

IDENTIFICACION

Título

Modelamiento *in vitro* por CRISPR/Cas9 de una mutación VUS encontrada en pacientes peruanas diagnosticadas con cáncer de mama hereditario

In vitro modelling with CRISPR/Cas9 of a VUS mutation found in Peruvian patients diagnosed with hereditary breast cancer

Nombre y apellido del autor o los autores

1. Solange Rosa Paredes Moscosso
2. Claudio Nicolas Villegas Llerena
3. Bernardo Esteban Quispe Bravo
4. Rodrigo Sánchez Macedo
5. Maria Luisa Guevara-Fujita
6. Ricardo Fujita*

Código ORCID de cada autor

1. 0000-0001-8461-2546
2. 0000-0002-2634-5504
3. 0000-0002-9361-7709
4. 0000-0003-4189-4963
5. 0000-0001-5457-231X
6. 0000-0002-9617-5109

Correo electrónico de cada autor

1. sparedesm@usmp.pe
2. cvillegasl1@usmp.pe
3. bquispeb@usmp.pe
4. rodrigo.sanchez@ombquality.com
5. mquevarag@usmp.pe
6. rfujitaa@usmp.pe

Afiliación institucional de los autores

Universidad de San Martín de Porres, Facultad de Medicina Humana, Instituto de Investigación, Centro de Investigación de Genética y Biología Molecular, Lima, Perú.

Resumen

El cáncer de mama es la neoplasia más común diagnosticada en mujeres alrededor del mundo y el segundo tipo de cáncer más frecuente en Perú. Hasta un 10% de casos son hereditarios, donde la mayoría corresponde a mutaciones en los genes supresores de tumor *BRCA1* y *BRCA2*. De dichas mutaciones, aproximadamente el 50% generan "Variantes Genéticas de Significado Indeterminado" (*Variants of Uncertain Significance*, VUS). Recientemente, hemos reportado la identificación de seis nuevas VUS

encontradas en familias peruanas con historia de cáncer hereditario, de las cuales, la variante c.5074+28T>A fue detectada en 07 familias. El objetivo de este estudio fue modelar dicha VUS en una línea celular de tejido mamario sano (SVCT) mediante la tecnología CRISPR/Cas9. Con este fin, las células SVCT fueron transfectadas por electroporación con los complejos de ARN guía y la enzima Cas9. Dicha transfección fue comprobada por citometría de flujo, el ensayo con la enzima T7EI y secuenciación Sanger. Como resultado, se obtuvieron clones (single-cell) por dilución limitante. El ADN de dichos clones fue evaluado por PCR y digestión enzimática. Los clones positivos fueron seleccionados y analizados por secuenciación Sanger, donde se identificó 01 clon portador de la VUS c.5074+28T>A. Asimismo, se modeló la variante patogénica c.5074+1G>A para usarla como control positivo en futuros estudios. De esta última variante, se obtuvieron 02 clones positivos en heterocigosis. Los resultados obtenidos sirven como prueba concepto para desarrollar una plataforma de modelado con CRISPR/Cas9 y el estudio de VUS de pacientes peruanas con cáncer de mama u otras enfermedades.

Abstract

Breast cancer is the most common neoplasia diagnosed in women around the world and the second most frequent in Peru. Up to 10% of cases are hereditary, most of which correspond to mutations in the *BRCA1* and *BRCA2* tumour suppressor genes. Of these mutations, approximately 50% generate Variants of Uncertain Significance (VUS). Recently, we have reported the identification of 06 new VUS found in Peruvian families with a history of hereditary cancer, of which the c.5074+28T>A variant was detected in 07 families. Therefore, the aim of this study was to model said VUS in a healthy breast tissue cell line (SVCT) using CRISPR/Cas9 technology. For this purpose, SVCT cells were transfected by electroporation with guide RNA complexes and Cas9 enzyme. Such transfection was verified by flow cytometry, the T7EI enzyme assay and Sanger sequencing. As a result, single-cell clones were obtained by limiting dilution, which were then expanded and cryopreserved. In parallel, the DNA of these clones was assessed by PCR and enzymatic digestion. The positive clones were selected and analysed by Sanger sequencing, where 01 clone carrying the VUS c.5074+28T>A was identified. Likewise, the pathogenic variant c.5074+1G>A was modelled to be used as a positive control in future studies. Of this last variant, 02 positive clones in heterozygosis were obtained. These results serve as a proof of concept for developing a cell modelling platform with CRISPR/Cas9 and the study of VUS from Peruvian patients with breast cancer or other diseases.

Palabras clave: Cáncer hereditario, VUS, pacientes peruanos, modelamiento *in vitro*, CRISPR/Cas9, línea celular SVCT.

Información sobre la contribución:

1. SRPM: Conceptualización, Escritura - Preparación del borrador original; Edición e Investigación.

2. CNVL: Conceptualización, Escritura - Revisar & Edición e Investigación.
3. BEQB: Investigación.
4. RSM: Investigación.
5. MLGF: Administración del proyecto y Supervisión .
6. RF: Conceptualización y Escritura - Revisar & Edición.

Información sobre conflicto de intereses: Los autores declaran no incurrir en ningún conflicto de interés.

Fuentes de financiamiento: Los autores agradecen el financiamiento de la Universidad de San Martín de Porres (Lima, Perú) y PROCIENCIA (a través del contrato N°153-2018-FONDECYT-BM-IADT-MU).

Agradecimientos: Los autores agradecen el estudio previo del Dr. José Luis Buleje Sono, trabajo sobre el cual se han sentado las bases de esta investigación. Al Dr. Joel de León por su apoyo en la implementación del laboratorio de cultivo celular y a los revisores anónimos por la revisión del borrador de este manuscrito que, indudablemente, contribuyen a su mejora.

Sobre ética en investigación y normas legales: Los autores declaran no haber incurrido en aspectos antiéticos ni haber omitido normas legales.

EL CUERPO DEL TRABAJO

1.1 Introducción

El cáncer de mama es la neoplasia más común diagnosticada en mujeres alrededor del mundo. Es también la principal causa de muerte por cáncer, con aproximadamente un millón de nuevos casos reportados y más de medio millón de muertes, según último reporte del GLOBOCAN realizado a nivel global en el 2018. Más aún, se estima que para el 2030, la carga de esta neoplasia se verá incrementada en 70% debido a cambios demográficos ¹.

La región de América Central y del Sur ha experimentado una transición epidemiológica y demográfica provocada por un rápido crecimiento económico, y actualmente padece una mayor carga de cáncer de mama. En

efecto, en el Perú, esta neoplasia es la segunda más común en mujeres. Sin embargo, ocupa el primer lugar en regiones como Lima y Arequipa ².

Un factor importante en el riesgo subyacente de cáncer de mama es la predisposición genética. Actualmente, se considera que hasta un 10% de todos los casos de esta enfermedad se deben a un alelo de susceptibilidad autosómica dominante ^{3; 4}. Entre ellos, alelos mutantes en los genes supresores de tumor *BRCA1* y *BRCA2* son los principales factores de susceptibilidad al cáncer de mama hereditario, los cuales predisponen al desarrollo precoz de tumores no solo de mama sino también de ovarios ⁵.

El secuenciamiento de siguiente generación (*Next-Generation Sequencing, NGS*) nos permite generar e interrogar data genómica a una escala sin precedentes. Sin embargo, el estudio de variantes genéticas en cohortes poco estudiadas -tales como la población peruana- tiende a resultar en la detección de una abundancia de variantes genéticas de significado incierto (*Variants with Uncertain Significance, VUS*).

Una mutación VUS es una alteración en la secuencia de genes, cuyo impacto sobre la función del producto del gen o sobre el riesgo de causar la enfermedad es desconocido ⁶, ya sea porque no se cuenta con suficiente evidencia que sea concluyente o porque dicha mutación nunca fue investigada. En muchos casos, las pruebas funcionales o su recurrencia en otras familias comprueban o descartan su causalidad en cáncer.

Según nuestros propios hallazgos, alrededor del 80% de variantes identificadas en pacientes con cáncer de mama hereditario corresponde a mutaciones tipo VUS (manuscrito en preparación). Dado que el impacto clínico de dichas variantes se desconoce, resulta apremiante la generación de modelos celulares que permitan su estudio *in vitro*.

En un estudio previo, nuestro centro de investigación (CIGBM) identificó 06 nuevas VUS en el gen *BRCA1* en 18 familias peruanas con diagnóstico de cáncer hereditario de mama y ovario ⁷. De estas, la variante intrónica c.5074+28T>A se encontró consistentemente en 07 familias. La frecuencia de esta variante en bases de datos poblacionales arrojó que no se trataba de un

polimorfismo. Estas características la convierten en una variante de interés. Por lo que, el objetivo de este estudio fue modelar, mediante la técnica CRISPR/Cas9, la VUS c.5074+28T>A en una línea celular de tejido mamario sano (SVCT).

1.2 Materiales y Métodos

1.2.1 Cultivo celular

Las células SVCT provienen del tejido mamario epitelial de una voluntaria del sexo femenino. Se caracterizan por ser inmortalizadas con el virus SV40. No son tumorigénicas en modelos murinos y expresan las proteínas de Keratina 7, 8 y 18. Para este estudio, la línea celular SVCT fue adquirida del repositorio *European Collection of Authenticated Cell Cultures* (ECACC), a través de Sigma-Aldrich (Merck). Se recibió un vial de células a -196C y se procedió a descongelar y cultivar según instrucciones del proveedor (https://www.sigmaaldrich.com/PE/es/product/sigma/cb_94122105).

En líneas generales, las células SVCT se deben mantener con el medio de cultivo DMEM suplementado con 10% de suero bovino fetal, previamente inactivado (heat inactivated foetal bovine serum), (HI-FBS, Gibco) y 1% de penicilina/estreptomicina (Pen/Strep, Gibco), en presencia de Insulina e Hidrocortisona. A este medio completo se le denominó "D10-SVCT". Cabe señalar que las células deben cultivarse en incubadoras a 37C y 5% de CO₂.

Los flasks de cultivo (SPL) utilizados variaron entre T25, T75 y T175, según el nivel de expansión requerido. Las placas de 96 y 6 pozos se utilizaron principalmente durante el período de screening de clones candidatos. En cualquier formato, estas células fueron re-cultivadas a 37C y 5% de CO₂ y una vez se alcanzaba el 80-90% de confluencia.

La viabilidad celular fue evaluada por tinción con Azul de Tripano (Gibco) y una cámara de Neubauer. Asimismo, los cultivos eran monitoreados periódicamente para detectar la presencia de contaminación por *Mycoplasma* a través del uso de Mycoplasma Detection PCR Kit (ThermoFisher).

1.2.2 Edición génica por CRISPR/Cas9

1.2.2.1 Diseño de ARN guías (sgRNAs) y secuencia molde (ssODN)

El sgRNA es el componente del sistema CRISPR/Cas9 que confiere la especificidad al guiar a la proteína Cas9 al sitio diana. Para el caso de los complejos RNP de nuestros experimentos, se usó un sgRNA modular, compuesto por dos cadenas de ARN; crARN (que contiene la secuencia de acoplamiento) y tracrARN (que provee el soporte para la interacción entre el gRNA y Cas9). Existen diversas herramientas online para el diseño y elección de moléculas gRNA, en nuestro caso, se utilizó las siguientes: Genetic Perturbation Platform, Broad Institute

(<https://portals.broadinstitute.org/gpp/public/gene/search>) y CRISPR design tool, MIT (<http://crispr.mit.edu/>).

El diseño de las secuencias ssODN para la introducción de las mutaciones tuvo como criterio el uso de secuencias asimétricas de ~90 pares de bases con modificaciones de fosforotioato en los extremos, de acuerdo con lo reportado por ⁸. Ambos componentes, gRNA y ssODN, fueron sintetizados por IDT (Integrated DNA Technologies, USA) y purificados usando procedimientos estándar.

1.2.2.2 Electroporación usando el Amaxa 4D Nucleofector (Lonza)

Electroporación de células SVCT (exp. piloto): Para identificar qué condiciones de electroporación eran las indicadas para las células SVCT, utilizando el sistema Amaxa 4D Nucleofector X Unit (Lonza), se utilizó el kit Amaxa™ 4D-Nucleofector™ Optimization Protocol for Cell Lines (Lonza). Se probaron 15 programas de nucleoporación y 03 soluciones de transfección distintas (SE, SF, SG). Como plásmido control, se utilizó el plásmido GFP+ incluido en el kit. Las células adherentes SVCT fueron preparadas y transfectadas según las instrucciones del fabricante.

Luego de 48h post-nucleoporación, se cosecharon las células, se lavaron dos veces con PBS y se distribuyeron en tubos de FACS (Corning), llegando a un volumen final de 500ul. Minutos antes de la adquisición de datos, se realizó una tinción con Yoduro de Propidio (PI, ThermoFisher) con el fin de distinguir entre células muertas (PI+) y vivas (PI-).

Las células fueron corridas utilizando un citómetro de flujo (Becton Dickinson, USA) de la empresa Cytometric. Los datos adquiridos fueron luego analizados utilizando Floreada.io, un software online gratuito para este tipo de data (<https://floreada.io/analysis>).

Transfección de SVCTs con sgRNAs

Los tubos de IDT conteniendo los crRNA (secuencias diseñadas) y el tracrRNA fueron centrifugados y su contenido disuelto siguiendo las instrucciones del proveedor (IDT). Ambos tipos de RNA fueron mezclados para permitir la formación de los ARN guías (sgRNAs). Luego de esto, se mezclaron los sgRNAs y la enzima Cas9 para formar las ribonucleoproteínas (RNPs).

Estandarización de sgRNAs (exp. piloto): Se transfirió el mix de RNPs, Electroporation Enhancer (IDT) y células al Amaxa 4D Nucleofector single cuvette (Lonza), asegurándose de que no quedaran burbujas. Se cerró la tapa cuidadosamente y se colocó las cubetas en el Amaxa 4D Nucleofector. Se utilizó el programa E0-100 y la solución SE.

Edición génica (exp. final): Una vez mezcladas las células con los RNPs, se agregó 4ul de la secuencia molde (ssODN, 100 uM). A dicho mix, se le agregó también el Electroporation Enhancer (IDT). Se mezcló por resuspensión y se transfirió el mix al Amaxa 4D Nucleofector single cuvette (Lonza). Luego se procedió según lo indicado líneas arriba (exp. piloto).

Después de la nucleofección, las células fueron resuspendidas en medio RPMI (Lonza), suplementado con 10% HI-FBS (Gibco) y 1% Pen/Strep (Gibco). A este medio completo se le denominó "R10". Las células en medio R10 fueron transferidas a eppendorfs de 1,5 ml y llevados a la incubadora (37C, 5% CO2) durante 10 min.

De cada eppendorf, se transfirieron células a placas de 96 pozos. Por eppendorf, se produjo un pocillo con células transfectadas en medio D10-SVCT suplementado con HDR Enhancer (IDT) y un pocillo con D10-SVCT sin suplemento. Las placas fueron llevadas a la incubadora y se cambió el medio de transfección luego de 12-24h. Luego de 72h, se cosecharon las células transfectadas y se expandieron a placas de 24 pozos.

1.2.2.3 Screening de clones (single-cell) de SVCT post CRISPR/Cas9

El screening de clones celulares modificados genéticamente se hizo a través de la metodología RFLP (Restriction Fragment Length Polymorphism). En nuestro caso, la enzima de restricción utilizada fue Nsil (New England Biolabs, USA) para la variante VUSCaMa c.5074+28T>A. La presencia de dicha variante crea un único sitio de restricción de Nsil, a diferencia del amplicón de PCR derivado de la secuencia wild-type (WT) que no presenta dicho sitio de corte. Asimismo, para la variante patogénica c.5074+1G>A, se utilizó la enzima de restricción Accl, la cual reconoce específicamente el nucleótido +1 de la secuencia WT. Una vez generada la variante, el sitio de corte desaparece, permitiendo así la distinción entre clones positivos y negativos para la modificación génica.

Similar a lo desarrollado en el caso de la evaluación de la eficiencia de corte del sistema CRISPR/Cas9, se extrajo el ADN de los clones celulares utilizando la solución de QuickExtract™ DNA Extraction Solution (Cambio, USA), según las instrucciones del fabricante. La solución fue añadida directamente sobre las células para su lisis. Luego de añadida la solución se deja incubar en la placa a temperatura ambiente por 5 minutos antes de someterla a un ciclo de PCR como sigue; 65°C por 15 minutos, vortex por 10 segundos, 95°C por 15 minutos y vortex por 10 segundos. Finalmente, se diluye el producto de la lisis celular con 5 volúmenes (150ul) de agua libre de DNAsas y RNAsas.

Una vez obtenido el DNA genómico de cada clon celular, se amplificó la región de interés (modificada por CRISPR/Cas9) mediante PCR y se digirió con la enzima de restricción correspondiente. Los productos digeridos fueron corridos en geles de agarosa 2% para su evaluación.

Según los resultados obtenidos del análisis RFLP, se eligieron clones candidatos para ser llevados a la siguiente fase: confirmación por secuenciación Sanger de las regiones modificadas. Con este fin, se amplificó la región de interés mediante PCR utilizando un kit de alta fidelidad, AmpliTaq Gold (ThermoFisher, USA). Una vez amplificados, los fragmentos de PCR fueron purificados utilizando el kit QIAquick PCR Purification (Qiagen, USA). Los productos purificados fueron secuenciados *in-house*. Los resultados del secuenciamiento Sanger fueron visualizados y alineados utilizando el programa informático Benchling (Benchling, USA).

1.3 Resultados

1.3.1 Variantes de un solo nucleótido (*Single Nucleotide Variants, SNVs*) seleccionadas para ser modeladas por CRISPR/Cas9

Dado que nuestro centro de investigación encontró la variante VUS c.5074+28T>A del gen *BRCA1* en 07 familias peruanas con diagnóstico de cáncer de mama hereditario, se decidió evaluar la posibilidad de modelarla *in vitro*. Con este fin, se evaluó su presencia en el ADN de 100 voluntarios sanos y en bases de datos poblacionales para ver si se trataba de un polimorfismo; sin embargo, no fue detectada (data no mostrada).

La variante VUS c.5074+28T>A, en adelante llamada variante VUSCaMa, se encuentra en el intrón 17, 28 nucleótidos después del exón 17 (**Fig. 1a**). Al analizar esta variante *in silico* utilizando el programa ESEFinder 3.0, se encontró que la VUSCaMa genera un sitio de reconocimiento (*motif*) de la proteína SRSF6 (**Fig. 1b**). Dicha proteína está involucrada en el splicing del mRNA y puede jugar un rol determinante en el splicing alternativo del mRNA (GeneCards). Más aún, el score del *motif* encontrado fue de 3.76, siendo el valor umbral de 2.67.

[Insertar Fig. 1 aquí]

Dado que, el fin último de este modelo es ser evaluado a través de estudios funcionales en una investigación posterior, se vio por conveniente modelar una variante patogénica. La mutación elegida fue la variante c.5074+1G>A, pues se encuentra también en el intrón 17 y, dado que afecta el primer nucleótido del intrón, afecta también el splicing del mRNA. Por esta razón y múltiples reportes de casos clínicos y esta mutación, ha sido clasificada como “patológica” (**Fig. 2**) por el ClinVar (<https://www.ncbi.nlm.nih.gov/clinvar/>), el repositorio de reportes de variantes genéticas y su relación con fenotipos clínicos, custodiado por el National Institutes of Health (NIH, USA).

[Insertar Fig. 2 aquí]

1.3.2 Nucleoporación de la línea celular de tejido mamario sano SVCT

La línea celular de tejido mamario sano "SVCT" fue utilizada como línea base para producir los nuevos modelos portadores de las variantes VUSCaMa (c.5074+28T>A) o Patológica (c.5074+1G>A). Dado que la mayoría de kits comerciales están estandarizados para líneas celulares cancerígenas, el programa y solución de electroporación utilizando el sistema Amaxa 4D-Nucleofector X-unit (Lonza) tuvo que ser determinado a través de experimentos de estandarización. Con este fin, se utilizó el kit Amaxa™ 4D-Nucleofector™ Optimization Protocol for Cell Lines (Lonza).

Luego de probar 45 condiciones: 15 programas y 03 soluciones de electroporación provistas por el fabricante, se encontró que la combinación del programa E0-100 y la solución SE resultaron en una viabilidad del 41% y una señal de GFP+ del 64% (**Fig. 3a**). La muerte celular fue evaluada por tinción con Yoduro de Propidio (Propidium Iodide, PI), donde una señal positiva para este compuesto indicaría permeabilidad de la membrana celular. La señal positiva de GFP indicó la introducción a la célula y expresión del vector control GFP+ utilizado.

Una vez determinadas las condiciones de electroporación, se electroporó las células SVCT con los complejos ribonucleoproteicos (RNPs), compuestos de los ARN guías candidatos y la proteína Cas9. Como control positivo, se utilizó nuevamente un vector GFP+. En esta oportunidad, la señal de GFP hallada fue de 84.66% de transfección (**Fig. 3b**).

[Insertar Fig. 3 aquí]

1.3.3 El ARN guía Sg3 tuvo la mejor eficiencia de corte mediante CRISPR/Cas9

Con el fin de alcanzar una mayor eficiencia de corte, se diseñaron 03 crRNAs, que resultaron en la generación de 03 ARN guías (sgRNAs). Se electroporaron 03 alícuotas de células SVCT, cada una con un sgRNA distinto y se dejaron crecer a 37C y 5% de CO₂. Luego de 48-72h, se extrajo el ADN de dichas células y se amplificaron productos de PCR correspondientes a la

región de interés del gen *BRCA1*. Utilizando el ensayo de la endonucleasa T7E1 del kit Genome Editing Detection (IDT), se digirieron los productos de PCR amplificados. La enzima T7E1, al ser una endonucleasa que corta al detectar un desemparejamiento (*mismatch*) en la secuencia nucleotídica, es capaz de indicar si se generó un corte por la enzima Cas9 que produjo, a su vez, una modificación en el ADN durante su reparación por empalme.

En la **Fig. 4a**, se muestran las digestiones de los controles del kit utilizado. El control negativo, al ser homodúplex, no genera corte. El control positivo (heterodúplex) muestra el corte de la enzima, que se traduce en 3 bandas. Se muestra también el producto de PCR de *BRCA1* del ADN de células SVCT no modificadas para referencia. Se incluye además los controles del kit CRISPR/Cas9 (IDT), correspondientes al gen *HPRT*. Finalmente, se muestran los resultados correspondientes a las células transfectadas con cualquiera de los 03 sgARNs. Se observa que en todas ellas hubo corte; sin embargo, no es posible determinar cuál de ellas fue la más eficiente. Esto puede deberse al tamaño de banda (221bp). Cabe indicar que el kit de la enzima T7E1 recomienda para este ensayo fragmentos de 600-900pb; de forma que sea más fácil discernir la acción de la enzima.

Como estrategia paralela, se generaron nuevos productos de PCR para su análisis por secuenciación Sanger. Los electroferogramas resultantes fueron analizados utilizando el software *TIDE: Tracking of Indels by Decomposition* (**Fig. 4b**). Este programa permite determinar la probabilidad de que se hayan generado *indels* (inserciones/delecciones) o modificaciones luego del corte por la enzima Cas9 y el número de nucleótidos que forman dichos indels. Asimismo, el software indica el porcentaje de secuencias aberrantes por cada nucleótido analizado. Esta información la presenta en un gráfico de barras.

Como se muestra en la **Fig. 4b**, el sg1 tuvo una eficiencia de corte del 23.4%. El sg2 alcanzó un 53.7%; mientras que el sg3 logró una eficiencia total de corte del 95.4%, donde el 84.2% de cortes generan una inserción de 01pb. De forma similar, la señal de secuencia aberrante se incrementa de forma notable a partir del sitio de corte esperado (posición nucleótido 87) y se observa una

clara diferencia de señal entre la secuencia control no modificada (barras negras) y la secuencia transfectada con el sg3 (barras verdes).

[Insertar Fig. 4 aquí]

1.3.4 Se generaron células SVCT portadoras de la variante VUS c.5074+28T>A y de la variante patogénica c.5074+1G>A

Una vez identificado el sgRNA con mejor eficiencia de corte, se procedió a realizar la transfección de células SVCT; esta vez incluyendo la secuencia molde (*single strand oligodeoxynucleotide donor*, ssODN) conteniendo la variante a modelar.

Al igual que en el experimento anterior, las células fueron transfectadas y cultivadas por 48-72h antes de extraer el ADN. Cabe señalar que en esta oportunidad, se cultivaron células en duplicado, de forma que una placa sirvió para expandir los cultivos; mientras que la otra placa fue utilizada para análisis.

Luego de expandir las células transfectadas a placas de 24 pozos, se realizó un cultivo de clones "*single-cell*" en placas de 96 pozos por dilución limitante. Como resultado, se obtuvieron 84 clones *single-cell* para la variante VUSCaMa y 34 clones para la variante patogénica. Una vez establecida su viabilidad fueron recultivados en placas de 96 pozos en duplicado con propósitos de expansión y análisis.

El ADN de todos los clones *single-cell* fue extraído y amplificado por PCR para la región de interés. Dichos amplicones fueron luego digeridos con la enzima Nsil para la variante VUSCaMa, pues el cambio de nucleótido crea el sitio de corte de dicha enzima (**Fig. 5a**). En el caso de la variante patogénica, su modelamiento elimina el sitio de corte de la enzima Accl.

En la **Fig. 5b**, se muestra la digestión de muestras provenientes de pacientes portadores de la variante VUSCaMa, a manera de control de corte de la enzima Nsil. La **Fig. 5c** muestra geles representativos del tamizaje realizado

a todos los clones *single-cell*. Aquellos que mostraron las bandas esperadas fueron llevados a la siguiente fase de análisis por secuenciación Sanger.

En la **Fig. 5d**, se observan los alineamientos de los clones portadores de las variantes modeladas. Se obtuvo 01 clon homocigoto para la variante VUSCaMa, mientras que se detectaron 02 clones heterocigotos para la variante patogénica. Un resumen de los clones generados, tamizados y finalmente identificados como positivos se muestra en la **Fig. 5e**.

[Insertar Fig. 5 aquí]

1.4 Discusión

La frecuencia de familias de cáncer de mama y ovario con mutaciones *missense*, supresiones cortas o duplicaciones varía entre el 15% y el 70% dependiendo de la población que se estudie ⁹⁻¹¹. La mayoría de las mutaciones en *BRCA1* y *BRCA2* es por cambios en el marco de lectura que resultan en proteínas no funcionales (Breast Cancer Information Core, base de datos BIC 2017). Más aún, las mutaciones que afectan las secuencias consenso de intrones en los sitios de empalme de los genes *BRCA* y las que crean sitios de corte y empalme ectópico representan el 15% de las mutaciones puntuales asociadas con cáncer de mama y de ovario ¹²⁻¹⁴. Como se comentó líneas arriba, existen también otro tipo de mutaciones de significado incierto que dan lugar a las llamadas “Variantes Genéticas de Significado Indeterminado” o VUS por sus siglas en inglés.

El objetivo de esta investigación fue modelar *in vitro* una variante VUS del gen *BRCA1*, encontrada en 07 familias con diagnóstico de cáncer hereditario de mama y ovario. En efecto, nuestro estudio logró identificar 01 clon homocigoto para la variante VUSCaMa. Adicionalmente, se detectó 02 clones heterocigotos para una variante patogénica ubicada en el mismo intrón de la variante VUSCaMa. Esto fue diseñado así para poder comparar lado a lado el efecto de ambas variantes en el splicing del mRNA, así como su papel en estudios funcionales como parte de estudios futuros.

Un factor importante en el diseño de sgRNAs es la posible generación de cortes en regiones no deseadas en el genoma del organismo diana, conocidos como efectos *off-target*. Las diversas herramientas usadas para el diseño y elección de gRNAs poseen algoritmos capaces de predecir que sitios en el genoma pueden sufrir cortes no específicos. Una vez identificados, estos sitios deben ser revisados y se debe comprobar que no han resultado modificados a consecuencia de la acción del sistema CRISPR/Cas9. Esto se debe hacer por medio de secuenciamiento tipo Sanger luego de la transfección y selección de clones genéticamente modificados. Una limitación de este estudio fue que no se logró analizar otras partes del genoma de los clones generados para controlar por dichas modificaciones off-target. Dada la accesibilidad creciente a secuenciación de genoma completo, se sugiere que futuros estudios contemplen el análisis global de la secuencia de ADN de los clones identificados.

En conclusión, este estudio es una prueba concepto de que es posible modelar variantes genéticas VUS identificadas en nuestra población para su posterior análisis. En nuestro conocimiento, esta investigación es la primera en su género en nuestro país, y a través de ella, se ha logrado implementar la metodología de edición genómica CRISPR/Cas9, así como las estrategias para el cultivo de células humanas con herramientas y capacidades locales. En el futuro, los modelos celulares generados podrán ser evaluados por análisis de expresión génica y proteica, y su efecto en el daño/reparación del ADN. Este conocimiento permitirá validarlas como biomarcadores con valor pronóstico y/o predictivo, contribuyendo al tratamiento de precisión de pacientes peruanas con cáncer de mama.

Literatura citada

1. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F. 2015. Cancer incidence and mortality worldwide: Sources, methods and major patterns in globocan 2012. *International journal of cancer Journal international du cancer*. 136(5):E359-386.
2. Pineros M, Ramos W, Antoni S, Abriata G, Medina LE, Miranda JJ, Payet E, Bray F. 2017. Cancer patterns, trends, and transitions in peru: A regional perspective. *Lancet Oncol*. 18(10):e573-e586.
3. Blackwood MA, Weber BL. 1998. Brca1 and brca2: From molecular genetics to clinical medicine. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 16(5):1969-1977.

4. Slavin TP, Maxwell KN, Lilyquist J, Vijai J, Neuhausen SL, Hart SN, Ravichandran V, Thomas T, Maria A, Villano D et al. 2017. The contribution of pathogenic variants in breast cancer susceptibility genes to familial breast cancer risk. *NPJ Breast Cancer*. 3:22.
5. Venkitaraman AR. 2002. Cancer susceptibility and the functions of *brca1* and *brca2*. *Cell*. 108(2):171-182.
6. Eccles BK, Copson E, Maishman T, Abraham JE, Eccles DM. 2015. Understanding of *brca* genetic results by breast cancer specialists. *BMC cancer*. 15:936.
7. Buleje J, Guevara-Fujita M, Acosta O, Huaman FDP, Danos P, Murillo A, Pinto JA, Araujo JM, Aguilar A, Ponce J et al. 2017. Mutational analysis of *brca1* and *brca2* genes in peruvian families with hereditary breast and ovarian cancer. *Mol Genet Genomic Med*. 5(5):481-494.
8. Renaud JB, Boix C, Charpentier M, De Cian A, Cochenne J, Duvernois-Berthet E, Perrouault L, Tesson L, Edouard J, Thinard R et al. 2016. Improved genome editing efficiency and flexibility using modified oligonucleotides with talen and crispr-cas9 nucleases. *Cell Rep*. 14(9):2263-2272.
9. Tereschenko IV, Basham VM, Ponder BA, Pharoah PD. 2002. *Brca1* and *brca2* mutations in russian familial breast cancer. *Hum Mutat*. 19(2):184.
10. Thorlacius S, Sigurdsson S, Bjarnadottir H, Olafsdottir G, Jonasson JG, Tryggvadottir L, Tulinius H, Eyfjord JE. 1997. Study of a single *brca2* mutation with high carrier frequency in a small population. *Am J Hum Genet*. 60(5):1079-1084.
11. El Saghir NS, Zgheib NK, Assi HA, Khoury KE, Bidet Y, Jaber SM, Charara RN, Farhat RA, Kreidieh FY, Decousus S et al. 2015. *Brca1* and *brca2* mutations in ethnic lebanese arab women with high hereditary risk breast cancer. *The oncologist*. 20(4):357-364.
12. Krawczak M, Reiss J, Cooper DN. 1992. The mutational spectrum of single base-pair substitutions in mrna splice junctions of human genes: Causes and consequences. *Human genetics*. 90(1-2):41-54.
13. Minucci A, Concolino P, De Bonis M, Costella A, Paris I, Scambia G, Capoluongo E. 2018. Preliminary molecular evidence associating a novel *brca1* synonymous variant with hereditary ovarian cancer syndrome. *Hum Genome Var*. 5:2.
14. Gutierrez-Enriquez S, Coderch V, Masas M, Balmana J, Diez O. 2009. The variants *brca1* *ivs6-1g>a* and *brca2* *ivs15+1g>a* lead to aberrant splicing of the transcripts. *Breast cancer research and treatment*. 117(2):461-465.

Fig. 1. Análisis *in silico* sugiere patogenicidad de la variante VUS c.5074+28T>A. a) Mapa de la variante VUS c.5074+28T>A, encontrada en 07 familias peruanas con cáncer de mama hereditario. La flecha rosada indica la ubicación de la variante en un producto de PCR (221bp) que abarca parte del intrón 17 y el exón 17 (flecha naranja) del gen *BRCA1*. Las flechas azul y gris corresponden a la ubicación de los *primers* utilizados. b) Análisis *in silico* de la variante VUS c.5074+28T>A utilizando el software ESEfinder 3.0. Los cuadros rojos tanto en la tabla como en el gráfico señalan la presencia de la variante en el *motif* de reconocimiento de la proteína SRSF6. La flecha roja indica la barra correspondiente al *motif* de SRSF6; mientras que la línea punteada denota el umbral de score para dicha proteína.

Fig. 2. Variante patogénica c.5074+21G>A. a) Mapa de la variante patogénica c.5074+21G>A. El producto de PCR utilizado para su estudio es el mismo que para la variante VUS. b) Resultado de la interpretación de la variante según ClinVar catalogándola como “Patogénica”.

Fig. 3. Nucleoporación de células SVCT con plásmido control GFP+. a) Estandarización del protocolo de electroporación de células SVCT utilizando el sistema Amaxa 4D-Nucleofector (Lonza) y un plásmido control GFP+. Se probaron 45 condiciones: 15 programas de electroporación y 03 soluciones “4D-Nucleofector” (data no mostrada). En el panel se muestran los *gates* utilizados para el Ctrl(-) y la condición Prog. E0-100/Solución SE, seleccionada para ser utilizada en experimentos sucesivos. Se incluye el plot de doble entrada para la tinción con PI (muerte celular) y GFP (electroporación). b) Electroporación de células SVCT para el experimento CRISPR/Cas9. Se muestran los histogramas del Ctrl(-) y las células electroporadas utilizando la condición Prog. E0-100/Solución SE del sistema Amaxa 4D-Nucleofector (Lonza). Las flechas indican los nombres de los ejes de los histogramas. GFP: Green fluorescent protein. PI: Yoduro de propidio.

Fig. 4. Comparación y evaluación de la eficiencia de 03 ARN guías por ensayo de la enzima T7E1 y por secuenciación Sanger. a) Ensayo de la enzima T7E1 utilizando el kit Alt-R Genome Editing Detection (IDT). Se amplificaron los productos de PCR correspondientes a: i) los controles homo y heterodúplex (kit T7E1), ii) *HPRT* en células SVCT electroporadas con controles del kit Alt-R CRISPR-Cas9, iii) *BRCA1* en células SVCT no electroporadas y en SVCT electroporadas con ARN guías sg1, sg2 o sg3. Dichos productos fueron digeridos con la enzima T7E1 y corridos en un gel de agarosa al 2%. Las flechas negras indican los tamaños de fragmento de las bandas observadas. b) Análisis de electroferogramas (secuenciación Sanger) utilizando el software *TIDE: Tracking of Indels by Decomposition*. Los productos de PCR correspondientes a las células SVCT no modificadas y SVCT electroporadas con sg1, sg2 o sg3 fueron evaluados por secuenciación Sanger. Los electroferogramas fueron luego analizados utilizando el software *TIDE: Tracking of Indels by Decomposition*. En el panel se muestra el espectro de indels y la señal de secuencia aberrante para los 03 ARN guías. Las barras rojas indican la alta probabilidad ($p < 0.001$) de que se dé el cambio o indel teniendo en cuenta el número de nucleótidos en el eje x. La línea azul punteada indica la posición del punto de corte. Las barras negras corresponden al % de secuencias aberrantes en la secuencia original; mientras que las verdes corresponden a la secuencia por la enzima Cas9 según los guías sg1, sg2 o sg3.

Fig. 5. Detección de clones de células SVCT portadores de las mutaciones c.5074+28T>A o c.5074+1G>A, generados mediante la técnica CRISPR/Cas9. a) Mapa de la secuencia wild-type (WT) y la secuencia portadora de la variante VUSCaMa (c.5074+28T>A). La presencia de esta última genera el sitio de corte de la enzima de restricción Nsil. b) Como control positivo de corte con dicha enzima, se digirieron muestras de pacientes portadores de la variante VUSCaMa. Las flechas rojas indican la presencia de bandas por corte enzimático. Como control negativo, se utilizó ADN proveniente de células SVCT digeridas o no con Nsil. c) Screening de clones SVCT (single-cell) por restricción enzimática. Se generaron clones tipo single-cell mediante cultivo celular por dilución limitante. Para la variante VUSCaMa, se obtuvieron 84 clones viables. Los productos de PCR resultantes fueron digeridos con la enzima Nsil y corridos en un gel de agarosa al 2%. Las flechas rojas señalan aquellos clones elegidos como candidatos para ser analizados por secuenciación Sanger. d) Alineamiento de secuencias *BRCA1* (WT) y clones modificados portadores de c.5074+28T>A (homocigoto) o c.5074+1G>A (heterocigoto). El nucleótido “Y” significa T o C. e) Gráfico de barras resumen de los clones single-cell generados, candidatos encontrados y clones exitosamente modificados para ambas variantes modeladas. Se incluye una tabla resumen donde se cuantifica este análisis.