

Informe final publicable de proyecto

Desarrollo y Validación de un panel de genes para identificar predisposición hereditaria a diversos tipos cáncer por NGS

Código de proyecto ANII: FSS_X_2022_1_173313

Fecha de cierre de proyecto: 20/03/2025

DUHAGON, María Ana (Responsable Técnico - Científico)

GELABERT LADÓ, María (Investigador)

CHÁVEZ GARCÍA, Santiago Roberto (Investigador)

MATHÓ PACIELO, Cecilia (Investigador)

OLIVEIRA RIZZO, Carolina Gabriela (Investigador)

SOTELO SILVEIRA, José Roberto (Investigador)

UNIVERSIDAD DE LA REPÚBLICA. ÁREA CIENCIAS DE LA SALUD (Institución Proponente) \\
MINISTERIO DE EDUCACIÓN Y CULTURA. INSTITUTO DE INVESTIGACIONES BIOLÓGICAS "CLEMENTE ESTABLE" \\
FACULTAD DE MEDICINA. FUNDACIÓN MANUEL PEREZ

Resumen del proyecto

Aproximadamente 5–10% de los pacientes con cáncer han heredado mutaciones que aumentan su riesgo, siendo común en jóvenes con antecedentes familiares. Detectarlos permite prevenir, orientar tratamientos y reducir costos. Aunque estas pruebas están recomendadas por guías internacionales, y en Uruguay posiblemente hay 1700 casos nuevos por año, rara vez son cubiertas por los prestadores de salud. La escasez de oferta de estos estudios a nivel nacionales y la heterogeneidad de la oferta internacional podría estar contribuyendo a este retraso. Este proyecto desarrolló una prueba para diagnosticar la mayoría de los cánceres hereditarios, que puede realizarse con equipamiento disponible en Uruguay. Esta empleó secuenciación de ADN de próxima generación (NGS). Se seleccionaron los 45 genes con mayor rendimiento clínico para estas patologías, se diseñaron los cebadores para amplificar sus secuencias de interés y construir librerías de secuenciación adaptando protocolos de AmpliSeq para el secuenciador IonGenestudio S5. Se ensayaron los protocolos usando ADN de muestras de referencia internacional para evaluar el desempeño del ensayo. Las lecturas obtenidas fueron analizadas bioinformáticamente para clasificar y priorizar las variantes génicas obtenidas. Se validaron parámetros analíticos fundamentales en diagnóstico in vitro como el umbral de profundidad, la sensibilidad, la especificidad, la precisión, la reproducibilidad y la cobertura. El ensayo alcanzó puntajes entre 99–100% para estos parámetros. Se identificaron las condiciones óptimas para el ensayo. Se registró el protocolo utilizado y las especificaciones técnicas de validación para facilitar su uso futuro y transferencia.

Este trabajo fue realizado por un equipo multicéntrico e interdisciplinario con experiencia previa en el área. En el mismo participaron investigadores jóvenes y también forma parte de una tesis de grado en biotecnología.

Ciencias Médicas y de la Salud / Biotecnología de la Salud / Tecnologías que involucran la identificación de ADN, proteínas y enzimas / Genética del Cáncer

Palabras clave: predisposición / cáncer / NGS /

Antecedentes, problema de investigación, objetivos y justificación.

Se han descrito más de 50 síndromes de susceptibilidad familiar al cáncer (PHC)(Brown et al. 2020). Se estima que alrededor del 10% de los pacientes con cáncer poseen una predisposición hereditaria al cáncer, que les confiere un riesgo aumentado de desarrollar la enfermedad en comparación con la población general. Esto implica que portan una mutación que incrementa su riesgo de desarrollar cáncer, la cual varía desde niveles moderados hasta probabilidades extremadamente elevadas (en algunos casos, del 50 al 90%). Identificar genéticamente a quienes portan estas mutaciones permite implementar estrategias preventivas efectivas, tanto en los pacientes como en sus familiares, lo cual contribuye a reducir la morbilidad y mortalidad asociadas al cáncer, así como costos de salud. Considerando la incidencia del cáncer en nuestro país (Ferlay et al. 2021), estimamos que cada año se diagnostican 1700 personas con cáncer con componentes hereditarios. Sin embargo, a la mayoría de estas personas no se les realizará un estudio genético por motivos posiblemente diversos, entre los que se incluye la falta de cobertura por los prestadores de salud en Uruguay.

La evaluación de la PHC requiere una caracterización precisa de la inactivación de genes que actúan como supresores tumorales y pueden presentar mutaciones a lo largo de toda su región codificante, lo que hace indispensable su secuenciación completa. Los estudios genéticos destinados a identificar mutaciones responsables de la predisposición hereditaria al cáncer (PHC) comenzaron con análisis centrados en genes

individuales. Posteriormente, se incorporaron paneles compuestos por varios genes relacionados con un mismo tipo tumoral. El cáncer de mama marcó el inicio de este enfoque, probablemente debido a su elevada incidencia y mortalidad, lo que motivó el desarrollo inicial de estudios dirigidos a uno o pocos genes específicos (Laduca et al. 2014). El cáncer de mama fue pionero en este sentido, posiblemente por su alta incidencia y mortalidad. Surgen así estudios de gene único o de pocos genes, que constituyen paneles pequeños focalizados para un único tipo de cáncer. Dado que algunos síndromes de PHD causan más de un tipo de cáncer, que la historia familiar de algunos pacientes es desconocida y también debido al gran desarrollo de las tecnologías de secuencias y el descenso de sus costos, se desarrollan luego paneles de genes más extendidos que incluyen más genes. Más aún, la heterogeneidad genética que caracteriza a los síndromes hereditarios de cáncer (Singh 2013), ha convertido a la secuenciación masiva (NGS) en una herramienta fundamental para su análisis e incorporación en el diagnóstico clínico. Actualmente, los paneles dirigidos basados en NGS son los más utilizados para identificar variantes genéticas asociadas a la PHC, constituyendo las indicaciones de prueba de NGS más solicitados luego de los de pruebas prenatales (casi equivalentes) en países como Estados Unidos (Phillips et al. 2018). Teniendo en cuenta la demanda potencial de nuestro país, nosotros pensamos que desarrollar una única prueba genética por NGS que abarque todos los genes que causan cáncer hereditario es más conveniente que el usar test diferentes para cada tipo específico de cáncer. Esta conveniencia no es solamente económica, en base a la inversión en síntesis e implementación (estrategias de mesa húmeda y seca), sino también considera la validación analítica y posteriormente clínica. Por último, si bien es posible secuenciar exomas o genomas completo de pacientes, esto no solo es más costos y lento que un ensayo focalizado como el que proponemos, sino que aumenta la complejidad del análisis de múltiples genes y variantes de significación inciertas y las implicaciones éticas que comportan (Soto et al. 2018).

La selección de los genes incluidos en un panel de predisposición hereditaria al cáncer (PHC) se basa principalmente en su capacidad diagnóstica, es decir, en la probabilidad de identificar variantes patogénicas en una población determinada de pacientes, lo cual está influido por la frecuencia con que se han reportado mutaciones en la literatura científica (Soto et al. 2018). Otro criterio fundamental es la solidez de la evidencia que respalda la asociación entre un gen alterado y la enfermedad, así como su relevancia. Si bien ampliar el número de genes puede incrementar las posibilidades de detectar variantes relevantes, también eleva el riesgo de encontrar resultados inciertos, especialmente en genes con baja penetrancia. Esto puede generar desafíos tanto en la elección de la prueba como en la interpretación de los resultados por parte del médico. Por esta razón, los paneles de tamaño intermedio —que actualmente incluyen entre 20 y 50 genes— suelen considerarse una alternativa más equilibrada y eficaz (Rubinstein et al. 2013).

Las pruebas genéticas de diagnóstico en el laboratorio clínico requieren satisfacer una serie de criterios analíticos para su utilización en diagnóstico, así como un tiempo de respuesta aceptable para las decisiones clínicas. Cada laboratorio clínico debe implementar el proceso NGS apropiado que incluya el procedimiento húmedo y el análisis bioinformático que cumpla con los estándares de calidad de las pruebas genéticas clínicas determinadas por el organismo regulador correspondiente (Rehm et al. 2013; Matthijs et al. 2016). Algunos de los parámetros importantes que deben validarse en los paneles de genes de NGS incluyen el umbral, la sensibilidad, la especificidad, la precisión y la cobertura. Aunque no existe una regla rápida y sencilla para determinar el número de muestras caracterizadas que se necesitan para la validación, cada laboratorio debe considerar el número de pacientes a analizar, el número y tipo de variantes examinadas en su panel objetivo para definirlo y los problemas específicos de la tecnología utilizada.

Nuestra propuesta buscó desarrollar un estudio de NGS específico para el diagnóstico de susceptibilidad hereditaria al cáncer agnóstico a origen tisular, validarla analíticamente y registrar detalladamente el procedimiento para que pueda ser implementado en cualquier laboratorio que disponga del equipamiento y

la capacitación. Para eso se seleccionaron muestras de ADN de referencia internacional recomendadas por las entidades internacionales que establecen las guías diagnósticas en el área (American Association of Molecular Pathology, American College of Medical Genetics, EuroGentest, European Society of Human Genetics). Además, se solicitó el asesoramiento a la unidad de protección de la propiedad intelectual de la UDELAR, para determinar si esta metodología es susceptible de protección y se revisó la normativa de calidad general para ensayos de laboratorio analítico (normas ISO (the International Organization for Standardization, ISO 9001:2015, ISO 17025:2017) directamente vinculadas a esta clase de ensayos.

Metodología/Diseño del estudio

La metodología de investigación se explica a continuación siguiendo las etapas consecutivas del ensayo.

I. Selección de regiones y diseño de amplicones:

Se utilizó un panel de genes similar al inicialmente propuesto, que fue revisado al inicio del proyecto. Se seleccionaron 45 genes y la región de interés (ROI) a secuenciar comprendió todos los exones y al menos 10 pares de bases de los intrones adyacentes. Para amplificarlos a partir del ADN genómico se diseñó un panel de Ion AmpliSeq™ usando la herramienta web Ion AmpliSeq™ Designer, que transforma listas de genes en grupos de cebadores para la construcción de bibliotecas específicas. Una vez obtenido el tamaño del panel, se estimó el costo por muestra para ajustar el número de ensayos.

II. Extracción de ADN.

Para la optimización y validación analítica del ensayo, se emplearon muestras comerciales de ADN humano proveniente del CEPH (Belmont et al. 2003), ampliamente utilizadas como material de referencia en validaciones de plataformas de secuenciación (Jennings et al., 2017). Además, se trabajó con una línea celular tumoral que pertenece al panel NCI-60 (DU-145, de cáncer de próstata)(Abaan et al. 2013) y se encuentran molecularmente caracterizadas, con diversas variantes conocidas en genes incluidos en el panel utilizado. Se extrajo ADN genómico de un cultivo celular de La línea DU145 y se adquirió la referencia de CEPH correspondiente al individuo 1347-02. Ambos ADN se analizaron espectrofotométricamente para determinar su cantidad y pureza.

III. Preparación de bibliotecas:

Se utilizaron al menos 20 ng de ADN genómico por muestra para la amplificación con el kit Ion AmpliSeq™. Los productos fueron parcialmente digeridos, adaptados, indexados con códigos Ion Xpress™, purificados con AMPure XP y cuantificados mediante Qubit 2.0 o Bioanalyzer.

IV. Secuenciación:

Las bibliotecas se ecualizaron y se generaron los moldes de secuenciación en el Ion Chef™. La secuenciación se realizó con el Ion GeneStudio™ S5 usando chips Ion 530™.

V. Análisis bioinformático de variantes:

Los datos se procesaron con Torrent Suite (Thermo Fisher Scientific) y se realizaron llamados de variantes tras el alineamiento al genoma GRCh38. Se complementó con herramientas como GATK, FastQC, bowtie2, SAMtools, BCFtools y ANNOVAR.

VI. Evaluación del desempeño del panel:

Se evaluaron parámetros clave del desempeño del panel, incluyendo la profundidad de cobertura, cobertura uniforme, sensibilidad, especificidad, positive percentage agreement (PPA), positive predictive

value (PPV) y reproducibilidad, de acuerdo con guías internacionales y protocolos previamente establecidos por nuestro grupo. Para ello, se realizaron secuenciaciones de las mismas muestras en réplicas técnicas bajo condiciones variables, considerando diferencias en el tipo de códigos de barras utilizados, construcción de librerías en paralelo o de forma independiente, corridas secuenciales o conjuntas, número total de lecturas por muestra y lotes distintos de reactivos. Esta estrategia permitió validar de forma robusta la consistencia y eficacia del panel desarrollado.

La secuenciación de estas muestras se llevó a cabo en el sistema Ion GeneStudio™ S5 disponible en el Instituto de Investigaciones Biológicas Clemente Estable. Esta plataforma permitió no solo la generación sino también el análisis de datos mediante el uso de chips escalables y protocolos automatizados.

VII. Registro del método. Se identificaron las condiciones óptimas del ensayo usando la mínima cantidad de lecturas necesarias para cubrir la ROI. Se escribió el protocolo detalladamente para que pueda ser realizado adecuadamente por un tercero.

Resultados, análisis y discusión

Hemos realizado una revisión bibliográfica extensa, así como un análisis comparativo de los paneles de genes causantes de cáncer hereditario publicados en revistas científicas u ofrecidos por empresas líderes en el área (Rubinstein et al. 2013) y seleccionamos un panel de 45 genes (ROI), dentro de los que se incluyen todos los genes de alto rendimiento diagnóstico en PHC y varios genes con rendimiento menor, pero utilidad clínica en cánceres específicos.

En base a los 45 genes seleccionados, diseñamos un panel de amplicones para amplificar las regiones codificantes y aproximadamente 10 pares de bases adyacentes. Usando estos cebadores multiplexados, amplificamos la ROI, construimos librerías de secuenciación y realizamos la secuenciación del ADN para las dos muestras de referencia internacional. A partir de las lecturas obtenidas determinamos que produciendo 1.000.000 de lecturas por muestra aseguramos leer más del 99% de la ROI a una profundidad de lecturas de al menos 30x, que es el mínimo actualmente recomendado para la detección de variantes genéticas germinales. Por otro lado, la uniformidad media de las lecturas sobre la ROI de este panel resultó ser de 97%, lo cual está dentro de lo reportado en la literatura (entre 85% y 98%). El comportamiento de cada uno de los pools de secuenciación necesarios para esta estrategia resultó similar, aunque sesgado levemente hacia el pool 2. Estos resultados indican que el ensayo tiene un excelente desempeño, siendo capaz de leer 99.78% de la región de interés a la profundidad requerida con un requerimiento de secuenciación más bajo que paneles comparables reportados en la literatura para paneles comparables.

A pesar de la alta cobertura, y como es habitual en este tipo de ensayos, algunas regiones no se leen bien independientemente de la profundidad de secuenciación. Dado que es fundamental identificarlas para indicarlas como limitación del estudio o para secuenciarlas por un método ortogonal y cubrir así el 100% de la ROI, se analizaron las lecturas para conocerlas. Las regiones no leídas corresponden a 6 amplicones (del total de 973) que impiden leer 483pb de la ROI (0.17% del total de 276270 pb). Estas regiones deberán ser secuenciadas por el método de Sanger o adaptar a una pequeña reacción multiplexada para agregar a los pools 1 y 2 previo a la construcción de la librería.

Las lecturas obtenidas de la secuenciación de las muestras de referencia fueron luego procesadas bioinformáticamente para identificar las variantes génicas en la ROI de cada muestra. Comparando nuestros resultados con los genomas de referencia de estas dos muestras calculamos la calidad analítica de la secuenciación usando este panel. Las variantes encontradas en el ROI corresponden en 100% a las variantes del genoma reportadas para la muestra, e incluyen tanto cambios de una base (SNVs) como delecciones de varias bases. Por lo tanto, la sensibilidad del ensayo es del 100%. Para calcular otros

parámetros analíticos del ensayo seguimos el método descrito por PipeMAGI(Marceddu et al. 2019). La especificidad y la precisión del ensayo resultaron también resultaron del 100%, la reproducibilidad de 99% y la repetitividad 98.8%.

Durante este proyecto evaluamos junto con expertos en el área, si este ensayo pudiera ser sujeto a algún modo de protección de la propiedad intelectual, tal como una patente de método. Concluimos que no sería una opción asequible ni favorable por costo/beneficio. Por este motivo, nos inclinamos por el registro escrito detallado de toda la metodología para permitir una transferencia tecnológica en el marco de acuerdos de trabajo futuros.

Este proyecto permitió la formación de recursos humanos, incluyendo especialmente una estudiante avanzada de doctorado y una estudiante de grado en biotecnología para la cual este proyecto constituye su tesis de grado.

Conclusiones y recomendaciones

Este proyecto consiguió diseñar, probar y validar analíticamente un método diagnóstico para la detección de mutaciones causantes de predisposición heredada al cáncer basado en la metodología de secuenciación de ADN por NGS.

Los resultados obtenidos durante el desarrollo de este proyecto evidencian que el panel AmpliSeq diseñado, que abarca 45 genes y 973 amplicones, logra una cobertura del 99,78% de la región de interés a una profundidad mínima de 30 lecturas por base con solo 1 millón de lecturas. Además, la uniformidad media alcanzada fue del 97%, ubicándose en el extremo superior de los valores reportados en la literatura especializada. Este excelente desempeño técnico es consistente entre los diferentes pools de secuenciación necesarios para la estrategia, aunque se observó un leve sesgo hacia el pool 2.

En términos de calidad analítica, el análisis de dos muestras de referencia internacional arrojó resultados concordantes al 100% con las variantes esperadas, tanto en el caso de variantes puntuales (SNVs) como de pequeñas delecciones. La sensibilidad, especificidad y precisión del ensayo resultaron ser del 100%, mientras que la repetitividad fue de 98,8%, conforme a los métodos de evaluación de PipeMAGI.

Como es habitual en este tipo de ensayos, algunas regiones no presentan una cobertura adecuada incluso aumentando la profundidad de secuenciación. En nuestro caso, se identificaron seis amplicones problemáticos que impiden cubrir 483 pb, es decir, el 0,17% de la región de interés. Estas regiones deben ser abordadas mediante métodos alternativos como secuenciación Sanger o incorporadas en un pequeño pool multiplexado adicional, integrable a los pools actuales.

En paralelo, se evaluó junto con expertos la posibilidad de proteger el desarrollo mediante una patente de método. Sin embargo, se concluyó que esta vía no era favorable ni costeable. Por tanto, se optó por el registro detallado de toda la metodología desarrollada a fin de facilitar futuras instancias de transferencia tecnológica y colaboración institucional.

En base a estos resultados, se recomienda completar la cobertura del 100% de la región de interés mediante secuenciación dirigida de las regiones no cubiertas, ya sea por Sanger o incorporándose a una reacción multiplexada. Asimismo, se sugiere avanzar con una validación clínica local ampliada en una cohorte de al menos 100 pacientes con sospecha de cáncer hereditario para evaluar el rendimiento diagnóstico y el costo-beneficio en el contexto nacional. Es también prioritario estandarizar los procedimientos mediante protocolos operativos claros y promover la capacitación de laboratorios públicos y privados que puedan implementar esta tecnología.

La incorporación del ensayo al Plan Integral de Atención en Salud (PIAS) del Ministerio de Salud Pública, tal vez con foco en pacientes de alto riesgo, sería clave para su implementación a nivel nacional de evaluarse como adecuado. Finalmente, se destaca la importancia de consolidar acuerdos de colaboración con de diagnóstico nacionales para poder realizar la transferencia tecnológica, lo que fortalecerá la soberanía

científica y tecnológica en el área de la genómica de nuestro país.

Este proyecto tiene el potencial de generar un impacto en el sistema nacional de salud, al permitir el acceso local al diagnóstico molecular de cáncer hereditario a través de un estudio de alto rendimiento, adecuadamente validado y protocolizado. Esto no solo reduciría los tiempos de respuesta y los costos asociados al envío de muestras al exterior, sino que también mejoraría la atención personalizada de los pacientes y consolidaría capacidades nacionales en secuenciación y análisis bioinformático, en consonancia con los objetivos estratégicos del Fondo Sectorial de Salud y la Agencia Nacional de Investigación e Innovación.

Productos derivados del proyecto

URI en repositorio de	Silo	Estado	Identificadores	Autores	Título	Tipo de producto
		En proceso		María Gelbart	Validación de un panel de genes de NGS para identificar predisposición hereditaria	Tesis de grado/monografías

Referencias bibliográficas

- Abaan OD, Polley EC, Davis SR, et al (2013) The exomes of the NCI-60 panel: A genomic resource for cancer biology and systems pharmacology. *Cancer Res* 73:4372–4382. <https://doi.org/10.1158/0008-5472.CAN-12-3342/659272/P/THE-EXOMES-OF-THE-NCI-60-PANEL-A-GENOMIC-RESOURCE>
- Belmont JW, Hardenbol P, Willis TD, et al (2003) The international HapMap project. *Nature* 426:789–796. <https://doi.org/10.1038/NATURE02168;KWRD=SCIENCE>
- Brown GR, Simon M, Wentling C, et al (2020) A review of inherited cancer susceptibility syndromes. *JAAPA* 33:10–16
- Ferlay J, Colombet M, Soerjomataram I, et al (2021) Cancer statistics for the year 2020: An overview. *Int J Cancer*
- Laduca H, Stuenkel AJ, Dolinsky JS, et al (2014) Utilization of multigene panels in hereditary cancer predisposition testing: Analysis of more than 2,000 patients. *Genetics in Medicine* 16:830–837. <https://doi.org/10.1038/GIM.2014.40>
- Marceddu G, Dallavilla T, Guerri G, et al (2019) PipeMagi: An integrated and validated workflow for analysis of NGS data for clinical diagnostics. *Eur Rev Med Pharmacol Sci* 12:6753–6765. https://doi.org/10.26355/EURREV_201908_18566
- Matthijs G, Souche E, Alders M, et al (2016) Guidelines for diagnostic next-generation sequencing. *Eur J Hum Genet* 24:2–5
- Phillips KA, Deverka PA, Hooker GW, Douglas MP (2018) Genetic Test Availability And Spending: Where Are We Now? Where Are We Going? *Health Aff* 37:710–716
- Rehm HL, Bale SJ, Bayrak-toydemir P, et al (2013) ACMG clinical laboratory standards for next-generation sequencing. *Genet Med* 15:733–747. <https://doi.org/10.1038/gim.2013.92.ACMG>
- Rubinstein WS, Maglott DR, Lee JM, et al (2013) The {NIH} genetic testing registry: a new, centralized database of genetic tests to enable access to comprehensive information and improve transparency. *Nucleic Acids Res* 41:D925–35
- Singh J (2013) The portal for rare diseases and orphan drugs. *J Pharmacol Pharmacother* 4:168–169
- Soto JL, Blanco I, Díez O, et al (2018) Documento de consenso sobre la implementación de la secuenciación

masiva de nueva generación en el diagnóstico genético de la predisposición hereditaria al cáncer. *Medicina (B Aires)* 151:80.e1--80.e10

Licenciamiento

Reconocimiento 4.0 Internacional. (CC BY)