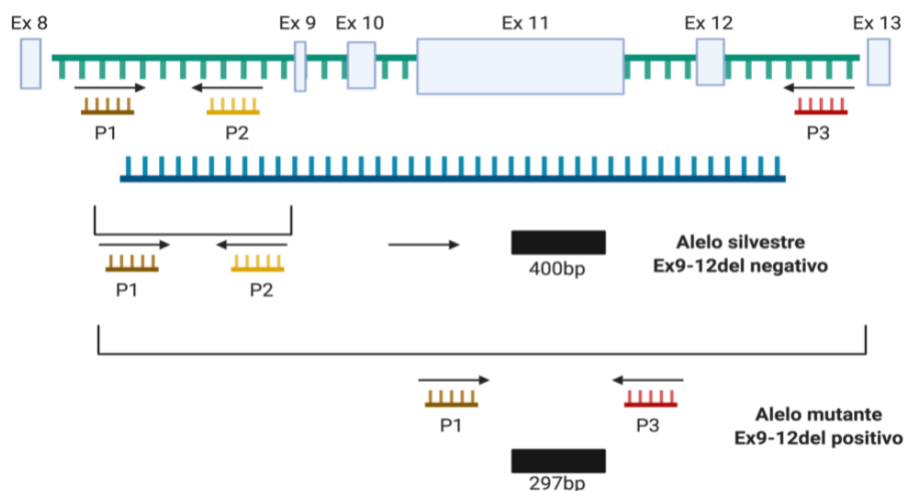


Figura 5: Diagrama que ilustra el diseño experimental para la detección de la ex9-12del (tomado y modificado de Quezada-Urban et al., 2018). En el diagrama se señalan los exones 8 a 13 (no a escala real) del gen BRCA1 y la posición de los 3 primers diseñados. Se muestra que al producirse una banda de 400 pb en una electroforesis, se considera un alelo silvestre. En cambio, si se produce una banda de 400 pb y otra banda de 297 pb, se reconoce que la muestra presenta la delección de los exones 9-12.

Estrategia de análisis bioinformático

Para el desarrollo de este trabajo, se probaron varios programas informáticos diseñados para analizar rearrreglos estructurales en muestras de línea germinal, como son: CNVkit, Panelcn.MOPS, ExomeDepth, DECoN, CNVNator, EXCAVATOR2, CANOES, CONTRA, por mencionar algunos. Dichos programas fueron seleccionados con base en estudios comparativos que evaluaron la eficiencia de detección de rearrreglos estructurales de varios programas, con lo que determinamos la lista de programas que se probó (Moreno-Cabrera *et. al.*, 2020; Gordeeva, *et. al.*, 2021)

Finalmente, en el estudio únicamente incluimos 3 programas: **CNVkit**, **Panelcn.MOPS** y **ExomeDepth**. Esta decisión se tomó por una serie de razones como son: la compatibilidad de los programas con el servidor que está a nuestra disposición, la facilidad o complejidad del uso y comprensión del código, el tiempo de ejecución del programa y el número de falsos positivos generados, además, por supuesto, de la compatibilidad de los programas con el tipo y calidad de información con la que contamos.

Dicho lo anterior, los programas seleccionados demostraron ser compatibles con el sistema e información a analizar, además de ser sencillos y rápidos de usar, y son lo suficientemente sensibles para detectar todos los controles positivos incluidos en este estudio.

CNVkit

En este trabajo se hizo uso del software *CNVkit* (Talevich, *et al.*, 2016) para identificar rearrreglos estructurales en los genes seleccionados del panel de secuenciación. Dicho software, basándose en la profundidad de secuenciación, determina el número de copias de un gen o región en particular, lo que permite establecer si existen o no rearrreglos estructurales en las regiones secuenciadas.

Para este estudio, se siguió el protocolo de análisis sugerido por los autores para analizar muestras germinales. Inicialmente el software integra los datos del archivo *bam* generado anteriormente junto con el archivo *bed* que contiene los locus que se secuenciaron (en este caso, las coordenadas genómicas y datos adicionales del panel de 143 genes) además de una referencia genética humana: *GRCh37 hg19*; además, incluye en el análisis de archivos *bam* de muestras normales (controles negativos). Esencialmente, el programa compara los datos de profundidad de la referencia y controles negativos con los datos proporcionados como muestras tumorales. La diferencia en cuanto a la profundidad de una región de la muestra tumoral con respecto a las muestras normales y la referencia indicaría un evento de cambio en el número de copias.

Al finalizar el análisis, se obtienen archivos segmentados por muestra, que incluyen información de las regiones analizadas con su coordenada genómica, los genes incluidos en dicha región, así como datos matemáticos como es el logaritmo base 2 con el que se representan las pérdidas o ganancias que se identificaron por región. Posteriormente se obtiene las estadísticas de calidad por segmento en cada muestra con los filtros *pi* (intervalo predictivo) y *ci* (intervalo de confianza), para posteriormente obtener los datos por cada gen y aplicar un filtro de 0.2 (mínimo necesario al cual los controles positivos se reportaban) para obtener los datos más confiables posibles y reducir al máximo el número de lecturas finales por gen.

ExomeDepth

ExomeDepth (Plagnol *et. al.*, 2012) es un modelo robusto para el análisis de la variación del número de copias a partir de datos de secuenciación de exomas. Este programa construye una referencia optimizada para maximizar la probabilidad de detectar rearrreglos estructurales en un set de muestras específico. Esto permite que el programa sea capaz de detectar las variaciones en el número de copias en diversos tipos de datos, aun cuando las variantes sean muy pequeñas (uno a dos exones), así como deleciones heterocigotas.

En este caso se siguió el protocolo sugerido por los autores.

Los archivos iniciales son: los bams de las muestras, los bams de los controles negativos, la referencia genómica formato fasta y el archivo bed que contiene las regiones objetivo de la secuenciación (estos mismos archivos se usan en los tres programas). El primer paso es generar una referencia específica para el set de muestras a analizar, para lo cual el programa toma en cuenta tanto los controles negativos como la referencia genómica general, además de los objetivos de secuenciación para obtener un panorama de las características de las muestras y su calidad.

Posteriormente se analizan las muestras empleando la referencia generada previamente para generar un archivo de lecturas final que reporta las lecturas obtenidas. En este archivo se tienen los datos de las variantes por exon y no por gen.

Finalmente se hace un filtrado manual eliminando las lecturas que se reportan como normales además de considerar otros aspectos como el efecto de lote (se discute a detalle posteriormente).

Panelcn.MOPS

Panelcn.MOPS (Povysil, *et. al.*, 2017) es un programa cuyo flujo de trabajo permite detectar CNVs en datos de secuenciación masiva por panel dirigido. La base matemática del programa se denomina *Copy Number estimation by a Mixture of Poisson* (Klambauer et al., 2012).

Al igual que *ExomeDepth* y *CNVkit*, los archivos iniciales son los archivos *bam* tanto de las muestras como de los controles negativos, la referencia en formato *fasta* y el archivo *bed* que contiene las regiones objetivo de la secuenciación.

Una de las ventajas principales de este programa, es que permite seleccionar uno o más genes en los que se desea enfocar el análisis, aun cuando la secuenciación se hizo por panel. Haciendo uso de esta característica, se seleccionaron únicamente los 14 genes de interés para este estudio, lo que reduce significativamente el número de lecturas en los resultados, aunque el filtrado posterior sigue siendo necesario. El resto de los parámetros se mantuvieron de acuerdo a lo sugerido por los autores.

Consideraciones de control de calidad de los datos

La profundidad de lectura por sí sola es un proxy insuficiente para el número de copias debido a los sesgos sistemáticos en la cobertura introducidos durante la preparación y secuenciación de la biblioteca. Para tener en cuenta cada uno de estos sesgos potenciales en la profundidad de la cobertura de lectura, los programas llevan a cabo la normalización a una referencia de una o más muestras normales representativas, con lo que los sesgos de cobertura sistemáticos pueden eliminarse en gran medida. Sin embargo, incluso después de la normalización a una referencia agrupada, los sesgos en la cobertura generalmente persisten en una muestra individual y aún deben eliminarse.

Los programas consideran cuatro principales factores por los cuales se puede presentar un sesgo en la lectura de y subsecuente análisis de los datos de secuenciación:

- Contenido GC
- Repeticiones
- Densidad de la región objetivo
- Sesgo por corrección computacional

Los cuáles son corregidos mediante métodos estadísticos y matemáticos propios de cada programa, así como el uso de las referencias genómicas y controles negativos incluidos en el análisis, lo que permite reportar resultados más certeros.

Filtrado

Aunque los programas son robustos, eficientes y fundamentados matemáticamente, se debe hacer un filtrado con el fin de reducir la variación en el análisis y el sesgo en cuanto al procesamiento bioinformático. El proceso de filtrado se hizo con el archivo finales que arroja cada programa. Inicialmente, se eliminaron los valores extremos atípicos para reducir la variación en el análisis. Posteriormente se eliminaron los resultados con un valor entero de 2, ya que esto indica que dicha región tiene las dos copias normales. Posteriormente se consideraron otros factores como el efecto de lote, que para efecto de este trabajo se define como “*la detección de la misma variante en el mismo gen en múltiples muestras de un mismo lote de secuenciación*”, ya que esto sugiere que se trata de un error en el proceso de secuenciación por lo que se consideran como falsos positivos. Finalmente se compararon los resultados arrojados por cada programa por muestra y únicamente se reportan las variantes que fueron detectadas por al menos 2 programas.

RESULTADOS

Controles

Para este estudio, se consideraron como controles positivos, aquellas muestras que conocemos que presentan la mutación fundadora en el gen *BRCA1 Ex9-12Del* de la población mexicana validada mediante PCR. Tanto el programa *CNVkit*, *ExomeDepth* y *Panelcn.MOPS* reportaron el **100%** de los controles positivos en los archivos de lectura finales, tanto para las muestras secuenciadas con *Illumina Miseq* como las secuenciadas con *Illumina Hiseq* (Figura 6).

De manera general, todos los programas fueron lo suficientemente sensibles para detectar la variante *ex9-12del* de los controles positivos, aunque casi no se reportan variantes que hayan detectado los 3 programas.

Análisis global

El panorama general de los resultados arroja que se detectaron 36 variaciones en el número de copias en un total de 34 muestras, que representan el 6.8% del total (considerando ambos grupos muestrales: *Illumina Miseq* e *Illumina Hiseq*).

Las variantes se detectaron en 12 de los genes de interés, siendo los más frecuentes *BRCA2* (1.4%, 7/500) * y *BRCA1* (1.2%, 6/500). Adicionalmente se detectaron CNVs en los siguientes genes, en menor proporción: *STK11* (1.2%, 6/500), *CDKN2A* (1%, 5/500), *ATM* (0.6%, 3/500), *MSH2* (0.6%, 3/500), *APC* (0.2%, 1/500), *CHEK2* (0.2%, 1/500), *PALB2* (0.2%, 1/500), *PMS2* (0.2%, 1/500), *PTEN* (0.2%, 1/500) y *TP53* (0.2%, 1/500).

Con respecto a la proporción entre amplificaciones y deleciones, se encontraron 17 deleciones y 17 amplificaciones.

No se detectó la presencia de la mutación fundadora *BRCA1 Ex9-12Del* en ninguna muestra que no fuera mexicana y tampoco se encontró esta variante en ninguna muestra que no se haya detectado antes mediante PCR, es decir, no se detectaron falsos positivos de esta variante.

A continuación, se describen los resultados por país a mayor detalle.

Guatemala

Con respecto a las muestras de Guatemala, únicamente contamos con 16 muestras de las cuales solo se detectó una variante en la muestra BR-23 en el gen *BRCA2* tratándose de una deleción detectada por dos de tres de los programas: *Panelcn.MOPS* y *ExomeDepth*. Así entonces, se detectaron CNVs en el **6.25%** de las muestras de Guatemala (Figura 6).

México

En el caso de las muestras de México, se analizaron dos grupos de muestras, las que secuenciaron con *Illumina Miseq* y las que se secuenciaron mediante *Illumina Hiseq*.

En el caso de las muestras que se secuenciaron con *Illumina Miseq*, se analizaron un total de 311 muestras. En este grupo, se detectaron un total de 13 variantes en 13 muestras, lo que representa el 4.18% del grupo muestral. La distribución y frecuencia de dichas variantes se detalla en la Figura 6.

El grupo de muestras de México que se secuenciaron mediante la plataforma *Illumina Hiseq* tuvo un total de 83 muestras, de las cuales se detectaron variantes en 11 de ellas (13.21%). En ninguna muestra se detectó más de una variante. Adicionalmente, cabe resaltar, que en las muestras con los identificadores IMSS04_150 e IMSS_124 se detectó una amplificación del gen *ATM* que fue reportada por los 3 programas.

Perú

Las muestras de Perú fueron secuenciadas mediante la plataforma *Illumina Hiseq*, siendo un total de 40 muestras. De dichas muestras, únicamente se reportan 6 variantes en 6 muestras, es decir, el 15%.

Colombia

Finalmente, con respecto a las muestras de Colombia (secuenciadas mediante *Illumina Hiseq*), de un total de 50 muestras se detectaron variantes en 7 muestras, lo que representa el 14% de las muestras con alguna variante.

Las muestras de Colombia son el único grupo en el que se reportan muestras con más de una variante: Col_545 con variantes en el gen *STK11* y *CDKN2A*, y Col_529 con variantes en *STK11* y *ATM*, siendo todas las variantes pérdidas en el número de copias. Así mismo, la pérdida en el gen *STK11* de la muestra Col_545 fue detectada por los 3 programas.

Generación de un código único para el análisis automatizado

Para facilitar el análisis haciendo uso de los 3 programas, se diseñó un código único mediante el lenguaje de programación *bash* capaz de reunir los códigos de los 3 programas sin importar que dos de ellos funcionen con R (*ExomeDepth* y *Panelcn.MOPS*) y uno con python (*CNVkit*). El programa funciona reuniendo los archivos iniciales necesarios para los 3 programas, es decir, los archivos bams de muestras y controles, el archivo bed de objetivos de secuenciación, el archivo fasta de la referencia genómica, etc. Y comienza el análisis con *CNVkit* para todas las muestras y exporta los archivos de resultados para continuar con los dos programas escritos en R y así analizar todas las muestras de manera automática con los 3 programas.

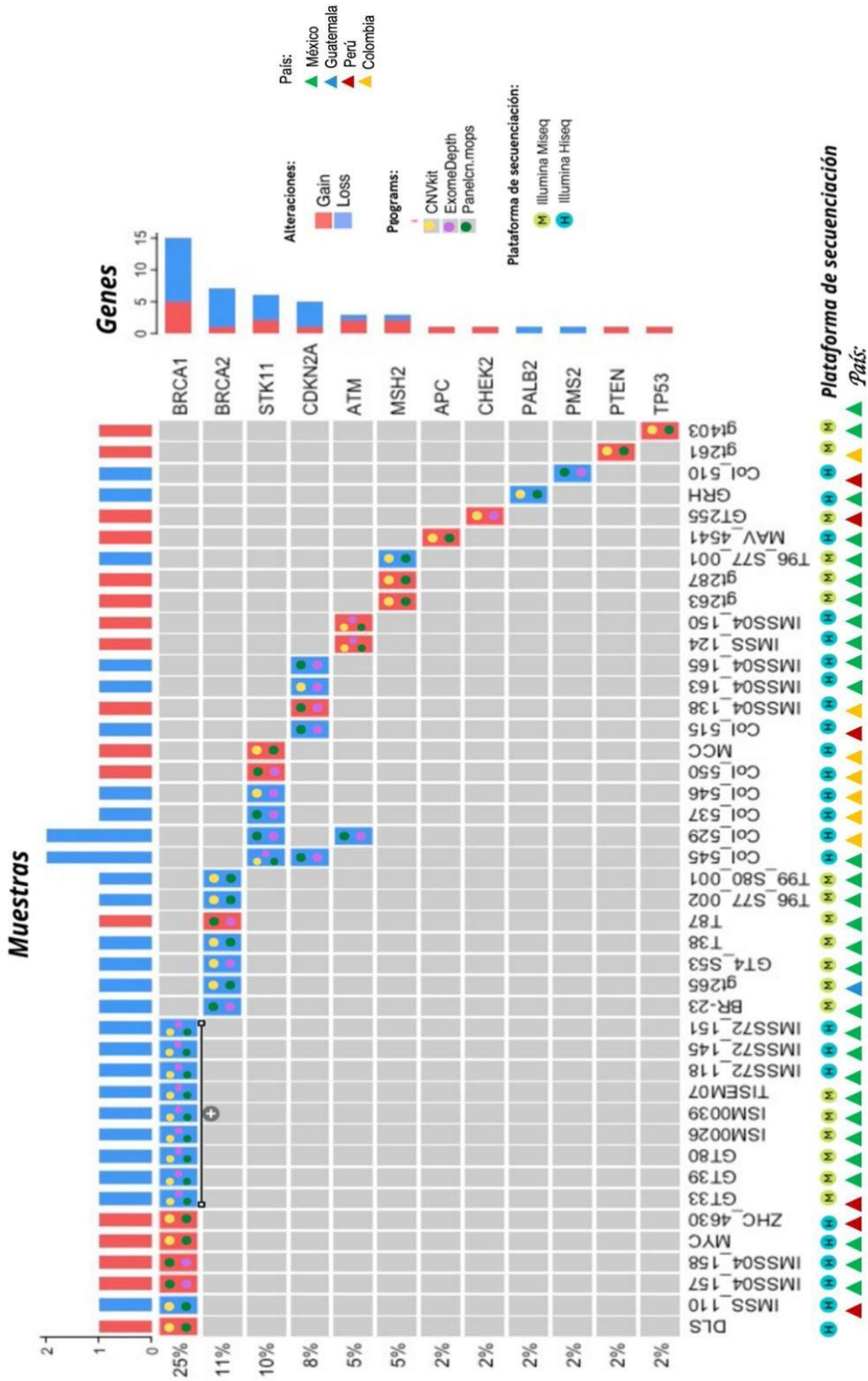


Figura 6. Gráfico de tipo *Oncoprint* que muestra la distribución de las variantes en el número de copias en las muestras y en el eje Y están los genes ordenados de forma descendente de acuerdo con el número de variantes que se detectaron en cada gen. En azul se indican las pérdidas y en rojo las ampliaciones. Los puntos de colores añadidos sobre cada variante indican los programas con los que se detectó dicha variante. En el eje X en la parte superior de la gráfica se muestra el número de variantes por muestra en el mismo eje X en la zona inferior se indica la plataforma de secuenciación con la que se secuenciaron las muestras. Así mismo, el país de origen de cada muestra se señala con un triángulo en la parte inferior. Además, se indican los controles positivos con un símbolo +.

DISCUSIÓN

El cáncer de mama es un tipo de cáncer que surge del tejido mamario que usualmente inicia en los ductos lactíferos, aunque en algunos casos, se presenta un cáncer de tipo *lobular*, que se origina en las glándulas que producen la leche materna. En casos raros, se pueden presentar *tumores filodes* o *angiosarcomas* en el seno. Además, el cáncer puede originarse de otros tejidos mamarios como *sarcomas* y *linfomas*, aunque estos no se consideran del todo como cáncer de mama (Veronesi *et al.*, 2017).

Particularmente, en la región de América Latina y el Caribe en el año 2018, el cáncer de mama (considerando ambos sexos y todas las edades) ocupa el primer lugar afectando a 2,288,283 personas, seguido por el cáncer de pulmón y colorrectal. En cuanto a la mortalidad, el primer lugar lo ocupa el cáncer de pulmón, que, en 2018, cobró la vida de 1,842,391 personas en América Latina. Sin embargo, cuando únicamente se consideran las estadísticas de incidencia y mortalidad de cáncer en mujeres, es claro que el primer lugar lo ocupa el cáncer de mama, que afectó a 2,288,283 mujeres y cobró la vida de 679,237 de ellas en 2018 (Bray, *et al.*, 2018).

El Síndrome Hereditario de Cáncer de Mama y Ovario (SHCMO) se define como un desorden hereditario aumenta significativamente el riesgo de padecer cáncer de mama u ovario (particularmente antes de los 50 años), con respecto a personas que no presentan el síndrome. La herencia de este genotipo es de tipo mendeliana, autosómica dominante. Sin embargo, del total de casos de cáncer de mama, entre el 5 y el 10% son a causa de mutaciones germinales hereditarias (Kida, Murai & Yamauchi, 2017). La mayoría de los estudios de los estudios genéticos del SHCMO se han concentrado en el análisis de variantes de nucleótido único, y aquellos que han evaluado LR lo han hecho principalmente en *BRCA1/2* o en genes únicos. Considerando que la tasa de LR en pacientes con SHCMO puede ser de alrededor del 7%, es posible que muchas pacientes no portadoras de LR no sean analizadas para la detección de estas alteraciones, resultando en falsos negativos (Kumaran *et al.*, 2017, Judkins *et al.*, 2012 & Mancini-DiNardo *et al.*, 2019).

Se conoce que en lo que respecta al SHCMO, los cambios en el número de copias tales como la *ex 9-12del* en México, pueden ser específicas para determinadas poblaciones y producen el mismo fenotipo de cáncer. Tener información con respecto a las variantes únicas de una población, así como aquellas variantes que son compartidas entre distintas poblaciones puede ser útil en el ámbito de los estudios genéticos clínicos para los profesionales de la salud y los pacientes. Además, estos datos pueden proporcionar información relevante para la consejería genética que se brinda a los pacientes que han sido diagnosticados con cáncer de mama y ovario de origen hereditario, al igual que sus familias.

Particularmente, en referencia a los análisis estadísticos e informáticos enfocados al estudio de rearrreglos estructurales en general enfrentan varios problemas que dificultan significativamente el análisis de estas variantes por lo que actualmente el estándar para detectar CNVs continúan siendo métodos experimentales como el *MLPA*.

De acuerdo con la revisión de la literatura, hasta el momento, este trabajo es el primero en analizar rearrreglos estructurales en muestras de línea germinal (n = 500) de pacientes de 4 países de América Latina (México, Perú, Colombia y Guatemala) en genes además de *BRCA1* y *BRCA2* y tomando como datos iniciales los archivos bam de secuenciación masiva por panel.

En el presente trabajo, los resultados indican que se detectaron 36 variaciones en el número de copias en un total de 34 muestras, que representan el 6.8% del total (considerando ambos grupos muestrales: Miseq e Hiseq), sin considerar las muestras que presentan la mutación fundadora *Ex9-12del*, lo cual corresponde con lo reportado en los trabajos de Judkins *et al.* (2012), y Mancini Dinardo, *et al.* (2019).

De manera general, todos los programas fueron lo suficientemente sensibles para detectar la variante *Ex9-12del* de los controles positivos, aunque casi no se reportan variantes que hayan detectado los 3 programas.

Con respecto a la proporción entre amplificaciones y deleciones, se encontraron 17 deleciones y 17 amplificaciones. Adicionalmente, con respecto a las plataformas de secuenciación, 13 de las muestras reportadas fueron secuenciadas mediante la plataforma Illumina Miseq, y 21 muestras fueron secuenciadas con Illumina Hiseq, lo cual es razonable al considerar las diferencias técnicas entre ambas plataformas y sus capacidades en cuanto a la profundidad y calidad de lectura (que es una de las bases más importantes para el análisis con los 3 programas empleados).

Considerando los antecedentes de este trabajo, los genes que presentaron variantes con mayor frecuencia fueron los genes *BRCA1* y *BRCA2*, como se reportó en los trabajos de Judkins *et al.* (2012) y Mancini Dinardo, *et al.* (2016) donde encontraron que la mayor proporción de rearrreglos estructurales se detectó en el gen *BRCA1* (26.4%) seguido de otros genes de alto riesgo como *PMS2* (11.7%), *CHEK2* (11.1%) y *MSH2* (8.9%). Adicionalmente, los autores mencionan que los resultados referentes al gen *PMS2* son de especial interés ya que es un gen que tiene una homología significativa con su pseudogen *PMS2CL* que puede causar errores en el proceso de análisis y secuenciación. Sin embargo, en el trabajo de Mancini Dinardo y colaboradores, también hacen énfasis en la prevalencia de los rearrreglos estructurales en los genes reportados, especialmente en el gen *STK11*. Los autores detectaron una prevalencia del 60.7% de rearrreglos estructurales en el gen *STK11* de todas las variantes patogénicas detectadas en este gen. De igual forma, en el presente trabajo, se encontró que uno de los genes en el que se detectaron más variantes es *STK11* (6 muestras). En este sentido, el gen *STK11* ha sido comúnmente estudiado en relación con otras enfermedades o síndromes distintos al cáncer de mama y ovario de tipo hereditario, como es el síndrome de Peutz-Jaeghers que se define como una enfermedad de herencia autosómica dominante que causa múltiples pólipos hemartomatosos en el estómago, intestino delgado y colon, además de lesiones cutáneas de alta pigmentación. Esta reportado que la mayoría (66-94%) de los casos parecen ser causados por una mutación en la línea germinal del gen supresor tumoral *STK11/LKB1* (serina/treonina cinasa 11). Los pacientes diagnosticados con este síndrome presentan un riesgo significativamente mayor de cánceres digestivos y no digestivos. Los cánceres digestivos comprometen el páncreas, el intestino delgado, el colon y el estómago. Los cánceres no digestivos incluyen los que afectan la mama, el pulmón, el útero, los ovarios y los testículos (Kaminsky, E., Kaul, V., Paschall, J. *et al.*, 2011). En este sentido, la alta prevalencia y frecuencia de CNVs en este gen, sugiere que puede tener una influencia importante en el desarrollo de cáncer de mama y ovario de origen germinal en pacientes de América Latina, por lo que considero importante que se continúe el estudio de rearrreglos estructurales en este gen, con un enfoque particular en el SHCMO, especialmente en las muestras de América del sur (Perú y Colombia) ya que la mayoría de las variantes reportadas son de muestras de estos países además de lo reportado por otros autores.

En cuanto a los aspectos técnicos del análisis, el uso de más de un programa, redujo el número de variantes reportadas e hizo el filtrado más sencillo y menos arbitrario. Adicionalmente, una complicación con respecto a los resultados obtenidos, es que los programas no generan los reportes escritos y estructurados de la misma forma, lo que puede dificultar la comparación de los resultados de cada programa, un claro ejemplo, es que los loci de las variantes pueden tener ligeras diferencias dependiendo del programa que se esté utilizando. Este trabajo sienta las bases para la implementación de un protocolo para la detección de rearrreglos estructurales en la línea germinal de pacientes con SHCMO a partir de datos de secuenciación masiva, sin la limitación que tienen otros métodos como el MLPA que no permiten analizar todas las regiones o genes de interés debido a la disponibilidad de kits y del precio de los mismos. Por lo anterior, el análisis computacional de CNV representa un avance importante en la detección de estas variantes de difícil identificación en un mayor número de genes y muestras, y a su vez representa una mejora en el diagnóstico molecular y la descripción epidemiológica de esta enfermedad.

CONCLUSIÓN:

1. Se detectaron variantes en el número de copias en el 6.8% de las muestras sin considerar la mutación fundadora *ex9-12del* y considerando todas las muestras por igual (aquellas secuenciadas mediante Illumina miseq e Illumina Hiseq).
2. Los programas empleados para el análisis fueron capaces de detectar la mutación fundadora *ex9-12del* en todas las muestras control positivo seleccionadas para el análisis.
3. En cuanto a las muestras por país, la frecuencia de CNVs se distribuye de la siguiente manera: México (Hiseq) = 11 muestras, México (Miseq) = 13 muestras, Guatemala = 1 muestra, Perú = 6 muestras y Colombia = 7 muestras.
4. Se detectaron variantes en 12 de los 14 genes de interés (reportando únicamente aquellas variantes detectadas por más de un programa), los cuales son: *BRCA2* (1.4%, 7/500), *BRCA1* (1.2%, 6/500), *STK11* (1.2%, 6/500), *CDKN2A* (1%, 5/500), *ATM* (0.6%, 3/500), *MSH2* (0.6%, 3/500), *APC* (0.2%, 1/500), *CHEK2* (0.2%, 1/500), *PALB2* (0.2%, 1/500), *PMS2* (0.2%, 1/500), *PTEN* (0.2%, 1/500) y *TP53* (0.2%, 1/500).
5. El protocolo de filtrado redujo de manera importante el número de variantes reportadas tanto para las muestras secuenciadas con Miseq y Hiseq.
6. El código único diseñado y aplicado en este trabajo, probó ser efectivo para este análisis, facilitando el uso de los 3 programas para analizar numerosas muestras, además de ser más amigable con los usuarios ya que reduce la necesidad de conocer ambos lenguajes de programación (R y python).

PERSPECTIVAS:

1. Una de las principales perspectivas para futuros trabajos centrados en el estudio de CNV en la población de América Latina es usar controles positivos de otras poblaciones, además de la de México, en el análisis. Con esto, se pueden obtener resultados más certeros para otras poblaciones.
2. Otra de las perspectivas a considerar es la validación experimental de las variantes detectadas en este trabajo mediante PCR, MLPA o microarreglos según sea más conveniente.
3. Analizar otras poblaciones en América Latina para obtener un panorama más completo en cuanto a las variaciones en el número de copias en pacientes con SHCMO.
4. Finalmente, otra perspectiva es la publicación del presente trabajo en una revista científica indexada, ya sea de forma aislada, o como parte complementaria de otro estudio.

BIBLIOGRAFÍA:

1. Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P. (2002). *Molecular biology of the cell* (4th ed.). Garland Science Taylor & Francis.
2. American Cancer Society. Cancer Statistics Center. <http://cancerstatisticscenter.cancer.org>. Accessed 11/21/2020.
3. American Society of Clinical Oncology. *Hereditary Breast and Ovarian Cancer*. Cancer.Net. Retrieved 21 January 2021, from <https://www.cancer.net/cancer-types/hereditary-breast-and-ovarian-cancer>.
4. Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., & Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*, 68(6), 394-424.
5. Del Valle, J., Feliubadaló, L., Nadal, M., Teulé, A., Miró, R., Cuesta, R., ... & Lázaro, C. (2010). Identification and comprehensive characterization of large genomic rearrangements in the *BRCA1* and *BRCA2* genes. *Breast cancer research and treatment*, 122(3), 733-743.
6. Ewald, I. P., Ribeiro, P. L., Palmero, E. I., Cossio, S. L., Giugliani, R., & Ashton-Prolla, P. (2009). Genomic rearrangements in *BRCA1* and *BRCA2*: A literature review. *Genetics and molecular biology*, 32(3), 437-446. <https://doi.org/10.1590/S1415-47572009005000049>
7. Fachal, L., Blanco, A., Santamariña, M., Carracedo, A., & Vega, A. (2014). Large Genomic Rearrangements of *BRCA1* and *BRCA2* among Patients Referred for Genetic Analysis in Galicia (NW Spain): Delimitation and Mechanism of Three Novel *BRCA1* Rearrangements. *PLOS ONE*, 9(3), e93306. <https://doi.org/10.1371/JOURNAL.PONE.0093306>
8. Ford, D., Easton, D. F., Stratton, M., Narod, S., Goldgar, D., Devilee, P., Bishop, D. T., Weber, B., Lenoir, G., Chang-Claude, J., Sobol, H., Teare, M. D., Struwing, J., Arason, A., Scherneck, S., Peto, J., Rebbeck, T. R., Tonin, P., Neuhausen, S., ... Zelada-Hedman, M. (1998). Genetic heterogeneity and penetrance analysis of the *BRCA1* and *BRCA2* genes in breast cancer families. The Breast Cancer Linkage Consortium. *American Journal of Human Genetics*, 62(3), 676-689. <https://doi.org/10.1086/301749>
9. Gordeeva, V., Sharova, E., Babalyan, K. *et al.* Benchmarking germline CNV calling tools from exome sequencing data. *Sci Rep* 11, 14416 (2021). <https://doi.org/10.1038/s41598-021-93878-2>
10. Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of cancer: The next generation. *Cell*, 144(5), 646-674. <https://doi.org/10.1016/J.CELL.2011.02.013/ATTACHMENT/3F528E16-8B3C-4D8D-8DE5-43E0C98D8475/MMC1.PDF>
11. Johnson, K. S., Conant, E. F., & Soo, M. S. (2021). Molecular subtypes of breast cancer: a review for breast radiologists. *Journal of Breast Imaging*, 3(1), 12-24.
12. Judkins, T., Rosenthal, E., Arnell, C., Burbidge, L. A., Geary, W., Barrus, T., Schoenberger, J., Trost, J., Wenstrup, R. J., & Roa, B. B. (2012). Clinical

- significance of large rearrangements in BRCA1 and BRCA2. *Cancer*, 118(21), 5210–5216. <https://doi.org/10.1002/CNCR.27556>
13. Kaminsky, E., Kaul, V., Paschall, J. *et al.* An evidence-based approach to establish the functional and clinical significance of copy number variants in intellectual and developmental disabilities. *Genet Med* **13**, 777–784 (2011). <https://doi.org/10.1097/GIM.0b013e31822c79f9>
 14. Key, T. J., Verkasalo, P. K., & Banks, E. (2001). Epidemiology of breast cancer. *Lancet Oncology*, 2(3), 133–140. [https://doi.org/10.1016/S1470-2045\(00\)00254-0](https://doi.org/10.1016/S1470-2045(00)00254-0)
 15. Kida, K., Murai, M., & Yamauchi, H. (2017). Diagnosis and Treatment of HBOC Syndrome by a Breast Surgical Oncologist. *Japanese Journal of Cancer and Chemotherapy*, 44(2), 111–115. Retrieved from <https://pubmed.ncbi.nlm.nih.gov/28223669/>
 16. Layde, P. M., Webster, L. A., Baughman, A. L., Wingo, P. A., Rubin, G. L., & Ory, H. W. (1989). The independent associations of parity, age at first full term pregnancy, and duration of breastfeeding with the risk of breast cancer. Cancer and Steroid Hormone Study Group. *Journal of Clinical Epidemiology*, 42(10), 963–973. [https://doi.org/10.1016/0895-4356\(89\)90161-3](https://doi.org/10.1016/0895-4356(89)90161-3)
 17. Liede, A., Karlan, B. Y., & Narod, S. A. (2004). Cancer risks for male carriers of germline mutations in BRCA1 or BRCA2: a review of the literature. *Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology*, 22(4), 735–742. <https://doi.org/10.1200/JCO.2004.05.055>
 18. Mancini-Dinardo, D., Judkins, T., Kidd, J., Bernhisel, R., Daniels, C., Brown, K., Meek, K., Craft, J., Holladay, J., Morris, B., & Roa, B. B. (2019). Detection of large rearrangements in a hereditary pan-cancer panel using next-generation sequencing. *BMC Medical Genomics*, 12(1), 1–9. <https://doi.org/10.1186/S12920-019-0587-3/FIGURES/5>
 19. Mazoyer S. Genomic rearrangements in the BRCA1 and BRCA2 genes (2005). *Hum Mutat.* 25:415–422.
 20. Moreno-Cabrera, J.M., del Valle, J., Castellanos, E. *et al.* Evaluation of CNV detection tools for NGS panel data in genetic diagnostics. *Eur J Hum Genet* 28, 1645–1655 (2020). <https://doi.org/10.1038/s41431-020-0675-z>
 21. National Comprehensive Cancer Network. (2020). Genetic/Familial High-risk assessment: breast, ovarian and pancreatic (V2.21) Retrieved from https://www.nccn.org/professionals/physician_gls
 22. Nielsen, F. C., van Overeem Hansen, T., & Sørensen, C. S. (2016). Hereditary breast and ovarian cancer: new genes in confined pathways. *Nature Reviews Cancer* 2016 16:9, 16(9), 599–612. <https://doi.org/10.1038/nrc.2016.72>
 23. Ossa Gomez, C. A., Achatz, M. I., Hurtado, M., Sanabria-Salas, M. C., Sullcahuaman, Y., Chávarri-Guerra, Y., ... & Ashton-Prolla, P. (2022). Germline Pathogenic Variant Prevalence Among Latin American and US Hispanic Individuals Undergoing Testing for Hereditary Breast and Ovarian Cancer: A Cross-Sectional Study. *JCO Global Oncology*, 8, e2200104.
 24. Palma, M. D., Domchek, S. M., Stopfer, J., Erlichman, J., Siegfried, J. D., Tigges-Cardwell, J., ... Nathanson, K. L. (2008). *The Relative Contribution of Point Mutations and Genomic Rearrangements in BRCA1*

- and BRCA2 in High-Risk Breast Cancer Families. Cancer Research, 68(17), 7006–7014. doi:10.1158/0008-5472.can-08-0599*
25. Perou CM, Sørli T, Eisen MB, et al. Molecular portraits of human breast tumours. *Nature* 2000;406(6797):747–752.
 26. Plagnol V., Curtis J., Michael Epstein, Kin Y. Mok, Emma Stebbings, Sofia Grigoriadou, Nicholas W. Wood, Sophie Hambleton, Siobhan O. Burns, Adrian J. Thrasher, Dinakantha Kumararatne, Rainer Doffinger, Sergey Nejentsev, A robust model for read count data in exome sequencing experiments and implications for copy number variant calling, *Bioinformatics*, Volume 28, Issue 21, 1 November 2012, Pages 2747–2754, <https://doi.org/10.1093/bioinformatics/bts526>
 27. Povysil, G., Tzika, A., Vogt, J., Haunschmid, V., Messiaen, L., Zschocke, J., Klambauer, G., Hochreiter, S., & Wimmer, K. (2017). panelcn.MOPS: Copy-number detection in targeted NGS panel data for clinical diagnostics. *Human mutation, 38(7)*, 889–897. <https://doi.org/10.1002/humu.23237>
 28. Qian, Y., Mancini-DiNardo, D., Judkins, T., Cox, H. C., Brown, K., Elias, M., Singh, N., Daniels, C., Holladay, J., Coffee, B., Bowles, K. R., & Roa, B. B. (2017). Identification of pathogenic retrotransposon insertions in cancer predisposition genes. *Cancer Genetics, 216–217*, 159–169. <https://doi.org/10.1016/J.CANCERGEN.2017.08.002>
 29. Quezada Urban, R., Díaz Velásquez, CE., et al. (2018). Comprehensive Analysis of Germline Variants in Mexican Patients with Hereditary Breast and Ovarian Cancer Susceptibility. *Cancers. 10(10)*, 361. <https://doi.org/10.3390/cancers10100361>
 30. Stuppia, L., Antonucci, I., Palka, G., & Gatta, V. (2012). Use of the MLPA assay in the molecular diagnosis of gene copy number alterations in human genetic diseases. *International Journal of Molecular Sciences, 13(3)*, 3245–3276. <https://doi.org/10.3390/IJMS13033245>
 31. Tang P, Tse GM. Immunohistochemical surrogates for molecular classification of breast carcinoma: a 2015 update. *Arch Pathol Lab Med* 2016;140(8):806–814.
 32. Tomasetti, C., Li, L., & Vogelstein, B. (2017). Stem cell divisions, somatic mutations, cancer etiology, and cancer prevention. *Science (New York, N.Y.), 355(6331)*, 1330–1334. <https://doi.org/10.1126/SCIENCE.AAF9011>
 33. Veronesi, U., Goldhirsch, A., Veronesi, P., Gentilini, O. D., & Leonardi, M. C. (Eds.). (2017). *Breast cancer: Innovations in research and management*. springer.
 34. Weitzel J.N., Lagos V.I., Herzog J.S., Judkins T., Hendrickson B., Ho J.S., Ricker C.N., Lowstuter K.J., Blazer K.R., Tomlinson G., et al (2007). Evidence for common ancestral origin of a recurring BRCA1 genomic rearrangement identified in high-risk Hispanic families. *Cancer Epidemiol Biomarkers Prev.*16:1615–1620.
 35. Wu, S., Zhu, W., Thompson, P., & Hannun, Y. A. (2018). Evaluating intrinsic and non-intrinsic cancer risk factors. *Nature Communications, 9(1)*. <https://doi.org/10.1038/S41467-018-05467-Z>