



UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY



Unidad Académica de
Genética, Facultad de
Medicina, UdelaR



Tesina de Grado

Licenciatura en Bioquímica

Identificación integrada de diversas alteraciones genómicas accionables en pan-cáncer mediante NGS

Agustina Sofía Pons Garcilazo

Tutora: Dra. Cecilia Mathó

Cotutor: Dr. Santiago Chávez

Laboratorio de Regulación de la Expresión Génica y Genómica de
Cáncer, Unidad Académica de Genética

Facultad de Medicina, UdelaR

14 de Noviembre, 2024

Montevideo, Uruguay

Índice

2. Resumen:	6
3. Introducción	8
3.1. <i>Cáncer</i>	8
3.1.1 Oncogenes y supresores tumorales	10
3.1.2. Inestabilidad de microsatélites	14
3.2. <i>Terapias dirigidas y mutaciones accionables:</i>	16
3.3. <i>Análisis genómico por secuenciación masiva:</i>	19
4. Objetivos:	22
5. Diseño de investigación y metodología:	23
5.1. <i>Selección de las regiones a estudiar y diseño de panel:</i>	23
5.2. <i>Extracción de ADN de muestras de tumores en FFPE:</i>	26
5.3. <i>Cuantificación y evaluación del ADN obtenido:</i>	28
5.4. <i>Preparación de librerías de secuenciación:</i>	29
5.5. <i>Secuenciación de próxima generación:</i>	31
5.6. <i>Análisis de los datos de secuenciación:</i>	32
5.6.1 Análisis automatizado con Ion Torrent Server	32
5.6.2. Análisis bioinformáticos con herramientas de uso libre.....	33
6. Resultados y Discusión:	35
6.1. <i>Obtención y evaluación de ADN</i>	35
6.1.1. Selección de kit de extracción	35
6.1.2. Cuantificación de ADN (Qubit vs NanoDrop).....	38
6.1.3. Control de Amplificación del ADN.....	41
6.2. <i>Optimización de secuenciación</i>	42
6.2.1. Preparación de librerías	42
6.2. <i>Evaluación y cuantificación de librerías:</i>	48
6.3. <i>Resultados primera secuenciación: Optimización</i>	50
6.4. <i>Desempeño del panel en muestras tumorales</i>	52
6.5. <i>Identificación de variantes y comparación utilizando muestras comerciales</i>	54
6.5.1. Aproximación a la detección del límite inferior de frecuencia alélica con este diseño experimental	57
6.6 <i>Desempeño muestras con diferente antigüedad</i>	58
7. Conclusiones y perspectivas	62
8. Aspectos Éticos	63

9. Financiación	63
10. Bibliografía	64
11. Anexo	70
<i>11.1 Protocolo de preparación de librerías</i>	<i>70</i>
<i>11.2 Comandos utilizados en el análisis bioinformático</i>	<i>77</i>

Agradecimientos:

En primer lugar, agradezco profundamente el apoyo de mi familia en todo mi recorrido académico. Gracias Grisel (mamá) y Marcos (papá) por apoyarme incondicionalmente y brindarme las herramientas que me permitieron construir y transitar de la mejor forma esta etapa de mi vida. Gracias Nati, Nico y Guille por ser pilares fundamentales en mi vida y su apoyo incondicional. Gracias por festejar mis logros como propios y ser siempre mi lugar seguro.

A Cecilia y Santiago, gracias por ser unos tutores excepcionales y grandes profesionales. Les agradezco la calidez en cada charla, la paciencia al enseñar y transmitir su conocimiento y ser tan comprensivos. Gracias por hacer de esta etapa final, una etapa de mucho conocimiento y crecimiento tanto personal como profesional.

A todos los que me acompañaron en el laboratorio, en la Unidad Académica de Genética y en el Departamento de Genómica del IIBCE, por su apoyo y colaboración.

A Santi, mi gran compañero por tu apoyo incondicional en cada momento y por ser mi gran motivación en todos estos años. Gracias por compartir siempre mis logros.

A mis amigas y amigos, gracias por acompañarme y escuchar tan pacientemente sobre mi recorrido. Por sus palabras de aliento y por estar al pendiente, siempre desde el lado de la admiración.

A mis compañeras de FCien, gracias por hacer mucho más divertido e interesante este recorrido por Facultad de Ciencias.

Finalmente, a la ANII por la financiación del proyecto.

Abreviaturas

Abs - Absorbancia

ADN - Ácido Desoxirribonucleico

FFPE –muestras de tumor Fijadas en Formalina y Embebidos en Parafina, del inglés *Formalin-fixed paraffin-embedded*

Kb- kilobases

min - Minutos

MSI - Inestabilidad de Microsatélites

NGS - Next-Generation Sequencing (Secuenciación de Próxima Generación)

nm - Nanómetros

PCR - Reacción en Cadena de la Polimerasa

pb- Pares de bases

2. Resumen:

La investigación abordada en este trabajo, apunta al desarrollo y validación de un panel de secuenciación de próxima generación (NGS) para la identificación de alteraciones genómicas accionables en cáncer. El objetivo principal es optimizar el diagnóstico molecular de tumores en Uruguay mediante la integración de un panel de 17 genes específicos y un panel de inestabilidad de microsatélites (MSI). Este enfoque busca no solo identificar mutaciones relevantes en oncogenes y genes supresores de tumores, sino también evaluar la inestabilidad de microsatélites, una característica importante en varios tipos de cáncer.

El diseño metodológico incluyó la extracción de ADN de muestras embebidas en parafina (FFPE), la preparación de librerías y la secuenciación mediante tecnología de secuenciación de próxima generación (NGS, por sus siglas en inglés) en el secuenciador Ion GeneStudio S5 (Thermo Fisher Scientific). En esta investigación se utilizaron tanto muestras de ADN comercial con mutaciones conocidas como muestras de pacientes reales, lo cual permitió evaluar la sensibilidad y especificidad del método en condiciones controladas y en un contexto clínico.

Para el procesamiento de los datos obtenidos a partir de la secuenciación se realizaron diversos análisis bioinformáticos, se emplearon sistemas de análisis disponibles que permiten el alineamiento de secuencias y la evaluación de la cobertura en las regiones objetivo. Estos análisis bioinformáticos son esenciales para cuantificar la profundidad de lectura en cada amplicón y asegurar que las mutaciones relevantes sean detectadas de manera precisa. Una cobertura adecuada garantiza que las regiones de interés estén bien representadas en la secuenciación, lo cual es crucial para la fiabilidad de los resultados obtenidos.

Los resultados preliminares indicaron que el panel propuesto es eficaz para detectar variaciones en genes de importancia clínica y para evaluar la inestabilidad de microsatélites. La validación en muestras de ADN comercial y de

pacientes permitió confirmar que se puede lograr una cobertura adecuada y un balance óptimo entre las lecturas de ambos paneles. Esta metodología es prometedora para su aplicación en el diagnóstico clínico del cáncer, facilitando un enfoque personalizado que optimiza recursos y mejora la precisión en la medicina genómica aplicada a la oncología en Uruguay.

Palabras clave: Cáncer, secuenciación de ADN, mutaciones accionables, NGS

3. Introducción

3.1. Cáncer

El cáncer es una enfermedad genética causada por la acumulación de mutaciones somáticas debido a errores en la replicación o daños en el ADN. Una mutación somática es un cambio en la secuencia de ADN de cualquiera de las células del cuerpo, excepto en las células germinativas. Cuando una mutación promueve la división celular o inhibe la apoptosis, la población de células prolifera, formando un tumor (Roy & Saikia, 2016).

Las células normales se dividen y mueren durante un periodo de tiempo programado. En cambio, la célula cancerosa “pierde” la capacidad para morir de forma programada y se divide casi sin límite, hasta formar una masa denominada tumor o neoplasia. Estos grupos de células cancerosas van acumulando alteraciones genéticas (en la secuencia de ADN) y epigenéticas (en la expresión de los genes), con las que adquieren habilidades extraordinarias que, en última instancia, los capacitan para diseminar a distancia o metastatizar (*¿Qué es el cáncer y cómo se desarrolla? - SEOM: Sociedad Española de Oncología Médica © 2019, s. f.*)

Las mutaciones pueden producirse a partir de errores en la replicación del ADN durante la división celular, la exposición a mutágenos o una infección viral. Cualquier fenómeno (“carcinógeno”) que dañe los genes en nuestras células – por ejemplo a través de la producción de los destructivos “radicales libres”- puede ocasionar cáncer. Pero para que la célula se torne cancerosa (“*transformación cancerosa*”) ha de ocurrir daño en varios genes causantes de cáncer en la misma célula. Es decir, los genes alterados tienen que formar parte de unos procesos característicos, o huellas, inherentes a la propia formación del cáncer: proliferación celular (o en sus inhibidores), resistencia a la muerte celular e inmortalidad, formación de vasos sanguíneos, capacidad de diseminación, reprogramación energética y evasión al sistema inmune. Debido a que cada uno

de estos procesos es enormemente complejo, con participación de muy numerosos mediadores que interactúan entre sí, puede llegarse a la misma enfermedad a través de mutaciones muy diversas.

En Uruguay, el cáncer es la causa de muerte más frecuente en personas menores a 70 años, y se calcula que en el período comprendido entre los años 2015 y 2019 fue responsable de 8.244 muertes al año en promedio. Los tipos de cáncer más frecuentes en Uruguay son el cáncer de mama, próstata, colorrectal y pulmón. En cuanto a la mortalidad, en mujeres el cáncer de mama es el tipo de cáncer con mayor tasa de mortalidad, seguido del cáncer de pulmón; mientras que, en los hombres, las estadísticas las lidera el cáncer de pulmón, seguido del colorrectal (Cáncer, s. f.). Estos aspectos clave pueden comprenderse mejor a través de la *Figura 1* que esquematiza las características distintivas tradicionales y las posibles nuevas adiciones en el desarrollo del cáncer.

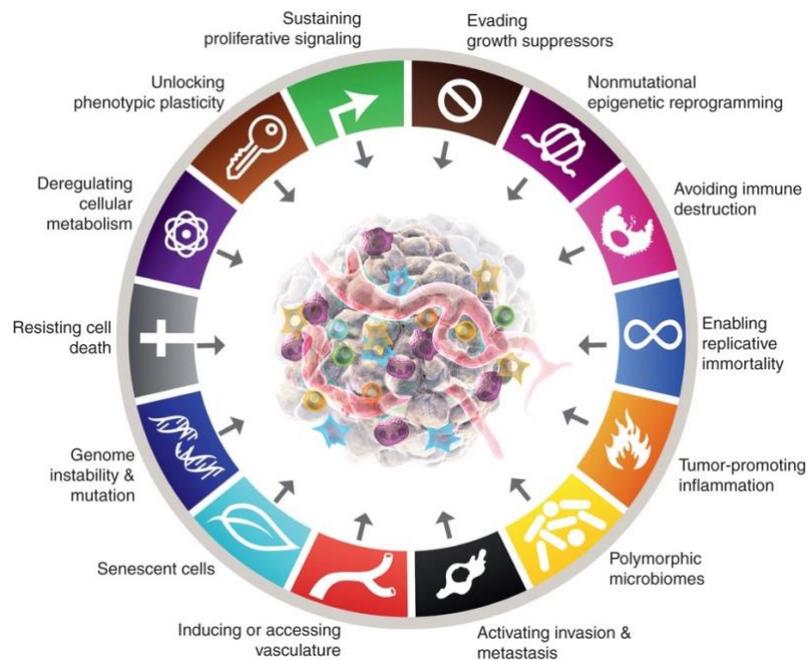


Figura 1: Representación de las características distintivas tradicionales y posibles nuevas adiciones del cáncer, incluyendo señales proliferativas sostenidas, evasión de supresores de crecimiento, reprogramación epigenética, evasión inmunitaria y microambiente inflamatorio, entre otros. Obtenido de (Hanahan, 2022).

La principal dificultad de esta enfermedad es que, a menudo se diagnostica en etapas avanzadas, lo que está asociado con un mal pronóstico y tratamiento. Sin embargo, el avance en el análisis genómico de los tumores ha permitido desarrollar tratamientos más prometedores, basados en terapias dirigidas que atacan alteraciones moleculares específicas (Karnes et al., 2014). Estos avances ofrecen nuevas esperanzas para los pacientes al mejorar la precisión del tratamiento y personalizar las intervenciones según las características del tumor.

3.1.1 Oncogenes y supresores tumorales

Un oncogén es un gen que sufrió una mutación y que tiene el potencial de causar cáncer. Antes de que un oncogén sufra la mutación se denomina protooncogén, y juega un papel en la regulación de la división celular normal. El cáncer puede ocurrir cuando un protooncogén sufre una mutación, que lo convierte en un oncogén y hace que las células se dividan y multipliquen en forma descontrolada. (*Oncogén*, s. f.-a). Este proceso puede observarse de manera gráfica en la *Figura 2*, que ilustra cómo una mutación en un oncogén conduce a una proliferación celular inadecuada

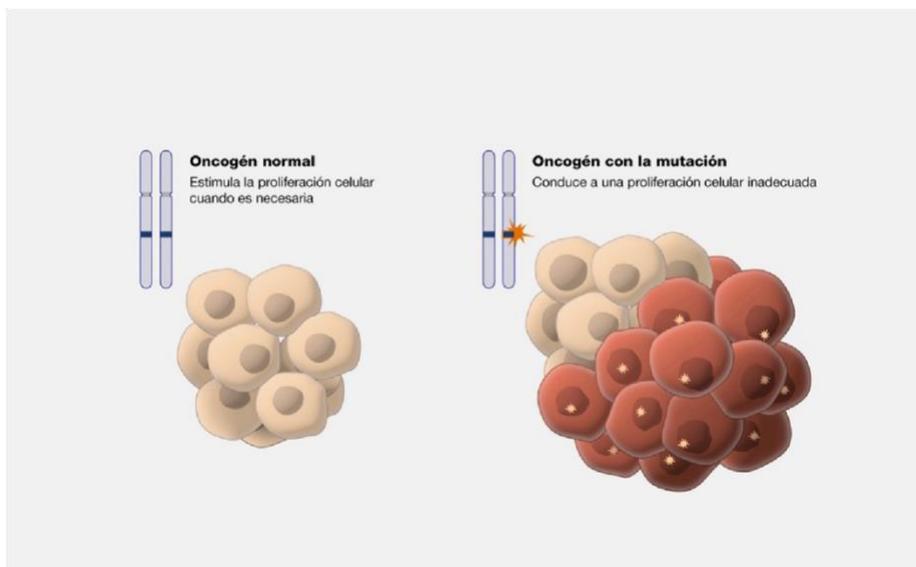


Figura 2: Ilustración de la estimulación de proliferación celular como consecuencia de una mutación en un oncogén. Imagen extraída de (Oncogén, s. f.-b)

Ejemplos de oncogenes incluyen mutaciones en el gen KRAS, que activan la proliferación celular en varios tipos de cáncer, como el colorrectal, pancreático y pulmonar (Mao et al., 2010), y el gen BRAF, cuya mutación V600E es común en el melanoma y otros cánceres, y es una diana de terapias dirigidas como vemurafenib y dabrafenib (Davies et al., 2002). La identificación de oncogenes mutados en tumores permite no solo entender mejor la biología del cáncer, sino también desarrollar tratamientos dirigidos que bloqueen las vías moleculares alteradas. Los inhibidores de tirosina quinasa, por ejemplo, actúan sobre mutaciones activadoras en el gen EGFR en cáncer de pulmón, lo que ha permitido mejorar las respuestas terapéuticas y la supervivencia en pacientes con este tipo de alteración (Pao & Chmielecki, 2010). La identificación de oncogenes accionables se ha convertido en un aspecto central de la oncología de precisión, ya que permite adaptar el tratamiento a las características genómicas del tumor, mejorando los resultados clínicos y reduciendo los efectos secundarios.

Los genes supresores de tumores son genes importantes que actúan dentro del genoma para regular varias funciones celulares. Estos genes se pueden clasificar ampliamente según su papel en el crecimiento celular/progresión del ciclo celular, la proliferación celular, los mecanismos de reparación del ADN y otras funciones cruciales de señalización celular, como la inducción de la apoptosis. Sin genes supresores de tumores funcionales, existe un alto riesgo de crecimiento celular desregulado, que es un mecanismo bien conocido para el desarrollo de cánceres (Sherr, 2004). Se han identificado mutaciones de pérdida de función en genes supresores de tumores en muchos tipos de cánceres, incluidos los de ovario, pulmón, colorrectal, cabeza y cuello, páncreas, útero, mama y vejiga. Incluso existen síndromes de cáncer familiares asociados con la pérdida de función, mutaciones de la línea germinal de genes supresores de tumores específicos, como el síndrome de Li-Fraumeni con la pérdida de TP53 (Joyce et al., 2024).

Muchos genes supresores de tumores han sido objeto de estudios y es probable que aún queden muchos más por descubrir. Los mecanismos de cada gen supresor de tumores y sus productos proteicos son complejos y están interrelacionados con otras vías de señalización celular, pero a continuación se muestran algunos mecanismos mejor conocidos:

El gen supresor de tumores TP53 también se conoce como el "guardián del genoma", ya que sirve para monitorear el estrés celular como la anoxia, identificar daños en el ADN o señalización inapropiada por oncoproteínas mutadas. El gen TP53 codifica la proteína p53, que controla la expresión de proteínas y su actividad en la detención del ciclo celular, la senescencia celular, la reparación del ADN y la apoptosis. La pérdida de p53 puede provocar una replicación celular continua a pesar del daño en el ADN y la imposibilidad de activar la muerte celular programada (Goh et al., 2011). En los casos en que el daño del ADN no es reparable, p53 induce eventos como la activación del gen BAX, que codifica una proteína proapoptótica que finalmente conduce a la apoptosis celular o senescencia. También actúa para inhibir el gen antiapoptótico BCL2 y estimula la liberación de citocromo c de las mitocondrias. El citocromo c activa las caspasas dentro de la célula responsables de su eventual degradación. Al igual que RB, p53 puede ser inactivado por oncoproteínas virales como la proteína E6 del VPH, revocando así los efectos antiproliferativos y otros efectos celulares importantes. La mayoría de los cánceres demuestran una mutación bialélica de pérdida de función en TP53. Los pacientes poco comunes con síndrome de Li-Fraumeni tienen una incidencia muy alta de una amplia variedad de cánceres, como cáncer de mama, sarcomas de tejidos blandos y óseos, y tumores cerebrales, ya que heredan una copia defectuosa de TP53 (Malkin, 2011; Stambolic et al., 1998; van Roy & Berx, 2008)

Comparando el perfil mutacional de oncogenes y genes supresores de tumor, vemos que estos presentan patrones de mutación distintos debido a sus mecanismos funcionales. Los oncogenes suelen mostrar mutaciones concentradas en zonas específicas denominadas zonas calientes o "hotspots",

ya que estas alteraciones puntuales permiten que la proteína adquiera una nueva función que promueve la proliferación celular descontrolada. Este fenómeno se observa en genes como KRAS, BRAF y EGFR, donde las mutaciones en sitios críticos activan constantemente las señales de crecimiento (Lito et al., 2016). Por el contrario, los genes supresores de tumores, como TP53 y RB1, tienden a experimentar mutaciones a lo largo de toda su secuencia codificante, ya que su inactivación no requiere una alteración en un sitio específico, sino que puede ocurrir en diversas regiones del gen, resultando en una pérdida de función general (Kandoth et al., 2013; Olivier et al., 2010). Esta diferencia en los patrones de mutación refleja la naturaleza funcional de estos genes en la oncogénesis y destaca la distinción entre mutaciones “gain-of-function” en oncogenes y “loss-of-function” en genes supresores de tumores.

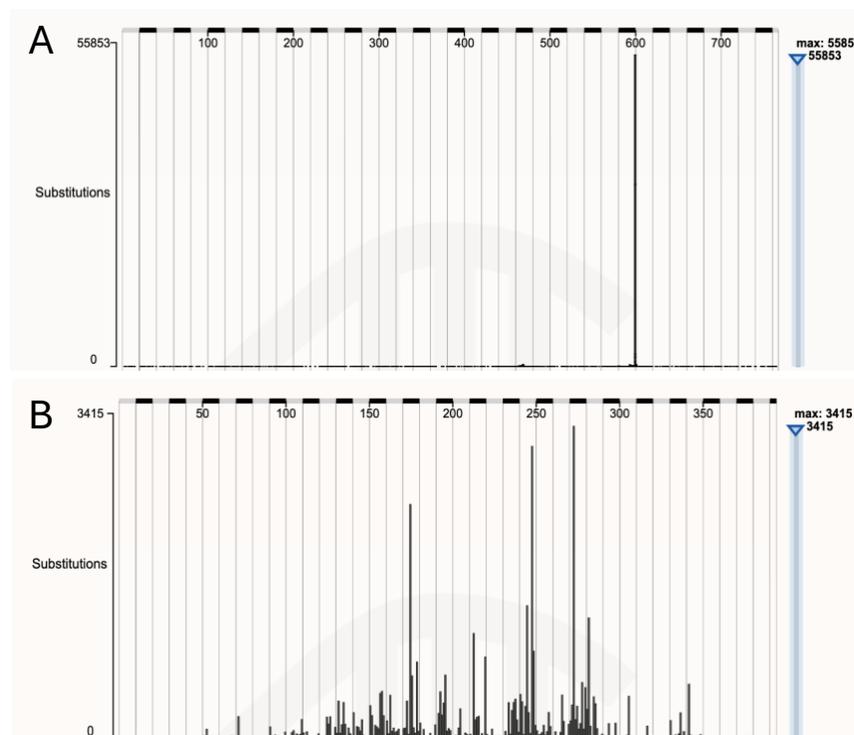


Figura 3: A: Perfil mutacional del oncogén BRAF, puntos críticos del genoma donde ocurren mutaciones del gen (“hotspots”) B: Perfil mutacional del gen supresor de tumor p53, mutaciones distribuidas a lo largo de toda la secuencia del gen. Imágenes obtenidas de (COSMIC | Catalogue of Somatic Mutations in Cancer, s. f.)

En este proyecto, trabajamos con una selección específica de oncogenes y genes supresores de tumores para evaluar su desarrollo y analizar la importancia de sus mutaciones en diversas muestras tumorales. Tanto los oncogenes como los genes supresores seleccionados para este estudio son aquellos que presentan mutaciones con terapias dirigidas disponibles, lo que permite explorar opciones terapéuticas específicas en función de las alteraciones genéticas identificadas.

3.1.2. Inestabilidad de microsatélites

La inestabilidad de microsatélites (MSI), tiene un rol importante en ciertos tipos de cáncer. Evaluar la presencia de MSI nos permite explorar alteraciones en los mecanismos de reparación del ADN y su impacto en la progresión tumoral.

Los microsatélites son secuencias cortas de ADN, por lo general formados por 1 a 4 pares de bases (pb), que se repiten múltiples veces en tándem en una ubicación genómica particular. Estas secuencias de ADN, habitualmente, no son codificantes. En ciertos tumores, es habitual encontrar inestabilidad de microsatélites. La inestabilidad de microsatélites (MSI) es una característica importante en diversos tipos de cáncer y se asocia a defectos en el sistema de reparación de errores de apareamiento (del inglés mismatch repair –MMR-). Este proceso de inestabilidad genética se observa cuando fallan las proteínas encargadas de corregir errores durante la replicación del ADN, dando lugar a variaciones en la longitud de los microsatélites, que son secuencias repetitivas en el ADN. La MSI es particularmente prevalente en cánceres colorrectal, gástrico y endometrial, y su detección puede ser crucial para identificar pacientes que podrían beneficiarse de terapias dirigidas, como el uso de inhibidores de puntos de control inmunitario (por ejemplo, pembrolizumab), aprobados para tumores con alta carga de mutaciones asociadas a MSI. La evaluación de MSI permite no sólo una mayor comprensión del comportamiento molecular del tumor, sino también una estratificación de pacientes que optimiza las opciones terapéuticas y mejora las tasas de respuesta clínica en pacientes con alta

inestabilidad genómica (Boland & Goel, 2010)(*Pembrolizumab - Instituto Nacional Del Cáncer*, 2014).

En el diagnóstico clínico, la detección de la inestabilidad de microsatélites (MSI) suele realizarse mediante la técnica de reacción en cadena de la polimerasa (PCR) en un conjunto de loci de referencia, como el panel Bethesda, que incluye marcadores estándar como BAT-25, BAT-26, D2S123, D5S346 y D17S250 (Boland & Goel, 2010). Este método permite comparar la longitud de las secuencias de microsatélites entre el tejido tumoral y el tejido normal del paciente para identificar cambios que indiquen MSI. Aunque esta técnica es rápida y menos costosa, tiene limitaciones importantes, como el análisis restringido a pocos marcadores, lo cual puede resultar insuficiente para detectar todas las variantes de MSI y limitar el diagnóstico en ciertos tipos de tumores (Vanderwalde et al., 2018). En contraste, la secuenciación de próxima generación (NGS) ofrece un enfoque más completo al permitir el análisis simultáneo de múltiples loci de microsatélites y otros genes relevantes en una sola prueba, optimizando así la precisión en la detección de MSI. Este estudio propone el uso de NGS para superar las limitaciones de los métodos convencionales y proporcionar una herramienta más robusta y efectiva para la caracterización molecular de tumores con alta inestabilidad genómica (Hause et al., 2016; Le et al., 2017).

La incorporación del estudio de inestabilidad de microsatélites (MSI) junto con un panel de genes en una única secuenciación de Next Generation Sequencing (NGS) representa un avance significativo en la caracterización molecular de tumores. Al integrar el análisis de MSI con un panel de genes, es posible no sólo evaluar la presencia de MSI, sino también identificar alteraciones genéticas en oncogenes y genes supresores de tumores relevantes para el desarrollo del cáncer y la respuesta a terapias dirigidas. Este enfoque combinado, habilitado por la tecnología NGS, facilita una evaluación comprensiva y simultánea de múltiples biomarcadores, optimizando el tiempo y los recursos, y promoviendo una medicina personalizada más eficaz.

3.2. Terapias dirigidas y mutaciones accionables:

El Instituto Nacional del Cáncer define la terapia dirigida como "un tipo de tratamiento contra el cáncer que se dirige a las proteínas que controlan cómo las células cancerosas crecen, se dividen y se propagan". El enfoque se denomina "dirigido" para distinguirlo de tratamientos, como la quimioterapia, que tienen un impacto más amplio. Por ejemplo, muchas formas de quimioterapia actúan contra las células que se dividen y forman nuevas células más rápidamente de lo normal. Esto hace que la quimioterapia sea eficaz contra las células cancerosas, porque su número aumenta más rápido que la mayoría de las células normales, pero también contribuye a los efectos secundarios de la quimioterapia, como la pérdida de cabello que se produce cuando la quimioterapia afecta a los folículos pilosos, que también se multiplican más rápido que la mayoría de las otras células del cuerpo (*Targeted Therapy for Cancer - NCI, 2014*).

A diferencia de la quimioterapia tradicional, la terapia dirigida actúa específicamente sobre proteínas presentes, generalmente, en la superficie de las células cancerosas que promueven su crecimiento y diseminación. La principal ventaja de este enfoque es que las terapias dirigidas tienen un efecto mínimo sobre las células sanas, reduciendo los efectos secundarios. Sin embargo, una limitación es que no todos los tipos de cáncer presentan las mutaciones genéticas que producen dichas proteínas. Por esta razón, se debe analizar el tumor de cada paciente para determinar si presenta una mutación accionable, lo cual se realiza mediante una biopsia y posterior análisis genómico de la misma (*Targeted Therapy for Cancer - NCI, 2014*).

Además del enfoque en terapias dirigidas, se ha incorporado el concepto de medicina de precisión, que implica la adaptación de los tratamientos médicos a las características individuales de cada paciente, considerando su perfil genético, el tipo específico de tumor y las mutaciones presentes en éste. Mientras que la terapia dirigida se focaliza en proteínas específicas que promueven el crecimiento y la diseminación de las células malignas, la medicina

de precisión amplía este enfoque al integrar un análisis exhaustivo de biomarcadores y factores moleculares que pueden influir en la respuesta terapéutica. Esto permite personalizar la estrategia de tratamiento, optimizando su efectividad y minimizando efectos secundarios adversos. En oncología, la medicina de precisión utiliza la información genética y molecular del tumor para seleccionar terapias que actúen específicamente sobre las mutaciones y alteraciones que impulsan la proliferación de células cancerosas (Ashley, 2016). Un componente central de este enfoque es la farmacogenómica, que estudia cómo las variaciones genéticas afectan la metabolización de los fármacos, permitiendo ajustar las dosis o seleccionar medicamentos basados en la capacidad metabólica del paciente. Variantes en genes como CYP2D6 y CYP2C19 pueden alterar la actividad enzimática, influyendo en la eficacia y seguridad de ciertos tratamientos (*Personalized medicine: Time for one-person trials - PubMed*, s. f.). Este enfoque integrado de medicina de precisión y terapias dirigidas está transformando el tratamiento oncológico, promoviendo una medicina personalizada y centrada en las características moleculares de cada paciente.

En Uruguay, las terapias dirigidas y medicamentos de alto costo pueden ser costeados por el Fondo Nacional de Recursos (FNR). Esta institución está regulada por la ley N°16.343 que funciona desde el año 1981, a partir de un decreto-ley. El objetivo de este fondo nacional es poder financiar en forma eficiente procedimientos de medicina altamente especializada y medicamentos de alto costo, comprendidos en las normativas de cobertura vigentes, para los usuarios del Sistema Nacional Integrado de Salud, contribuyendo a la mejora de la calidad y la sostenibilidad económico-financiera del sistema. En el caso de los medicamentos, el FNR negocia y compra los medicamentos para los tratamientos bajo cobertura, que son dispensados en su mayoría en las Farmacias de los prestadores integrales en donde están afiliados los pacientes. («¿Qué es el Fondo Nacional de Recursos?», s. f.).

En la normativa reciente (FNR), se ha ampliado la cobertura para incluir una serie de fármacos de terapia dirigida, diseñados para tratar diferentes tipos de cáncer en función de las características moleculares específicas de los tumores. Por ejemplo, la última inclusión de Pembrolizumab el cual es particularmente útil en pacientes con alta expresión de PD-L1 ya que bloquea la interacción entre PD-1 y PD-L1, permitiendo que el sistema inmunológico reconozca y ataque con mayor eficacia las células cancerosas («Nuevas normativas de cobertura», s. f.).

Otros fármacos disponibles que han demostrado alta eficiencia en terapias dirigidas incluyen Dabrafenib y Trametinib, los cuales se utilizan en pacientes con melanoma metastásico que presentan mutaciones en el gen BRAF, específicamente la mutación “hotspot” V600E. Estos tratamientos combinan inhibidores de BRAF y MEK, lo que ha mejorado significativamente la supervivencia de los pacientes con este tipo de alteración. De manera similar, las mutaciones activadoras del gen EGFR en el cáncer de pulmón de células no pequeñas (CPCNP) han llevado al desarrollo de inhibidores de la tirosina quinasa como Osimertinib, Erlotinib, Gefitinib y Afatinib, los cuales bloquean la señalización anormal que impulsa el crecimiento tumoral (Soria et al., 2018).

La gran cantidad de terapias dirigidas en rápido crecimiento puede categorizarse según sus respectivos efectos sobre una o más capacidades distintivas, como se ilustra en los ejemplos presentados en la *figura 4*. De hecho, la eficacia observada de estos medicamentos representa, en cada caso, una validación de una capacidad particular: si una capacidad es verdaderamente importante para la biología de los tumores, su inhibición debería obstaculizar el crecimiento y la progresión tumoral. Observamos que la mayoría de los medicamentos contra el cáncer dirigidos a características distintivas que se han desarrollado hasta la fecha han sido diseñados para atacar objetivos moleculares específicos que, de una u otra forma, facilitan capacidades particulares. Esta especificidad de acción se ha considerado una virtud, ya que presenta actividad inhibitoria contra un objetivo con, en principio, relativamente menos efectos fuera del objetivo y, por lo tanto, menos toxicidad no específica (Hanahan & Weinberg, 2011).

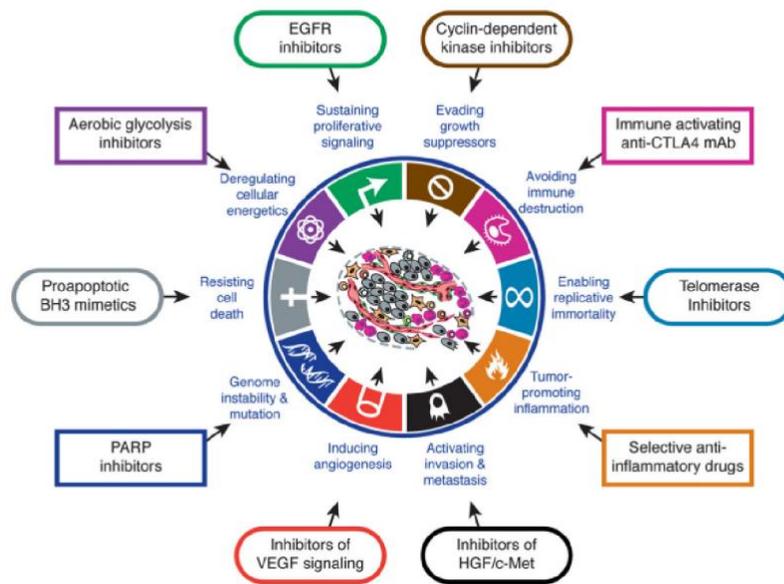


Figura 4: Medicamentos que interfieren con cada una de las capacidades adquiridas necesarias para el crecimiento y la progresión tumoral. Los medicamentos enumerados son solo ejemplos ilustrativos; existe un extenso pipeline de fármacos candidatos con diferentes objetivos moleculares y modos de acción en desarrollo para la mayoría de estas características distintivas. Obtenido de (Hanahan & Weinberg, 2011)

El creciente número de terapias dirigidas disponibles para distintas mutaciones específicas resalta la importancia de conocer el perfil molecular de cada tumor. La detección precisa y oportuna de estas alteraciones es fundamental para seleccionar el tratamiento adecuado para cada paciente. En este contexto, contar con una técnica unificadora y eficaz para analizar estas mutaciones no solo optimiza el manejo clínico, sino que también facilita el acceso a tratamientos personalizados, mejorando así el pronóstico y la calidad de vida de los pacientes con cáncer. La adopción de estas tecnologías representa un paso crucial hacia una medicina oncológica más precisa y accesible en Uruguay.

3.3. Análisis genómico por secuenciación masiva:

Tradicionalmente, el diagnóstico molecular del cáncer basado en la detección de cambios de un nucleótido o pequeñas inserciones o deleciones en uno o varios

genes se basaba en la secuenciación Sanger. Debido al creciente número de genes implicados en el cáncer, actualmente esta técnica resulta ardua y costosa. Además, requiere una gran cantidad de ADN tumoral, en ocasiones difícil de conseguir, y no permite la detección de cambios en el ADN de poblaciones subclonales. El desarrollo de la secuenciación NGS ha supuesto una revolución en el abordaje del diagnóstico oncológico y ha proporcionado una información sobre la biología molecular tumoral más completa y accesible a la práctica clínica. Gracias al desarrollo de estas nuevas tecnologías, junto con la implementación de programas bioinformáticos de análisis de datos, la identificación de alteraciones patogénicas y de nuevos genes relacionados con el desarrollo de una enfermedad es ahora más fácil, rápida y tiene una mejor relación costo-efectividad (Calabria et al., 2016).

A pesar de los avances en las tecnologías de secuenciación, la secuenciación completa del genoma sigue siendo costosa y, por lo tanto, menos accesible para su implementación en la práctica clínica de rutina (Mardis, 2013). Esta limitación económica ha impulsado el desarrollo de enfoques más específicos, como la secuenciación de paneles de genes, que permite analizar únicamente aquellas regiones del genoma que tienen una alta probabilidad de ser clínicamente relevantes para el paciente. De esta manera, se optimizan los recursos al focalizarse en genes o regiones genómicas específicas que han demostrado tener un impacto directo en el pronóstico o tratamiento del cáncer (Murtaza et al., 2013; Zehir et al., 2017)

La secuenciación con una gran profundidad de lectura en estas regiones seleccionadas es esencial para detectar alteraciones en subclones tumorales, lo cual aporta una sensibilidad diagnóstica superior en comparación con técnicas de menor resolución. Este enfoque justifica la elección metodológica del presente trabajo, donde se emplea un panel de genes de alta relevancia clínica y se prioriza una secuenciación de alta profundidad, optimizando así la relación costo-beneficio y facilitando una aplicación práctica en el diagnóstico oncológico. Además, esto permite una implementación más viable en la práctica clínica al

enfocarse en alteraciones genéticas específicas que son directamente accionables en términos terapéuticos.

La creación de un panel de genes óptimo para evaluar alteraciones genómicas en tumores de diferentes orígenes tisulares, sumado al estudio de la inestabilidad de microsatélites, generalmente se realiza de manera fragmentada, lo que implica la participación de varios laboratorios y un mayor volumen de muestras. La capacidad de analizar todas estas regiones en una única secuenciación sigue siendo un desafío que se abordará en este trabajo de tesis. Adicionalmente, la interpretación de los resultados representa otro reto considerable, debido al aumento en el número de genes evaluados. Es crucial limitar los genes incluidos a aquellos que han demostrado tener una relevancia clínica, así como realizar revisiones periódicas de las variantes detectadas y monitorear a los portadores para esperar nuevos hallazgos que aporten más evidencia científica para su futura clasificación clínica (Rosado-Jiménez et al., 2023).

4. Objetivos:

Objetivo general: Validación de un panel de genes que permite detectar alteraciones genéticas accionables en muestras de tumores en parafina (FFPE) mediante secuenciación masiva (NGS)

Objetivos específicos:

- Evaluar y seleccionar un kit de extracción de ADN desde tumores en parafina (FFPE) que permita obtener material adecuado para ser utilizado en secuenciación masiva.
- Determinar las condiciones de amplificación y combinación de dos paneles independientes para obtener una única librería de secuenciación por muestra.
- Evaluar el desempeño del panel tanto en muestras comerciales como en muestras tumorales reales de archivo.

5. Diseño de investigación y metodología:

A nivel esquemático se presenta el flujo de trabajo en la *figura 5*. En breve, la metodología empleada en este estudio abarcó desde la extracción del ADN, hasta la interpretación de los resultados obtenidos del análisis por secuenciación de nueva generación (NGS). El protocolo de análisis integral de la muestra incluye la extracción de ADN, cuantificación y evaluación de la calidad del ADN obtenido, preparación de bibliotecas, secuenciación y análisis bioinformático.

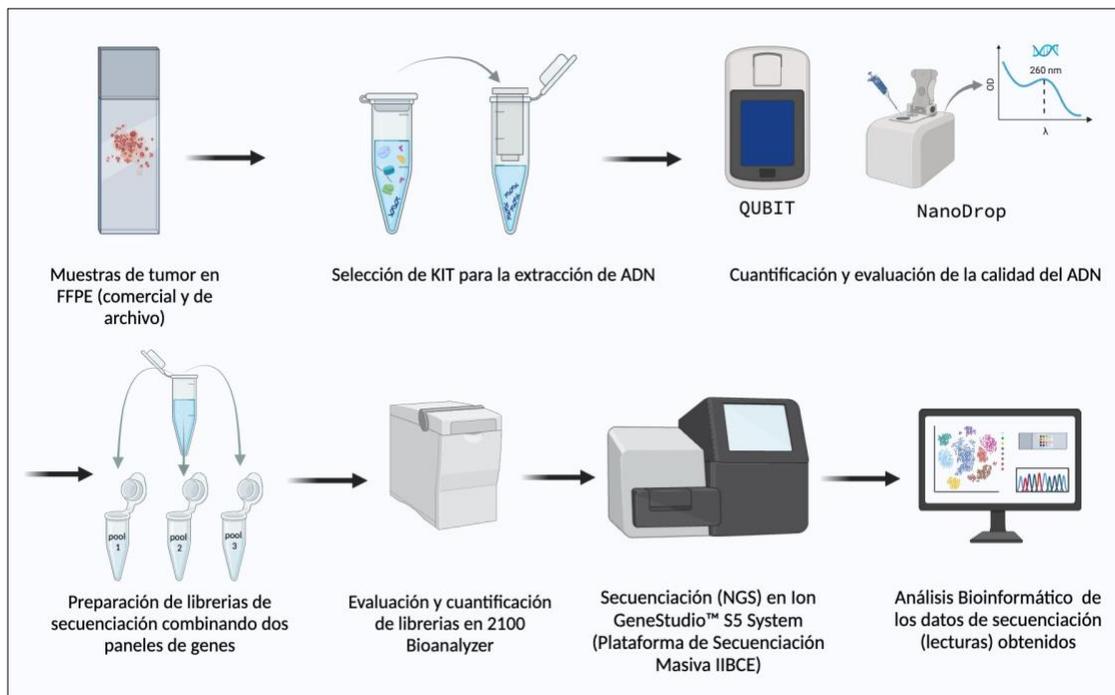


Figura 5: Esquema de la metodología empleada para el análisis de muestras de tumor en FFPE y posterior análisis mediante NGS.

5.1. Selección de las regiones a estudiar y diseño de panel:

La selección de genes y regiones específicas para la secuenciación se realizó en base a una revisión exhaustiva de la bibliografía, con el objetivo de identificar genes que contienen mutaciones accionables, es decir, alteraciones genéticas que tienen relevancia clínica y pueden guiar la selección de tratamientos

específicos en pacientes con cáncer. En algunos genes candidatos se evaluarán toda la región codificante debido a la alta diversidad de mutaciones que suelen presentar, como ocurre comúnmente en los genes supresores de tumor. En otros casos, se analizarán únicamente regiones conocidas como "*hotspots*" o sitios recurrentes de mutación, que han demostrado tener implicancia en la proliferación tumoral y en la respuesta a terapias dirigidas, como sucede con los oncogenes. Los genes analizados a nivel de *hotspots* se encuentran marcados con un asterisco en la *tabla 1*.

Con el fin de optimizar recursos y realizar un análisis integral, se emplearon dos paneles distintos para estudiar tanto los genes blanco como la inestabilidad de microsatélites. El primer panel, de diseño personalizado (OAD239222), incluye un total de 17 genes seleccionados por su implicancia en diversos tipos de cáncer y la disponibilidad de terapias dirigidas para sus mutaciones. Este panel involucra 926 amplicones, agrupados en dos grupos (pools) para mejorar la eficiencia de la amplificación y tiene un tamaño de 59,45kb. El segundo panel es un panel comercial, denominado *Ion AmpliSeq™ Microsatellite Instability Research Panel* de ThermoFisher Scientific, que fue diseñado específicamente para la detección de inestabilidad de microsatélites (MSI). Este panel utiliza cebadores específicos para amplificar 76 regiones de microsatélites (amplicones), optimizando así la detección de inestabilidad en múltiples loci genómicos. Este panel cubre 6,5 kb y los amplicones generados están ubicados en los cromosomas del 1 al 20 y en el cromosoma X. Identificar la inestabilidad de microsatélites en tumores permite el tratamiento con Pembrolizumab y Nivolumab.

La utilización de estos dos paneles permite una caracterización exhaustiva de las muestras tumorales, cubriendo tanto las alteraciones genéticas accionables en genes críticos como la evaluación de MSI en una sola prueba. El uso del panel personalizado asegura una alta especificidad para mutaciones relevantes en la clínica, mientras que el panel comercial de MSI proporciona un análisis robusto y estandarizado de la inestabilidad genómica. La combinación de estos

enfoques garantiza una cobertura óptima y permite realizar un análisis detallado y económicamente rentable en condiciones similares a las de un entorno de diagnóstico clínico.

La lista detallada de los genes incluidos en el panel de genes personalizado (OAD239222) se presenta en la *tabla 1*.

Tabla 1: Genes seleccionados para su estudio en el panel de genes. Se incluye el origen tisular en el cual suelen estar mutados y el fármaco dirigido a una mutación en ese gen. Oncogenes marcados con un asterisco (*), se secuencian zonas “hotspot”. (GeneCards - Human Genes | Gene Database | Gene Search, s. f.).

<i>GEN DIANA</i>	<i>TIPO DE CÁNCER</i>	<i>AGENTES TERAPÉUTICOS</i>
ATM	Cáncer de mama, cáncer de ovario	Doxorubicin, Rucaparib, Cyclophosphamide
BARD1	Cáncer de mama	Carboplatin, Docetaxel, Trastuzumab, Olaparib
BRAF*	Melanoma, cáncer de tiroides, cáncer de pulmón	Vemurafenib, Dabrafenib, Trametinib, Cobimetinib, Encorafenib, Binimetinib
BRCA1	Cáncer de mama, cáncer de ovario, cáncer de próstata	Olaparib, Rucaparib, Talazoparib
BRCA2	Cáncer de mama, cáncer de ovario, cáncer de próstata	Olaparib, Rucaparib, Talazoparib
BRIP1	Cáncer de ovario	Rucaparib
CDK12	Cáncer de próstata	Nivolumab, Olaparib, Pembrolizumab, Abiraterone
CDKN2A	Melanoma, cáncer de páncreas, cáncer de pulmón	Palbociclib, Ribociclib, Abemaciclib
CHEK1	Cáncer de mama, cáncer de ovario	Cisplatin, Etoposide, Gemcitabine, Floxuridine
CHEK2	Cáncer de mama	Fostamatinib, Rucaparib, Floxuridine, Gefitinib, Olaparib
EGFR*	Cáncer de pulmón, cáncer colorrectal	Erlotinib, Gefitinib, Afatinib, Osimertinib, Cetuximab, Panitumumab, Necitumumab
FANCL	Cánceres relacionados con anemia de Fanconi	Adenosine phosphate 24

KRAS*	Cáncer colorrectal, cáncer de pulmón, cáncer de páncreas	Sotorasib, Adagrasib, Cetuximab, Panitumumab
MET*	Cáncer de pulmón	Crizotinib, Capmatinib, Tepotinib
NRAS*	Melanoma, cáncer de tiroides	Carboplatin, Cetuximab, Cisplatin, Dabrafenib, Gemcitabine
PALB2	Cáncer de mama, cáncer de páncreas	Olaparib
PIK3CA	Cáncer de mama, cáncer colorrectal	Alpelisib, Everolimus

5.2. Extracción de ADN de muestras de tumores en FFPE:

Las muestras de ADN utilizadas en este estudio fueron seleccionadas para representar condiciones que simulan un entorno de laboratorio clínico y que permiten evaluar el desempeño de los paneles de secuenciación en la detección de mutaciones accionables. Se incluyeron tanto muestras comerciales de control, adquiridas de Horizon Discovery, como muestras archivadas de pacientes reales obtenidas del Banco de Tumores del Hospital Central de las Fuerzas Armadas, las cuales están embebidas en bloques de parafina (FFPE). Las muestras comerciales presentan mutaciones conocidas con frecuencias alélicas definidas, lo que permite evaluar la sensibilidad y especificidad del método en condiciones controladas y estandarizadas. Las muestras comerciales empleadas con sus números de catálogo se presentan en la tabla 2. Esto es especialmente relevante para validar la detección de variantes genéticas con importancia clínica. Por otro lado, las muestras de archivo representan un modelo clínico real, con antigüedades y calidades de ADN variables que pueden reflejar las limitaciones y desafíos del procesamiento de tejidos en FFPE en entornos de diagnóstico. Esta combinación de muestras controladas y clínicas asegura una validación integral del panel en condiciones equivalentes a las esperadas para el empleo de esta tecnología en la etapa final.

Tabla 2: Muestras comerciales control utilizadas para la validación del panel

<i>Código de catálogo</i>	<i>Nº de genes mutados</i>	<i>Genes mutados incluidos en nuestro panel</i>	<i>Frecuencia alélica</i>
HD125	1	1: EGFR	5%
HD810	5	4: BRCA1, BRCA2, BRIP1, BARD1	7.5% -100%
HD832	151	10: BARD1, BRAF, BRCA1, BRCA2, CDK12, EGFR, KRAS, MET, NRAS, PIK3CA	1% - 92.5%

Para obtener el ADN de estas muestras se ensayaron dos kits de extracción de ADN específicos para muestras en parafina (FFPE): QIAamp DNA FFPE kit (Cat #56404 Qiagen, USA) y Cobas® DNA Sample Preparation kit (M/N: 05985536190, Roche). Estos kits son ampliamente utilizados para este fin con buen rendimiento (Bonnet et al., 2018) (Kumar et al., 2016a). Luego de la evaluación y en base a la calidad del ADN obtenido se terminó optando por trabajar con el kit de QIAamp de aquí en adelante. Los protocolos de extracción de ambos kits comparten principios fundamentales, como el uso de xileno y etanol para quitar la parafina de las muestras, seguidos de una lisis enzimática con proteinasa K y una incubación a alta temperatura para liberar el ADN y revertir parcialmente los enlaces inespecíficos producidos por la formalina. Sin embargo, estos kits presentan diferencias importantes en los métodos de captura y purificación del ADN. El kit Cobas® utiliza un sistema de unión a fibras de vidrio, donde el ADN se retiene en un filtro durante los pasos de centrifugación, mientras que el QIAamp se basa en una membrana de sílica en columnas MinElute, ofreciendo una mayor flexibilidad en el volumen de elución final. Además, el QIAamp permite la automatización del protocolo en sistemas QIAcube, facilitando el procesamiento de múltiples muestras de manera simultánea, lo que podría representar una ventaja en términos de eficiencia en laboratorios de alto rendimiento. Esta comparación entre ambos kits permitió evaluar cuál ofrece mejor rendimiento en términos de cantidad y pureza del ADN extraído, factores relevantes para los análisis posteriores de este trabajo.

A su vez, ensayamos 15 muestras tumorales de archivo, provenientes del Banco de Tumores del Hospital Central de las Fuerzas Armadas.

5.3. Cuantificación y evaluación del ADN obtenido:

Para verificar que el ADN obtenido por el kit es adecuado para utilizar NGS, se procederá a determinar la presencia de contaminantes en la muestra, utilizando espectrofotometría de microvolúmenes (NanoDrop 1000 Spectrophotometer, ThermoFisher Scientific). En primer lugar, se determinarán las relaciones de absorbancia 260/280 y 260/230, las cuales nos informan sobre la pureza de la extracción. El cociente esperado para estas muestras es de ~1.8 para 260/280 y en el rango 2.0 – 2.2 para la relación 260/230.

A pesar de que NanoDrop también nos informa la concentración de la muestra, para obtener este valor utilizamos un método más específico para estas muestras: la fluorimetría utilizando kits de Qubit (ThermoFisher Scientific). Este método se basa en el uso de sondas fluorescentes para determinar la concentración de ácidos nucleicos y proteínas en una muestra. Cada sonda es específica para un tipo de molécula: ADN, ARN o proteína. Estas sondas tienen una fluorescencia despreciable al encontrarse libres en solución. Sin embargo, al unirse con sus moléculas objetivo: ADN, ARN o proteína, las sondas se vuelven intensamente fluorescentes (Kumar et al., 2016b). Más específicamente, se utilizó el kit Qubit dsDNA HS assay (ThermoFisher Scientific) con los tubos correspondientes (Qubit™ Assay Tubes, Cat #: Q32856).

Para evaluar la integridad del ADN extraído de muestras embebidas en parafina (FFPE), se realizó una electroforesis en gel de agarosa al 2%, comparando los resultados obtenidos con datos reportados en la bibliografía de otros estudios que trabajan con ADN de muestras en FFPE. Posteriormente, se llevaron a cabo dos PCR de prueba: una con cebadores que amplifican fragmentos de 195 pares de bases y otra para un fragmento de 153 pb. Estas PCR de control tienen como

objetivo confirmar que, a pesar de la fragmentación característica del ADN obtenido a partir de muestras en FFPE, el material conserva suficiente integridad para ser amplificado en productos de esta longitud.

Tabla 3: Lista de cebadores utilizados para esta PCR

<i>Cebadores</i>	<i>Secuencia (5'->3')</i>	<i>Largo (pb)</i>	<i>Tm Teórica (Primer blast NCBI)</i>	<i>Tm experimental</i>	<i>Producto (pb)</i>	<i>Región que amplifica</i>
<i>Cebador 1 sentido (F)</i>	<i>TAGAGGGACTGACGGGTGAT</i>	<i>20</i>	<i>59,37</i>	<i>62</i>	<i>153</i>	<i>Exón 7 de MUTYH</i>
<i>Cebador 1 antisentido (R)</i>	<i>CTCCTGGGTCCTACCCTCC</i>	<i>20</i>	<i>60,69</i>	<i>62</i>	<i>153</i>	
<i>Cebador 2 sentido (F)</i>	<i>GTGGTATTGCAGGCCTCTGT</i>	<i>20</i>	<i>60,04</i>	<i>62</i>	<i>195</i>	<i>Exón 3 de MUTYH</i>
<i>Cebador 2 antisentido (R)</i>	<i>TGAGGAGTTAGGGTGGAGGG</i>	<i>20</i>	<i>59,95</i>	<i>62</i>	<i>195</i>	

5.4. Preparación de librerías de secuenciación:

Para la preparación de librerías de secuenciación empleamos el kit Ion AmpliSeq Library Kit Plus, un kit comercial optimizado para la preparación rápida y eficiente de librerías de amplicones mediante paneles de Ion AmpliSeq. En la *figura 6* se presenta un esquema del procedimiento empleado. Adaptamos el protocolo estándar proporcionado por el fabricante para ajustarlo a las necesidades específicas de nuestro estudio. Una descripción detallada del protocolo modificado se encuentra en el anexo (sección 11.1). En términos generales, cada pool de cebadores se amplifica mediante una PCR individual; luego, los productos de PCR se combinan en proporciones equimolares, permitiendo la secuenciación conjunta de ambos paneles. Posteriormente, se realiza una digestión con la enzima FuPa, y se incorporan barcodes y adaptadores. Finalmente, se efectúa una amplificación por PCR adicional. La librería resultante es cuantificada utilizando Qubit, y su calidad se evalúa mediante el análisis en Bioanalyzer.

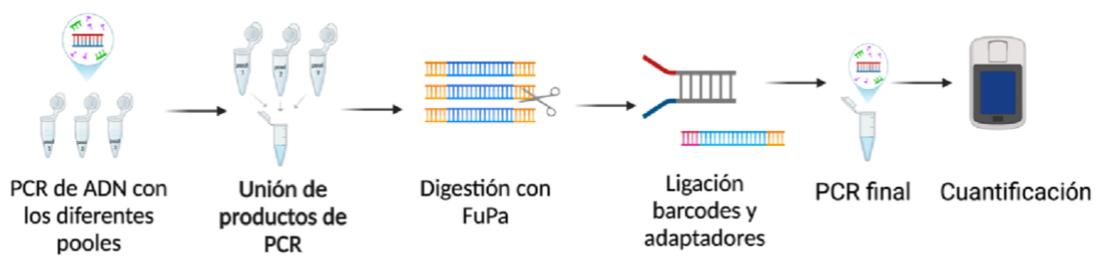


Figura 6: Esquema de la metodología empleada para la elaboración de librerías de secuenciación (BioRender, s. f.)

El kit Ion AmpliSeq Library Kit Plus permite realizar reacciones de PCR en múltiples sitios de manera escalable, con capacidad para amplificar simultáneamente desde 12 hasta 24,000 regiones genómicas objetivo en un solo pocillo. Gracias a esta capacidad, es posible preparar librerías para secuenciación de próxima generación a partir de cantidades ínfimas de ADN inicial, incluso tan baja como 1 ng, lo cual es especialmente ventajoso cuando se trabaja con muestras escasas o fragmentadas, como aquellas embebidas en parafina (FFPE) (Ion AmpliSeq™ Library Kit Plus, s. f.).

Entre las características destacadas del kit Ion AmpliSeq Library Kit Plus se encuentran la preparación de librerías altamente multiplexadas, la posibilidad de realizar la amplificación de miles de regiones objetivo con una cobertura uniforme y alta especificidad. También se incluye la eliminación de las secuencias de cebadores durante la preparación de la librería, lo que permite el análisis preciso de las regiones objetivo. El kit emplea un formato en placas que facilita el manejo de muestras, permitiendo compatibilidad con laboratorios de alto rendimiento y protocolos automatizados.

Adicionalmente, el kit incluye la opción de realizar multiplexado con códigos de barras mediante los adaptadores de código de barras Ion Xpress o IonCode, permitiendo la combinación de múltiples muestras en un solo chip de secuenciación. Esto optimiza tanto el tiempo como el costo de las corridas de secuenciación y asegura comparaciones precisas entre muestras. La preparación de librerías mediante el Ion AmpliSeq Library Kit Plus resulta en un

producto listo para la preparación de molde en el sistema Ion Chef, seguido de secuenciación (*Ion AmpliSeq™ Library Kit Plus*, s. f.).

Adaptamos el protocolo de Ion AmpliSeq™ para realizar la secuenciación unificada del panel de genes y del panel de inestabilidad de microsatélites (MSI). Para lograr esto fue necesario combinar los productos de PCR generados por ambos paneles. Establecimos ciertos supuestos en la mezcla de los productos de PCR, con el objetivo de asegurar que, al secuenciar la mezcla conjunta, se logre una cobertura y profundidad adecuada para ambos paneles, de forma que se pueda garantizar un análisis equilibrado y confiable de todas las regiones de interés en simultáneo.

Una vez obtenidas las librerías, dada su baja concentración, fueron evaluadas utilizando electroforesis automatizada, en un 2100 Bioanalyzer Instrument (Agilent Technologies) utilizando el High Sensitivity DNA Kit. Este método nos permite separar y estimar el tamaño de fragmentos de ADN que poseen concentraciones en el orden de pico y nanogramos, utilizando solamente 1 µL de muestra. Es necesario utilizar este equipo ya que evaluar el tamaño de los amplicones utilizando una electroforesis convencional en gel de agarosa 2% consumiría la totalidad de la librería producida.

5.5. Secuenciación de próxima generación:

Para realizar la secuenciación (NGS), se utilizó el sistema Ion GeneStudio S5 (Thermo Fisher Scientific) disponible en la plataforma de secuenciación masiva del Instituto de Investigaciones Biológicas Clemente Estable (IIBCE). Este equipo permite obtener datos de alta calidad optimizando los tiempos de procesamiento y ofreciendo una gran precisión en la identificación de variantes genéticas (*Ion GeneStudio System Features - UY*, s. f.).

Para la secuenciación, se seleccionó el chip Ion 530, diseñado para proporcionar una alta capacidad de datos en un formato compacto, lo cual resulta adecuado para estudios que requieren una cobertura de profundidad media a alta. Este chip permite procesar hasta 850 flujos, lo cual es ideal para este tipo de

proyectos, asegurando dar lugar a lecturas del largo adecuado para el análisis de genes específicos o regiones del genoma (*Ion 530™ Chip Kit*, s. f.). Los datos obtenidos de la secuenciación fueron procesados y analizados en el software de análisis de Thermo Fisher Scientific (Torrent Server), que ofrece herramientas estandarizadas para las tareas generales de análisis de datos de NGS. Esto permitió una detección precisa de las variantes genómicas presentes en nuestras muestras. Esta metodología garantiza un alto nivel de precisión en la identificación de variantes, incluso en muestras FFPE, conocidas por su degradación inherente.

5.6. Análisis de los datos de secuenciación:

5.6.1 Análisis automatizado con Ion Torrent Server

Para el análisis de los datos de secuenciación se utilizó el software Ion Torrent Server, que permite realizar un procesamiento automatizado de las secuencias obtenidas, evaluar la cobertura y variantes en las regiones objetivo. En primer lugar, el software realiza el alineamiento o mapeo de las lecturas sobre el genoma humano de referencia que se le indica. Posteriormente, se aplicó el plugin de cobertura, que cuantifica la profundidad y uniformidad de las lecturas en cada región del panel, generando conteos que expresan la cantidad de lecturas que corresponden a cada amplicón. Este análisis es fundamental para corroborar que se obtenga una cobertura suficiente en las áreas de interés, detectando cualquier zona con baja profundidad que podría comprometer la calidad del análisis.

A continuación, se procedió con el llamado de variantes, un proceso automatizado que identifica cambios en las secuencias de ADN, como SNPs (polimorfismos de nucleótido único) e indels (inserciones y deleciones). Se utilizó el Ion Torrent Server, con el plugin Variant caller y seleccionando la opción de Somatic, low stringency. Este análisis da como resultado un archivo de variantes, específicamente en formato vcf (Variant Call Format). Este archivo se realiza comparando las secuencias obtenidas para cada muestra con el genoma de

referencia lo que da lugar a un listado de las diferencias observadas. También indica para las variantes detectadas, información sobre su calidad y profundidad de cobertura, permitiendo una evaluación detallada de las mutaciones presentes en la muestra.

Además, se incluyó un análisis de inestabilidad de microsatélites (MSI) cuya ejecución queda sujeta a la disponibilidad de recursos y calidad de los datos. En conjunto, este flujo automatizado en Ion Torrent Server permite un análisis integral de los datos de secuenciación, proporcionando información clave sobre la calidad de las lecturas, la presencia de variantes y la estabilidad de las secuencias en las regiones de interés.

5.6.2. Análisis bioinformáticos con herramientas de uso libre

Para el análisis bioinformático de las lecturas de secuenciación, utilizamos herramientas de uso libre que permitieron diversas manipulaciones y análisis de los datos. En particular nos centramos en la evaluación de la cobertura en las regiones objetivo. Gran parte de este trabajo se realizó utilizando programas de uso libre que corren en línea de comando en un servidor con Sistema Operativo Linux. El detalle del conjunto de comandos se presenta en el Anexo (sección 11.2) Primero, las lecturas de secuenciación fueron alineadas al genoma humano de referencia (versión hg19) utilizando Bowtie2, una herramienta de mapeo rápida y precisa. Este paso permitió identificar la ubicación exacta de cada lectura en el genoma, generando un archivo de mapeo (formato SAM) que contiene información sobre la posición de cada lectura y la calidad de la alineación. Este alineamiento es fundamental para asignar la correspondencia entre las lecturas obtenidas y las regiones genómicas donde se generaron. Posteriormente, los archivos de mapeo generados en formato SAM se convirtieron al formato binario BAM, lo cual permite una compresión y almacenamiento más eficiente de los datos. Una vez en formato BAM, el archivo fue ordenado según la posición de las lecturas en el genoma, optimizando su

organización para análisis posteriores. Además, se generó un archivo índice en formato BAI, lo que permite acceder rápidamente a regiones específicas del genoma en el archivo BAM, permitiendo la observación de las áreas de interés durante el análisis de variantes y cobertura en visualizadores genómicos.

Luego, se utilizó el paquete Bedtools para calcular la cobertura de cada región de interés. Esta herramienta cuantifica el número de veces que cada región o amplicón fue cubierto por las lecturas obtenidas, generando métricas que indican la profundidad de cobertura en cada región. Este análisis es clave para evaluar la eficiencia del panel de secuenciación, ya que permite verificar que cada amplicón haya recibido una cobertura adecuada para poder determinar variantes en esas regiones. En conjunto, estos pasos nos proporcionaron un análisis bioinformático exhaustivo que asegura tanto la calidad de las alineaciones como la cobertura en las regiones objetivo.

Por otro lado, una vez obtenido el análisis de variantes a partir del Servidor de Ion Torrent, se procedió a realizar una identificación de las variantes en los 17 genes mediante herramientas de uso libre online. Utilizando el archivo .vcf obtenido de Ion Torrent este fue anotado utilizando wANNOVAR (wANNOVAR, s. f.), lo que dio lugar a una tabla para cada muestra que presenta las variantes halladas y toda la información anotada sobre ellas, tales como frecuencia alélica en 1000 genomas y en gnomAD (Genome Aggregation Database), calidad de la lectura (Phred score), así como las predicciones *in silico* utilizando varias herramientas tales como SIFT, Polyphen2 entre otras.

6. Resultados y Discusión:

6.1. Obtención y evaluación de ADN

6.1.1. Selección de kit de extracción

Para evaluar y seleccionar el kit de extracción de ADN más adecuado para las muestras comerciales y de archivo embebidas en parafina, realizamos en primera instancia la extracción de dos muestras de tumorales de archivo. Cada muestra fue tratada utilizando un kit de extracción diferente: el QIAamp DNA FFPE Kit y el Cobas® DNA Sample Preparation Kit. Para poder cuantificar el ADN extraído y validar la presencia de contaminantes, utilizamos los métodos de cuantificación Qubit (método fluorimétrico Qubit™ Assay Tubes, Cat #: Q32856) y Nanodrop (NanoDrop 1000 Spectrophotometer, ThermoFisher Scientific). Esta comparación directa nos permitió identificar cuál de los kits de extracción ofrece un rendimiento superior en términos de calidad y cantidad de ADN extraído, lo cual nos sirvió para poder elegir el método más adecuado y efectivo para nuestras muestras de estudio. La guía utilizada para cada extracción fue la brindada por el comerciante de cada kit comercial (*HB-0353-004_HB_QA_DNA_FFPE_0220_WW*, s. f.)(*Cobas DNA protocol*, s. f.).

Las muestras 1 y 3 corresponden al mismo tejido de tumor, al igual que las muestras 2 y 4, que también provienen de un único tumor.

Tabla 4: Comparación resultados de extracción de ADN con kit Qiagen vs Cobas.

Nº de muestra	Kit de extracción	Volumen final (µL)	Abs 260nm	Relación Abs 260/280nm	Relación Abs 260/230nm	Concentración ADN NanoDrop (ng/µL)	Concentración ADN Qubit (ng/µL)	Cantidad ADN (ng) obtenido
1	Cobas	100	0,612	1,71	1,57	30,6	2,54	254
2	Cobas	100	1,256	1,81	1,83	62,8	6,12	612
3	Qiagen	30	3,193	1,90	2,16	159,7	8,98	269
4	Qiagen	30	2,516	1,91	2,17	125,8	5,66	170

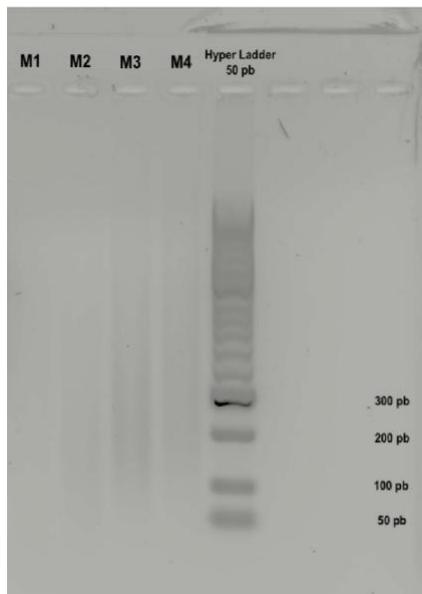
En primer lugar, evaluamos la relación de absorbancia 260/280 nm, que es indicativa la presencia de proteínas contaminantes (con máximo de absorción a 280nm) en nuestra muestra de ADN (máximo de absorción a 260nm). Al comparar los resultados obtenidos para las muestras 1 y 3 y las muestras 2 y 4, se observa que la relación de absorbancia 260/280 muestra mejores valores en las extracciones realizadas con el kit de Qiagen. Este kit logra ratios entre 1.8 y 2.0, lo cual es indicativo de un ADN de pureza adecuada según los estándares establecidos.

Además, evaluamos la relación de absorbancia 260/230 nm, lo que nos informa sobre la presencia de contaminantes como sales, fenoles u otros compuestos orgánicos que tienen un máximo de absorción en el rango de los 230 nm. Los resultados en este caso también favorecen al kit de Qiagen, con valores comprendidos entre 2.0 y 2.2. Estos resultados sugieren que el kit de Qiagen es superior en términos de pureza del ADN extraído, tanto en la eliminación de proteínas como de otros contaminantes.

Al observar la concentración de ADN obtenida con cada kit, notamos que los valores reportados por NanoDrop y Qubit para la misma muestra no son comparables, ya que generalmente NanoDrop informa concentraciones hasta diez o 20 veces mayores que Qubit. Esta diferencia se discutirá en la próxima sección. Si se consideran los valores obtenidos por Qubit para las muestras 1 y 3, así como 2 y 4, podemos ver que el kit QIAamp de Qiagen proporciona una

mayor concentración de ADN en comparación con el kit Cobas. Sin embargo, al evaluar la cantidad total de ADN extraído, el kit Cobas resulta superior, logrando un mayor rendimiento en cantidad de ADN. De acuerdo con la bibliografía revisada, otros estudios también muestran que el kit Cobas permite extraer una mayor cantidad de ADN, mientras que el kit QIAamp de Qiagen mantiene una mayor calidad, con un mayor porcentaje de muestras en el rango ideal de pureza (ratio 260/280 entre 1.8 y 2.0). Esto sugiere que el kit Cobas es adecuado para aplicaciones que requieren altas cantidades de ADN, mientras que el kit QIAamp es preferible cuando se prioriza la pureza del ADN extraído (Sarnecka et al., 2019).

Considerando tanto nuestros resultados como la evidencia bibliográfica, concluimos que el kit QIAamp de Qiagen ofrece un mejor rendimiento en términos de pureza, por lo que utilizaremos este kit en futuras extracciones de ADN. Además, una vez extraído el ADN, realizamos una electroforesis en gel de agarosa al 2% para evaluar su integridad. En la *figura 7* se muestran los resultados de esta electroforesis.



***Figura 7:** Electroforesis en gel de agarosa 2%, muestras 1 -4 ADN extraído con kits comercial, como marcador de peso molecular utilizamos el Hyper Ladder™ 50pb.*

Observamos que las muestras 3 y 4 presentan una mayor intensidad en el gel de agarosa versus las muestras 1 y 2. También, vemos que el ADN aparece como una banda difusa, sin bandas definidas. Este mismo resultado ha sido reportado por otros investigadores, quienes señalan que el ADN en muestras embebidas en parafina suele estar altamente fragmentado (Einaga et al., 2017). Al someterlo a electroforesis en gel de agarosa, este tipo de ADN genera un perfil característico con un patrón difuso en lugar de bandas discretas nítidas.

6.1.2. *Cuantificación de ADN (Qubit vs NanoDrop)*

De aquí en adelante, la extracción de ADN tanto de muestras comerciales como de muestras de archivo se realizó utilizando el kit comercial QIAamp de Qiagen.. Aunque el protocolo permite un rango de volumen de elución final entre 30 y 100 μL , se optó por eluir con 100 μL con el objetivo de maximizar el rendimiento del ADN obtenido, garantizando así suficiente material para los análisis posteriores.

La cuantificación y evaluación de la pureza del ADN se llevaron a cabo mediante los equipos Qubit y NanoDrop. Los resultados obtenidos para las distintas muestras analizadas se presentan en la *Tabla 5*.

Tabla 5: Cuantificación y validación de ADN extraído con kit QIAamp de Qiagen

Nº muestra	Absorbancia 260 nm	Relación Abs 260 / 280 nm	Relación Abs 260 / 230 nm	Concentración ADN NanoDrop (ng/ μL)	Concentración ADN Qubit (ng/ μL)
1	0,612	1,71	1,57	30,6	2,54
2	1,256	1,81	1,83	62,8	6,12
3	3,193	1,90	2,16	159,7	8,98
4	2,516	1,91	2,17	125,8	5,66
5	0,793	2,14	1,84	39,7	5,2
6	0,779	1,98	1,88	38,9	3,46
7	2,02	1,85	2,22	101	10
8	0,482	1,97	2,30	24	2,74
9	0,191	1,77	1,09	9,6	2,06

10	0,504	1,99	1,10	25,2	5,4
11	0,939	1,93	1,85	46,9	21,6
12	1,394	1,84	1,84	69,7	20,6
13	0,538	1,96	1,71	26,9	2,02
14	2,164	2,01	2,20	108,2	26,0
15	2,909	2,12	0,32	46,4	3,94
16	2,869	2,31	0,37	53,3	7,26
17	4,434	2,04	1,03	228,2	61,6
18	0,153	1,71	1,55	11,9	5,76
19	0,108	2,71	0,47	2,5	1,07
20	0,331	1,83	1,78	29,5	12,5
21	0,497	1,78	1,68	41,7	27,2
22	0,165	1,76	1,14	9,4	4,47
23	0,157	1,32	0,67	5,3	5,3

Al comparar las concentraciones de ADN obtenidas mediante NanoDrop y Qubit, se observa que NanoDrop reporta valores siempre superiores a los de Qubit, en ocasiones hasta 10 o 20 veces mayores para la misma muestra. Esta discrepancia puede atribuirse a la capacidad de NanoDrop para detectar no solo ADN, sino también contaminantes como ARN, proteínas y compuestos orgánicos residuales de la parafina o del proceso de extracción, lo que conduce a una sobreestimación de la concentración de ADN. Por el contrario, Qubit utiliza sondas fluorescentes altamente específicas para ADN de doble cadena, lo que permite una cuantificación más precisa y menos susceptible a interferencias de contaminantes. La distribución de los valores de concentración obtenidos por los dos métodos de cuantificación utilizados se observa en la *figura 8*:

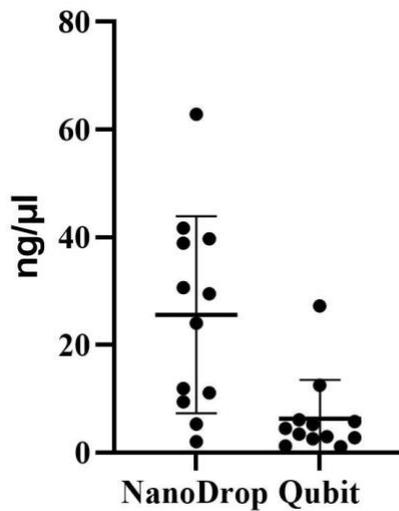


Figura 8: Comparación de valores obtenidos al cuantificar ADN por NanoDrop y Qubit

Esta diferencia en las mediciones ha sido documentada en estudios previos. Por ejemplo, en el artículo “DNA Qualification Workflow for Next Generation Sequencing of Histopathological Samples” (Simbolo et al., 2013) donde encontraron que NanoDrop tiende a sobreestimar la concentración de ADN en comparación con Qubit, especialmente en muestras de ADN degradado o contaminado. También, el estudio de Kumar y colaboradores (Kumar et al., 2016c), reportó que las concentraciones de ADN medidas por NanoDrop eran consistentemente más altas que las obtenidas por Qubit, atribuyendo esta diferencia a la detección inespecífica de contaminantes por parte de NanoDrop.

En el contexto de muestras embebidas en parafina, es común que el ADN esté fragmentado y presente contaminantes residuales del proceso de fijación y embebido. Estos factores pueden influir en las mediciones de absorbancia de NanoDrop, resultando en una sobreestimación de la concentración de ADN. Por lo tanto, para obtener una cuantificación más precisa del ADN en este tipo de muestras, es recomendable utilizar métodos como Qubit, que ofrecen mayor especificidad y sensibilidad, minimizando la interferencia de los contaminantes y proporcionando resultados más fiables para aplicaciones posteriores (Masago et al., 2021; Sarnecka et al., 2019). Por lo tanto, los valores de concentración

que consideramos de aquí en adelante para la preparación de librerías fueron los que obtuvimos por Qubit.

6.1.3. Control de Amplificación del ADN

Como control adicional para confirmar que el ADN es apto para amplificación mediante PCR durante la preparación de las librerías, realizamos dos PCR de control empleando cebadores disponibles en el laboratorio. Estas reacciones de control incluyeron amplicones de diferentes tamaños: uno de 195 pares de bases y otro de 153 pares de bases. Aunque las regiones amplificadas por estos cebadores no coinciden con las regiones de interés del estudio, es decir que no están incluidas en el panel de secuenciación, su uso resulta fundamental para evaluar la calidad del ADN en la muestra y su capacidad de amplificación.

La obtención de amplicones mediante estas PCR de prueba confirma la integridad suficiente del material obtenido y sugiere la ausencia de inhibidores o contaminantes que pudieran obstaculizar el proceso de amplificación. En la Figura 8 se presenta la electroforesis en gel de agarosa al 2% de los productos de estas reacciones, lo cual proporciona una visualización clara de los resultados obtenidos.

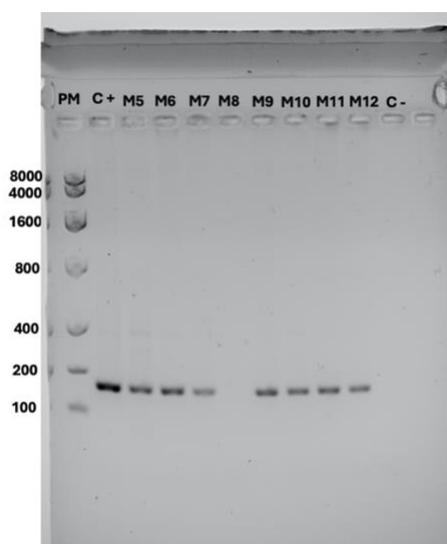


Figura 9: Electroforesis en gel de agarosa 2%, productos de PCR. Productos de PCR de 153 pb. PM: Marcador de peso molecular: KAPA Express Ladder.

Los resultados obtenidos indicaron que el ADN extraído es efectivamente apto para amplificación, ya que logramos generar amplicones tanto de 195 y de 153 pares de bases en nuestras pruebas. Teniendo en cuenta que los amplicones generados en nuestro panel de genes tienen un tamaño promedio de 175 pares de bases, mientras que los correspondientes al panel de inestabilidad de microsatélites alcanzan 86 pares de bases vemos que el ADN obtenido podría ser amplificado en nuestras regiones de interés. Este control confirma que el ADN obtenido en las muestras presenta una calidad adecuada y debería ser amplificable por los cebadores utilizados para generar las librerías de secuenciación.

6.2. Optimización de secuenciación

6.2.1. Preparación de librerías

Como mencionamos anteriormente, contamos con dos paneles distintos para la secuenciación: un panel de 17 genes y otro de inestabilidad de microsatélites (MSI). Con el fin de optimizar costos y evitar la necesidad de preparar 2 librerías y secuenciar la misma muestra dos veces (una vez con cada panel), optamos por combinar los productos de PCR de ambos paneles antes de la preparación de las librerías. Para ello, adaptamos el protocolo recomendado en la guía de Ion AmpliSeq Library Kit Plus User (TFS- AssetsLSGmanualsMAN0017003_IonAmpliSeqLibraryKitPlus_UG.pdf, s. f.), distribuyendo el ADN obtenido en las mezclas de reacción de los tres pools como si fueran de un mismo panel (ilustración figura 10). Para luego combinar los productos de PCR obtenidos en cada reacción (figura 11). Dado que el fabricante no recomienda combinar productos de PCR de diferentes paneles, implementamos una serie de pruebas para optimizar la proporción de cada producto en la mezcla, buscando una combinación que asegure una cobertura balanceada de lecturas.

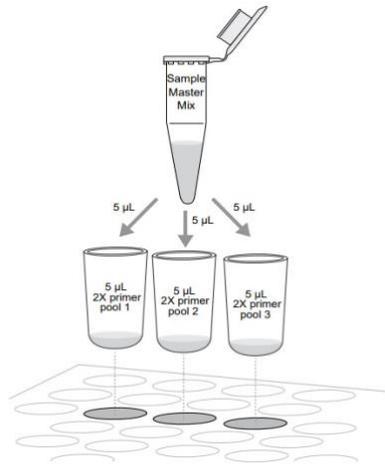


Figura 10: Ilustración de la distribución de ADN realizada en tres pools. En nuestro caso el pool 1 y 2 corresponden al panel de 17 genes y el pool 3 corresponde al panel MSI. Obtenida de Ion AmpliSeq Library Kit Plus User.

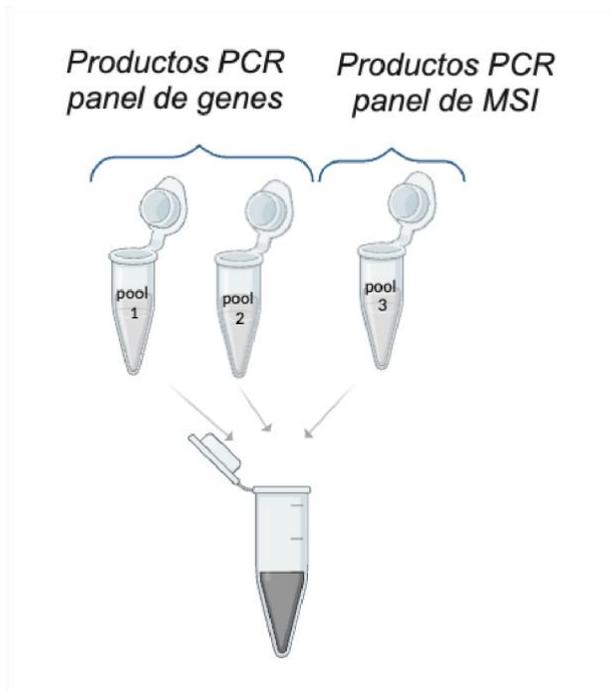


Figura 11: Ilustración de la combinación de los productos de PCR a partir de los dos paneles utilizados. Obtenida de (BioRender, s. f.)

En la primera secuenciación realizada, utilizamos una muestra de control de ADN comercial de alta calidad y probamos diferentes proporciones de mezcla

para determinar el volumen adecuado de cada producto de PCR y lograr así un equilibrio en la profundidad de lecturas entre los dos paneles.

Primera suposición de la relación de mezcla de volúmenes de producto de PCR a emplear:

Utilizando el Ion chip 530TM, que permite una carga máxima de 40 muestras y rinde 15 millones de lecturas, calculamos que cada muestra podría obtener aproximadamente 375,000 lecturas, ecuación 1.

$$\frac{15.000.000 \text{ (lecturas)}}{40 \text{ (muestras)}} = 375.000 \quad Ec \text{ (1)}$$

Para el panel de genes, tiene un tamaño de 59,45 kb y estimamos una cobertura deseada de 5000 lecturas por amplicón (de tamaño 175 pb), lo que nos dio un requerimiento de alrededor de 1.700.000 lecturas para este panel, *ecuación 2*.

$$\frac{59.450 \text{ (pb)} \times 5000 \text{ (lecturas)}}{175 \text{ (pb)}} = 1.700.000 \quad Ec \text{ (2)}$$

Por otro lado, el panel de MSI tiene un tamaño de 6,5kb y estimando una cobertura deseada de 5000 lecturas por amplicón (de tamaño 85 pb) necesitaría aproximadamente 400,000 lecturas para alcanzar la misma cobertura. Esto nos llevó a una relación ideal de 4.25:1 en la profundidad necesaria entre el panel de genes y el de MSI, *ecuación 3 y 4*.

$$\frac{6.500 \text{ (pb)} \times 5000 \text{ (lecturas)}}{85 \text{ (pb)}} = 400.000 \quad Ec \text{ (3)}$$

$$\frac{1.700.000}{400.000} = 4.25 \quad Ec \text{ (4)}$$

Dado que buscamos una mezcla equimolar de ambos paneles, ajustamos los volúmenes de los productos de PCR para aproximarnos a esta relación. Además, considerando que el 60% de las lecturas del panel MSI suelen perderse

por problemas de calidad (esta información fue brindada por el fabricante del kit utilizado), incrementamos ligeramente el volumen del producto de MSI en la mezcla final, añadiendo 10 µL del producto de PCR del panel de genes y 5 µL del panel de MSI.

Segunda suposición de la relación de mezcla de volúmenes de producto de PCR a emplear:

Para realizar una segunda suposición de mezcla equimolar entre los dos paneles, tuvimos en cuenta el tamaño de cada panel. Teniendo el panel de genes 59,45 kb y el panel de MSI 6,5 kb, la relación por tamaño únicamente de ambos paneles es de 9.1 :1, *ecuación 5*.

$$\frac{59,45 \text{ kb}}{6.5 \text{ kb}} = 9.1 \quad Ec (5)$$

Teniendo en cuenta esta relación, colocamos 10 µL proveniente de los productos de PCR del panel de genes y 1,1 µL del panel de MSI.

Otro factor evaluado fue el número de ciclos en la PCR de cada panel. El panel de genes, que amplifica cerca de 460 amplicones por pool, requiere 16 ciclos para ADN de buena calidad y 19 ciclos para ADN de baja calidad. Por otro lado, el panel de MSI genera 76 amplicones y requiere 19 ciclos para ADN de buena calidad y 22 para ADN de baja calidad.

Esta información se detalla en la *tabla 6*, obtenida del manual Ion AmpliSeq™ Library Kit Plus.

Tabla 6: Ciclos recomendados por fabricante según calidad de ADN, tabla obtenida del manual del Ion AmpliSeq™ Library Kit Plus.

Primer pairs per pool	Recommended number of amplification cycles (10 ng DNA, 3,000 copies)		Anneal/extend time
	High quality DNA	Low quality DNA (FFPE DNA or cfDNA)	
12–24	21	24	4 minutes
25–48	20	23	4 minutes
49–96	19	22	4 minutes
97–192	18	21	4 minutes
193–384	17	20	4 minutes
385–768	16	19	4 minutes
769–1,536	15	18	8 minutes
1,537–3,072	14	17	8 minutes
3,073–6,144	13	16	16 minutes
6,145–24,576	12	15	16 minutes

Dado que el ADN utilizado inicialmente es de origen comercial y, por lo tanto, de alta calidad, probamos unificar ambos paneles en 19 ciclos. Además, evaluamos la posibilidad de realizar PCRs separadas para cada panel (16 ciclos para el panel de genes y 19 para el de MSI), comparando también distintas proporciones de ADN en estas pruebas. Estos ensayos adicionales permitieron evaluar el impacto tanto de la proporción de volúmenes de PCR como del número de ciclos, ayudando a optimizar la eficiencia de amplificación y la calidad de las lecturas en ambos paneles.

Finalmente, antes de realizar esta primera secuenciación, evaluamos la cantidad de ADN necesaria para detectar las mutaciones presentes en la muestra tumoral. Probamos muestras con 30 ng de ADN y otras con 15 ng, lo cual nos permitió reducir a la mitad el uso de reactivos de librería y optimizar el consumo del kit. Nuestro objetivo principal es mantener la rentabilidad económica mediante distintas pruebas que aseguren una combinación equimolar, una cobertura adecuada y la profundidad de lectura necesaria, sin comprometer la calidad de la información obtenida.

En esta primera secuenciación, se utilizó una única muestra de ADN comercial de alta calidad (Control DNA (from Individual 1347-02) N° de catálogo #403062

(Thermo Fisher Scientific). A continuación, se describen las características de cada muestra preparada, las cuales se encuentran resumidas en la *tabla 7*.

- Muestras 1 y 3: En estas muestras, los productos de PCR se mezclaron según el número de lecturas esperadas para cada panel. Ambas muestras tienen la misma proporción de mezcla de productos de PCR, pero difieren en la cantidad total de ADN empleado. Esto nos permitió evaluar cómo la cantidad de ADN influye en la eficiencia de la secuenciación y en la profundidad obtenida.
- Muestra 2: Aquí, la mezcla de productos de PCR se realizó en base al tamaño total de cada panel. Al ajustar las proporciones considerando el tamaño en kilobases, buscamos equilibrar la contribución de cada panel en la secuenciación final y asegurar una cobertura uniforme.
- Muestra 4: Se utilizó la misma proporción de mezcla de productos de PCR que en las muestras 1 y 3. Sin embargo, en este caso aplicamos un número de ciclos diferencial durante la PCR: se realizaron 19 ciclos para el panel de genes y 16 ciclos para el panel de MSI. Esta variación nos permitió investigar el impacto del número de ciclos en la amplificación de cada panel y optimizar las condiciones para obtener una profundidad adecuada sin sobreamplificar ninguno de los productos.
-

Tabla 7: Detalles de la combinación en microlitros de los productos de PCR de los diferentes paneles, número de ciclos aplicados y cantidad de ADN utilizado en la primera secuenciación.

Muestra	Volumen de producto de PCR (μ L)			Cantidad ADN (ng)	Nº ciclos PCR
1	10	10	5	30	19 unificando ciclos
2	10	10	1,1	30	19 unificando ciclos
3	5	5	2,5	15	19 unificando ciclos
4	5	5	2,5	15	16 / 19

6.2. Evaluación y cuantificación de librerías:

Una vez completada la preparación de las librerías de secuenciación, realizamos un análisis de control de calidad utilizando electroforesis automatizada (Bioanalyzer). Este paso es fundamental para confirmar la presencia de amplicones de los tamaños esperados, asegurando que el proceso de amplificación y preparación de librerías se haya realizado correctamente. El Bioanalyzer permite visualizar y cuantificar los fragmentos de ADN, verificando que las bandas correspondan a los tamaños específicos definidos en el diseño del panel. Este control de calidad previo a la secuenciación es esencial para evitar problemas en la lectura y asegurar que las librerías están correctamente preparadas para obtener resultados confiables. En las *figuras 12 y 13* se muestran las primeras librerías realizadas como ejemplo de cómo se visualizan los resultados arrojados por el equipo 2100 Bioanalyzer Instrument.

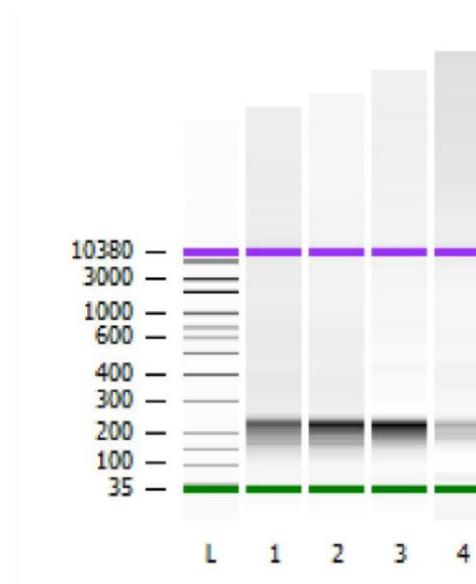


Figura 12: Resultados obtenidos de la electroforesis automatizada realizada en 2100 Bioanalyzer Instrument para la primera secuenciación.

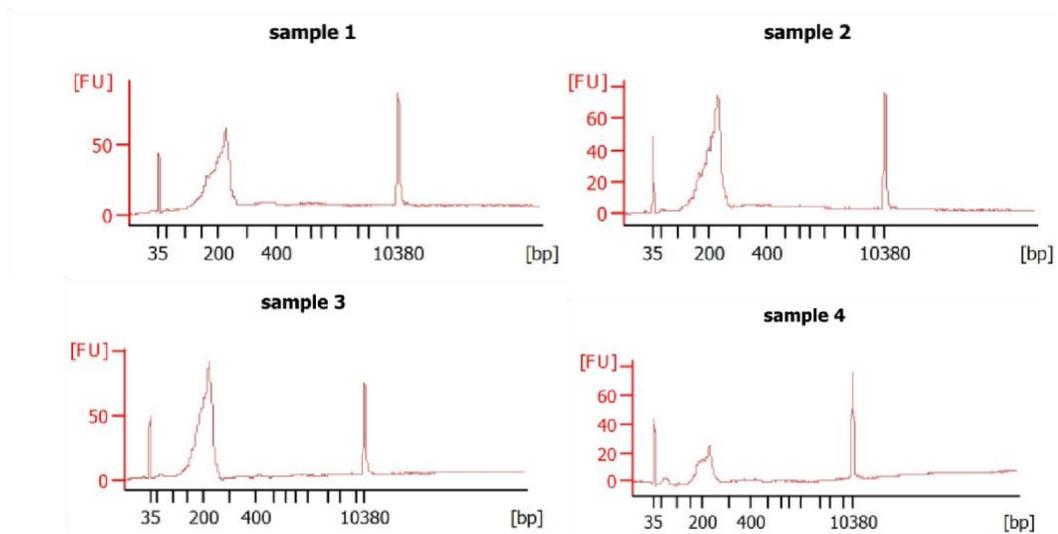


Figura 13: Distribución de tamaño de los amplicones obtenidos. En el eje de las X se muestra el tamaño en pb y en el eje de las Y la intensidad de fluorescencia en unidades de fluorescencia (FU). Para cada muestra se obtiene un pico a 35 pb y otro a 10380 pb los cuales corresponden a los marcadores de peso inferior y superior internos de cada carril. La flecha indica la distribución de amplicones obtenida cuyo rango de tamaño se encuentra entre 100 y 280 pb.

Los resultados obtenidos del análisis de control de calidad mediante el 2100 Bioanalyzer Instrument (Figuras 11 y 12) muestran una distribución de amplicones que corresponde con las expectativas para la primera secuenciación. En la Figura 11, se observa la electroforesis automatizada de las muestras 1 a 4, evidenciando fragmentos claramente definidos, lo que sugiere una correcta preparación de las librerías y la integridad de los amplicones. Las bandas en las posiciones esperadas indican que los fragmentos de ADN generados están en el rango de tamaños adecuado, asegurando que las regiones objetivo fueron amplificadas de manera efectiva.

En la *Figura 12*, se presenta la distribución detallada de los amplicones para cada muestra, donde se observan picos de ADN en el entorno de 200 pb, tamaño concordante con los diseños del panel. Este perfil es característico de amplicones generados para secuenciación en paneles de genes específicos, y la aparición de picos nítidos en todas las muestras sugiere que no hubo contaminación ni amplificación inespecífica significativa. La intensidad de los

picos es uniforme entre las muestras, lo que indica una preparación homogénea y una cantidad de ADN comparable entre ellas.

Estos resultados son consistentes con estudios previos que indican que amplicones de aproximadamente 200 pb son óptimos para secuenciación en plataformas de Ion Torrent, utilizando el chip 530. En conjunto, estos datos confirman que las librerías están correctamente preparadas y que los amplicones generados cumplen con los requisitos de tamaño y pureza, lo que asegura una secuenciación eficiente y confiable en pasos posteriores.

6.3. Resultados primera secuenciación: Optimización

Para evaluar las diferentes condiciones de las muestras secuenciadas en la primera instancia, analizamos el promedio de profundidad de lectura por amplicón. La profundidad de lectura es un indicador fundamental de la cobertura y fiabilidad del panel, ya que asegura que las regiones de interés estén suficientemente representadas para una detección precisa de variantes. Este análisis fue realizado tanto para el panel de 17 genes como para el panel de inestabilidad de microsatélites (MSI).

En la *Tabla 8* se presentan los valores de la media de profundidad por amplicón para ambos paneles, y los resultados gráficos se muestran en la *Figura 14*. En esta figura, se observa la distribución de la cobertura promedio por amplicón en cada panel y corrida, lo cual permite visualizar las variaciones en profundidad de lectura entre el panel de 17 genes y el panel de MSI en cada secuenciación. Este enfoque nos facilita identificar diferencias en la profundidad alcanzada para cada panel y optimizar las condiciones de secuenciación para asegurar una representación equilibrada y precisa de los amplicones en futuros experimentos.

Tabla 8: Análisis de secuenciación, valor de la media de profundidad por amplicón para el panel de genes y MSI:

Muestra	Media panel 17 genes	Media panel MSI
1	983	199
2	1005	45
3	1019	197
4	878	1426

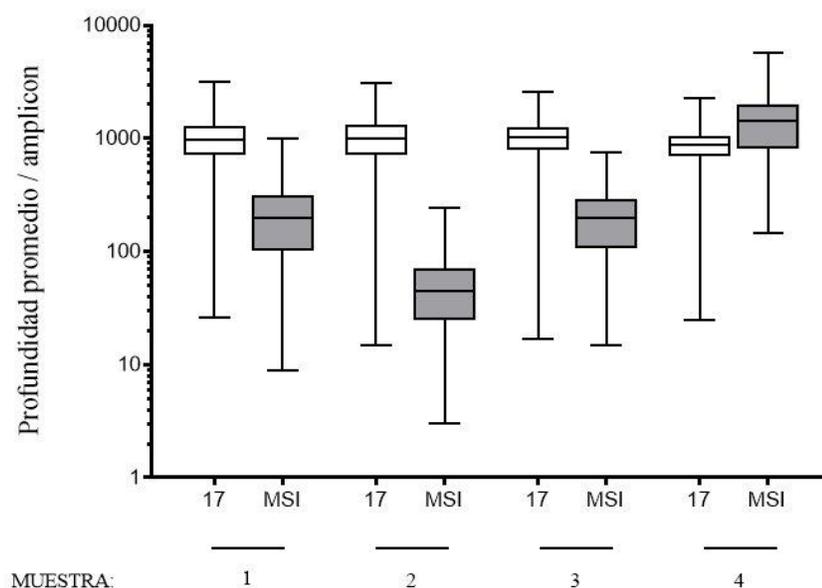


Figura 14: Evaluación de profundidad por amplicón de las muestras 1, 2, 3 y 4. 17: refiere al panel de 17 genes MSI: refiere al panel de inestabilidad de microsatélites

Al observar la profundidad promedio obtenida para el panel de 17 genes, notamos que las diferentes muestras presentan una profundidad relativamente similar. La muestra 3, en particular, muestra una mayor profundidad de lectura; esta muestra fue preparada con una relación de volumen de producto de PCR del doble para el panel genes y un mismo número de ciclos de PCR para los dos paneles. En contraste, la corrida 4 fue la que presentó la menor profundidad en el panel de genes.

En cuanto a la cobertura para el panel de MSI, la corrida 4 se destaca con una profundidad significativamente mayor en comparación con las otras corridas.

Esto podría atribuirse al uso de un número de ciclos diferenciales en la PCR de esta muestra, donde se realizaron 19 ciclos para el panel de MSI y 16 ciclos para el panel de genes. Esta diferencia en el número de ciclos podría estar sugiriendo que un mayor número de ciclos para el panel de genes resulta en una mayor captación de lecturas en la secuenciación, reduciendo así las lecturas para el panel de MSI.

Otro aspecto relevante es que el uso de 15 ng de ADN en las muestras 3 y 4 no parece afectar negativamente la profundidad ni la cobertura de los amplicones, lo que indica que esta cantidad es suficiente para obtener buenos resultados en ambos paneles.

En conclusión, los resultados de esta primera secuenciación sugieren varias observaciones importantes: es posible trabajar con 15 ng de ADN y mantener una buena profundidad de lectura; el uso de ciclos de PCR separados para cada panel mejora la proporción de lecturas entre MSI y genes; y la proporción de mezcla 5:5:2.5 resulta efectiva. Estos hallazgos ofrecen una base sólida para optimizar futuras secuenciaciones manteniendo una cobertura balanceada y rentable para ambos paneles.

6.4. Desempeño del panel en muestras tumorales

Tras las primeras conclusiones obtenidas en la secuenciación inicial, realizamos una segunda secuenciación que incluyó tanto muestras de pacientes reales como ADN comercial con mutaciones y frecuencias alélicas conocidas. Para esta fase, utilizamos una proporción de mezcla de productos de PCR de 5:5:2.5 y aplicamos un ciclado diferencial para cada pool de cebadores, siguiendo un diseño optimizado para equilibrar la cobertura entre los paneles de genes y de inestabilidad de microsatélites (MSI).

Como parte de la evaluación de esta secuenciación, analizamos la profundidad de lectura obtenida para cada amplicón, es decir, el número de veces que cada región del ADN fue leída durante la secuenciación. Los resultados de esta

evaluación se presentan en la *figura 15*, donde se observa la media de la profundidad de lectura para cada amplicón.

En primer lugar, vemos la distribución de la profundidad de los amplicones (*figura 15 - A*) donde podemos observar que la mayoría de ellos alcanzó una profundidad suficiente, con una variación moderada entre los diferentes amplicones. Esta uniformidad es crucial para asegurar una representación homogénea en el panel de genes, indicando que la proporción de mezcla utilizada resultó adecuada para lograr una cobertura equilibrada.

En el segundo gráfico, correspondiente al panel de MSI, observamos la profundidad de lectura para cada amplicón en ese panel. La mayoría de los amplicones se encuentran en un rango de profundidad aceptable, aunque algunos presentan variaciones más amplias, lo que podría estar relacionado con la complejidad de las regiones repetitivas evaluadas en el panel de MSI.

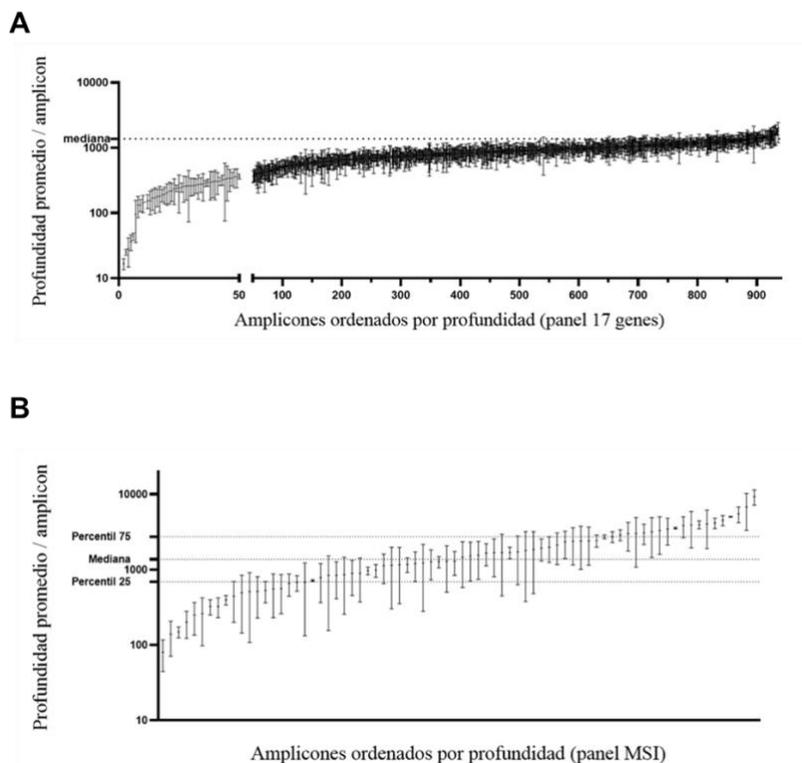


Figura 15: Distribución de la profundidad de lectura por amplicón en la segunda secuenciación. A: Distribución de profundidad para amplicones del panel de 17 genes. B: Distribución de profundidad para amplicones del panel de MSI.

Estos resultados sugieren que la combinación de productos de PCR en la proporción 5:5:2.5, junto con el ciclado diferencial, fue eficaz para obtener una cobertura adecuada en ambos paneles. La media de profundidad alcanzada en cada amplicón demuestra que esta metodología permite una secuenciación confiable tanto para genes específicos como para regiones de inestabilidad de microsatélites, lo cual es fundamental para el diagnóstico molecular en el contexto clínico de cáncer.

6.5. Identificación de variantes y comparación utilizando muestras comerciales

Luego del procesamiento de los datos, obtenemos el número de lecturas en las cuales se detectó la mutación y la cantidad de lecturas totales para ese amplicón. El resultado de los análisis también nos arroja la frecuencia alélica de la mutación, la ubicación genómica entre otros datos.

A modo de control en primera instancia, analizamos los datos obtenidos para las muestras comerciales y comparamos las variantes detectadas en nuestro panel de genes y MSI, con las mutaciones teóricas proporcionadas por los fabricantes de las muestras de ADN comercial.

La *tabla 9* presenta los resultados de esta comparación para una de las muestras de ADN comercial (#HD832, Horizon Discovery), destacando tanto la frecuencia alélica esperada como la frecuencia alélica obtenida en nuestro análisis.

Tabla 9: Comparación de variantes detectadas en el panel de secuenciación con frecuencias alélicas esperadas en muestras de ADN comercial (#HD832)

Cromosoma	Gen	Variante	Frecuencia alélica esperada	Frecuencia alélica obtenida
7	BRAF	p.V600E	10,7%	10,05%
7	EGFR	p.G719S	24,5%	23,75%
7	EGFR	p.L858R	2,80%	4,05%
7	EGFR	p.Q787Q	15,0%	11,50%

12	KRAS	p.G13D	15,0%	13,07%
12	KRAS	p.G12D	6,30%	5,84%
7	MET	p.A1357A	7,0%	6,14%
7	MET	p.L238fs*25	7,0%	7,25%
1	NRAS	p.Q61K	12,5%	12,10%
3	PIK3CA	p.E545K	8,80%	7,30%
3	PIK3CA	p.H1047R	17,5%	17,58%

La detección de variantes en genes relevantes, como BRAF, EGFR, KRAS, MET, NRAS y PIK3CA, confirma la precisión de nuestro panel de secuenciación para identificar mutaciones en genes de interés clínico. Observamos que la frecuencia alélica obtenida en la mayoría de los casos es similar a la frecuencia teórica esperada, con algunas variaciones menores. Por ejemplo, en el caso de la variante p.G719S del gen EGFR, la frecuencia alélica esperada era de 24.5%, mientras que obtuvimos un valor muy cercano, de 23.75%. Estas coincidencias indican que el panel es adecuado para detectar variantes de baja frecuencia con un alto grado de precisión, lo cual es fundamental para estudios en los que pequeñas diferencias en la frecuencia alélica pueden tener implicaciones clínicas significativas.

La *tabla 10* presenta los resultados de esta comparación para otra de las muestras de ADN comercial (#HD810, Horizon Discovery), destacando tanto la frecuencia alélica esperada como la frecuencia alélica obtenida en nuestro análisis.

Tabla 10: Comparación de variantes detectadas en el panel de secuenciación con frecuencias alélicas esperadas en muestras de ADN comercial (#HD810)

Cromosoma	Gen	Variante	Frecuencia alélica esperada (%)	Frecuencia alélica obtenida (%)
17	BR CA1	S1634G**	7,5	8,4
17	BR CA1	K1183R	7,5	9
17	BR CA1	K820E	7,5	8,9
17	BR CA1	R1443STOP	32,5	34,3
17	BR CA1	D435Y	7,5	8,3
17	BR CA1	P871L	15	17,6
13	BR CA2	N289H	7,5	13,6
13	BR CA2	V2466A	100	99,5
13	BR CA2	N991D	7,5	9,3
13	BR CA2	K1691fs	32,5	32,8
13	BR CA2	N1784fs	40	34,6
13	BR CA2	D1420Y	32,5	31
13	BR CA2	I2675fs	10	16,1
17	BRI P1	S919P	100	99,6
2	BA RD1	R378S	42,5	46,7

Para ambas muestras, se observan pequeñas desviaciones en algunas variantes. Estas diferencias pueden atribuirse a factores técnicos, como la eficiencia de la amplificación o el número de ciclos de PCR aplicados, o a la calidad del ADN en las muestras comerciales. A pesar de estas variaciones, los resultados generales reflejan una alta concordancia entre las frecuencias alélicas esperadas y las obtenidas, lo que valida el uso de nuestro panel de secuenciación para la identificación precisa de mutaciones en muestras clínicas.

6.5.1. Aproximación a la detección del límite inferior de frecuencia alélica con este diseño experimental

Con el objetivo de determinar el límite inferior de frecuencia alélica detectable con este panel, es que ensayamos una muestra comercial secuenciando la misma librería en dos secuenciaciones diferentes con profundidades diferentes. Para ello, utilizamos la muestra de ADN comercial (#HD125, Horizon Discovery), la cual posee 5% de frecuencia alélica del cambio C > T en la posición 55249071 del cromosoma 7 (hg19), lo que produce una mutación en el gen EGFR con cambio en la secuencia aminoacídica (p.T790M). Los resultados obtenidos se muestran en la *tabla 11*.

Tabla 11: Resultados obtenidos de secuenciar la misma librería de muestra con diferente profundidad

Lecturas totales	Largo promedio de las lecturas (pb)	Lecturas de la mutación	Lecturas totales para el sitio	Frecuencia Alélica experimental
2498632	111	158	2960	5,3%
597878	112	25	656	3,8%

Observando los resultados obtenidos en la *tabla 11*, vemos que al dar una profundidad mayor a la secuenciación obtenemos una frecuencia alélica experimental de 5,3% la cual se aproxima al valor esperado. Por otra parte, cuando la misma librería es secuenciada asignándole una profundidad de 4 veces menor aún es detectable la mutación presente en este tejido.

La secuenciación a una profundidad cercana a 500K de lecturas es un escenario similar al esperado en un contexto de desarrollo del test teniendo en cuenta el costo por muestra. Observamos que al bajar la profundidad asignada a la muestra obtenemos una frecuencia alélica experimental menor a la esperada. Esta desviación es esperada ya que al asignar un número menor de lectura se pierda precisión en la frecuencia alélica esperada. Igualmente, destacamos la detección de alteraciones en un tejido con muy baja frecuencia mutacional.

Estos resultados respaldan la capacidad de nuestro panel para ser utilizado en el diagnóstico molecular, proporcionando una herramienta confiable para la detección de mutaciones en genes clave involucrados en la oncogénesis. La buena correlación entre las frecuencias esperadas y obtenidas sugiere que el sistema es adecuado para aplicaciones clínicas y que puede detectar variantes con una frecuencia alélica cercana a los valores teóricos proporcionados por los fabricantes

6.6 Desempeño muestras con diferente antigüedad

En esta fase del análisis, evaluamos el desempeño del panel de secuenciación en muestras tumorales reales de pacientes de diferente antigüedad de extracción del tumor. Los resultados de la secuenciación obtenidos para estas muestras tumorales y comerciales se presentan en la *tabla 12*.

Tabla 12: Resultados de secuenciación: muestras analizadas de diferente antigüedad de extracción

<i>Muestra</i>	<i>Descripción</i>	<i>Antigüedad / origen de la muestra</i>	<i>Largo promedio de las lecturas (pb)</i>
<i>ADN control CEPH</i>	<i>ADN genómico comercial</i>	<i>- / en solución</i>	<i>114</i>
<i>2</i>	<i>ADN tumoral de paciente</i>	<i>> 1 año / FFPE</i>	<i>89</i>
<i>3</i>	<i>Muestra real de paciente</i>	<i>> 1 año / FFPE</i>	<i>76</i>
<i>5</i>	<i>ADN comercial #HD832</i>	<i>- / FFPE</i>	<i>113</i>
<i>6</i>	<i>ADN tumoral de paciente</i>	<i>> 1 año / FFPE</i>	<i>88</i>
<i>7</i>	<i>ADN tumoral de paciente</i>	<i>> 1 año / FFPE</i>	<i>78</i>
<i>8</i>	<i>ADN comercial #HD125</i>	<i>- / FFPE</i>	<i>111</i>

10	<i>ADN tumoral de paciente</i>	<i>< 1 año / FFPE</i>	<i>114</i>
11	<i>ADN tumoral de paciente</i>	<i>< 1 año / FFPE</i>	<i>115</i>
12	<i>ADN tumoral de paciente</i>	<i>< 1 año / FFPE</i>	<i>112</i>
13	<i>ADN tumoral de paciente</i>	<i>< 1 año / FFPE</i>	<i>104</i>
14	<i>ADN tumoral de paciente</i>	<i>< 1 año / FFPE</i>	<i>108</i>
15	<i>ADN tumoral de paciente</i>	<i>< 1 año / FFPE</i>	<i>108</i>
16	<i>ADN tumoral de paciente</i>	<i>> 1 año / FFPE</i>	<i>79</i>
17	<i>ADN tumoral de paciente</i>	<i>< 1 año / FFPE</i>	<i>106</i>
18	<i>ADN tumoral de paciente</i>	<i>> 1 año / FFPE</i>	<i>96</i>
20	<i>ADN tumoral de paciente</i>	<i>> 1 año / FFPE</i>	<i>83</i>
21	<i>ADN tumoral de paciente</i>	<i>> 1 año / FFPE</i>	<i>71</i>
22	<i>ADN tumoral de paciente</i>	<i>< 1 año / FFPE</i>	<i>111</i>

En primer lugar, observamos que el tamaño promedio de lecturas es menor en muestras tumorales de mayor antigüedad de extracción (> 1 año de extracción de ADN). También, vemos que las muestras que fueron extraídas hace menos de un año, presentan un largo promedio similar al que tienen las muestras control de ADN comercial. Las muestras extraídas recientemente tienen un largo promedio de lecturas de 109,7 pb y las muestras extraídas hace más de un año tienen un largo promedio de 82,5 pb. Estos resultados muestran que hay una diferencia

significativa en el largo de lectura obtenido entre muestras antiguas y recientes, con un p-valor de 0.0003 por el test de T de Mann Whitney.

Este resultado puede atribuirse a la fragmentación del ADN, que es significativamente más pronunciada en muestras FFPE de mayor antigüedad. Esto podría explicarse por el deterioro y la mayor fragmentación del ADN debido a los efectos de la fijación con formalina y el tiempo de almacenamiento prolongado. La formalina genera entrecruzamientos y daños químicos en las moléculas de ADN, lo que con el tiempo reduce la integridad del material genético. Este proceso impacta negativamente en la calidad de los fragmentos obtenidos, resultando en lecturas más cortas en las muestras más antiguas.

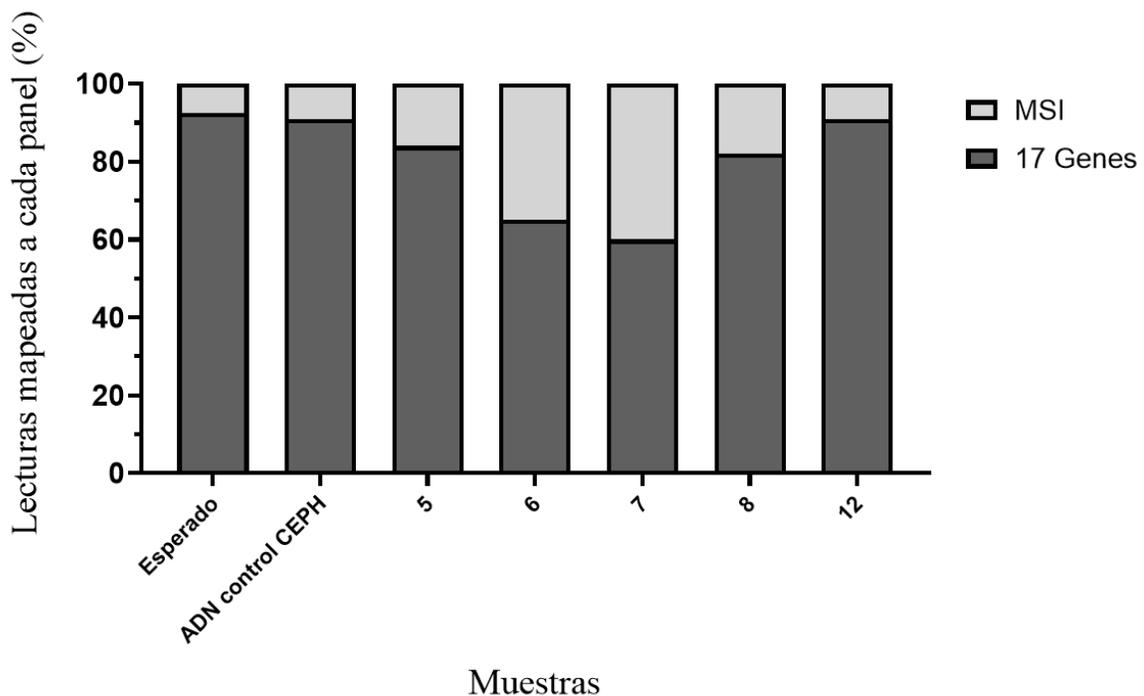


Figura 16: Porcentaje de lecturas mapeadas en los paneles de 17 genes y de inestabilidad de microsatélites (MSI) en muestras de diferente antigüedad. Esperado: Distribución esperada de las lecturas según la cantidad de amplicones de cada panel (926 para el panel de genes y 76 para el panel de MSI)

Observamos que las muestras más antiguas presentan un mayor porcentaje de lecturas mapeadas en el panel de MSI en comparación con el panel de 17 genes. Dado que el panel de MSI se compone de amplicones más cortos

(promedio de 86 pb) en comparación con el panel de 17 genes (promedio de 175 pb), el ADN fragmentado en muestras antiguas es más compatible con el tamaño de los amplicones de MSI, facilitando su amplificación y mapeo.

Este hallazgo sugiere que la antigüedad y el estado de fragmentación del ADN pueden influir en la eficiencia de mapeo en función del tamaño de los amplicones en cada panel. Específicamente, las muestras antiguas, al contener fragmentos de ADN más cortos, podrían favorecer el rendimiento del panel de MSI.

Estas observaciones son importantes para la interpretación de resultados en estudios que involucren muestras de ADN almacenadas por periodos prolongados, ya que indica que la eficiencia de detección puede variar según el panel utilizado y la calidad del ADN en función de su antigüedad.

7. Conclusiones y perspectivas

En esta investigación, logramos establecer una proporción óptima entre los dos paneles evaluados, permitiendo realizar una única secuenciación por muestra que brinda una cobertura y profundidad adecuadas para cada tipo de análisis. Asimismo, nuestro panel corroboró las variantes génicas presentes en el ADN de referencia, mostrando frecuencias comparables a las reportadas por sus fabricantes. Esto valida el correcto desempeño de nuestro panel en condiciones experimentales controladas.

Por otra parte, fue posible utilizar este test para evaluar muestras de tumor (FFPE) de pacientes obteniendo coberturas y profundidades similares a las obtenidas en las muestras comerciales.

Para futuras investigaciones, es fundamental ampliar la validación del panel utilizando un mayor número de muestras tumorales para confirmar su aplicabilidad en un contexto clínico. Además, se recomienda determinar el umbral mínimo de frecuencia alélica detectable en función de las lecturas totales obtenidas, lo cual optimizará la precisión en la identificación de variantes de baja frecuencia. Será también necesario llevar a cabo un análisis exhaustivo de las regiones con baja cobertura para descartar la presencia de mutaciones relevantes en estas áreas y mejorar la interpretación de los resultados. Finalmente, se plantea realizar el análisis de inestabilidad de microsatélites tanto con Ion Reporter como con diferentes herramientas de uso libre y evaluar el desempeño de ambas estrategias para determinar el estado de inestabilidad para cada muestra.

8. Aspectos Éticos

Este proyecto ha sido aprobado por el CECISH (Comité Ético y Científico de Investigación en Seres Humanos) de la Dirección Nacional de Sanidad de las Fuerzas Armadas en febrero de 2021.

9. Financiación

Esta tesina se desarrolló en el contexto del Proyecto de investigación aplicada financiado por el Fondo María Viñas, Agencia Nacional de Investigación e Innovación (FMV_3_2020_1_162775), responsable científica Cecilia Mathó titulado “Desarrollo de un test para la identificación integrada de múltiples tipos de alteraciones genómicas accionables en pan-cáncer mediante NGS.”

10. Bibliografía

- Ashley, E. A. (2016). Towards precision medicine. *Nature Reviews. Genetics*, 17(9), 507-522.
<https://doi.org/10.1038/nrg.2016.86> BioRender. (s. f.). Recuperado 28 de octubre de 2024, de <https://app.biorender.com/illustrations/6708662775799373e9451f15>
- Boland, C. R., & Goel, A. (2010). Microsatellite instability in colorectal cancer. *Gastroenterology*, 138(6), 2073-2087.e3. <https://doi.org/10.1053/j.gastro.2009.12.064>
- Bonnet, E., Moutet, M.-L., Baulard, C., Bacq-Daian, D., Sandron, F., Mesrob, L., Fin, B., Delépine, M., Palomares, M.-A., Jubin, C., Blanché, H., Meyer, V., Boland, A., Olasso, R., & Deleuze, J.-F. (2018). Performance comparison of three DNA extraction kits on human whole-exome data from formalin-fixed paraffin-embedded normal and tumor samples. *PLoS One*, 13(4), e0195471. <https://doi.org/10.1371/journal.pone.0195471>
- Boyiadzis, M. M., Kirkwood, J. M., Marshall, J. L., Pritchard, C. C., Azad, N. S., & Gulley, J. L. (2018). Significance and implications of FDA approval of pembrolizumab for biomarker-defined disease. *Journal for Immunotherapy of Cancer*, 6(1), 35. <https://doi.org/10.1186/s40425-018-0342-x>
- Calabria, I., Pedrola, L., Berlanga, P., Aparisi, M. J., Sánchez-Izquierdo, D., Cañete, A., Cervera, J., Millán, J. M., & Castel, V. (2016). El nuevo reto en oncología: La secuenciación NGS y su aplicación a la medicina de precisión. *Anales de Pediatría*, 85(5), 273.e1-273.e7. <https://doi.org/10.1016/j.anpedi.2016.05.006>
- Cáncer, C. H. de L. contra el. (s. f.). *Comisión Honoraria de Lucha contra el Cáncer*. Comisión Honoraria de Lucha contra el Cáncer. Recuperado 24 de octubre de 2024, de <https://www.comisioncancer.org.uy/home>
- Cobas DNA protocol. (s. f.).
- Colomer, R., Miranda, J., Romero-Laorden, N., Hornedo, J., González-Cortijo, L., Mouron, S., Bueno, M. J., Mondéjar, R., & Quintela-Fandino, M. (2023). Usefulness and real-world outcomes of next generation sequencing testing in patients with cancer: An observational study on the impact of selection based on clinical judgement. *eClinicalMedicine*, 60. <https://doi.org/10.1016/j.eclinm.2023.102029>
- COSMIC | *Catalogue of Somatic Mutations in Cancer*. (s. f.). Recuperado 2 de noviembre de 2024, de <https://cancer.sanger.ac.uk/cosmic>
- Dankner, M., Rose, A. A. N., Rajkumar, S., Siegel, P. M., & Watson, I. R. (2018). Classifying BRAF alterations in cancer: New rational therapeutic strategies for actionable mutations. *Oncogene*, 37(24), 3183-3199. <https://doi.org/10.1038/s41388-018-0171-x>
- Davies, H., Bignell, G. R., Cox, C., Stephens, P., Edkins, S., Clegg, S., Teague, J., Woffendin, H.,

- Garnett, M. J., Bottomley, W., Davis, N., Dicks, E., Ewing, R., Floyd, Y., Gray, K., Hall, S., Hawes, R., Hughes, J., Kosmidou, V., ... Futreal, P. A. (2002). Mutations of the BRAF gene in human cancer. *Nature*, 417(6892), 949-954. <https://doi.org/10.1038/nature00766>
- Einaga, N., Yoshida, A., Noda, H., Suemitsu, M., Nakayama, Y., Sakurada, A., Kawaji, Y., Yamaguchi, H., Sasaki, Y., Tokino, T., & Esumi, M. (2017). Assessment of the quality of DNA from various formalin-fixed paraffin-embedded (FFPE) tissues and the use of this DNA for nextgeneration sequencing (NGS) with no artifactual mutation. *PLoS One*, 12(5), e0176280. <https://doi.org/10.1371/journal.pone.0176280>
- GeneCards—Human Genes | Gene Database | Gene Search*. (s. f.). Recuperado 3 de noviembre de 2024, de <https://www.genecards.org/>
- Goh, A. M., Coffill, C. R., & Lane, D. P. (2011). The role of mutant p53 in human cancer. *The Journal of Pathology*, 223(2), 116-126. <https://doi.org/10.1002/path.2784>
- Hanahan, D. (2022). Hallmarks of Cancer: New Dimensions. *Cancer Discovery*, 12(1), 31-46. <https://doi.org/10.1158/2159-8290.CD-21-1059>
- 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>
- Hause, R. J., Pritchard, C. C., Shendure, J., & Salipante, S. J. (2016). Classification and characterization of microsatellite instability across 18 cancer types. *Nature Medicine*, 22(11), 1342-1350. <https://doi.org/10.1038/nm.4191>
- HB-0353-004_HB_QA_DNA_FFPE_0220_WW*. (s. f.).
- Ion 530™ Chip Kit*. (s. f.). Recuperado 3 de noviembre de 2024, de <https://www.thermofisher.com/order/catalog/product/A27764>
- Ion AmpliSeq™ Library Kit Plus*. (s. f.). Recuperado 31 de octubre de 2024, de <https://www.thermofisher.com/order/catalog/product/4488990>
- Ion GeneStudio System Features—UY*. (s. f.). Recuperado 3 de noviembre de 2024, de <https://www.thermofisher.com/ht/en/home/life-science/sequencing/next-generation-sequencing/instruments/ion-genestudio-system/features.html>
- Joyce, C., Rayi, A., & Kasi, A. (2024). Tumor-Suppressor Genes. En *StatPearls*. StatPearls Publishing. <http://www.ncbi.nlm.nih.gov/books/NBK532243/>
- Kandoth, C., McLellan, M. D., Vandin, F., Ye, K., Niu, B., Lu, C., Xie, M., Zhang, Q., McMichael, J. F., Wyczalkowski, M. A., Leiserson, M. D. M., Miller, C. A., Welch, J. S., Walter, M. J., Wendl, M. C., Ley, T. J., Wilson, R. K., Raphael, B. J., & Ding, L. (2013). Mutational landscape and significance across 12 major cancer types. *Nature*, 502(7471), 333-339. <https://doi.org/10.1038/nature12634>
- Kumar, D., Panigrahi, M. K., Suryavanshi, M., Mehta, A., & Saikia, K. K. (2016a). Quantification of DNA

Extracted from Formalin Fixed Paraffin-Embedded Tissue Comparison of Three Techniques:

Effect on PCR Efficiency. *Journal of Clinical and Diagnostic Research: JCDR*, 10(9), BC01. <https://doi.org/10.7860/JCDR/2016/19383.8407>

Kumar, D., Panigrahi, M. K., Suryavanshi, M., Mehta, A., & Saikia, K. K. (2016b). Quantification of DNA

Extracted from Formalin Fixed Paraffin-Embedded Tissue Comparison of Three Techniques:

Effect on PCR Efficiency. *Journal of Clinical and Diagnostic Research: JCDR*, 10(9), BC01-

BC03. <https://doi.org/10.7860/JCDR/2016/19383.8407>

Kumar, D., Panigrahi, M. K., Suryavanshi, M., Mehta, A., & Saikia, K. K. (2016c). Quantification of DNA

Extracted from Formalin Fixed Paraffin-Embedded Tissue Comparison of Three Techniques:

Effect on PCR Efficiency. *Journal of Clinical and Diagnostic Research: JCDR*, 10(9), BC01-

BC03. <https://doi.org/10.7860/JCDR/2016/19383.8407>

Le, D. T., Durham, J. N., Smith, K. N., Wang, H., Bartlett, B. R., Aulakh, L. K., Lu, S., Kemberling, H., Wilt, C., Luber, B. S., Wong, F., Azad, N. S., Rucki, A. A., Laheru, D., Donehower, R., Zaheer, A., Fisher, G. A., Crocenzi, T. S., Lee, J. J., ... Diaz, L. A. (2017). Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. *Science (New York, N.Y.)*, 357(6349), 409-413. <https://doi.org/10.1126/science.aan6733>

Lito, P., Solomon, M., Li, L.-S., Hansen, R., & Rosen, N. (2016). Allele-specific inhibitors inactivate mutant KRAS G12C by a trapping mechanism. *Science (New York, N.Y.)*, 351(6273), 604-608. <https://doi.org/10.1126/science.aad6204>

Loong, H. H., Kwan, S.-C. S., Mok, T. S.-K., & Lau, Y.-M. (2018). Therapeutic Strategies in EGFR Mutant Non-Small Cell Lung Cancer. *Current Treatment Options in Oncology*, 19(11), 58. <https://doi.org/10.1007/s11864-018-0570-9>

Malkin, D. (2011). Li-fraumeni syndrome. *Genes & Cancer*, 2(4), 475-484.

<https://doi.org/10.1177/1947601911413466>

Mao, C., Qiu, L.-X., Liao, R.-Y., Du, F.-B., Ding, H., Yang, W.-C., Li, J., & Chen, Q. (2010). KRAS mutations and resistance to EGFR-TKIs treatment in patients with non-small cell lung cancer: A meta-analysis of 22 studies. *Lung Cancer*, 69(3), 272-278. <https://doi.org/10.1016/j.lungcan.2009.11.020>

Mardis, E. R. (2013). Next-generation sequencing platforms. *Annual Review of Analytical Chemistry*

(Palo Alto, Calif.), 6, 287-303. <https://doi.org/10.1146/annurev-anchem-062012-092628>

Masago, K., Fujita, S., Oya, Y., Takahashi, Y., Matsushita, H., Sasaki, E., & Kuroda, H. (2021).

- Comparison between Fluorimetry (Qubit) and Spectrophotometry (NanoDrop) in the Quantification of DNA and RNA Extracted from Frozen and FFPE Tissues from Lung Cancer Patients: A Real-World Use of Genomic Tests. *Medicina (Kaunas, Lithuania)*, 57(12), 1375. <https://doi.org/10.3390/medicina57121375>
- Murtaza, M., Dawson, S.-J., Tsui, D. W. Y., Gale, D., Forshew, T., Piskorz, A. M., Parkinson, C., Chin, S.-F., Kingsbury, Z., Wong, A. S. C., Marass, F., Humphray, S., Hadfield, J., Bentley, D., Chin, T. M., Brenton, J. D., Caldas, C., & Rosenfeld, N. (2013). Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA. *Nature*, 497(7447), 108-112. <https://doi.org/10.1038/nature12065>
- Nuevas normativas de cobertura. (s. f.). *Fondo Nacional de Recursos*. Recuperado 3 de noviembre de 2024, de https://www.fnr.gub.uy/noticia_slide/nuevas-normativas-de-cobertura/
- Olivier, M., Hollstein, M., & Hainaut, P. (2010). TP53 mutations in human cancers: Origins, consequences, and clinical use. *Cold Spring Harbor Perspectives in Biology*, 2(1), a001008. <https://doi.org/10.1101/cshperspect.a001008>
- Oncogén*. (s. f.-a). Recuperado 28 de octubre de 2024, de <https://www.genome.gov/es/geneticsglossary/Oncogen>
- Oncogén*. (s. f.-b). Recuperado 28 de octubre de 2024, de <https://www.genome.gov/es/genetics-glossary/Oncogen>
- Pao, W., & Chmielecki, J. (2010). Rational, biologically based treatment of EGFR-mutant non-smallcell lung cancer. *Nature Reviews Cancer*, 10(11), 760-774. <https://doi.org/10.1038/nrc2947>
- Pembrolizumab—Instituto Nacional del Cáncer* (nciglobal,ncienterprise). (2014, septiembre 18). [pdqDrugInfoSummary]. <https://www.cancer.gov/about-cancer/treatment/drugs/pembrolizumab>
- Personalized medicine: Time for one-person trials—PubMed*. (s. f.). Recuperado 2 de noviembre de 2024, de <https://pubmed.ncbi.nlm.nih.gov/25925459/>
- ¿Qué es el cáncer y cómo se desarrolla? - SEOM: Sociedad Española de Oncología Médica © 2019*. (s. f.). Recuperado 24 de octubre de 2024, de <https://seom.org/informacion-sobre-el-cancer/que-es-el-cancer-y-como-se-desarrolla?start=2>
- ¿Qué es el Fondo Nacional de Recursos?* (s. f.). *Fondo Nacional de Recursos*. Recuperado 26 de octubre de 2024, de <https://www.fnr.gub.uy/que-es-el-fondo-nacional-de-recursos/>
- Rosado-Jiménez, L., Mestre-Terkemani, Y., García-Aliaga, Á., Marín-Vera, M., Macías-Cerrolaza, J. A., Sarabia-Meseguer, M. D., García-Hernández, M. R., Zafra-Poves, M., Sánchez-Henarejos, P., Peña, F. A. de la, Alonso-Romero, J. L., Noguera-Velasco, J. A., & Ruiz-Espejo, F. (2023). Variantes genéticas recurrentes y priorización de variantes de

significado clínico desconocido asociadas al síndrome de cáncer de mama y ovario hereditario en familias de la Región de Murcia. *Advances in Laboratory Medicine*, 4(3), 288.

<https://doi.org/10.1515/almed-2023-0032>

Roy, P. S., & Saikia, B. J. (2016). Cancer and cure: A critical analysis. *Indian Journal of Cancer*, 53(3),

441-442. <https://doi.org/10.4103/0019-509X.200658>

Sarnecka, A. K., Nawrat, D., Piwowar, M., Ligęza, J., Swadźba, J., & Wójcik, P. (2019). DNA extraction from FFPE tissue samples—A comparison of three procedures. *Contemporary Oncology (Poznan, Poland)*, 23(1), 52-58. <https://doi.org/10.5114/wo.2019.83875>

Schuh, A., Dreau, H., Knight, S. J. L., Ridout, K., Mizani, T., Vavoulis, D., Colling, R., Antoniou, P.,

Kvikstad, E. M., Pentony, M. M., Hamblin, A., Protheroe, A., Parton, M., Shah, K. A., Orosz, Z., Athanasou, N., Hassan, B., Flanagan, A. M., Ahmed, A., ... Taylor, J. C. (2018). Clinically actionable mutation profiles in patients with cancer identified by whole-genome sequencing. *Cold Spring Harbor Molecular Case Studies*, 4(2), a002279. <https://doi.org/10.1101/mcs.a002279>

Sherr, C. J. (2004). Principles of tumor suppression. *Cell*, 116(2), 235-246.

[https://doi.org/10.1016/s0092-8674\(03\)01075-4](https://doi.org/10.1016/s0092-8674(03)01075-4)

Simbolo, M., Gottardi, M., Corbo, V., Fassan, M., Mafficini, A., Malpeli, G., Lawlor, R. T., & Scarpa, A. (2013). DNA qualification workflow for next generation sequencing of histopathological samples. *PLoS One*, 8(6), e62692. <https://doi.org/10.1371/journal.pone.0062692>

SITUACION EPIDEMIOLOGICA DEL URUGUAY EN RELACION AL CANCER. (s. f.). Recuperado 21 de octubre de 2024, de <https://www.comisioncancer.org.uy/Ocultas/SITUACION-EPIDEMIOLOGICA-DEL-URUGUAY-EN-RELACION-AL-CANCER-uc311>

Soria, J.-C., Ohe, Y., Vansteenkiste, J., Reungwetwattana, T., Chewaskulyong, B., Lee, K. H.,

Dechaphunkul, A., Imamura, F., Nogami, N., Kurata, T., Okamoto, I., Zhou, C., Cho, B. C.,

Cheng, Y., Cho, E. K., Voon, P. J., Planchard, D., Su, W.-C., Gray, J. E., ... FLAURA Investigators. (2018). Osimertinib in Untreated EGFR-Mutated Advanced Non-Small-Cell Lung Cancer. *The New England Journal of Medicine*, 378(2), 113-125.

<https://doi.org/10.1056/NEJMoa1713137>

Stambolic, V., Suzuki, A., de la Pompa, J. L., Brothers, G. M., Mirtsos, C., Sasaki, T., Ruland, J., Penninger, J. M., Siderovski, D. P., & Mak, T. W. (1998). Negative regulation of PKB/Akt-dependent cell survival by the tumor suppressor PTEN. *Cell*, 95(1), 29-39. [https://doi.org/10.1016/s0092-8674\(00\)81780-8](https://doi.org/10.1016/s0092-8674(00)81780-8)

Targeted Therapy for Cancer—NCI (nciglobal,ncicenterprise). (2014, agosto 15). [cgvArticle].

<https://www.cancer.gov/about-cancer/treatment/types/targeted-therapies>

TFS-AssetsLSGmanualsMAN0017003_IonAmpliSeqLibaryKitPlus_UG.pdf. (s. f.). Recuperado 10 de noviembre de 2024, de https://assets.thermofisher.com/TFS-Assets%2FLSG%2Fmanuals%2FMAN0017003_IonAmpliSeqLibaryKitPlus_UG.pdf

van Roy, F., & Berx, G. (2008). The cell-cell adhesion molecule E-cadherin. *Cellular and Molecular*

Life Sciences: CMLS, 65(23), 3756-3788. <https://doi.org/10.1007/s00018-008-8281-1>

Vanderwalde, A., Spetzler, D., Xiao, N., Gatalica, Z., & Marshall, J. (2018). Microsatellite instability

status determined by next-generation sequencing and compared with PD-L1 and tumor mutational burden in 11,348 patients. *Cancer Medicine*, 7(3), 746-756.

<https://doi.org/10.1002/cam4.1372>

Wannovar. (s.f.). WANNONAR: Annotate and prioritize variants for clinical and functional studies. Recuperado de <https://wannovar.wglab.org/>

Yan, L., & Zhang, W. (2018). Precision medicine becomes reality-tumor type-agnostic therapy.

Cancer Communications (London, England), 38(1), 6. [https://doi.org/10.1186/s40880-018-](https://doi.org/10.1186/s40880-018-0274-3)

0274-3

Zehir, A., Benayed, R., Shah, R. H., Syed, A., Middha, S., Kim, H. R., Srinivasan, P., Gao, J., Chakravarty, D., Devlin, S. M., Hellmann, M. D., Barron, D. A., Schram, A. M., Hameed, M., Dogan, S., Ross, D. S., Hechtman, J. F., DeLair, D. F., Yao, J., ... Berger, M. F. (2017). Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of

10,000 patients. *Nature Medicine*, 23(6), 703-713. <https://doi.org/10.1038/nm.433>

11. Anexo

11.1 Protocolo de preparación de librerías

1. Amplificación de los genes de interés:

El protocolo que se muestra debajo fue adaptado a partir del Apéndice B del Manual del IonAmpliSeq Library Kit Plus. Se siguió el protocolo de 3 pool panels at 2X concentration, debido a que los cebadores tienen esta característica y se muestra el ejemplo utilizando 15 ng de ADN genómico obtenido de FFPE como partida y la mitad de volúmenes de los reactivos de librerías sugeridos por el fabricante.

Preparación de la Master Mix: En tubos eppendorf Lo-Bind

5 X Ion Ampliseq HiFi Mix	3.5 µl
ADN	≤5 µl (15 ng finales) (conc. mínima 3 ng/µl)
Nuclease free water	hasta 8.5 µl volumen final

A partir de esta *Master mix* se ponen 2.5 µl en cada tubo de PCR (3 tubos) y se agregan a éstos 2.5 µl de pool de cebadores 1 y 2.5 µl de pool de cebadores 2 y 2.5 µl de pool de cebadores para MSI.

Programa de ciclado

1 Ciclo	Activación de la enzima	99°C	2 minutos
19 Ciclos	Desnaturalización	99°C	15 segundos
	Anneal & extend	60°C	4 minutos
Hold	-	10°C	∞

(*) Los 19 Ciclos están asociados al número de *cebadores en los pools*. Ver tabla debajo.

Los pooles 1 y 2 requieren 16 ciclos o 19 ciclos (dependiendo si es ADN de buena calidad o baja calidad, respectivamente.) El pool 3 que es el panel de MSI llevaría 19 o 22 ciclos según la tabla extraída de la guía:

Primer pairs per pool	Recommended number of amplification cycles (10 ng DNA, 3,000 copies)		Anneal/extend time
	High quality DNA	Low quality DNA (FFPE DNA or cfDNA)	
12-24	21	24	4 minutes
25-48	20	23	4 minutes
49-96	19	22	4 minutes
97-192	18	21	4 minutes
193-384	17	20	4 minutes
385-768	16	19	4 minutes
769-1,536	15	18	8 minutes
1,537-3,072	14	17	8 minutes
3,073-6,144	13	16	16 minutes
6,145-24,576	12	15	16 minutes

En todos los casos dicen de partir de 10 ng que es lo que hicimos porque partimos de 30 ng y lo repartimos en 3 tubos. Al utilizar la mitad del kit de librería utilizamos en algunos casos 15 ng de ADN total de muestra.

--- **STOPPING POINT** --- Estos productos de PCR se pueden guardar a 10°C overnight (heladera) o a -20°C por más tiempo.

2) Digestión parcial de los amplicones

Juntar los tubos de las 3 PCRs correspondientes a cada pool antes de agregar la FuPa, las proporciones de cada una van a variar, el volumen final debería

ser 15 μ l. Colocamos 5 μ l de pool 1 y 2 y 2.5 μ l del pool de MSI, completando con 2.5 μ l para llegar a los 15 μ l.

Se agregan 1.5 μ l de FuPa a cada tubo y pipetear suavemente 5 veces cada tubo para homogeneizar (hacerlo en frío)

Programa de Ciclado:

50°C - 10 min

55°C - 10 min

60°C - 20 min

10°C - hold (hasta 1 hr máximo)

3) Ligar adaptadores a amplicones

a. Dilución de adaptadores (1:4)

Ion P1 adapter	1 μ l
Ion Xpress Barcode	1 μ l
Nuclease free water	2 μ l

Total 4 μ l - Se toman 3 μ l para cada tubo del mismo paciente.

b. Ligación:

Switch Solution	3 μ l
Barcode dilution (de parte a)	1.5 μ l
Amplicones digeridos	16.5 μ l
Total	42 μl

Agregar 1.5 µl de ADN ligasa, cada tubo queda con 22.5 µl, realizar vortex y spin.

Programa de ciclado PCR:

22°C - 30 min

68°C - 5 min

72°C - 5 min

10°C - hold hasta 24 hs

--- **STOPPING POINT** --- Estos productos se pueden guardar a 10°C overnight (heladera) o a -20°C por más tiempo.

4) Purificación con AMPURE.

1. Preparar etanol 70% suficiente para el numero de muestras (se necesitan realizar 2 lavados de 300 µl cada uno por muestra, calcular entonces 620 µl por muestra)
2. Poner en lo-bind tubes todo el volumen del paso anterior (22.5 µl) más 33.75 µl de Agencourt AMPure XP Reagent (1,5 X sample volume) pipetear arriba y abajo 5 veces y vortexear. (Este es el último punto donde escalamos por tener 3 pooles).
3. Incubar las muestras a temperatura ambiente por 5 minutos (mientras tanto poner a descongelar el low TE).
4. *Spinear* por 2 segundos si es necesario y quedaron gotas en las paredes, no pelletear las beads!!

5. Poner los tubos en un rack magnético e incubar un mínimo de 2 minutos o hasta que la solución se aclare.
6. Sacar cuidadosamente y descartar el sobrenadante sin distorsionar el pellet.
7. Agregar ~200-300 μ l de etanol 70%, asegurarse que las beads queden totalmente cubiertas.
8. Girar rápido el tubo en el rack magnético (2 vueltas)
9. Sacar cuidadosamente y descartar el sobrenadante sin distorsionar el pellet (no sacar del rack magnético)
10. Agregar ~200-300 μ l de etanol 70%, asegurarse que las *beads* queden totalmente cubiertas.
11. Girar rápido el tubo en el rack magnético (2 vueltas)
12. Sacar cuidadosamente y descartar el sobrenadante sin distorsionar el pellet (no sacar del rack magnético)
13. Asegurarse que no quede etanol en el tubo, sacando bien con pipeta. Si es necesario *spinear* y sacar con pipeta p20 el resto de etanol.
14. Dejar secar 5 minutos en el rack magnético con la tapa abierta para que se evapore el etanol, sin sobre-secar. Tiene que haber perdido el brillo y estar opaco.

5) Amplificación de las librerías

1. Sacar los tubos del magneto y adicionarle 25 μ l de 1X Library Amp Mix y y 1 μ l de 25X *Library Amp Primers*.

2. Vortexear bien la mezcla y breve centrifugación.
3. Poner los tubos en el magneto, y esperar por al menos 2 minutos.
4. Transferir cuidadosamente ~25 μ l de sobrenadante a un tubo de PCR sin distorsionar el pellet.

Programa de ciclado PCR:

98°C- 2 min 1 ciclo

98°C- 15 seg }
64°C- 1 min } 5 ciclos

10°C- hold hasta 24 hs

6) Purificación de las librerías amplificadas

6.1 Primera ronda de purificación con AMPure al 0.5 X:

1. Centrifugar los tubos post PCR
2. Agregar 12.5 μ l (0.5X) a cada tubo que contiene ~25 μ l de librería y mezclar bien.
3. Incubar la muestra por 5 min a temperatura ambiente.
4. Poner los tubos en un rack magnético e incubar un mínimo de 5 minutos o hasta que la solución se aclare.
5. Cuidadosamente transferir el sobrenadante de cada pocillo a un tubo wppendorf ADN lo-bind.

6.2 Segunda ronda de purificación con AMPure al 1.2 X:

1. Agregar 30 μ l (1.2X del volumen inicial de librería post-amplificación) a los tubos lo-bindl pipetear arriba y abajo 5 veces y vortexear.

2. Incubar las muestras a temperatura ambiente por 5 minutos
3. *Centrifugar brevemente Spinear* por 2 segundos si es necesario y quedaron gotas en las paredes, no pelletear las beads!!
4. Poner los tubos en un rack magnético e incubar un mínimo de 2 minutos o hasta que la solución se aclare.
5. Sacar cuidadosamente y descartar el sobrenadante sin distorsionar el pellet.
6. Agregar ~200-300 µl de etanol 70%, asegurarse que las beads queden totalmente cubiertas.
7. Girar rápido el tubo en el rack magnético (2 vueltas)
8. Sacar cuidadosamente y descartar el sobrenadante sin distorsionar el pellet (no sacar del rack magnético)
9. Agregar ~200-300 µl de etanol 70%, asegurarse que las *beads* queden totalmente cubiertas.
10. Girar rápido el tubo en el rack magnético (2 vueltas)
11. Sacar cuidadosamente y descartar el sobrenadante sin distorsionar el pellet (no sacar del rack magnético)
12. Dejar secar 5 minutos en el rack magnético con la tapa abierta para que se evapore el etanol, sin sobre-secar
13. Sacar el tubo del rack magnético y agregar 30 µl de low TE para dispersar las beads magnéticas, subiendo y bajando volumen más de 5 veces.
14. Incubar a temperatura ambiente por lo menos 2 minutos.
15. Colocar el tubo en el rack magnético, esperar al menos dos minutos y recoger el sobrenadante, el cual contiene la librería.
16. Evaluar la cantidad de la librería utilizando Qubit y la distribución de fragmentos obtenidos utilizando Bioanalyzer.

11.2 Comandos utilizados en el análisis bioinformático

A continuación, se listan los programas utilizados y el comando ejemplo que se utilizó para cada uno de ellos. Se presenta en *Itálicas* los nombres genéricos de los archivos. Los comandos se presentan en un orden cronológico donde, generalmente, el archivo de salida de un programa se carga como entrada en el siguiente y así sucesivamente.

1. Bowtie2.

Utilizamos este programa para mapear nuestras lecturas al genoma humano. Previo a utilizar bowtie2 se debe indexar el genoma humano de referencia (archivo fasta, en este caso del hg19) lo que nos produce 4 archivos que quedan bajo el prefijo que nosotros le asignemos en el paso de indexado. Ese prefijo es el que se debe aportar al momento de realizar el comando de mapeo que presentamos a continuación.

```
bowtie2 -p4 --local -x ruta_a_referencia/prefijo_referencia  
archivo_lecturas.fastq -S archivo_mapeo.sam
```

2. Samtools

Se utilizó la suite de programas conocida como Samtools para manipular los archivos de mapeo que genera el mapeador (archivo de texto plano .sam). En breve, el primer programa genera un archivo de mapeo binario (.bam) que es de menor tamaño y da lugar a una consulta mucho más rápida para los siguientes programas. Luego, empleando el programa *sort*, los datos de mapeo de cada lectura se ordenan según las coordenadas del genoma. Por último, el programa *index* construye un índice que es necesario en el caso de que se quiera inspeccionar los mapeos en un visualizador genómico (Por ejemplo, *IGV viewer*)

2a. Samtools view

```
samtools view -bS archivo_mapeo.sam > archivo_mapeo_binario.bam
```

2b. Samtools sort

```
samtools sort archivo_mapeo_binario.bam -o  
archivo_mapeo_binario_ordenado.bam
```

2c. Samtools index

```
samtools index -b archivo_mapeo_binario_ordenado.bam -o  
archivo_mapeo_binario_ordenado.bam.bai
```

3. BedTools counts

Previo a realizar los conteos preparamos un archivo de anotación (archivo .gff) a partir de los archivos de regiones (archivos .bed) que Ampliseq nos provee para cada panel. Para esto utilizamos comandos de manipulación de archivos de texto de linux para compaginar la información de ambos archivos. Luego usamos una herramienta de conversión de BED a GFF (*BED-to-GFF converter*) del Servidor Online de Galaxy (https://usegalaxy.org/root?tool_id=bed2gff1).

La asignación de las lecturas a cada amplicón no resulta trivial en el caso de exones grandes donde existen amplicones contiguos que presentan solapamiento. Luego de evaluar varias alternativas, consideramos que aplicar las opciones *-f 0.90* y *-F 0.90* de modo alternativo (opción *-e*) nos ofrecía resultados aceptables para seguir trabajando. En particular la opción *-f 0.90* exige que la lectura en cuestión cubra el 90% de la región (amplicon) donde se quiere asignar. La opción *-F 0.90* exige que el 90% de la lectura esté incluida en la región en cuestión para ser considerada. Al incluir la opción *-e*, esto significa que se van a contar las lecturas que cumplan una u otra de las condiciones (*-f* o *-F*). Lo que observamos en nuestros datos es que esta configuración minimiza los conteos por duplicado para amplicones que presentan mucho solapamiento, al tiempo que

esto recupera muchas lecturas cortas incompletas que son comunes para el caso de los amplicones del panel de inestabilidad y también es común que se produzcan cuando la calidad del ADN no es óptima.

```
bedtools coverage -counts -e -f 0.90 -F 0.90 -a
archivo_amplicones.gff -b archivo_mapeo_binario_ordenado.bam >
archivo_conteo.txt
```

4. Exportar a archivos xls y gráficas en GraphPad

Los archivos de conteo obtenidos son archivos de texto plano tabulares y presentan para cada amplicon el número de lecturas asignado al mismo. Compilamos en una hoja de cálculo los conteos obtenidos para los amplicones en cada una de nuestras muestras de secuenciación para llevar adelante los análisis. Se realizaron gráficos y análisis estadísticos utilizando el software GraphPad Prism v8.0.2.