

Maestría en Biotecnología

Facultad de Ciencias

**DESARROLLO Y VALIDACIÓN DE UN
PROTOCOLO DE BAJO COSTO PARA LA
DETECCIÓN DE MUTACIONES SOMÁTICAS EN
LOS ONCOGENES *KRAS*, *NRAS* Y *EGFR* POR
NGS**

Valentina Russo

Orientador: **PhD. Juan Martín Marqués**

Co-orientadora: **PhD. Astrid Agorio**

Mayo de 2019

AGRADECIMIENTOS

En esta meta que me propuse fueron muchas las personas que han estado a mi lado brindándome su apoyo.

En primer lugar destaco a mis compañeros de trabajo porque cada uno aportó su granito de arena para ayudarme a que este proyecto saliera adelante.

Tuve la oportunidad de conocer a mucha gente valiosa en los cursos en los que participé con los cuales he seguido compartiendo charlas y experiencias.

Estoy eternamente agradecida a mi familia que ha estado siempre para mí. Papá, mamá, Agus, no lo hubiera logrado sin ustedes.

Carlos, sos mi cable a tierra y mi apoyo incondicional, gracias por apoyarme y quererme tanto.

A mis tutores, Juan agradezco la generosidad de que compartas conmigo lo mucho que sabes. Astrid, sos mi gran mentora, lo das todo con pasión y sin condiciones. He aprendido mucho de ambos.

Gracias a cada uno por ayudarme a lograr lo que se resume a continuación...

RESUMEN

El cáncer es una de las principales causas de muerte en todo el mundo. Muchos enfoques terapéuticos contra esta enfermedad utilizan terapias dirigidas al receptor del factor de crecimiento epidérmico (EGFR). EGFR participa en cascadas de señalización intracelulares implicadas en el desarrollo y progresión tumoral. Mutaciones somáticas activantes en el gen *EGFR* se asocian con un aumento significativo de la respuesta a los fármacos inhibidores de la actividad tirosin-kinasa de EGFR (EGFR-TKIs), utilizados normalmente en cáncer de pulmón. Por otro lado, mutaciones en los oncogenes que codifican para las proteínas KRAS y NRAS, protagonistas de la vía de EGFR aguas abajo, provocan la activación constitutiva de la vía, independientemente de la actividad de EGFR y, debido a esto, generan baja respuesta a las terapias con anticuerpos monoclonales anti-EGFR, utilizadas normalmente en cáncer colorrectal. La determinación de mutaciones somáticas en estos genes es importante para predecir la eficacia terapéutica de estas drogas y poder identificar aquellos pacientes que se beneficiarán de las mismas. La técnica comúnmente utilizada para la detección de mutaciones somáticas es la secuenciación Sanger, que cuenta con una sensibilidad del 15 al 25% [1-4]. Sin embargo, los avances tecnológicos de los últimos años generaron una técnica de secuenciación masiva, NGS (Next Generation Sequencing), la cual permite detectar las variaciones genómicas que se dan en el ADN de un tumor, de forma más automatizada que Sanger, en varias muestras a la vez y para diferentes genes blanco, en un único experimento, y con una sensibilidad entre el 2-10% [5, 6]. El objetivo de este trabajo fue generar un protocolo de NGS utilizando la plataforma *Ion Torrent PGM™*, que fuera eficiente, rápido y con un menor costo que el planteado por el proveedor de forma que sea competitivo con Sanger; así como optimizar la detección de variantes en las secuencias generadas, utilizando como modelo mutaciones somáticas en los oncogenes *KRAS*, *NRAS* y *EGFR*.

Los resultados relevantes de este trabajo fueron la optimización de la extracción de ADN a partir de tejido FFPE (*Formalin-fixed paraffin-embedded*), la creación de un método de trazabilidad de muestras mediante STRs (*Short Tandem Repeats*), la adaptación y puesta a punto de una alternativa para la construcción de bibliotecas utilizando el kit *NEBNext® Fast DNA Library Prep Set for Ion Torrent™* a partir de amplicones y por último, la validación del sistema al que se llegó. También se evaluaron tres *software* para el análisis de secuencias: *Torrent Suite™* (*Thermo Fisher Scientific*), *Sophia DMD®* (*Sophia Genetics*) y *CLC Genomics Workbench* (*Qiagen*), y se comparó su *performance*. Con el sistema de NEB utilizado se logró una disminución de costos de 50% con respecto al sistema de Ampliseq y de 12% con respecto a la secuenciación Sanger, con una base de 8 muestras procesadas en paralelo, 2 veces por semana. En lo que respecta a la validación de la técnica se obtuvo un 99,5% de especificidad, 12% de sensibilidad (en muestras con 50 o más copias de genoma), 100% de repetibilidad y 100% de exactitud. Gracias a esta reducción en los costos se puede plantear a futuro la posibilidad de realizar un cambio de tecnología sin encarecer el diagnóstico final, siempre que exista un mínimo de 15 muestras por semana en el laboratorio.

PALABRAS CLAVES

Cáncer colorrectal, cáncer de pulmón, *KRAS*, *NRAS*, *EGFR*, NGS, validación.

ÍNDICE

1. INTRODUCCIÓN	10
1.1. CÁNCER	10
1.1.1. <i>Cáncer colorrectal</i>	11
1.1.2. <i>Cáncer de pulmón</i>	12
1.2. MEDICINA DE PRECISIÓN	13
1.2.1. <i>Vía de señalización mediada por EGFR</i>	14
1.3. PRUEBAS DIAGNÓSTICAS ACOMPAÑANTES PARA TRATAMIENTOS ANTI-EGFR Y TKI 20	
1.3.1. <i>Secuenciación de nueva generación (NGS)</i>	21
2. OBJETIVO	29
2.1. OBJETIVO GENERAL	29
2.2. OBJETIVOS ESPECÍFICOS	29
3. METODOLOGÍA	30
3.1. MUESTRAS.....	30
3.1.1. <i>Muestras clínicas</i>	30
3.1.2. <i>Controles de calidad</i>	30
3.1.3. <i>ADNs con mutaciones conocidas</i>	30
3.1.4. <i>ADN de sangre para referencia</i>	31
3.2. EXTRACCIÓN DE ADN DE TEJIDO FFPE	31
3.3. CUANTIFICACIÓN POR PCR EN TIEMPO REAL (QPCR) UTILIZANDO SYBR GREEN®.....	32
3.3.1. <i>Curva estándar para cuantificación de ADN</i>	32
3.3.2. <i>Curva estándar para cuantificación de bibliotecas</i>	32
3.3.3. <i>Reacción de qPCR</i>	32
3.4. CUANTIFICACIÓN DE ADN O BIBLIOTECAS POR QUBIT	33
3.5. AMPLIFICACIÓN POR PCR	34
3.5.1. <i>Cebadores para KRAS, NRAS y EGFR</i>	34
3.5.2. <i>Cebadores para amplificar microsatélites</i>	35
3.5.3. <i>Reacción de PCR</i>	35
3.6. ELECTROFORESIS EN GEL DE AGAROSA	36
3.7. SECUENCIACIÓN SANGER	36
3.8. MICROSATÉLITES POR ELECTROFORESIS CAPILAR.....	38
3.8.1. <i>Amplificación de STRs</i>	38
3.8.2. <i>Electroforesis capilar</i>	39
3.9. CONSTRUCCIÓN Y CUANTIFICACIÓN DE BIBLIOTECAS	39
3.9.1. <i>Protocolo Ampliseq</i>	39
3.9.2. <i>Protocolo NEB adaptado</i>	40

3.10. AMPLIFICACIÓN CLONAL, ENRIQUECIMIENTO Y SECUENCIACIÓN EN ION TORRENT PGM™	40
3.11. ANÁLISIS BIOINFORMÁTICO.....	41
4. RESULTADOS Y DISCUSIÓN	43
4.1. TEJIDO TUMORAL FIJADO CON FORMALDEHÍDO Y EMBEBIDO EN PARAFINA (FFPE)43	
4.1.1. Extracción de tejido FFPE.....	44
4.1.2. Evaluación de la extracción de ADN de tejido FFPE mediante PCR en tiempo real	48
4.1.3. Evaluación del rendimiento de las extracciones de ADN en muestras de FFPE utilizando el protocolo de Qiagen adaptado.....	50
4.1.4. Trazabilidad de los ADNs extraídos utilizando microsatélites	51
4.2. EVALUACIÓN Y PUESTA A PUNTO DE ALTERNATIVAS AL PROTOCOLO DEL PROVEEDOR PARA LA DETECCIÓN DE MUTACIONES SOMÁTICAS UTILIZANDO LA PLATAFORMA ION TORRENT PGM™	57
4.2.1. Construcción de bibliotecas	60
4.2.2. Cuantificación de bibliotecas.....	75
4.2.3. Análisis de secuencias y comparación de software alternativos	80
4.2.4. Análisis de costos.....	90
4.3. VALIDACIÓN.....	94
4.3.1. Rango de trabajo informable.....	94
4.3.2. Especificidad analítica (sustancias interferentes)	97
4.3.3. Sensibilidad analítica (límite de detección)	99
4.3.4. Precisión (repetibilidad).....	100
4.3.5. Exactitud.....	103
4.3.6. Resumen validación	108
4.3.7. Análisis global de los datos obtenidos tras la validación	109
5. DISCUSIÓN GENERAL.....	112
6. CONCLUSIÓN	119
7. BIBLIOGRAFÍA	120
8. ANEXOS	128
8.1. EJEMPLO DE CONSENTIMIENTO INFORMADO PARA ESTUDIO DE LOS GENES KRAS Y NRAS	128
8.2. FRECUENCIAS ALÉLICAS DE LA SOCIEDAD ARGENTINA DE GENÉTICA FORENSE ASOCIACIÓN CIVIL (SAGF) 2017.....	129
8.3. RESULTADOS DEL LÍMITE DE DETECCIÓN DEL PANEL KN REALIZADO PARA 50 Y 500 COPIAS.....	130

ABREVIACIONES

- ΔR_n : variación de la señal de fluorescencia generada en cada ciclo de la reacción de PCR
- €: moneda Euro
- A: Adenina
- ADN o DNA: Ácido desoxirribo-nucleico o en inglés *DeoxyriboNucleic Acid*
- ADNdc o dsDNA: ADN doble hebra o en inglés *double strand DNA*
- ARMS: sistema de amplificación de mutaciones refractarias, del inglés *Amplified Refractory Mutation System*
- ASB-PCR: PCR de alelo específica con un agente bloqueador, del inglés *Allele-Specific PCR with a Blocking reagent*.
- BAM: Formato binario de alineamiento, del inglés *Binary Alignment-map Format*.
- barcode: código de barras
- BRAF: del inglés *v-raf murine sarcoma viral oncogene homolog B*
- C: Citosina
- CCR: Cáncer colorrectal
- CCRm: Cáncer colorrectal metastásico
- CDT: Pruebas diagnósticas acompañantes, del inglés *Companion Diagnostic Test*
- chip: circuito integrado
- chr: cromosoma
- cluster: racimo
- CN: control negativo
- CNA: alteración del número de copias, del inglés *Copy Number Alteration*
- CODIS: base de datos de ADN de EEUU, del inglés *Combined DNA Index System*
- CT: ciclo umbral, del inglés *cycle threshold*
- dNTP: deoxinucleótido tri fosfato
- EC: Electroforesis capilar
- EEUU: Estados Unidos
- EGF: Factor de crecimiento epidérmico, del inglés *Epidermal GroWTh Factor*
- EGFR: Receptor del factor de crecimiento epidérmico, del inglés *Epidermal GroWTh Factor Receptor*
- FAP: Poliposis adenoatosa familiar, del inglés *Familial adenomatous polyposis*
- FDA: del inglés *Food and Drug Administration*
- FFPE: Fijado con formaldehído y embebido en parafina, del inglés *Formaldehyde-fixed paraffin-embedded*
- Flow: flujo de dNTPs en la secuenciación
- FN: falso negativo
- FP: falso positivo
- Fwd: directo, del inglés *Forward*
- G: Guanina

- GDP: Guanosina difosfato
- GEF: Factor intercambiador de nucleótidos de guanina, del inglés *Guanine nucleotide Exchange Factors*
- GTP: Guanosín trifosfato
- HB-EGF: factor de crecimiento tipo EGF de unión a heparina, del inglés *heparine binding-EGF*
- HNPCC: Síndrome de Lynch, del inglés *Hereditary nonpolyposis colorectal cancer*
- Horizon: empresa *Horizon Diagnostic*
- hotspot: punto caliente
- HRAS: *Harvey RAS*
- HRM: Curva de disociación de alta resolución, del inglés *High Resolution Melt*
- INDEL: inserción deleción
- ISP: perlas del inglés *Ion Sphere™ Particle*
- KRAS: *Kristen RAS*
- LB: Límite del blanco
- LD: Límite de detección
- MAB: Anticuerpo monoclonal, del inglés *Monoclonal Antibody*
- MAPK: Proteína-cinasas activadas por mitógenos, del inglés *Mitogen-Activated Protein Kinase*
- MEGA: *Molecular Evolutionary Genetics Analysis*
- MNP: Polimorfismo de múltiples nucleótidos, del inglés *Multiple Nucleotide Polymorphism*
- MSP: Ministerio de Salud Pública
- nº: número
- nd: no detectable
- NEB: *New England Biolabs*
- NGS: Secuenciación de nueva generación, del inglés *Next Generation Sequencing*
- NRAS: Neuroblastoma RAS
- NSCLC: carcinoma de células de pulmón no pequeñas, del inglés *not small cells lung carcinoma*
- OMS: Organización Mundial de la Salud
- PAN: porcentaje de acuerdo negativo
- Panel KN: panel que comprende los genes *KRAS* y *NRAS*
- PAP: porcentaje de acuerdo positivo
- pb: pares de bases
- PCR: Reacción en cadena de la polimerasa, del inglés *Polymerase Chain Reaction*
- PD: poder de discriminación
- PGM: *Personal Genome Machine*
- PIK3CA: fosfatidilinositol 3-quinasa catalítica alfa
- *pool*: conjunto
- PTEN: fosfatidilinositol-3,4,5-trisfosfato 3-fosfatasa

- qPCR: PCR cuantitativa en tiempo real, del inglés *quantitative PCR*
- R²: coeficiente de determinación
- RAS: Rat Sarcoma virus
- ROX: carboxy-X-rhodamine
- RRHH: Recursos Humanos
- Rv: reverso
- SAGF: Sociedad Argentina de Genética Forense Asociación Civil
- SD: Desvío estándar, del inglés *Standard Deviation*
- SNP: Polimorfismo de nucleótido simple, del inglés *Single Nucleotide Polymorphism*
- STD: estándar
- STR: repetidos cortos en tándem, del inglés *Short Tandem Repeat*
- SYBR Green: Cianina asimétrica
- T: Timina
- Template: plantilla
- TGF α : Factor de crecimiento transformante α , del inglés *transforming growth factor- α*
- TKI: Inhibidor Tirosin Kinasa, del inglés *Tyrosine Kinase Inhibitor*
- Tm: Temperatura disociación, del inglés *melting temperature*
- TMAP: del inglés *Torrent Mapping Alignment Program*
- TVC: del inglés *Torrent Variant Caller*
- VN: verdadero negativo
- VP: verdadero positivo
- VPTP: Valor predictivo técnico positivo
- WES: Secuenciación del exoma completo, del inglés *Whole Exome Sequencing*
- WGS: Secuenciación del genoma completo, del inglés *Whole Genome Sequencing*
- WT: normal, del inglés *Wild-Type*

1. INTRODUCCIÓN

1.1. Cáncer

El cáncer es una de las principales causas de muerte en todo el mundo, ocupando el segundo lugar luego de las enfermedades cardio-vasculares. Debido al aumento en el control de las enfermedades infecciosas gracias al progreso médico en su prevención y tratamiento, así como también a las mejoras generales en la calidad de vida, la expectativa de vida tuvo un aumento sustantivo. Como consecuencia, esto dio lugar a que, sobre la segunda mitad del siglo XX, emergiera un nuevo perfil epidemiológico en el cual las enfermedades crónicas y degenerativas, especialmente la patología cardiovascular y el cáncer, se constituyeron en las primeras causas de muerte a nivel mundial [7]. Según valores de la OMS (Organización Mundial de la Salud), en el año 2015, 8,8 millones de personas murieron a causa del cáncer, siendo este valor el 13% de las muertes en el mundo [8, 9]. Se estima que el número de casos nuevos de cáncer pasará de alrededor de 14 millones en 2012 a más de 20 millones en 2030, y que casi dos tercios de esos casos ocurrirán en los países menos desarrollados. El último reporte de la OMS muestra que más del 60% de todos los nuevos casos de cáncer se registra en África, Asia y América Central y del Sur, regiones de bajos o medianos recursos [8, 10].

De acuerdo con la definición establecida por la *American Cancer Society*, el cáncer es un grupo de enfermedades caracterizado por el crecimiento descontrolado y propagación de células anormales. Si este proceso no es controlado, puede resultar en la muerte del paciente [11]. Es muy importante tener en cuenta que el cáncer no es una única enfermedad sino un conjunto variado de enfermedades, es por ello que el análisis epidemiológico de la situación es necesariamente complejo en tanto que debe atender a este espectro variado de enfermedades [12]. Actualmente se encuentran definidos más de 200 tipos distintos de cáncer, cada uno con síntomas y formas de diagnóstico y tratamiento diferentes. Los cuatro tipos más comunes de cáncer son: pulmón, mama, colorrecto (CCR) y próstata, siendo 4 de cada 10 diagnósticos de alguno de estos tipos (40% de todos los casos de cáncer globales), según datos de 2018 del Proyecto Globocan (base de datos de relevamiento de cáncer a cargo de la Asociación Internacional de Registros del Cáncer, IACR) [9]. En la Figura 1 se observa el porcentaje de nuevos casos de cáncer para ambos sexos en el total de países estudiados por la OMS [13].

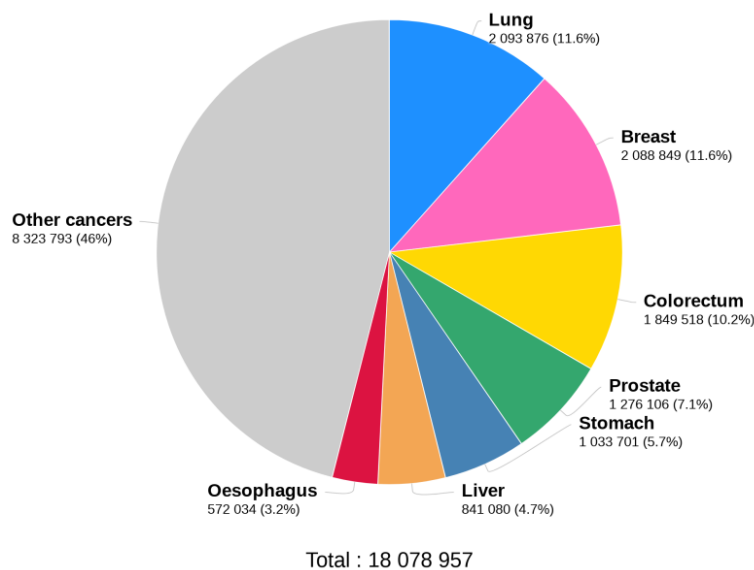


Figura 1: Número de nuevos casos para los distintos tipos de cáncer a nivel mundial para ambos sexos. Figura extraída de Globalcan 2018.

En Uruguay, según datos del Ministerio de Salud Pública (MSP), las muertes por cáncer constituyen aproximadamente un cuarto (24,4%) del total de las defunciones registradas en cada año (estas cifras corresponden al año 2016). En base a los datos del Registro Nacional de Cáncer en el período 2010-2014, se han registrado 79.716 casos incidentes de tumores malignos, esto representa un promedio de unos 15.943 casos nuevos cada año. En lo referente a la mortalidad, se han registrado 39.902 muertes por cáncer, lo cual representa un promedio de 7.980 muertes por año [12]. Las tasas estandarizadas por edad de incidencia del cáncer para ambos sexos en Uruguay están muy próximas al conjunto de los países más desarrollados, sin embargo, las tasas de mortalidad muestran relaciones más desfavorables [12]. El cáncer más frecuente en Uruguay, juntando ambos sexos, es el cáncer de mama (unos 1860 casos anuales, de los cuales 1840 son mujeres) seguido de cerca por el cáncer colorrectal (unos 1830 casos anuales). Les siguen el cáncer de próstata (1450 casos anuales) y el de pulmón (1400 casos anuales) [14].

1.1.1. Cáncer colorrectal

El CCR es uno de los cánceres más comunes tanto en el hombre como en la mujer, representando casi el 10% del total de los cánceres en el mundo. Es el tercer cáncer más frecuente entre los hombres (luego de los cánceres de próstata y pulmón) y el segundo más frecuente en mujeres [7]. Los cánceres del colon, recto-sigmoides y recto suelen agruparse en las estadísticas internacionales por la similitud en la patología y la epidemiología. En su mayoría

involucran a los mismos segmentos del intestino y los mismos factores de riesgo, lo que hace difícil, al momento del diagnóstico, saber con certidumbre dónde comenzó un tumor que se extendió a segmentos vecinos [7, 12].

La enfermedad se produce cuando las células del colon o recto crecen y se multiplican sin control, dañando e invadiendo el tejido circundante e interfiriendo con la función normal del órgano. Esta patología se produce de forma progresiva, comenzando por lesiones menores, como pólipos o lesiones displásicas, hasta adenomas, carcinomas intraepiteliales y carcinomas infiltrantes, llegando a generar metástasis a distancia. El colon ocupa los primeros 180 cm del intestino grueso mientras que el recto junto con el conducto anal se ubican en los últimos 25 cm del mismo. Alrededor de un 70% de los CCRs se encuentran en la primera porción del intestino grueso mientras que el resto aparece en el recto [15, 16].

En cuanto a los tipos de cáncer de colon, más del 95% son adenocarcinomas, a su vez, 20% de los cánceres de colon se heredan o están asociados con un historial sólido de cáncer de colon en la familia. Los principales tipos de cáncer de colon que se heredan incluyen: el síndrome de cáncer colorrectal hereditario no polipósico o HNPCC (también llamado síndrome de Lynch) que representa entre el 5-7% de los cánceres de colon; y la poliposis adenomatosa familiar (FAP) que causa cientos a miles de pólipos en el tracto gastrointestinal [16]. A nivel molecular existen dos orígenes diferentes: uno de ellos se encuentra asociado a la inestabilidad cromosómica (85%), mientras que el otro grupo está asociado a la presencia de inestabilidad de microsatélites o a la alteración de genes específicos de reparación del ADN (8-12%) [17].

El CCR es un grave problema en el espectro sanitario del Uruguay, las tasas de incidencia están entre las más altas del mundo y los resultados expresados por las tasas de mortalidad muestran un control más limitado de la enfermedad que aquellos obtenidos en países más desarrollados [18].

1.1.2. Cáncer de pulmón

El cáncer de pulmón es el tumor que causa una mayor mortalidad tanto en hombres como en mujeres. Si bien la mayoría de los casos están relacionados con el tabaquismo, cada vez hay más diagnósticos entre los no fumadores, especialmente en las mujeres. La mayoría de los cánceres de pulmón se diagnostican después de que la enfermedad se ha diseminado. Como resultado, la tasa de supervivencia a cinco años para los cánceres de pulmón es solo del 19% [19, 20].

Los cánceres de pulmón se clasifican en dos grupos principales de acuerdo con el tipo de células cancerosas que forman el tumor: cáncer de pulmón de células no pequeñas y cáncer de pulmón de células pequeñas. Existen diferencias significativas en el pronóstico y tratamiento para cada categoría. Cada uno de ellos crece y se disemina de un modo distinto y se trata de una forma diferente. El tratamiento también depende del estadio o de qué tan avanzado se encuentre [19]. El cáncer de pulmón de células no pequeñas (NSCLC) es el tipo más común de cáncer de pulmón y representa aproximadamente el 85% de los cánceres de pulmón. Surge de las células epiteliales de los pulmones, un tipo de célula que recubre la superficie de los órganos, incluidas las vías respiratorias. Estos cánceres tienden a comenzar como nódulos solitarios. A medida que crecen, pueden invadir las estructuras circundantes o diseminarse (hacer metástasis) a los ganglios linfáticos dentro del tórax, así como a órganos distantes. Hay varios tipos de cáncer de pulmón de células no pequeñas, los más comunes son: adenocarcinomas (comienza en las células glandulares que recubren los alvéolos), carcinoma de células escamosas (comienza en células epiteliales escamosas delgadas y planas que recubren las vías respiratorias de los pulmones) y carcinoma de células grandes (tipo menos común de cáncer de pulmón de células no pequeñas) [20]. El cáncer de pulmón de células pequeñas constituye aproximadamente el 15% de los cánceres de pulmón y casi siempre es causado por fumar tabaco. A menudo comienza en las porciones más centrales del pecho. También crecen y se propagan rápidamente a otras partes del cuerpo, incluidos los ganglios linfáticos [19, 20].

En Uruguay, el cáncer de pulmón es el segundo tipo de cáncer más frecuente en hombres, pero continúa siendo la primera causa de muerte por cáncer en este sexo. En mujeres, es actualmente el cuarto cáncer más frecuente (luego del cáncer de mama, colorrecto y cervix) [21].

1.2. Medicina de precisión

El diagnóstico correcto del cáncer es esencial para poder establecer un tratamiento adecuado y eficaz, porque cada tipo de cáncer requiere un protocolo específico que puede abarcar una o más modalidades, tales como la cirugía, la radioterapia o la quimioterapia [8]. Tradicionalmente, el cáncer se ha clasificado en función del tipo de tumor, ubicación anatómica, etapa en la que se encuentra o por terapias previas. Por este motivo, una persona diagnosticada de cáncer, usualmente recibía el mismo tratamiento que otras personas que tenían el mismo tipo y estadio de cáncer. Sin embargo, sucedía que individuos diferentes podían responder distinto al mismo tratamiento. Después de décadas de investigación, los científicos comprendieron que los tumores de los pacientes presentan cambios genéticos que causan que el tumor crezca y se disemine, y que en un mismo tipo de cáncer pueden encontrarse diferentes cambios genéticos dependiendo de la persona. Por ello, actualmente se

sabe que es más preciso diferenciar cada tipo de cáncer en función del tipo de alteraciones en el ADN [22]. Esta nueva era en el tratamiento contra el cáncer, mediante la "**medicina de precisión**", está llevando a tratamientos cada vez más focalizados. Con la evaluación genética de tumores se revelan las alteraciones en el ADN que promueven el crecimiento del cáncer. Conforme se comprenden estas variaciones, aumenta la capacidad de proveer mejores terapias de tratamiento para atacar de forma específica los cambios en el perfil genético del tumor [23].

El perfil de los genes que se encuentran mutados en cáncer de pulmón cambia mucho en función del tabaquismo. El tabaco influye en la carga mutacional y en el perfil de genes mutados. En pacientes fumadores la mutación más frecuente se da en el gen *KRAS*, mientras que en los no fumadores se da en el gen *EGFR*. Si se encuentra la alteración molecular y se trata al paciente de una manera dirigida, se obtiene una supervivencia global de más de 3 años, en contraposición a la quimioterapia convencional no dirigida que no da una supervivencia superior a los 12-15 meses [24, 25]. Es importante tener en cuenta que a pesar de que las mutaciones en *KRAS* son las más frecuentes, hoy en día no existe ningún fármaco dirigido a ellas (la detección de mutaciones en este gen solo sirve para determinar un mal pronóstico). Sin embargo, hay muchos tratamientos dirigidos a mutaciones en *EGFR*, y los pacientes presentan diferentes evoluciones [24].

En lo que refiere al CCR metastásico (CCRM), su gran incidencia y la alta tasa de mortalidad que provoca, ha desencadenado el desarrollo y aplicación de distintos protocolos terapéuticos. Muchos de ellos se basan en agentes quimioterapéuticos combinados con anticuerpos monoclonales contra blancos oncogénicos. Los anticuerpos más comúnmente utilizados como tratamiento son específicos contra EGFR (cetuximab y panitumumab) [5, 17].

1.2.1. Vía de señalización mediada por EGFR

El receptor EGFR es una glicoproteína transmembrana compuesta por un dominio extracelular aminoterminal para unión a ligandos, una hélice transmembrana hidrófoba, un dominio citoplasmático que contiene un dominio tirosina quinasa y una región carboxiterminal que contiene residuos de tirosina y elementos reguladores del receptor. El gen que codifica para esta proteína en humanos es EGFR (ubicación GRCh37/hg19: chr7:55,086,714-55,324,313, 7p11.2) [26-30]. Dicha proteína tiene 134kD y 1210 aminoácidos [30].

La unión de ligandos específicos, tales como el factor de crecimiento transformante (TGF α), el factor de crecimiento epidérmico (EGF), el factor de crecimiento tipo EGF de unión a heparina (HB-EGF), la anfirregulina, la betacelulina o la epirregulina, a la porción extracelular N-terminal

del receptor EGFR induce la dimerización del mismo (formando homodímeros o heterodímeros con otro miembro de la familia de EGFR humana), y activa el dominio quinasa intracelular. El resultado es la auto-fosforilación de residuos de tirosina en cada unidad del dímero, activando una cascada de señalización intracelular de eventos que incluyen el reclutamiento y la fosforilación de varios factores intracelulares (Figura 2). En ausencia de ligando, el EGFR permanece en un estado basal en el cual su conformación suprime la actividad quinasa y se evita la formación de dímeros [26-29].

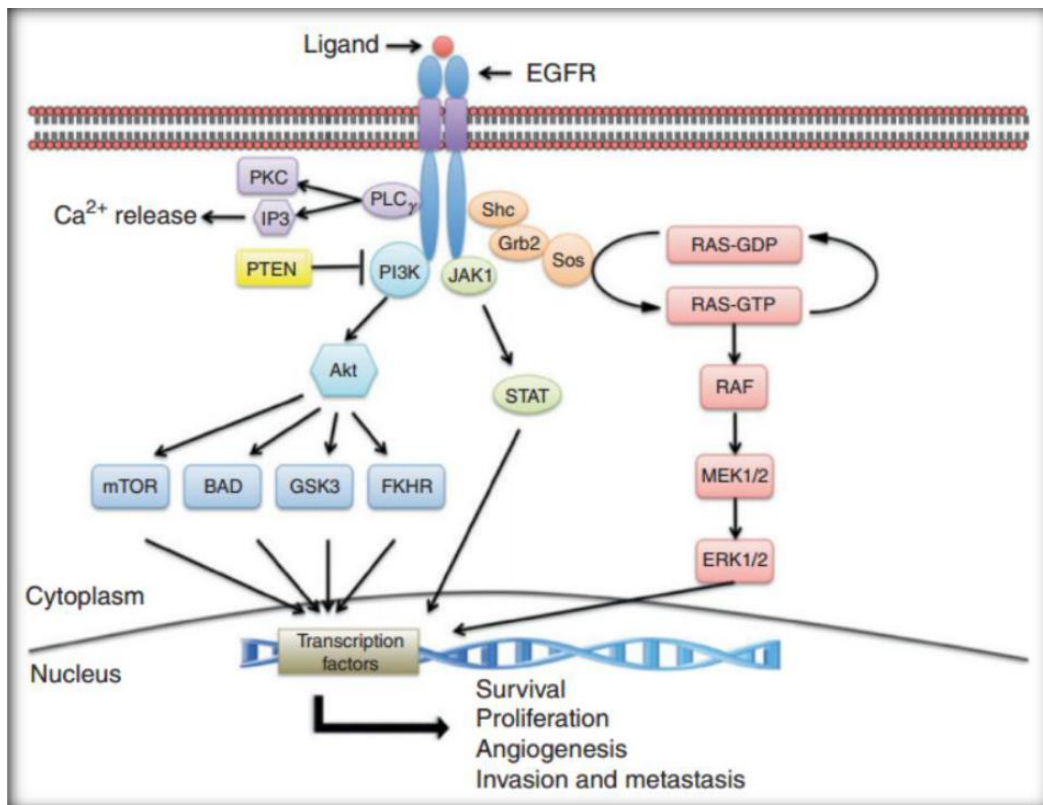


Figura 2: Vías de señalización mediadas por EGFR. En rojo: vía RAS-RAF-MAPK; en azul: vía PI3K-Akt; en verde: vía JAK-STAT; y en violeta: vía PLC γ . Figura extraída de Markman, et al. (2010).

Una vez producida la activación de la proteína EGFR, la señal puede dirigirse por cuatro vías de señalización hacia el interior de la célula: 1) RAS-RAF-MAPK, la cual regula la proliferación celular, migración, adhesión, supervivencia y reparación del ADN; 2) PI3K-Akt, el cual controla diferentes procesos biológicos, tales como el crecimiento y la proliferación, la activación transcripcional y la apoptosis; 3) JAK-STAT, la cual regula la división celular, la motilidad, invasión, adhesión, la reparación celular y la supervivencia, y por último la vía de PLC γ que regula señalización mediada por Calcio. En la Figura 2 se presentan las distintas vías de señalización mediadas por EGFR [17, 27, 29, 31].

La existencia de mutaciones en el gen *EGFR* puede alterar la expresión o actividad de la proteína codificada. Existen mutaciones que llevan a la sobre-expresión de EGFR (conocidas como de regulación positiva) o a la hiperactividad de la proteína, las cuales se han asociado con un gran número de cánceres. Los altos niveles de la proteína EGFR en tumores se han relacionado con una enfermedad agresiva, un mal pronóstico, una disminución de la supervivencia, una mala respuesta a la terapia y al desarrollo de resistencia a los agentes citotóxicos [17, 29, 32]. De las mutaciones conocidas del dominio EGFR tirosina quinasa, más del 90% ocurren como deleciones cortas en el exón 19 o como mutaciones puntuales en el exón 21. Estas mutaciones pueden dar como resultado la activación constitutiva de las vías de transducción de señales, lo que lleva a la proliferación celular o antiapoptosis, independientemente de la presencia de ligando extracelular. Otras mutaciones menos comunes ocurren en los exones 18 y 20. En consecuencia, la presencia de mutaciones en *EGFR* y su identificación como un oncogén ha hecho que la proteína codificada sea el objetivo de varias terapias contra el cáncer [33].

Las terapias usadas actualmente para tratar pacientes con mutaciones en *EGFR* incluyen anticuerpos monoclonales (MAb) y drogas llamadas inhibidores de tirosina quinasa (TKIs). Los anticuerpos, incluyendo cetuximab (Erbix[®]) y panitumumab (Vectibix[®]), se unen a la porción extracelular de EGFR para evitar que los ligandos activen los receptores. Gefitinib (Iressa[®]) y erlotinib (Tarceva[®]), son TKIs que se unen a la parte interior de EGFR e inhiben el proceso de activación. Se ha demostrado que estos fármacos son mejores que la quimioterapia tanto en tasa de respuesta (70-80% vs 35-40%) como en supervivencia libre de de progresión (1 año vs 4-5 meses), aunque en ningún estudio se demostró beneficio en supervivencia global [34].

En la actualidad existen TKIs de segunda y tercera generación, que poseen mayor capacidad de inhibir la proteína en la porción con variantes provocadas por las mutaciones en *EGFR*. Los inhibidores de segunda generación también inhiben a la porción *wild type* (WT) del receptor, lo que hace que sean más eficaces pero que tengan más efectos secundarios. Los inhibidores de tercera generación no inhiben a la porción WT del receptor y por lo tanto no presentan tantos efectos secundarios. Por otra parte, se sabe que los pacientes tratados con estos inhibidores de primera generación dejan de responder al cabo de un año aproximadamente. Esto es debido a una nueva variante adquirida (T790M) que genera resistencia a los fármacos. Los inhibidores de segunda y tercera generación tienen más capacidad de inhibir a esta variante de resistencia. Así, los inhibidores de segunda generación proporcionan una mayor supervivencia (16 meses) que los de primera generación pero con más efectos secundarios. La farmacéutica *Astra Zeneca* desarrolló un medicamento dirigido a esa nueva variante en este tipo de pacientes, Osimertinib [35, 36].

1.2.1.1. Oncogenes de la vía RAS-RAF-MAPK

Debido a la falta de respuesta a los tratamientos contra EGFR, se hicieron estudios que demuestran que la presencia de mutaciones en el gen *KRAS* están asociadas a dicha falta de respuesta. El uso de los fármacos cetuximab y panitumumab en pacientes con CCRm con *KRAS* no mutado (*WT*) prolonga la supervivencia libre de progresión y la supervivencia global comparados con el mejor tratamiento de soporte. Es por ello que las pruebas sobre el gen *KRAS* se han convertido en obligatorias para determinar la estrategia terapéutica más adecuada [1, 17, 37-39].

Dentro de la familia de genes *RAS* se encuentra el gen *KRAS* (ubicación GRCh37/hg19: chr12:25,357,723-25,403,870, 12p12.1), *NRAS* (ubicación GRCh37/hg19: chr1:115,247,085-115,259,515, 1p13.2) y *HRAS* (ubicación GRCh37/hg19: chr11:532,242-537,287, 11p15.5) [40-42]. Se encuentran en cromosomas diferentes y tienen una expresión variable en distintos tejidos, pero sus secuencias, estructuras y funciones son muy similares. Estos genes codifican para las proteínas RAS, las cuales median la señalización producida mediante la activación de algunos receptores de membrana. Las tres proteínas *RAS* presentan un 85% de identidad a nivel de aminoácidos. Este nivel de conservación no es uniforme a lo largo del gen, por lo que se pueden identificar cuatro regiones distintas. La primera región (entre los aminoácidos 1 al 85) es la más conservada y las tres proteínas son idénticas. La segunda región (siguientes 80 aminoácidos) muestra un 85% de identidad entre las proteínas *RAS*. Las últimas dos regiones son más variables entre las proteínas. La similitud de estas proteínas y la presencia de pseudogenes pueden causar dificultades en la amplificación y el análisis bioinformático de las variantes [43]. Las proteínas RAS inactivas están unidas a GDP (Guanosina difosfato), en el momento en que un factor intercambiador de nucleótidos de guanina (GEF) favorece la formación de GTP-RAS (Guanosina trifosfato-RAS), la proteína se activa. Luego, dicho GTP es hidrolizado por la actividad GTPasa intrínseca de las proteínas RAS, provocando la inactivación de la proteína RAS [17, 29, 37].

El gen *KRAS* presenta dos copias en el genoma humano, denominadas *KRAS1* y *KRAS2*. El ARNm codificado por *KRAS2* es el principal y tiene una longitud de 5,5 kb. El gen *KRAS1* es un pseudogen derivado de *KRAS2* por un *splicing* alternativo [44]. *KRAS2* (de aquí en más *KRAS*) codifica para una proteína humana de 21,6kD con 189 aminoácidos [41]. Al igual que otros miembros de la familia *RAS*, la proteína *KRAS* es una GTPasa que participa de forma temprana en muchas vías de transducción de señales. Dicha proteína se encuentra normalmente unida a las membranas celulares debido a la presencia de un grupo isoprenilo en su C-terminal. Su principal función es actuar como un interruptor molecular de encendido/apagado. Una vez que se enciende, recluta y activa las proteínas necesarias para la propagación de la señal mediada por EGFR y de señales de otros receptores. Para ello, se une

a GTP en el estado activo y posee una actividad enzimática intrínseca que escinde el fosfato terminal del nucleótido, convirtiéndolo en GDP. Tras la conversión de GTP a GDP, la proteína KRAS se encuentra desactivada [17, 29, 37, 38].

La proteína KRAS es codificada por un oncogén, por lo que una sustitución de un sólo nucleótido puede generar un cambio de aminoácido y provocar una mutación activante. Cuando se presentan determinadas mutaciones en el gen *KRAS*, la actividad GTPasa de la proteína queda bloqueada y permanece constitutivamente activa y unida a GTP, independientemente de la señal proveniente de EGFR. En la Figura 3 se muestra la comparación entre la vía RAS-RAF-MAPK de una célula con funcionamiento normal de la vía de EGFR (Fig. 3A), una célula con funcionamiento normal de la vía de EGFR a la cual se le aplican los MAb cetuximab o panitumumab (Fig. 3B), y una célula con el gen *KRAS* mutado donde se genera la activación de la vía intracelular, independientemente de EGFR e insensible al MAb usado (Fig. 3C).

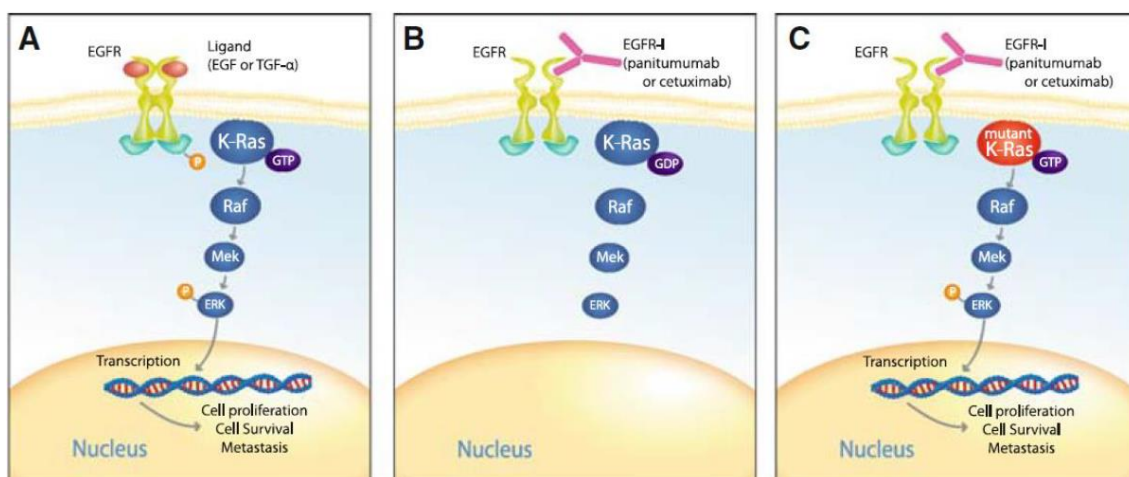


Figura 3: Ruta de señalización mediada por EGFR. (A) Célula con funcionamiento normal de la vía de EGFR; (B) inhibición del EGFR por Cetuximab o Panitumumab, detiene la señalización corriente abajo, (C) mutación en *KRAS* provoca la activación por unión constitutiva a GTP y permite la señalización corriente abajo independientemente de la acción de Cetuximab o Panitumumab. Figura extraída de Siddiqui, Piperdi (2010) [45].

Mutaciones somáticas activantes en el gen *KRAS* suelen ocurrir en los codones 12 (70 a 80%) y 13 (15 a 20%) del exón 2 [38]. Las restantes mutaciones se localizan en los exones 3, en un 2,5% (codones 59 y 61) y 4 en un 3,2% (codones 117 y 146). Dichas mutaciones pueden provocar una baja respuesta a las terapias [37, 39, 46]. Aproximadamente el 25% de los pacientes con adenocarcinoma de pulmón son positivos para una mutación en el gen *KRAS*. Reciben tratamientos basados en platino, como el cisplatino, que han sido efectivos para prolongar el tiempo de supervivencia de muchos pacientes con cáncer de pulmón con mutaciones activantes en *KRAS* [47]. En los CCRm, aproximadamente el 40% de los casos

presentan mutaciones en el gen *KRAS*, permitiendo que el restante porcentaje de pacientes respondan eficientemente a las terapias con anticuerpos anti-EGFR. Sin embargo, entre el 40 al 60% de los pacientes con *KRAS WT* no responden a las mismas. Este problema es debido a la presencia de mutaciones en otros genes de la vía RAS-RAF-MAPK, cuyos cambios provocan una deficiencia en la acción de las terapias, como es el caso del gen *NRAS* [48].

El gen *NRAS* es un oncogén miembro de la familia de genes *RAS* que codifica para una proteína de 21kD, NRAS, con 189 aminoácidos, cuya función es similar a la proteína KRAS [40]. Como todas las proteínas RAS, tiene la capacidad de unirse tanto a GTP como a GDP, presenta actividad GTPasa y su función normal es como una proteína G reguladora involucrada en el control normal del crecimiento celular [38]. Al igual que en el gen *KRAS*, en *NRAS* existen mutaciones activantes en los exones 2 (codones 12 y 13), 3 (codones 59 y 61) y 4 (codones 117 y 146). Las mutaciones en *NRAS* no promueven tanto la proliferación celular y la supresión de la diferenciación, sino más bien inducen la sobreexpresión de las funciones anti-apoptóticas. Las mutaciones en este gen se encuentran en el 3-5% de los pacientes con CCRm y se ha demostrado una reducción de la respuesta a cetuximab y panitumumab en dichos pacientes [38]. Por ello se han incorporado estas mutaciones en los estudios para saber si conviene o no administrar las drogas anti-EGFR.

Por otro lado, se encuentra el gen *BRAF* el cual codifica para una quinasa serin-treonina y es el efector que se encuentra por debajo de las proteínas RAS en la vía RAS-RAF-MAPK. Al igual que con los genes anteriores se ha asociado la presencia de mutaciones en el gen *BRAF* (10-15%) con una falta de respuesta a tratamientos con anticuerpos monoclonales anti-EGFR. Dichas mutaciones generan que la vía de señalización a través de MAPK (*Mitogen Activated Protein Kinase*) se vea constitutivamente activada [38].

Más allá de los tres genes nombrados, existen otros genes subyacentes al receptor de EGFR que pueden estar constitutivamente activados por mutaciones hasta en el 10% de los casos, como es el caso del gen *PIK3CA* (*phosphatidylinositol-4,5-bisphosphate 3-kinase*) o por pérdida de función, como es el caso del gen *PTEN* (*fosfatidilinositol-3,4,5-trisfosfato 3-fosfatasa*), hasta en el 40% de los casos [49].

De todos los genes mencionados que afectan la sensibilidad a los tratamientos anti-EGFR y TKi, se evalúan principalmente *KRAS*, *NRAS* y *EGFR*.

1.3. Pruebas diagnósticas acompañantes para tratamientos anti-EGFR y TKi

Las pruebas diagnósticas acompañantes (CDT) con fines terapéuticos son estudios que ayudan a los médicos a decidir qué tratamientos recetar, adaptándolos específicamente al paciente. Estos estudios permiten identificar a los pacientes que tienen más probabilidades de beneficiarse del producto terapéutico, conocer a los pacientes que podrían estar en mayor riesgo de presentar reacciones adversas al tratamiento, monitorear la respuesta al tratamiento para poder ajustarlo y lograr una mejor seguridad y eficacia, e identificar a los pacientes para los que el producto terapéutico sea efectivo. Un gran reto para quienes realizan el diagnóstico es poder desarrollar estas pruebas durante la generación del fármaco de interés, logrando así un desarrollo conjunto del tratamiento con la prueba específica que aporta información para su uso seguro [50, 51].

Existen distintas metodologías que pueden ser usadas para detectar mutaciones somáticas activantes en los genes *EGFR*, *KRAS* y *NRAS*, todas ellas presentan ventajas y desventajas. Se debe tener en cuenta que el diagnóstico sobre estos genes está limitado por dos factores: primero, las células tumorales con el gen mutado se encuentran normalmente en minoría con respecto a las células tumorales con el gen *WT* y las células no tumorales; segundo, las muestras utilizadas para la detección de mutaciones provienen de tejido tumoral, al cual se le realiza un proceso de fijación con formalina (FFPE, *Formalin-fixed paraffin-embedded*), donde, dependiendo del proceso, la integridad del ADN puede verse seriamente comprometida [37].

La técnica considerada como “*Gold Standard*” para la detección de mutaciones es la secuenciación Sanger. Esta técnica permite secuenciar completamente un fragmento específico de ADN posibilitando la detección de cambios puntuales (SNP, *Single Nucleotide Polymorphism*) e INDELS (inserciones-deleciones). La secuenciación Sanger es una técnica específica pero su sensibilidad es relativamente baja (15-25%) para detectar mutaciones que pueden estar presentes en baja proporción en muestras de tejidos FFPE. Además, cuando se realiza a gran escala, la secuenciación Sanger consume mucho tiempo y recursos humanos; y la heterogeneidad intratumoral aumenta más la dificultad en la interpretación de los datos [1-4].

Existen otras técnicas alternativas para detectar SNPs basadas en otros métodos moleculares. Entre ellas está HRM (*High Resolution Melt*), pirosecuenciación y varios métodos basados en PCR cuantitativa en tiempo real (qPCR). La qPCR es una variante de la reacción en cadena de la polimerasa (PCR) utilizada para amplificar y simultáneamente cuantificar el producto de la amplificación de ADN. También sirve para identificar productos de ADN o amplicones específicos a partir de su temperatura de disociación (T_m) [52]. Comercialmente existe el kit *TheraScreen*, el cual utiliza cebadores ARMS (*Amplified Refractory Mutation System*) y sondas *Scorpion* para detectar las siete mutaciones más comunes del gen *KRAS* en los codones 12 y

13. Este método presenta mayor sensibilidad que Sanger siendo capaz de detectar mutaciones en *KRAS* cuando las mismas se encuentran en un 5% del ADN total [1, 53]. Otra de las metodologías basadas en qPCR utiliza las sondas *TaqMan* combinadas con ASB-PCR (*Allele-Specific PCR with a Blocking reagent*) cuya sensibilidad llega al 1% [54]. Para *EGFR* se encuentra la prueba cobas® EGFR Mutation Test v2 que se basa en PCR en tiempo real. La misma permite identificar 42 mutaciones en los exones 18, 19, 20 y 21 del gen *EGFR*, incluida la mutación de resistencia T790M. Este kit tiene un nivel de detección de mutaciones de al menos 5% [55].

El gran inconveniente de los ensayos de qPCR es que son específicos del blanco molecular, por lo que solo pueden detectar *hotspot* mutacionales para los cuales se diseñó originalmente el ensayo, pero no otras mutaciones en el ADN del tumor. Cuando se deben analizar múltiples genes, tanto la secuenciación de Sanger como la qPCR requieren una reacción de amplificación por cada amplicón estudiado, necesiándose en cada reacción una cantidad de ADN específica, por lo que puede no ser factible debido a la disponibilidad limitada de muestras de tejido FFPE [4].

Los avances tecnológicos de los últimos años han desarrollado una nueva técnica de secuenciación masiva conocida como NGS (*Next Generation Sequencing*) la cual permite secuenciar un gran número de bases por ensayo. Utilizando esta técnica se pueden detectar todos los tipos de variaciones genómicas en un único experimento, incluyendo SNPs y/o INDELS. Además, un gran número de amplicones pueden amplificarse en un solo tubo de reacción, así como secuenciar posteriormente todos ellos en un solo ensayo [56].

1.3.1. Secuenciación de nueva generación (NGS)

La tecnología de NGS ha desempeñado un papel esencial en la comprensión de las vías genéticas involucradas en cáncer. En comparación con los métodos anteriores de secuenciación, presenta numerosas ventajas. Principalmente, este es un método de alto rendimiento, ya que permite la secuenciación simultánea y dirigida de múltiples regiones genómicas en múltiples muestras para detectar distintas mutaciones en un único ensayo. Otra ventaja importante, en la secuenciación rutinaria de ADN de tumores, es el tiempo de generación del resultado reducido así como la automatización, lo que acorta mucho el tiempo de emisión del resultado así como limita posibles errores humanos en el proceso. Además, un análisis de NGS requiere una baja cantidad de ADN en contraste con los métodos de secuenciación tradicionales [5]. Utilizando esta técnica se pueden detectar variaciones genéticas (conocidas, tipo *hotspot* y variaciones *de novo*) que se dan típicamente en un tumor, en un experimento único, incluyendo variantes como mutaciones puntuales o deleciones pequeñas menores a 15pb, con una sensibilidad entre el 2-10% [5, 6].

Mediante NGS se puede realizar la secuenciación del genoma completo (WGS), la secuenciación del exoma completo (WES) o la secuenciación dirigida. WGS permite la secuenciación de todo el genoma, lo que requiere una gran cantidad de ADN. Para detectar con precisión mutaciones somáticas, es posible que se necesite una cobertura de secuenciación de 100 a 200 veces más que para detectar variantes germinales, lo que es muy costoso [4]. WES se enfoca en las regiones codificantes (exones) de un genoma, aproximadamente un 2,5% del genoma humano, para descubrir variantes raras o comunes asociadas con un trastorno o fenotipo. WES reduce costos y tiempo en comparación con WGS. Los métodos más comunes se basan en la hibridación mediante sondas de oligonucleótidos para "capturar" fragmentos de ADN dirigidos, enriqueciendo así en secuencias exónicas [5, 6]. Alternativamente, la secuenciación dirigida, centrada en una selección de genes de interés para una enfermedad específica, es más precisa y accesible en términos de tiempo y costo para aplicaciones clínicas en los laboratorios. Por ello este abordaje se utiliza más en diagnóstico.

El flujo de trabajo de NGS está constituido por diferentes pasos, desde la extracción del ácido nucleico hasta la anotación de variantes, como se muestra en la Figura 4. Los distintos protocolos para realizar las etapas del flujo de trabajo de NGS distinguen a una tecnología de otra y determina el tipo de datos producidos desde cada plataforma. Actualmente hay dos empresas principales que ofrecen plataformas NGS: *Illumina* y *Thermo Fisher Scientific*. Cada una de las plataformas utiliza diferentes métodos químicos de secuenciación y detección de señales [5, 57-59].

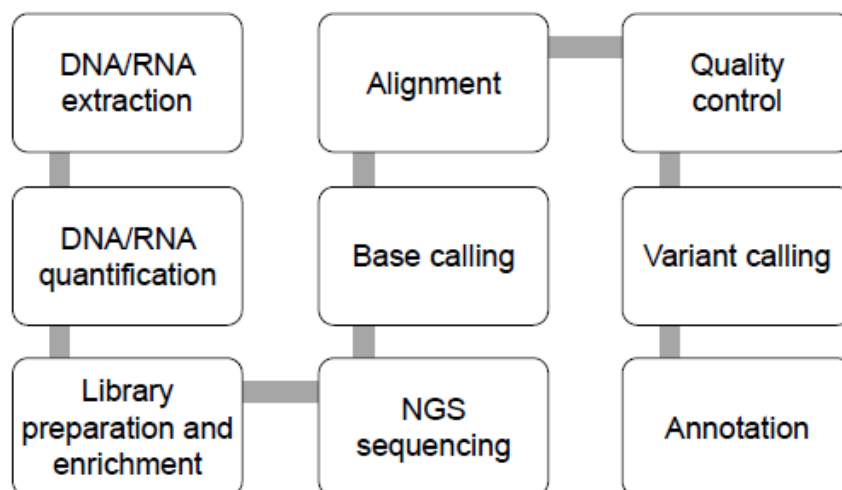


Figura 4: Flujo de trabajo de NGS: extracción de ADN/ARN; cuantificación; preparación de bibliotecas y enriquecimiento; secuenciación; asignación de bases; alineamiento; control de calidad; llamado de variantes y anotación de variantes. Figura extraída de Serrati, et al. (2016).

Las plataformas *Illumina* se caracterizan por realizar la amplificación clonal sobre la misma superficie en la que se va a llevar a cabo la secuenciación, es lo que se conoce como PCR en puente. Los cebadores complementarios a los dos adaptadores se encuentran unidos covalentemente a la superficie, y el ADN sintetizado se curva haciendo puentes entre los adaptadores. Al final se genera una amplificación en forma de *cluster* (racimo). Los secuenciadores *Illumina* utilizan cuatro nucleótidos marcados con fluorescencia y una cámara digital para visualizar el nucleótido marcado incorporado. Para cada nucleótido incorporado se genera una imagen que se procesa posteriormente para deducir la secuencia de nucleótidos. Esta metodología se caracteriza por el uso de nucleótidos marcados que bloquean de forma reversible la elongación de la cadena. De este modo, tras la detección de la incorporación del fluoróforo, y la eliminación del mismo, es posible continuar con un nuevo ciclo de adición de un nuevo nucleótido marcado. Es lo que se conoce como secuenciación por síntesis. La tasa de error estimada para la tecnología *Illumina* es del 0,4% [60-63].

Por otro lado, *Thermo Fisher Scientific* utiliza nucleótidos sin marcar y un sistema de detección que no es óptico. En su lugar, se basa en el hecho de que la adición de un dNTP a una cadena de ADN libera un ion hidrógeno. La secuenciación por síntesis se realiza en pocillos microscópicos interconectados con un chip semiconductor. Después de la incorporación de los nucleótidos uno por uno, los protones liberados dan como resultado un cambio en el pH, medido por el chip semiconductor. Previo el ADN se amplifica clonalmente en esferas microscópicas. La tasa de error estimada para la tecnología Ion Torrent es 1,8%-1,9%, principalmente en la detección de homopolímeros [62, 63]. Esta metodología es la que se utiliza en este trabajo, por lo que se detalla a continuación.

1.3.1.1. Tecnología Ion Torrent™

El primer instrumento de *Ion Torrent™*, *Ion Personal Genome Machine™* (PGM), fue el secuenciador de nueva generación más económico del mercado. Dada su capacidad de salida y la velocidad de las corridas (2 horas), se ha orientado hacia genomas más pequeños y secuenciación dirigida. El proveedor ofrece tres modelos de *chip* con diferentes capacidades, por un lado está el chip 314 donde se puede llegar a secuenciar 1 millón de lecturas, chip 316 6 millones de lecturas y chip 318 11 millones de lecturas aproximadamente [64].

En 2015, se lanzó el *Ion S5™* y el *Ion S5™XL*. Los sistemas emplean la tecnología de semiconductores Ion Torrent en múltiples formatos de chip que le permiten ejecutar múltiples aplicaciones de investigación, como la secuenciación de paneles genéticos, ensayos personalizados, genomas pequeños, exomas y transcriptomas, todo en un solo instrumento, así como identificación de mutaciones somáticas, rubro en el cual *Thermo Fisher Scientific* y su método de amplificación ha mostrado algunas ventajas frente a su competidor [65].

El flujo de trabajo de la tecnología *Ion Torrent* consta de cuatro pasos principales: construcción de bibliotecas, amplificación clonal, secuenciación y análisis [64]. El primer paso en el flujo de trabajo es la **construcción de bibliotecas**. Si bien hay una serie de opciones disponibles, el proceso es parecido y generalmente implica tomar fragmentos de un tamaño uniforme (generalmente 200-400pb) para luego ligarle covalentemente adaptadores que servirán para la secuenciación. Los fragmentos pueden provenir de la fragmentación del ADN genómico o de la amplificación por PCR de regiones específicas del genoma. Los adaptadores pueden tener unido un código de barras que permita diferenciar a las muestras (*barcode*) tras la secuenciación en paralelo de múltiples muestras. En la Figura 5 se esquematiza el proceso de construcción de bibliotecas.

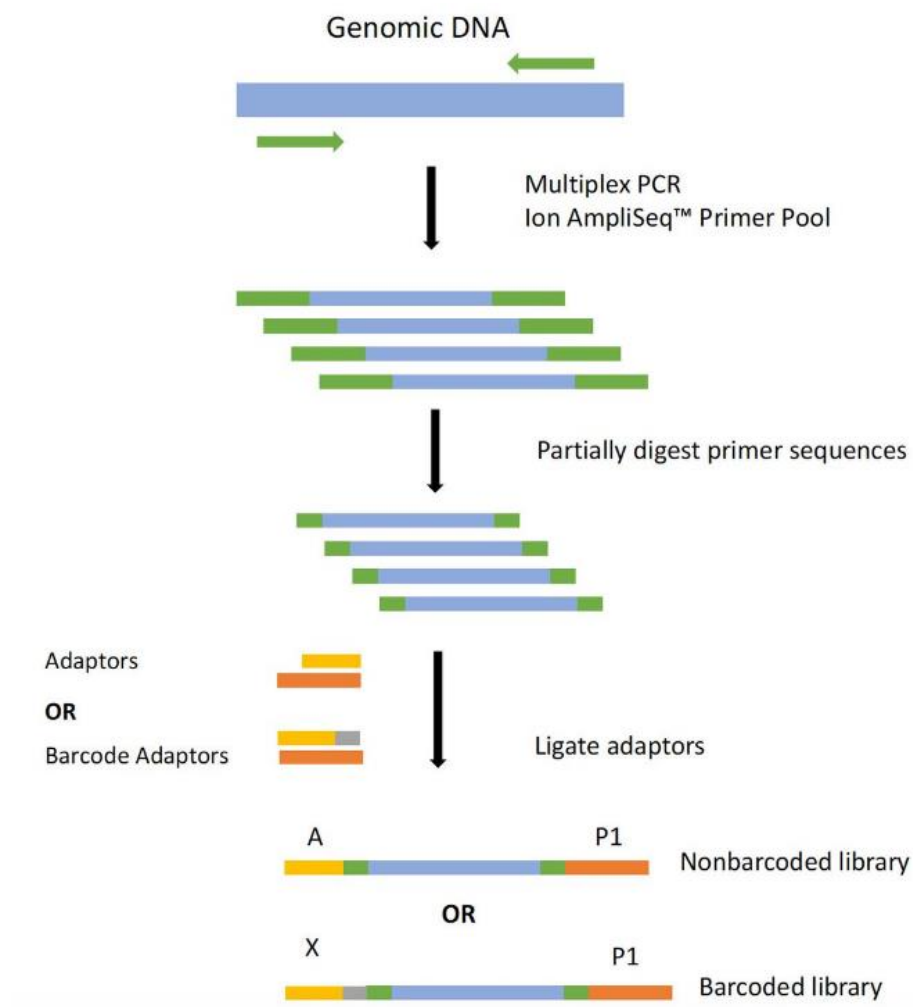


Figura 5: Figura que esquematiza el proceso de construcción de bibliotecas de la tecnología *Ion Torrent*.
Figura extraída de Floragenex 2018 [66].

Los fragmentos generados durante la preparación de la biblioteca se unen a perlas y se amplifican utilizando PCR en emulsión (**amplificación clonal**). Las perlas recubiertas con cebadores complementarios se mezclan con una solución acuosa diluída que contiene los fragmentos a secuenciar junto con los reactivos de PCR necesarios. Esta solución se mezcla luego con aceite para formar una emulsión de microgotas. La concentración de fragmentos se mantiene lo suficientemente baja como para que cada microgota contenga sólo un fragmento por cada perla (o posiblemente ninguna, pero casi nunca más de un fragmento). La amplificación clonal de cada fragmento se lleva a cabo dentro de las microgotas (Figura 6). Después de la amplificación, la emulsión se “rompe” (por extracción orgánica y centrifugación) y las perlas amplificadas se enriquecen en un gradiente de glicerol.

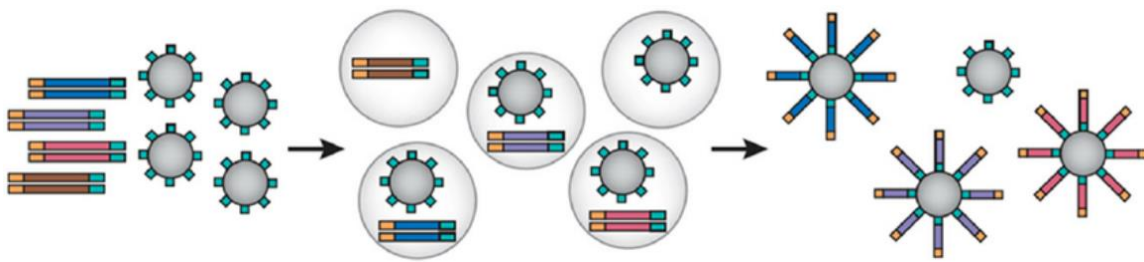


Figura 6: Proceso de amplificación clonal. Las perlas recubiertas con cebadores complementarios se mezclan con una solución acuosa diluída que contiene los fragmentos a secuenciar junto con los reactivos de PCR necesarios. Figura extraída de Wonyong K (2012) [67].

Lo que realmente caracteriza a los sistemas de *Ion Torrent* es su tecnología de **secuenciación**. En ella, se traduce directamente la información codificada químicamente (A, C, G, T) en información digital (0,1) en un chip semiconductor. El resultado es una tecnología de secuenciación simple, rápida, rentable y escalable [68].

Cuando un nucleótido se incorpora a una hebra de ADN mediante una polimerasa, se libera un ion hidrógeno como producto secundario (Figura 7).

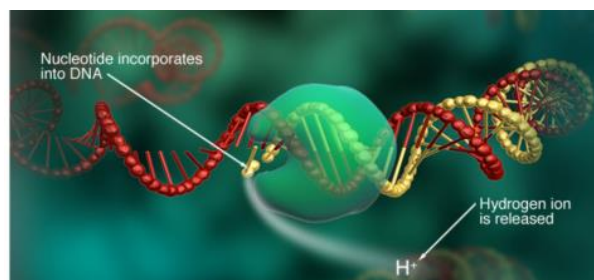


Figura 7: En esta figura se encuentra representada la incorporación de un nucleótido en una cadena de ADN y la liberación de un ion hidrógeno en consecuencia [68].

Si hay dos bases idénticas en la cadena de ADN, el nivel de protones liberados será el doble, el voltaje será doble y el chip registrará dos bases idénticas. Debido a que se trata de detección directa (sin escaneo, sin cámaras, sin luz), cada incorporación de nucleótidos se registra en segundos (Figura 8).

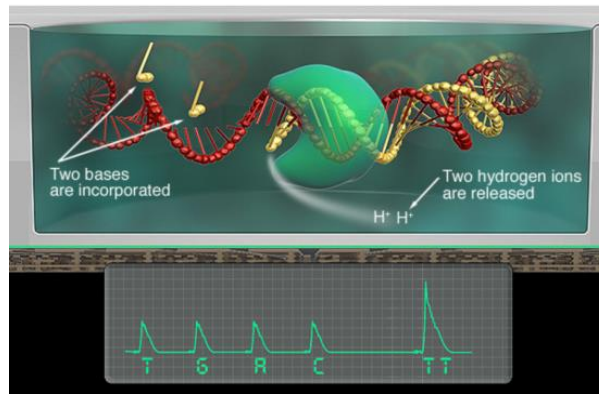


Figura 8: En esta figura se encuentra representada la actividad de la polimerasa liberando dos protones tras incorporar dos T consecutivas (arriba). Abajo esta representado el registro de las incorporaciones de los dNTPs en donde G es Guanina, A es Adenina, C es Citocina y las dos T son dos Timinas [68].

Por ejemplo, si se agrega un nucleótido, una C, a una cadena de ADN, se liberará un ion hidrógeno (se incorpora 1C por cada uno de los fragmentos alrededor de la perla). La carga de ese ion cambiará el pH de la solución, el medidor de pH detectará a la base y pasará directamente de la información química a la información digital. El secuenciador *Ion Personal Genome Machine™* luego inunda secuencialmente el chip con un nucleótido diferente tras otro. Si el siguiente nucleótido que inunda el chip no coincide, no se registrará ningún cambio de voltaje y por lo tanto, no se llamará una base (Figura 9). Si en vez de entrar una sola T entran 2 T habrá un cambio de pH proporcional y el algoritmo de asignación de bases detecta la entrada de 2T.

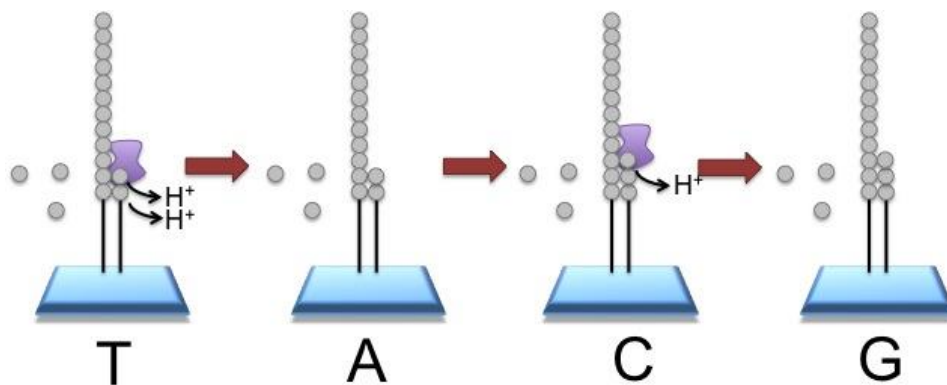


Figura 9: Ejemplo de incorporación de bases en una cadena de ADN [69].

La gran limitación de esta tecnología es la detección de homopolímeros del mismo nucleótido (por ejemplo, TTTTT) ya que si se incorporan múltiples nucleótidos del mismo tipo esto resulta en un mayor cambio de pH y una señal electrónica proporcionalmente mayor. Las señales generadas a partir de un número alto de repeticiones son difíciles de diferenciar de las repeticiones de un número similar pero diferente; por ejemplo, los homopolímeros de longitud 7 son difíciles de diferenciar de los de longitud 8 [70]. En el correr de los años los algoritmos de detección de variantes han ido mejorando y actualmente se detecta cierto ruido en polímeros por arriba de 8 o 9 pb. Los secuenciadores semiconductores *Ion Torrent* producen una longitud de lectura promedio de aproximadamente 400 nucleótidos por lectura.

En la última etapa se realiza el **análisis de datos** de NGS. Este se puede dividir en cuatro operaciones principales: asignación de base (del inglés *base calling*), alineación de lecturas, identificación de variantes y anotación de variantes. Una gran cantidad de programas comerciales, de código abierto y desarrollados por laboratorios están disponibles para cada uno de estos pasos. Es importante tener en cuenta que las principales variantes de secuencia (SNP, MNP INDEL) requieren un enfoque computacional diferente para una identificación específica. La alineación de las lecturas de secuencias que contienen INDELS son técnicamente complejas, y requieren algoritmos diseñados específicamente para la tarea [71, 72].

El sistema *Ion Torrent* incluye un programa que alinea las lecturas generadas contra el genoma humano (hg19) utilizando TMAP (*Torrent Mapping Alignment Program*) generando un archivo BAM [73]. Para la identificación de variantes utiliza el *plug in* TVC (*Torrent Variant Caller*) el cual detecta polimorfismos de un sólo nucleótido (SNP), polimorfismos de múltiples nucleótidos (MNP), inserciones, deleciones y sustituciones de bloque. Además *Ion Torrent* ofrece el *software* "*Torrent Suite*" [74], que actúa como la interfaz principal para una serie de funciones básicas. También disponen de la interfaz "*Ion Reporter*" que sirve para una variedad de soluciones de análisis de código abierto [75].

1.3.1.2. Costos de la tecnología de NGS

La tecnología de NGS surge en un contexto donde las preocupaciones sobre los costos de la atención médica son cada vez mayores, donde el objetivo es poder guiar la prescripción de terapias dirigidas muy costosas. Los grandes avances tecnológicos en NGS han permitido una rápida difusión de las aplicaciones de NGS en entornos clínicos. Sin embargo, aún no está claro si se pueden implementar estas tecnologías en el diagnóstico a un costo accesible [76-78].

Según van Nimwegen *et al* [76], en el año 2016 el costo por muestra de ADN humano para realizar un WGS era de 1669€, 792€ para WES y 333€ para secuenciación dirigida. WES y la secuenciación dirigida son alternativas considerablemente más bajas que WGS. Sin embargo, esto no implica que estos abordajes NGS deban ser preferidos en la práctica clínica, ya que esto debe basarse en el compromiso entre los costos y la utilidad clínica esperada del abordaje elegido.

Un estudio realizado por Marino *et al.* [77] en 2018 muestra que el costo total promedio de NGS para los paneles de genes dirigidos es de aproximadamente 607€ (± 207) para estudios de genética somática y 550€ (± 140) para el análisis en línea germinal. Según este estudio, los consumibles son el controlador de costo más alto del proceso de secuenciación, representando un 48% del costo total en genética somática y el 41% en oncogenética en línea germinal. Los costos de equipo y *software* representan un 8% y 9% para análisis somáticos y de línea germinal, respectivamente, y los costos laborales un 14% y 18%, respectivamente. Finalmente, los costos adicionales fueron entre 30% y 32% del total. Con estos datos queda claro que uno de los mayores costos proviene de la compra de los consumibles utilizados para la secuenciación.

Por otro lado, la utilización de NGS tanto para análisis clínicos como en investigación ha hecho aumentar exponencialmente la cantidad de datos de secuencia generados. Los datos de secuencia que se pueden descargar de las bases de datos son aparentemente gratuitos, pero los costos surgen en la necesidad de recursos de análisis y almacenamiento computacional, así como en la capacitación necesaria para manejar e interpretar los datos. Es por ello que los nuevos paradigmas para el almacenamiento y análisis de datos son cada vez más importantes. Se ha generado una mayor demanda de los recursos computacionales y el conocimiento necesario para manejar los datos de secuencia. Es de vital importancia que a medida que aumenta la cantidad de datos de secuenciación, estos datos no se almacenen simplemente, sino que se organicen de manera escalable, fácil e intuitiva [79].

2. OBJETIVO

2.1. Objetivo general

El objetivo de este trabajo es generar un protocolo alternativo de NGS, utilizando la plataforma *Ion Torrent PGM™*, que sea eficiente, confiable, rápido y con un menor costo que el protocolo original planteado por el proveedor de forma que sea competitivo con Sanger; así como optimizar el análisis de variantes de las secuencias generadas, utilizando como modelo mutaciones somáticas en los oncogenes *KRAS*, *NRAS* y *EGFR* que permiten tomar conductas personalizadas a nivel terapéutico.

2.2. Objetivos específicos

- 2.2.1. Optimizar la extracción y cuantificación de ADN a partir de tejido FFPE.
- 2.2.2. Diseñar un método de trazabilidad de muestras.
- 2.2.3. Generar y poner a punto una alternativa al protocolo del proveedor para las etapas de mayor costo en la secuenciación por NGS utilizando la plataforma *Ion Torrent PGM™*.
- 2.2.4. Validar a nivel diagnóstico el protocolo desarrollado en el objetivo 2.2.3. para la detección de mutaciones somáticas en los oncogenes *KRAS*, *NRAS* y *EGFR*.
- 2.2.5. Comparar el desempeño de diferentes programas bioinformáticos en la detección de mutaciones somáticas en los oncogenes *KRAS*, *NRAS* y *EGFR*.

3. METODOLOGÍA

3.1. Muestras

3.1.1. Muestras clínicas

Se utilizaron 161 muestras provenientes de biopsias de tejidos tumorales (cáncer de pulmón o colorrecto) fijadas en formol e incluidas en parafina, proceso denominado FFPE (*Formalin-fixed paraffin-embedded*). Todas las muestras procedieron de pacientes del laboratorio Genia quienes firmaron un consentimiento informado donde dice que "la misma podrá ser utilizada por dicho laboratorio para validaciones metodológicas u otros fines científicos estadísticos respetando siempre y en todo la normativa aplicable". Ver anexo 8.1.

3.1.2. Controles de calidad

Se utilizó 10 muestras de tejido FFPE para la detección de mutaciones somáticas en el gen *EGFR* provenientes de la organización EMQN, control de calidad del cual participa el Laboratorio Genia. EMQN ofrece esquemas de Evaluación de Calidad Externa (EQA) o Pruebas de Competencia (PT) acreditadas según la norma ISO 17043 que están diseñados para probar todo el proceso analítico de un laboratorio de diagnóstico molecular, incluida la capacidad de interpretar datos a la luz de la información clínica suministrada con una referencia, y para producir un informe claro y preciso [80].

También se utilizó 12 muestras FFPE del control de calidad de la organización UK NEQAS para la detección de mutaciones somáticas en los genes *KRAS* y *NRAS*. UK NEQAS es una organización del Reino Unido que tiene como objetivo garantizar un estándar de EQA adecuado, receptivo y de alta calidad para los laboratorios clínicos [81].

Las muestras de controles de calidad son tejidos FFPE artificiales con más de 50% de células neoplásicas. Todas las muestras traen una indicación clínica del paciente ficticio y datos del mismo.

3.1.3. ADNs con mutaciones conocidas

Se utilizaron ADNs de referencia con mutaciones en los genes *KRAS* y *NRAS* provistos por la empresa *Horizon Diagnostic* (desde aquí referidas como "Horizon"). En la Tabla 1 se

especifican cada uno de los ADN de referencia mediante su número de catálogo, detallando las mutaciones que contienen en forma heterocigota. Estos ADN tienen una concentración de 5ng/ul y la mutación se encuentra en un 50%.

Tabla 1: ADN de referencia con mutaciones en los genes *KRAS* y *NRAS* provistos por la empresa *Horizon Diagnostic*. ADN utilizados en los estudios para la validación de la técnica.

Catálogo	Gen	Mutación
HD378	<i>KRAS</i>	c.34G>A p.Gly12Ser
HD379	<i>KRAS</i>	c.34G>C p.Gly12Arg
HD380	<i>KRAS</i>	c.34G>T p.Gly12Cys
HD374	<i>KRAS</i>	c.35G>A p.Gly12Asp
HD376	<i>KRAS</i>	c.35G>C p.Gly12Ala
HD377	<i>KRAS</i>	c.35G>T p.Gly12Val
HD375	<i>KRAS</i>	c.38G>A p.Gly13Asp
HD632	<i>KRAS</i>	c.183A>C p.Gln61His
HD411	<i>KRAS</i>	c.436G>A p.Ala146Thr
HD373	<i>KRAS</i>	WT
HD190	<i>NRAS</i>	c.35G>T p.Gly12Val
HD539	<i>NRAS</i>	c.181C>A p.Gln61Lys
HD534	<i>NRAS</i>	c.182A>G p.Gln61Arg
HD524	<i>NRAS</i>	c.182A>T p.Gln61Leu
HD529	<i>NRAS</i>	c.183A>T p.Gln61His
HD372	<i>NRAS</i>	WT

3.1.4. ADN de sangre para referencia

Se utilizó una muestra de sangre como control de ADN genómico. La extracción se realizó con el kit *DNeasy® Blood & Tissue* (Qiagen, #69504). Se llevó a una concentración de 0,1ng/ul. El mismo fue denominado con el código "ADN ctrl". Además se utilizaron otras muestras de sangre durante el trabajo que se extrajeron del mismo modo y se llevaron a una concentración de 0,2ng/ul.

3.2. Extracción de ADN de tejido FFPE

Para la extracción de las muestras de tejido FFPE se utilizó el kit *QIAamp DNA FFPE Tissue Kit* (Qiagen, #56404) con modificaciones en el protocolo (se especifican en el punto 4.1.1). Se utilizó aceite mineral (Sigma, #M8410) para la etapa de desparafinado de los tejidos FFPE.

3.3. Cuantificación por PCR en tiempo real (qPCR) utilizando SYBR Green®

El SYBR Green es un colorante fluorescente de uso común que se une a las moléculas de ADN de doble cadena. Cuando el SYBR green se encuentra unido al ADN doble cadena emite fluorescencia. Esto permite que la lectura de fluorescencia al final de cada ciclo sea proporcional al producto de doble cadena formado. De esta forma es posible medir de forma indirecta la cantidad de producto de doble cadena formado al final de cada ciclo en tiempo real. El reactivo Power SYBR® Green PCR Master Mix también incluye una versión patentada del colorante ROX™, una referencia pasiva interna, para normalizar las fluctuaciones de fluorescencia no relacionadas con la PCR y minimizar la variabilidad entre pocillos que puede resultar de varias causas, como el error de pipeteo y la evaporación de la muestra [82-84].

3.3.1. Curva estándar para cuantificación de ADN

Se utilizó un ADN comercial (*Quantifiler™ Human DNA Standard 200ng/ul, Thermo Fisher Scientifics #4343895*) el cual se diluyó un cuarto para obtener una concentración de 50ng/μl. A partir de esa dilución se realizaron diluciones al tercio de forma seriada hasta una concentración de 0,02ng/μl obteniéndose una curva estándar de ADN para su utilización en la técnica qPCR. Los puntos de la curva que se utilizaron fueron 5,56ng/μl, 1,85ng/μl, 0,62ng/μl, 0,21ng/μl, 0,07ng/μl y 0,02ng/μl.

3.3.2. Curva estándar para cuantificación de bibliotecas

Se utilizó una biblioteca control de *E.Coli* DH10B [85] con un tamaño definido de 190bp y una concentración de 68pM, la cual viene en el kit *Ion Library TaqMan™ Quantitation Kit (Thermo Scientific #4468802)*. Se hicieron tres diluciones seriadas al décimo desde la biblioteca pura, obteniéndose: 6,80, 0,68 y 0,068 pM.

3.3.3. Reacción de qPCR

Para la cuantificación de ADN, las muestras extraídas en el punto 3.2 fueron diluidas al medio de forma seriada partiendo de una dilución inicial de 1/5 hasta una dilución 1/160. Los cebadores utilizados fueron hTERT163 Fwd (5'-TGCCCTAGTGGCAGAGACAA-3') y Rv (5'-ATTTGAACAGGCCTCGCACC-3'). El tamaño del amplicón generado es de 163pb. La mezcla de reacción para la qPCR se compuso de *Power SYBR® Green PCR Master Mix 1X (Thermo Fisher Scientifics #4367659)*, cebadores Fwd y Rv a 0,8μM, 2μl de la dilución de la muestra de

ADN y agua ultra pura (*UltraPure™ Distilled Water*, Invitrogen #10977-015) para llegar a un volumen final de 12,5µl. Dicha mezcla se realizó en campana para PCR (con tratamiento de luz UV previo a la manipulación del ADN). Para cada ensayo se utilizó un control negativo (CN) con agua ultra pura y la curva estándar de ADN genómico puro (punto 3.3.1).

Para la cuantificación de bibliotecas se utilizaron las bibliotecas puras, diluidas 1/10 y 1/100. Los cebadores utilizados fueron los cebadores universales que utiliza el Library Equalizer™ Kit para amplificar bibliotecas (*Thermo Fisher Scientific*, #4482298) o los cebadores que trae el Ion AmpliSeq™ Library Kit 2.0 para re-amplificar bibliotecas (*Thermo Fisher Scientific* #4480441) o los cebadores que trae el kit NEBNext® Fast DNA Library Prep Set for Ion Torrent™ (*New England BioLabs® Inc.* #E6270L). La mezcla de reacción para la qPCR se compuso de: *Power SYBR® Green PCR Master Mix 1X* (*Thermo Fisher Scientific* #4367659), cebadores a 0,8µM (cada uno), 2µl de biblioteca y agua ultra pura (*UltraPure™ Distilled Water*, Invitrogen #10977-015) para llegar a un volumen final de 12,5µl. Dicha mezcla se realizó en campana para PCR (con tratamiento de luz UV previo a la manipulación del ADN). Para cada ensayo se amplificó la curva estándar de *E.Coli* DH10B (punto 3.3.2).

Para la qPCR se utilizó el equipo “7500 Real time PCR System” de *Thermo Fisher Scientific*. Las condiciones de ciclado fueron:

50°C 2'; 95°C 10'; [95°C 15'', 60°C 1'*] x40; 95°C 15''; 60°C 1'; 95°C 30'' (Ramp Rate 1%*); 60°C 15''
*Lectura de fluorescencia (Data Collection)

El análisis de los datos se realizó a través del *software HID Real-Time PCR Analysis Software v1.1* (*Thermo Fisher Scientific* #A24664).

3.4. Cuantificación de ADN o bibliotecas por Qubit

Para la cuantificación de ADN o bibliotecas por Qubit se utilizó el kit *Qubit™ dsDNA HS Assay* (*Thermo Fisher Scientific* # Q32854). Para dicha metodología se utilizaron 199ul de buffer (*Qubit® dsDNA HS Buffer*) y 1ul de reactivo colorimétrico (*Qubit® dsDNA HS Reagent*). Luego a 198ul de la solución se le adicionaron 2ul de muestra de ADN o biblioteca a cuantificar. El equipo utilizado fue "*Qubit® 2.0 fluorometer*" de Invitrogen.

Este tipo de cuantificación utiliza una sonda que es altamente selectiva para el ADN doble cadena (ADNdc o dsDNA), incluso a bajas concentraciones. La sonda fluoresce sólo cuando esta unida a su molécula diana y permite medir concentraciones de 10 pg/µL a 100 ng/µL. El ensayo se realiza a temperatura ambiente y la señal es estable durante 3 horas. Los

contaminantes comunes como sales, nucleótidos libres, solventes, detergentes o proteínas son bien tolerados en el ensayo [86].

3.5. Amplificación por PCR

Se realizó la amplificación por PCR de los exones 2, 3 y 4 del gen *KRAS*, exones 2, 3 y 4 del gen *NRAS* y exones 18, 19, 20 y 21 del gen *EGFR*. En conjunto con la amplificación de los exones se realizó la amplificación de microsatélites para evaluar la trazabilidad de las muestras. Se trabajó con dos sub paneles (mezcla de cebadores Fwd y Rv para la amplificación de un conjunto de exones), uno de ellos para evaluar los genes *KRAS* y *NRAS* (panel KN) y el otro para evaluar el gen *EGFR* (panel EGFR). A cada uno de ellos se les incorporaron los cebadores para la amplificación de microsatélites.

3.5.1. Cebadores para *KRAS*, *NRAS* y *EGFR*

Los cebadores se compraron a *Thermo Fisher Scientific* utilizando el servicio que ofrece para realizar cebadores compatibles con *Ion AmpliSeq Library kit 2.0 (Ion AmpliSeq™ Designer)* [87]. La secuencia utilizada fue la misma que plantea el panel *Ion AmpliSeq Colon and Lung Research Panel v2* [88] para la amplificación de los genes *KRAS*, *NRAS* y *EGFR*. El panel KN contuvo a los cebadores: *KRAS_Ex2 Fwd* y *Rv*, *KRAS_Ex3 Fwd* y *Rv*, *KRAS_Ex4 Fwd* y *Rv*, *NRAS_Ex2 Fwd* y *Rv*, *NRAS_Ex3 Fwd* y *Rv* y *NRAS_Ex4 Fwd* y *Rv*. El panel EGFR contuvo a los cebadores: *EGFR_Ex18 Fwd* y *Rv*, *EGFR_Ex19 Fwd* y *Rv*, *EGFR_Ex20 Fwd* y *Rv* y *EGFR_Ex21 Fwd* y *Rv*. En la Tabla 2 se especifican las regiones exónicas evaluadas para los tres genes y el nombre de cada juego de cebadores.

Tabla 2: Regiones evaluadas de los genes *KRAS*, *NRAS* y *EGFR*. Se detalla el nombre del par de cebadores utilizado para cada región, la posición inicial y final de la región evaluada, y el tamaño del amplicón generado (incluyendo cebadores).

Nombre cebador	Nombre cebador en el panel	Chr	Pos. inicial	Pos. final	Tamaño amplicón (pb)
KRAS_Ex2	CHP2_KRAS_1	12	25398186	25398304	172
KRAS_Ex3	CHP2_KRAS_2	12	25380260	25380364	156
KRAS_Ex4	CHP2_KRAS_3	12	25378549	25378658	167
NRAS_Ex2	CHP2_NRAS_1	1	115258689	115258774	134
NRAS_Ex3	CHP2_NRAS_2	1	115256504	115256584	126
NRAS_Ex4	Hotspot_1.17678	1	115252190	115252305	166
EGFR_Ex18	CHP2_EGFR_4	7	55241635	55241729	134
EGFR_Ex19	CHP2_EGFR_5	7	55242411	55242540	172
EGFR_Ex20	CCP_EGFR_6	7	55248974	55249100	162
EGFR_Ex21	CHP2_EGFR_8	7	55259507	55259628	163

3.5.2. Cebadores para amplificar microsatélites

Para la trazabilidad de las muestras se diseñaron cebadores para regiones de los STRs (*Short Tandem Repeat*): TPOX, CSF1PO, TH01, D13S818, D16S539 y D5S818, para ello se utilizó el servicio que ofrece *Thermo Fisher Scientific* para el diseño de cebadores compatibles con el *Ion Ampliseq Library kit 2.0 (Ion AmpliSeq™ Designer)* [87]. En la Tabla 3 se especifican las seis regiones de microsatélites evaluadas y el nombre del par de cebadores utilizado para cada región.

Tabla 3: Regiones de los seis STRs evaluados. Se detalla el nombre del par de cebadores utilizado para cada región, el tipo de repetido, la posición inicial y final cada región evaluada, y el rango de amplificación.

Nombre	Repetido	Chr	Pos. Inicial	Pos. Final	Rango de amplificación (pb)
TPOX	[AATG] _n	2	1493370	1493531	255-302
CSF1PO	[AGAT] _n	5	149455824	149456018	312-360
TH01	[AATG] _n	11	2192234	2192434	150-195
D13S317	[TATC] _n	13	82722058	82722281	168-207
D16S539	[GATA] _n	16	86386212	86386435	258-310
D5S818	[AGAT] _n	5	123111165	123111378	111-157

3.5.3. Reacción de PCR

Se utilizaron dos enzimas *Taq* Polimerasas para la amplificación múltiple por PCR de los exones más los STRs: *5X Ion Ampliseq™ HiFi Mix* del kit *Ion Ampliseq Library kit 2.0 (Thermo Fisher Scientific #4475345)* y la enzima *Platinum® Multiplex PCR Master Mix 2X (Thermo Fisher Scientific #4464269)*.

La mezcla de reacción utilizando la primera enzima (desde ahora también nombrada como "HiFi") se compuso de *Ion Ampliseq™ HiFi Mix 1X*, cebador para exones de 0,2μM (cada uno), cebador para STRs 0,03μM (cada uno), ADN cuantificado por qPCR (0,2ng-2ng de ADN) y agua ultra pura csp 50μl (*UltraPure™ Distilled Water, Invitrogen #10977-015*). El ciclado utilizado para la enzima HiFi fue: 2' 99°C; [15" 99°C; 4' 60°C] 27x.

Los componentes de la mezcla de reacción para la segunda enzima (desde ahora también nombrada como "Platinum") fueron: *Platinum® Multiplex PCR Master Mix 1X (Thermo Fisher Scientific #4464269)*, cebador para exones de 0,2μM (cada uno), cebador para STRs 0,03μM (cada uno), ADN cuantificado por qPCR (0,2ng-2ng de ADN) y agua ultra pura csp 50μl (*UltraPure™ Distilled Water, Invitrogen #10977-015*). El ciclado utilizado para la enzima Platinum fue: 2' 95°C; [30" 95°C; 4' 60°C] 33x.

La temperatura de hibridación de cebadores se consideró en base a la T_m de los cebadores utilizados [89, 90]. Se mantuvo la etapa unificada de hibridación de cebadores-extensión a 60°C en el ciclado con Platinum al igual que lo recomendado para la enzima HiFi.

Para la reacción de PCR se utilizó el termociclador “*Veriti™ 96-Well Thermal Cycler*” de *Thermo Fisher Scientific*s.

3.6. Electroforesis en gel de agarosa

Se observaron los productos de las PCRs por electroforesis en geles de agarosa (*UltraPure™ Agarose, Invitrogen #16500-500*) al 3% en buffer TBE 1X (*UltraPure™ 10X TBE Buffer, Invitrogen #15581-044*). Los ácidos nucleicos se visualizaron mediante *GelRed (GelRed™ Nucleic Acid Gel Stain 10,000X in Water, Biotium #41003)*, el cual se diluyó a una concentración final de 1X. Como marcador de peso molecular se utilizó *O'GeneRuler Ultra Low Range DNA Ladder, Ready-to-Use 10-300 bp (Thermo Scientific #SM1223)*, cargando 0,6µg en el pocillo. Para sembrar las muestras en el gel se usó buffer de carga (*6X Orange DNA Loading Dye, Thermo Scientific #R0631*) a una concentración final de 1X.

La cuba de electroforesis utilizada fue “*ENDURO™ Gel XL Electrophoresis System*” de *Labnet International, Inc.* a un voltaje de 8Volts/cm de gel. Para visualizar los geles se utilizó un transiluminador UV de *Labnet International, Inc.*

3.7. Secuenciación Sanger

Para la secuenciación Sanger de cada uno de los exones se realizaron PCRs independientes para cada gen, por un lado los exones 2, 3 y 4 del gen *KRAS*, por otro lado los exones 2, 3 y 4 del gen *NRAS* y por último los exones 18, 19, 20 y 21 del gen *EGFR*.

Los componentes de la mezcla de reacción fueron: *Platinum® Multiplex PCR Master Mix 2X (Thermo Fisher Scientific #4464269)*, cebadores a 0,26µM (cada uno) (Tabla 4), ADN cuantificado por qPCR (entre 0,2 y 2 ng de ADN) y agua ultra pura csp 12,5µl (*UltraPure™ Distilled Water, Invitrogen #10977-015*). El ciclado utilizado fue: 2' 95°C; [15" 95°C; 30" 60°C; 15" 72°C] x36; 10' 72°C. Para la reacción de PCR se utilizó el termociclador “*Veriti™ 96-Well Thermal Cycler*” de *Thermo Fisher Scientific*s.

Tabla 4: Nombre de los cebadores utilizados en la secuenciación Sanger y tamaño del amplicón generado.

Nombre cebador	Tamaño amplicón (pb)
KRAS ex2 Fwd y Rv	172
KRAS ex3 Fwd y Rv	156
KRAS ex4 Fwd y Rv	167
KRAS ex2 Fwd y Rv	134
KRAS ex3 Fwd y Rv	126
KRAS ex4 Fwd y Rv	166
EGFR ex18 Fwd y Rv	134
EGFR ex19 Fwd y Rv	172
EGFR ex20 Fwd y Rv	162
EGFR ex21 Fwd y Rv	163

Los productos de la PCR se purificaron mediante un tratamiento con 0,26U de fosfatasa alcalina (FastAP Thermosensitive Alkaline Phosphatase, *Thermo Fisher Scientific* 1U/ul # EF0651) y 0,2 U de exonucleasa I (Exonuclease I, *Thermo Fisher Scientific*, # EN0581) sobre 5µl de producto de PCR, en un volumen final de 6,8µl. Esta etapa se realizó para eliminar el exceso de dNTPs y cebadores, respectivamente. La fosfatasa alcalina es una hidrolasa que elimina dos grupos fosfatos en posición 5' de ADN y dNTPs, liberando un Pi. La exonucleasa I (ExoI) cataliza la reacción de degradación de ADN de cadena simple en dirección 3'-5'. Dicha reacción necesitó una incubación de 60' a 37°C y posterior inactivación de las enzimas 15' a 85°C. El producto de la PCR purificado se utilizó para la reacción de secuenciación Sanger que se realizó con el kit *BigDye® Terminator v3.1* (*Thermo Fisher Scientific* #4337455), los cebadores a 10µM, 1µl del producto de PCR purificado y agua ultra pura (UltraPure™ Distilled Water, Invitrogen #10977-015) para llegar a un volumen de 6µl. Para la reacción de secuenciación Sanger se utilizó el termociclador "VERITI®96" de Applied Biosystems. Las condiciones de ciclado fueron: 3' 95°C; [15" 94°C; 10" 56°C; 4' 72°C] x28; 4' 72°C.

El volumen total de los productos de la secuenciación se purificó BigDye XTerminator® Purification Kit (*Thermo Fisher Scientific* #4376484) siguiendo las recomendaciones del proveedor. El producto purificado se corrió en el secuenciador capilar ABI3500xL Genetic Analyzer (*Thermo Fisher Scientific*).

Los resultados de la secuenciación Sanger se analizaron con el software *MEGA 6.0* (Molecular Evolutionary Genetics Analysis) [91].

3.8. Microsatélites por electroforesis capilar

Para evaluar la trazabilidad de las muestras analizadas, se realizó la amplificación y posterior electroforesis capilar de los STRs: TPOX, CSF1PO, TH01, D13S818, D16S539, D5S818 y amelogenina.

3.8.1. Amplificación de STRs

Se utilizaron cebadores marcados, previamente diseñados por el Laboratorio Genia (el diseño no se especifica), para la amplificación de STRs (Tabla 5). La mezcla de cebadores se equilibró para tener señal similar en todos los marcadores.

Tabla 5: Cebadores marcados utilizados en la amplificación de los STRs para la posterior electroforesis capilar.

Nombre	Fluorocromo
hTPOX_F	OH
hTPOX_R	NED
hCSF1PO_F	VIC
hCSF1PO_R	OH
hTH01_F	6-FAM
hTH01_R	OH
hD13S317_F	OH
hd13S317_R	VIC
hD16S539_F	OH
hD16S539_R	VIC
hD5S818_F	OH
hD5S818_R	VIC
hAmelogenin_F	NED
hAmelogenin_R	OH

Los componentes de la mezcla de reacción para la PCR de STRs fueron: *Platinum® Multiplex PCR Master Mix 1X (Thermo Fisher Scientifics #4464269)*, cebadores a 0,4µM (cada uno) (Tabla 5), 0,2ng de ADN y agua ultra pura (*UltraPure™ Distilled Water, Invitrogen #10977-015*) para llegar a un volumen de 15µl. La misma fue realizada en campana para PCR con tratamiento de luz UV previo a la manipulación del ADN.

Para la reacción de PCR se utilizó el termociclador "Veriti™ 96-Well Thermal Cycler" de *Applied Biosystems*. El ciclado utilizado fue: 2' 95°C; [10'' 95°C; 1' 60°C, 20'' 60°C] 34x.

3.8.2. Electroforesis capilar

Los productos de la PCR de STRs se corrieron en el secuenciador capilar "3500xL Genetic Analyzer for Human Identification" de Applied Biosystems. Para ello, se mezcló 9µl de Hi-Di™ Formamide (Applied Biosystems #4311320) con 0,5µl de escalera interna GeneScan 600LIZ Size standard (Applied Biosystems # 4408399). Tras vortexear y centrifugar, se alicuotaron 9µl de dicha mezcla en una placa apta para el secuenciador capilar (MicroAmp Optical 96-well Reaction Plate With Barcode (Applied Biosystems #4306737) y se le agregó 1µl del producto de PCR en cada pocillo.

Los datos generados en la electroforesis capilar se analizaron por el software GeneMapper® Software 5 (Thermo Fisher Scientifics #4475073).

3.9. Construcción y cuantificación de bibliotecas

3.9.1. Protocolo Ampliseq

La construcción de bibliotecas con el Ion AmpliSeq™ Library Kit 2.0 (desde aquí en adelante nombrado bibliotecas Ampliseq) se realizó siguiendo la guía de uso "Ion AmpliSeq™ Library Kit 2.0 USER GUIDE" [92]. Se utilizó el kit Ion AmpliSeq™ Library Kit 2.0-96LV (Thermo Scientific #4480441), los adaptadores con barcode Ion Xpress™ Barcode adapter 1-96 (Thermo Scientific #4474517) y el sistema de purificación con Agencourt AMPure XP (Beckman Coulter #A63882).

Los ADNs extraídos y cuantificados, se amplificaron utilizando la enzima 5X Ion Ampliseq™ HiFi Mix del kit Ampliseq Library 2.0 como se indica en el apartado 3.5.3. Los productos de la PCR se digirieron con la enzima FuPA presente en el kit Ion AmpliSeq™ Library Kit 2.0-96LV incubando en termociclador a 50°C 10', 55°C 10' y 60°C 20'. Luego, se realizó la ligación de los adaptadores con barcode utilizando las 96 opciones del Ion Xpress™ Barcode adapter 1-96. Para ello se incubó a 22°C 30'y 72°C 10'. Las bibliotecas generadas se purificaron utilizando el sistema de purificación por perlas magnéticas de Agencourt AMPure XP a 1,5X realizando dos lavados de las perlas con etanol 70% (Etanol Absoluto, Dorwil #D010-03-03) y se eluyeron en 20µl de agua ultrapura (UltraPure™ Distilled Water, Invitrogen #10977-015). Las bibliotecas purificadas se cuantificaron utilizando el kit Ion Library TaqMan™ Quantitation Kit (Thermo Scientific #4468802) siguiendo la guía de uso "Ion AmpliSeq™ Library Kit 2.0 USER GUIDE" [92].

3.9.2. Protocolo NEB adaptado

Para la construcción de bibliotecas con el protocolo adaptado en este trabajo (desde aquí en adelante nombrado bibliotecas NEB), se utilizó el kit de *NEBNext® Fast DNA Library Prep Set for Ion Torrent™* (New England BioLabs® Inc. #E6270L), los adaptadores con *barcode Ion Xpress™ Barcode adapter 1-96* (Thermo Scientific #4474517) y el sistema de purificación con *Agencourt AMPure XP* (Beckman Coulter #A63882).

Los ADNs extraídos y cuantificados, se amplificaron utilizando la enzima *Platinum® Multiplex PCR Master Mix 2X* como se indica en el apartado 3.5.3. Los productos de PCR se purificaron utilizando el sistema de purificación por perlas magnéticas de *Agencourt AMPure XP* a 1,8X realizando dos lavados de las perlas con etanol 70% (Etanol Absoluto, Dorwil #D010-03-03) y se eluyeron en 20µl de agua ultrapura (*UltraPure™ Distilled Water, Invitrogen #10977-015*). Para las siguientes etapas de construcción de bibliotecas se utilizó el kit *NEBNext® Fast DNA Library Prep Set for Ion Torrent™* siguiendo las recomendaciones del manual de uso "*NEBNext® Fast DNA Library Prep Set for Ion Torrent™ Instruction Manual*" [93]. A los productos de PCR purificados (50ng) se les realizó la etapa de reparación de los extremos y fosforilación 5' en una reacción con *NEBNext End Repair Reaction Buffer*, *NEBNext End Repair Enzyme Mix* (con el fin de mantener el buffer requerido en esta etapa y que puede influir en la siguiente) y agua ultra pura para llegar a un volumen de 15µl. La incubación para esta etapa fue de 20' a 25°C y 10' a 70°C. Luego, se realizó la ligación de los adaptadores utilizando *T4 DNA Ligase Buffer*, *Bst 2.0 WarmStart DNA Polymerase*, T4 DNA Ligase, 0,5µl del adaptador con barcode correspondiente, 15µl del producto de la reacción de *END Repair* y agua suficiente para completar 25µl de volumen final. La incubación para esta etapa fue de 15' a 25°C y 5' a 65°C. Los adaptadores no están fosforilados en el extremo 5' lo que minimiza la formación de dímeros de adaptadores. Dado que las ligasas requieren de un fosfato en 5' para la ligación, la ausencia de fosfato en el extremo 5' de los adaptadores resulta en una discontinuidad en una de las cadenas de la doble hebra tras la ligación al producto de PCR (en inglés: *nick DNA*). Esta discontinuidad se repara utilizando *Bst 2.0 WarmStart™ DNA Polymerase* [94]. El producto de esta etapa es la biblioteca. La misma se purifica utilizando *Agencourt AMPure XP 1,5X* y se eluye en 20µl de agua ultrapura. Las bibliotecas purificadas se cuantificaron utilizando *Qubit* (ver apartado 3.4).

3.10. Amplificación clonal, enriquecimiento y secuenciación en *Ion Torrent PGM™*

Estas etapas se realizaron siguiendo la guía de uso del proveedor "*Ion PGM™ Hi-Q™ View OT2 Kit*" [95], para lo cual se utilizó el kit *Ion PGM™ OT2 Supplies* (Thermo Fisher Scientific

#A27744), el kit *Ion PGM™ Hi-Q™ View OT2 Reagents* (Thermo Fisher Scientific #A29811) y el kit *Ion PGM™ Hi-Q™ OT2 Solutions* (Thermo Fisher Scientific #A27742). Para la secuenciación se utilizaron los *chip Ion 314™ Chip Kit v2 BC* (Thermo Fisher Scientific #4488144) o *Ion 316™ Chip Kit v2 BC 9* (Thermo Fisher Scientific #4488149) y el kit de secuenciación utilizado fue *Ion PGM™ Hi-Q™ Sequencing Kit* (Thermo Fisher Scientific #A25592, el cual contiene *Ion PGM HiQ Seq dNTP*, *Ion PGM HiQ Seq Reagent*, *Ion PGM Hi-Q Seq Solution* y *Ion PGM Seq Supplies*).

El secuenciador utilizado fue: *Ion Personal Genome Machine™* (PGM™) *System* de Thermo Fisher Scientific y el número de flujos de dNTPs utilizado fue de 500.

3.11. Análisis bioinformático

Las secuencias generadas por el *Ion Torrent PGM™* se alinearon contra el genoma de referencia hg19 (*Assembly GRCh37* disponible en la base de datos NCBI). La ubicación cromosómica de los genes evaluados es la siguiente: *EGFR* en 7p12, *KRAS* en 12p12.1 y *NRAS* en 1p13.2. Las secuencias de referencia (*GenBank*) utilizadas fueron: NM_005228.4 para *EGFR*, NM_004985.3 para *KRAS* y NM_002524.4 para *NRAS*.

Los programas y archivos utilizados para el análisis de secuencias fueron los siguientes:

1. Alineamiento de secuencias contra el genoma humano (hg19):

- *Torrent Suite™ software mediante el programa TMAP (Torrent Mapping Alignment Program, versión 5.10, Thermo Fisher Scientific) [74]*
- *CLC Genomics Workbench 12 software (versión de prueba, Qiagen) [96]*

2. Detección de variantes:

- *IT Variant Caller (Thermo Fisher Scientific)[97]*
- *Sophia DDM® (Sophia Genetics) [98]*

Utilizando los archivos:

- *Archivo “ColonLungV2.20140523.designed.bed” (todas las variantes)*
- *Archivo “ColonLungV2.20150730.hotspot.bed” (variantes hotspot)*

4. Análisis de los microsatélites en NGS:

- *HID_STR_Genotyper [99].*

5. Visualización del alineamiento de las lecturas y la presencia de variantes contra el genoma de referencia:

- *Integrative genome viewer (IGV) (Broad Institute) [100]*

Clasificación de variantes:

- *Software Ingenuity Pathway Analysis (Qiagen) [101]*
- *Base de datos COSMIC (Sanger Institute Catalogue of Somatic Mutations in Cancer) [102]*

Nomenclatura de las variantes encontradas:

- *HGVS (Human Genome Variation Society)[103]*

4. RESULTADOS Y DISCUSIÓN

El objetivo principal de este trabajo fue implementar una alternativa al protocolo de NGS (utilizando la plataforma *Ion Torrent PGM™*) planteado por el proveedor, *Thermo Fisher Scientific*, que presentara una reducción de costos. La aplicación de esta alternativa se planteó para la detección de mutaciones somáticas en los oncogenes *KRAS*, *NRAS* y *EGFR*, pudiéndose extender a otros estudios. Este trabajo se realizó en el Laboratorio Genia con los materiales y equipos disponibles en el mismo, lo cual fue una condicionante a la hora de explorar diferentes opciones técnicas.

4.1. Tejido tumoral fijado con formaldehído y embebido en parafina (FFPE)

Para el análisis de variantes somáticas en tumores de colorrecto y pulmón, se debe analizar el ADN proveniente de tejido tumoral. La mejor fuente de ADN para este tipo de estudios es el tejido fresco congelado ya que a partir de él se logra obtener una buena cantidad de ADN para ser analizado. Sin embargo, la mayoría de las muestras clínicas se conservan en bloques de parafina para el examen histomorfológico y para preservar el tejido por más tiempo. Este modo de preservación del tejido tumoral conocido como FFPE, implica que el tejido sea fijado con formaldehído y embebido en parafina, lo cual permite preservar la estructura del mismo por medio de enlaces cruzados entre las proteínas [104]. Los tejidos FFPE resultan la principal fuente de tejido para los estudios moleculares.

Como consecuencia de la fijación del tejido con formaldehído, el ADN extraído de muestras FFPE es de mala calidad. El ADN se encuentra normalmente fragmentado, parcialmente desnaturizado y con presencia de enlaces cruzados entre el ADN y las proteínas, o entre ADN y ADN. A su vez, durante la fijación, se producen modificaciones químicas de las bases del ADN, como la depuración y la desaminación de las citosinas, lo que lleva a cambios de nucleótidos del tipo G>A o C>T [105]. Es por este motivo que el tejido FFPE se considera un material de partida difícil de trabajar para la mayoría de las técnicas moleculares debido a la calidad generalmente baja del ADN y ARN extraídos [105]. La mala calidad del ADN extraído se agrava por múltiples factores durante el proceso de fijación como lo es el tiempo entre la adquisición del tejido y el inicio de la fijación, la composición y el pH de los reactivos utilizados, el tiempo y la temperatura de la fijación y, el tiempo y condiciones de almacenamiento. No existe un método estándar de fijación del tejido, y por lo tanto, el material disponible para los estudios moleculares puede variar mucho en su calidad (debido a las condiciones de manejo del tejido antes, durante y después de la fijación).

En este proyecto se trabajó con muestras de tejido FFPE provenientes de pacientes con cáncer de colorrecto y pulmón para la detección de mutaciones somáticas en los oncogenes *KRAS*, *NRAS* y *EGFR*. Debido a la dificultad que presenta trabajar con las muestras de tejido FFPE fue importante utilizar un método de extracción adecuado que permita obtener suficiente ADN para el protocolo de NGS. Por ello se trabajó sobre mejoras del método de extracción y se utilizó un método de cuantificación sensible que diera una idea del nivel de degradación del ADN. Además se ideó un método de trazabilidad útil para seguir la identidad de las muestras hasta la obtención del resultado final (que permitiera evidenciar intercambio de muestras en el caso de que ocurriera).

4.1.1. Extracción de tejido FFPE

La calidad del ADN obtenido de tejido FFPE depende del método de extracción elegido y de todos los factores del tejido FFPE previamente mencionados. Dependiendo de los requisitos de la aplicación molecular posterior en términos de integridad, pureza y cantidad del ADN, un método de extracción puede ser elegido entre varios métodos que van desde los protocolos internos hasta los kits comerciales. En el laboratorio Genia se ha utilizado por varios años el kit *QIAamp DNA FFPE Tissue Kit de Qiagen* para la extracción de tejidos FFPE para su uso en protocolos de diagnóstico tumoral [106]. Es por este motivo que se comenzó por evaluar este kit como método de extracción para utilizar en este trabajo.

Al revisar la bibliografía, se constató que muchos laboratorios han invertido esfuerzos en optimizar protocolos de extracción de ADN para tejidos FFPE [105, 107, 108]. La Figura 10 muestra la comparación de cinco kits comerciales en cuanto al tiempo de extracción, rendimiento de la extracción de ADN y eficiencia de la PCR cuando se utilizan los ADNs extraídos. Uno de los kit con mayor eficiencia según los datos de Muñoz-Cadavid *et al* (2010) (Figura 10) es el kit *QIAamp DNA FFPE Tissue Kit* de Qiagen, el cual tiene un rendimiento similar al kit de TaKaRa, tanto en el rendimiento de la extracción como en la eficiencia de la PCR posterior [107].

Method	Time (h:s)	DNA quality (%) ^a	PCR efficiency (%) ^b
QIAamp DNA FFPE Tissue Kit (Qiagen)	6:00	89 (72/81)	90.3 (65/72)
TaKaRa Dexpat (Takara Bio Inc.)	2:30	91.4 (74/81)	93.2 (69/74)
PureLink Genomic DNA Mini Kit (Invitrogen)	6:00	83 (67/81)	83.6 (56/67)
WaxFree DNA (TrimGen Genetic Diagnostics)	5:00	79 (64/81)	57.8 (37/64)
QuickExtract FFPE DNA Extraction Kit (Epicenter Biotechnologies)	3:50	60.5 (49/81)	59.2 (29/49)

^a Quality of DNA detection (calculated as the number of housekeeping genes PCR positive divided by the total number of samples tested).

^b Efficiency of detection of fungal DNA (calculated as the number of samples positive by any panfungal PCR assay divided by the number of samples housekeeping gene PCR positive).

Figura 10: Comparación de cinco kits comerciales de extracción de tejido FFPE en tiempo de realización, calidad de ADN (%) y eficiencia del ADN extraído en la PCR (%). Figura extraída de C. Muñoz-Cadavid *et al.* (2010).

En otro estudio donde se valoran diferentes técnicas de extracción (Janecka *et al* (2015)), comparan ocho kits comerciales para la extracción de ADN de tejido FFPE y concluyen que uno de los kits más adecuados para el aislamiento de ADN de alta calidad y cantidad aceptable es *QIAamp DNA FFPE Tissue Kit* de Qiagen. En esta comparación, se evalúan distintos parámetros: rendimiento, pureza, eficiencia de la PCR y aspectos técnicos de los distintos procedimientos. En cuanto al rendimiento existe un kit basado en una tecnología de extracción por perlas magnéticas (*MagMAX FFPE DNA Isolation Kit from Ambion*) que es el que logra mayor cantidad de ADN extraído (9,65µg), el resto va desde 0,97 a 6,20µg. En lo que refiere a la pureza de las muestras extraídas, sólo el kit de Qiagen permite el aislamiento de ADN puro, disminuyendo al máximo el nivel de contaminación por proteínas (midiendo la relación A_{260}/A_{280}). Con respecto a la calidad de ADN para la PCR posterior, Qiagen es uno de los kit que logra mejores resultados [108].

Con estos datos se puede concluir que el kit de Qiagen está dentro de los mejores kits de extracción de tejidos FFPE disponibles actualmente. Por este motivo, se decidió continuar trabajando con el kit y se realizaron modificaciones para optimizar su uso en las instalaciones del Laboratorio Genia, en base a datos encontrados en la bibliografía.

El protocolo *QIAamp DNA FFPE Tissue* de Qiagen consiste en tres etapas fundamentales: desparafinado del tejido FFPE con Xylol, disgregación de las membranas con Proteinasa K y recuperación del ADN mediante columnas con membrana de sílice. Para adaptar el protocolo al Laboratorio Genia, se identificaron tres puntos sobre los cuales intervenir. El primero de ellos fue la **utilización de Xylol** [109], un producto inflamable de alta toxicidad que afecta al sistema nervioso central. Este producto se utiliza al comienzo de la extracción y sirve para desparafinar y liberar el tejido. Considerando que en el laboratorio no se cuenta con una campana de extracción de gases, se buscó la manera de eliminar la etapa con Xylol (y su posterior limpieza con etanol 100%) para sustituirla por otra forma de desparafinar el tejido FFPE. Para ello se probaron dos alternativas: **(a) eliminación mecánica de la parafina**. El objetivo de este cambio fue eliminar el uso de un compuesto químico para quitar la parafina. Se probó calentar a 60°C por 20 minutos las láminas de vidrio conteniendo los cortes de tejido parafinado, colocándolas sobre un termobloque previamente precalentado (Cardoso 2016, comunicación personal) [110-112]. Con esto se logró fundir la parafina y luego se traspasó el tejido tumoral, raspando el contenido con un bisturí desde la lámina al tubo. Posteriormente se continuó con el protocolo del proveedor en la etapa de disgregación de las membranas con Proteinasa K. Este método tuvo varios inconvenientes, uno de ellos fue en la etapa de unión del ADN a las columnas ya que se bloquearon las columnas, teniendo que realizar centrifugaciones por más tiempo para que todo el líquido pasara a través de las mismas. Por otro lado, al comparar el rendimiento de este método con el tradicional según estudio por PCR en tiempo real, se observó una disminución del mismo. Esta baja, se atribuyó a la ineficiencia del traspaso del tejido al tubo. Por lo que se decidió descartar esta opción dada la dificultad en la manipulación,

el riesgo de pérdida de material y la complejidad de trabajar con muchas muestras en paralelo (algo habitual en los laboratorios de diagnóstico). **(b) Sustitución del Xylol por aceite mineral** [113, 114]. En este caso se combinó la utilización de un compuesto orgánico con calor para desparafinar el tejido FFPE. La prueba consistió en traspasar el tejido parafinado (de igual modo que se indica en el protocolo de Qiagen [106]) a un tubo conteniendo aceite mineral. Para desparafinar se vortexeó el tubo por un minuto y se probó calentar a dos temperaturas diferentes: 60°C y 90°C por 30 minutos. En ambos casos se logró liberar el tejido y la parafina quedó disuelta en el aceite mineral. Se continuó con el protocolo del proveedor en la etapa de disgregación de las membranas con Proteinasa K. Tras esta etapa el ADN quedó en la fase acuosa. La comparación del rendimiento de este método con el tradicional mediante PCR en tiempo real mostró ser equivalente, con aceite mineral se extrajo buena cantidad de ADN al igual que con la extracción utilizando Xylol. Además el aceite mineral eliminó eficientemente la parafina, independientemente de que la temperatura de incubación fuera 60°C o 90°C, por lo que se decidió incorporar este cambio al protocolo con una incubación a 60°C por 30 minutos.

Otro de los puntos que se evaluó en el proceso de extracción fue el rendimiento del kit de Qiagen. A pesar de que el rendimiento del kit es bueno comparado con otros métodos (ver Figura 10), fue necesario optimizar el rendimiento según las condiciones de trabajo para compensar la fragmentación del ADN en las muestras FFPE. Con el fin de aumentar el rendimiento se probó realizar la extracción partiendo de distintos tamaños de región tumoral (Figura 11) esperando obtener una concentración de ADN de 0,02ng/μl como mínimo (0,2ng en 10μl de ADN extraído, que es el equivalente a 54 genomas aproximadamente) [115].

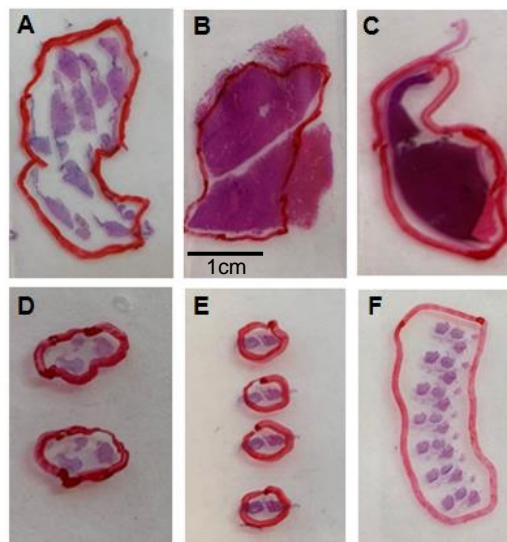


Figura 11: Ejemplo de láminas de tejido FFPE con diferentes áreas de región tumoral (marcadas con línea roja) que llegan al Laboratorio Genia. Las láminas A, B y C corresponden a tejido tumoral de CCR, y las láminas D, E y F corresponden a tejido tumoral de pulmón.

Siguiendo las recomendaciones del protocolo de Qiagen [106] en donde se indica utilizar un mínimo de 3 secciones (0,9 cm²) cada una con un grosor de hasta 10 µm, se probaron extracciones de una misma muestra ensayando diferentes áreas de tejido tumoral. Según la disponibilidad de tejido de la muestra que llega al Laboratorio Genia se comparó la extracción de 1, 2 y 4 cm². Se realizó la prueba para 3 muestras y se cuantificaron las mismas por PCR en tiempo real (como lo indica el apartado 3.3). Los resultados (Tabla 6) mostraron que a partir de 2cm² la extracción presentó un rendimiento adecuado para continuar con el protocolo de trabajo (0,2ng de ADN). Con 4cm² el rendimiento fue superior pero es importante destacar que una gran parte de las muestras que llegan para ser analizadas no cuentan con un área tumoral tan grande, sobre todo las muestras de tejido de pulmón, que normalmente son pequeñas (ver Figura 11).

Tabla 6: Comparación de la extracción de ADN a partir de diferentes áreas de tejido tumoral FFPE para tres muestras. El valor de la cuantificación corresponde a la dilución seleccionada (primera dilución sin inhibición).

Muestra	ng/µl de ADN para 1cm ²	ng/µl de ADN para 2cm ²	ng/µl de ADN para 4cm ²
FFPE_1	0,020	0,47	0,91
FFPE_2	0,008	0,28	1,05
FFPE_3	0,016	0,34	1,52

Una vez que se estableció el área mínima de tejido para obtener un rendimiento adecuado en la extracción se decidió buscar una solución al problema de los inhibidores de la PCR presentes en las muestras FFPE. La naturaleza de los inhibidores no se conoce exactamente, pero se sabe que están asociados al proceso de fijación y parafinado de las muestras. Con el fin de mejorar la calidad del ADN extraído se realizaron diferentes pruebas de lavados con etanol intentando eliminar el inhibidor sin reducir el rendimiento de la extracción de ADN. Tras realizar varias pruebas se vio que el etanol no eliminaba el inhibidor ya que el mismo acompañaba al ADN en la elución de la columna, evidenciado por la persistencia de la inhibición de la PCR tras realizar los lavados. Ninguna de las pruebas permitió reducir el inhibidor sin reducir la cantidad de ADN extraído. Se desconoce la procedencia del inhibidor presente en las muestras FFPE pero con las pruebas antes mencionadas se dedujo que el/los inhibidor/es son hidrosolubles. Un posible origen de la inhibición puede ser la presencia de un exceso de ADN fragmentado de pequeño tamaño que co-purifique con el ADN de mayor tamaño que sirve para amplificar.

En resumen, el protocolo de extracción de ADN de tejido FFPE utilizado en este trabajo fue el *QIAamp DNA FFPE Tissue Kit* al cual se le hizo una modificación sustancial sustituyendo el Xylol por aceite mineral. Este cambio permitió trabajar sin campanas de extracción de gases, disminuyendo el riesgo del operario por exposición a un compuesto tóxico. A su vez, el uso de aceite mineral permitió eliminar los lavados con etanol que se hacen posterior a la utilización

del Xylol (para lavar dicho compuesto). Para mejorar el rendimiento de la extracción se seleccionó una área mínima de tejido tumoral a ser extraído (2cm²), en aquellos casos en donde no se alcance dicho límite, se debería solicitar nuevos cortes de tejido. En cuanto a la inhibición presente en algunas muestras de ADN extraídas de tejido FFPE, no se encontró una solución que eliminara el inhibidor sin perder ADN.

4.1.2. Evaluación de la extracción de ADN de tejido FFPE mediante PCR en tiempo real

El ADN extraído de muestras provenientes de tejido FFPE, debió ser evaluado previo a ser utilizado en el proceso de NGS. Esta etapa es de suma importancia ya que es necesario saber la cantidad y calidad del ADN que se obtiene luego de la extracción. Según estudios reportados previamente [116], el tipo de tecnología utilizada en la cuantificación del ADN condiciona el resultado de la evaluación de la calidad del ADN extraído de tejido FFPE. El laboratorio Genia cuenta con dos metodologías de cuantificación, PCR en tiempo real (que utiliza el equipo “7500 Real time PCR System” de *Thermo Fisher Scientific*) y Qubit® (que utiliza el kit *Qubit™ dsDNA HS Assay* de *Thermo Fisher Scientific*). Ambos métodos son buenos para determinar la cantidad de ADN extraído pero la PCR en tiempo real (qPCR), además, permite detectar inhibidores de la PCR. Debido a la ventaja que presenta la qPCR se decidió utilizar esta técnica para evaluar las extracciones de ADN, utilizando el reactivo *Power SYBR® Green PCR Master Mix* [82].

Los cebadores utilizados para la cuantificación fueron diseñados respetando ciertos requisitos: amplificación de un único amplicón, única temperatura de disociación (T_m), alta eficiencia, y que amplificaran un producto con tamaño similar al producto de las reacciones de PCR que se hacen posteriormente para el diagnóstico. Se diseñaron *in silico* el par de cebadores hTERT163 [117] que amplifican un fragmento de 163pb del gen de la telomerasa transcriptasa inversa humana (de copia única en el genoma). El tamaño del fragmento amplificado fue de suma importancia ya que las muestras provenientes de tejidos FFPE poseen ADN muy degradado. Por lo tanto, para poder cuantificar la cantidad de “ADN amplificable”, o sea el ADN posible de ser utilizado en la etapa posterior (amplificación de genes de interés: *KRAS*, *NRAS* y *EGFR*), con un tamaño entre 100 y 200pb, se necesitaron cebadores capaces de amplificar fragmentos de un tamaño similar. Cebadores que amplificasen fragmentos de menor tamaño sobre-estimarían la cantidad de ADN “amplificable” [71].

Un punto clave en la evaluación del ADN de tejido parafinado fue detectar posible inhibición en la qPCR. Dicha inhibición puede ser debida a la presencia de inhibidores, posiblemente productos extraídos conjuntamente con el ADN en la etapa de extracción. Los mismos podrían actuar capturando ácidos nucleicos, inhibiendo la polimerasa o interfiriendo con las condiciones

físico-químicas de la reacción de qPCR. Para ello, se realizaron diluciones seriadas al medio de cada muestra, desde una dilución 1/5 a una dilución 1/160. A medida que disminuye la concentración de ADN al medio, debe aumentar el C_T en un valor de 1 entre una dilución y la siguiente. Cuando hay inhibición esta relación no se cumple, por esto se consideró no inhibidas aquellas diluciones de la muestra en la que se cumplía que la relación entre su C_T y el de la dilución siguiente fuera de 1. En la Figura 12 se presenta un ejemplo de una muestra FFPE en donde se observa inhibición en las primeras diluciones. Recién entre la dilución 1/40 y la dilución 1/80 se vio una diferencia de 1 ciclo, permitiendo seleccionar la dilución 1/40 (0,9ng/ μ l) para su uso en las etapas posteriores.

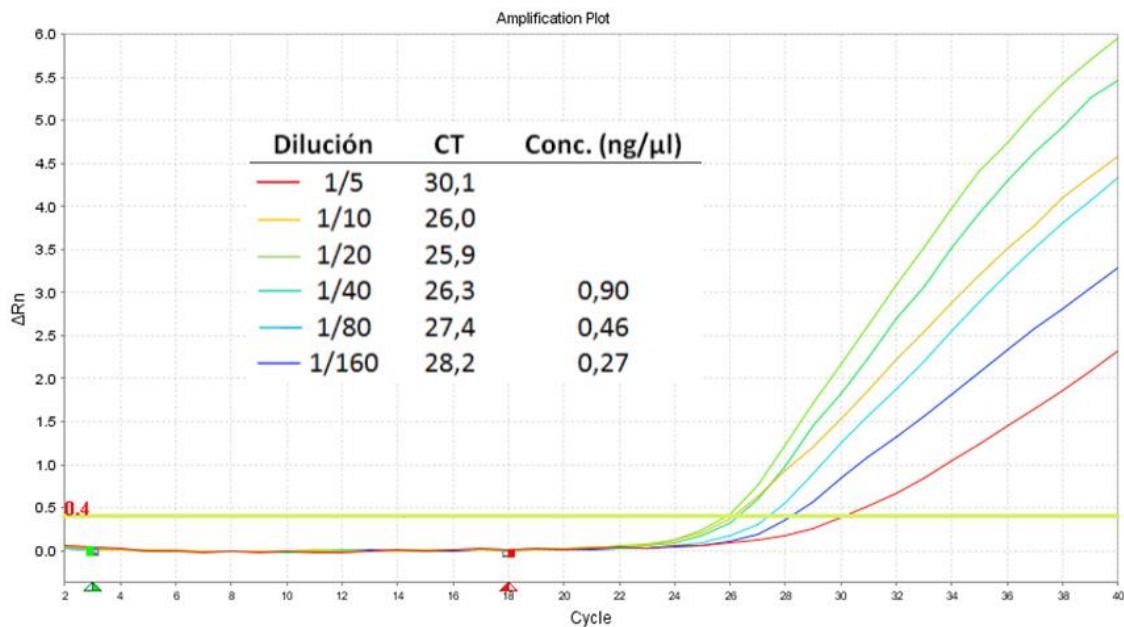


Figura 12: Gráfica ΔR_n vs. ciclos de muestra de ADN proveniente de tejido FFPE que presenta inhibición en las primeras diluciones.

Además de evaluar la inhibición para seleccionar la dilución correcta de la muestra, fue importante tener en cuenta la concentración mínima de ADN necesaria en este estudio. El límite de concentración establecido en este trabajo para detectar una mutación heterocigota somática fue de 0,02ng/ μ l si se realiza una PCR con 10 μ l de ADN (0,2ng de "ADN amplificable"), lo que corresponde a 54 copias del genoma [115]. Por lo tanto la elección de la dilución tuvo que tener en cuenta tanto el nivel de inhibición como la concentración de ADN, manejando un equilibrio en donde la inhibición fuera baja sin perder de vista el límite mínimo de "ADN amplificable" si el ADN se encuentra degradado.

En el contexto de un laboratorio de diagnóstico clínico como es el caso del Laboratorio Genia, aquellas muestras con menor cantidad de ADN que el límite mínimo establecido se deben re-

extraer y volver a evaluar el ADN, y si no se consigue una mejora, se deben pedir nuevos cortes del taco original marcados por el anatomopatólogo. En el caso de no tener suficiente tejido tumoral, se rechaza la muestra y no se puede emitir un diagnóstico.

4.1.3. Evaluación del rendimiento de las extracciones de ADN en muestras de FFPE utilizando el protocolo de Qiagen adaptado

Una vez extraídas 179 muestras de tejido FFPE utilizando el protocolo de Qiagen adaptado, se realizó una evaluación del desempeño de la extracción de ADN. Para ello se hizo un análisis diferencial de las muestras según distintas características: grupo A, 36 muestras provinieron de cortes de tejido FFPE fijados recientemente (menos de un mes de fijación); grupo B, 125 muestras tenían un tiempo de fijación desconocido (variable); y grupo C, 18 muestras provinieron de controles de calidad (ver apartado 3.1.2).

En el caso de las muestras recientemente fijadas (grupo A), se logró extraer una cantidad de ADN mayor o igual a 0,2ng en 10µl finales en el 97% de los casos (sólo se rechazó una muestra). Estos resultados coincidieron con lo reportado en la bibliografía para tejidos fijados recientemente ya que a menor tiempo de fijación y almacenamiento del material FFPE mejor es la calidad y cantidad del ADN extraído [105, 107, 118, 119].

En el grupo B los resultados fueron más variados, hubo un 13% de rechazos de muestras por tener menos de 50 copias de genoma (menos de 0,2ng en 10µl) distribuyéndose el resto en varios rangos de cantidad de ADN. Un 50% de las muestras tuvieron más de 500 copias. Las grandes variaciones en la cantidad de ADN de estas muestras se pudieron deber a los diferentes procesos de fijación que tuvieron y del tiempo de almacenamiento (desconocido).

En lo que refiere al grupo C se obtuvo buena cantidad de ADN en todas las muestras. No hubo ningún rechazo de muestras. Por el contrario, un 70% de las extracciones tuvieron más de 500 copias de genoma (o sea, más de 1,86ng de ADN en 10µl). Esto, probablemente, fue debido a que los tejidos de los controles de calidad son tejidos FFPE artificiales y con gran cantidad de células tumorales.

A continuación, se realizó un análisis global de las mismas. Si se observa la Figura 13 donde se representa a través de un histograma el porcentaje de muestras extraídas distribuidas por rangos de número de copias de genoma, un 9,5% de las muestras tuvieron menos de 50 copias (menos de 0,2ng en 10µl) y se rechazaron, mientras que el 90,5% restante fue viable para su utilización en el diagnóstico. De las muestras aceptables (mayor o igual a 0,2ng en

10µl) se pudieron observar dos rangos de copias de genoma, las que tuvieron entre 50 y 499 copias (35,8%), y las que tuvieron de 500 copias en adelante (54,7%).

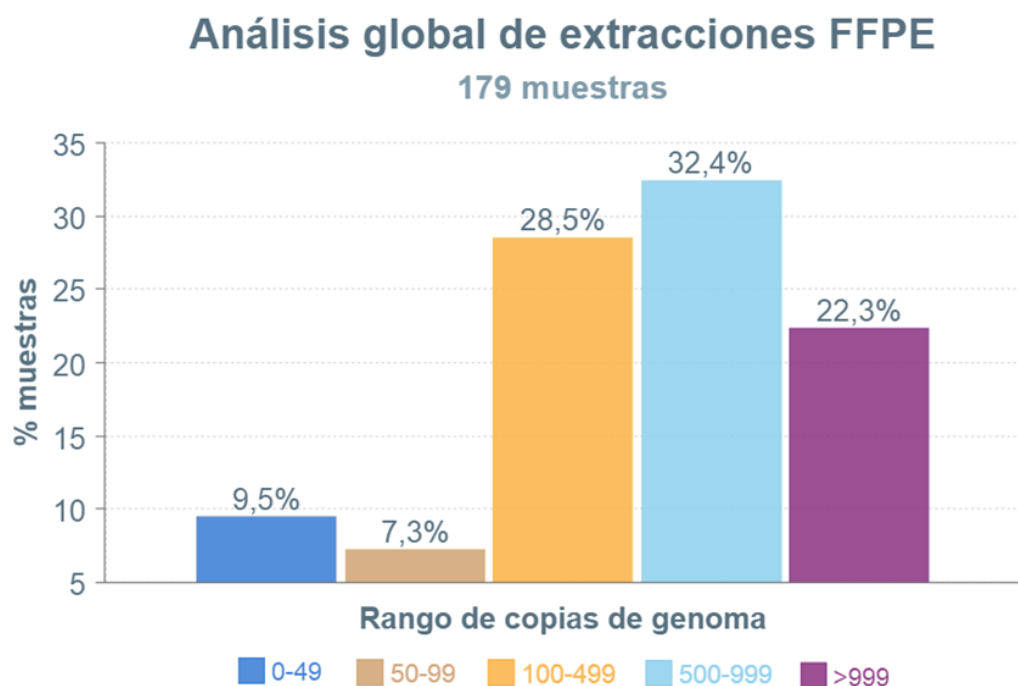


Figura 13: Histograma que representa el análisis del rendimiento de 179 muestras realizadas en esta tesis con el protocolo de Qiagen adaptado. Distribución de muestras en función del rango de número de copias de genoma en 10µl de extracción (volumen máximo utilizado en la PCR).

Una vez que las muestras se extrajeron y el ADN fue evaluado mediante PCR en tiempo real, se procedió a utilizar ese ADN en la técnica de NGS. Para tener forma de seguir el ADN desde este punto hasta la etapa de resultados de NGS, se ideó un método de trazabilidad de los ADNs extraídos que se describe a continuación.

4.1.4. Trazabilidad de los ADNs extraídos utilizando microsatélites

La trazabilidad de una muestra desde que ingresa al laboratorio hasta que se entrega el informe final al paciente es crucial en un laboratorio de diagnóstico genético. Normas como la ISO17025 y la ISO15189 recomiendan que un laboratorio de diagnóstico tenga un sistema que pueda monitorear todos los pasos a los que se someten las muestras biológicas para mantener su integridad y evitar el etiquetado incorrecto, la mezcla accidental y la contaminación de los tubos [120, 121]. Todos estos posibles problemas no son deseables en un laboratorio de diagnóstico, en particular, al planificar estudios genéticos a gran escala que no están completamente automatizados (como es el caso del protocolo de NGS en este trabajo), por lo que es recomendado dentro de las buenas prácticas de laboratorio en NGS verificar la identidad de la muestra mediante métodos alternativos [122, 123].

Los microsatélites, específicamente los STRs (repetidos cortos en tándem), se han utilizado ampliamente para identificación humana y se pueden aplicar en el laboratorio clínico para la identificación de muestras [124]. El Laboratorio Genia tiene más de 25 años de experiencia en el análisis de STRs para la identificación humana, genética forense y resolución de paternidades. Tomando ventaja de esta experiencia, se ideó un método de trazabilidad de muestras basado en el análisis de marcadores moleculares multialélicos. Para ello se realizó el análisis de STRs mediante PCR múltiple, posterior electroforesis capilar (EC) utilizando el equipo 3500/3500xL Genetic Analyzer de Thermo Fisher Scientifics y el análisis de los datos utilizando el software GeneMapper® ID-X [125]. Para que estos datos sirvieran en el protocolo de trazabilidad de las muestras se debieron comparar con los mismos STRs pero analizados tras el proceso de amplificación y secuenciación por NGS. Por este motivo, es que se planteó amplificar STRs junto con la amplificación de los genes de interés (*KRAS*, *NRAS* y *EGFR*) en el protocolo de NGS.

La base de datos nacional de ADN de EEUU conocida como CODIS (por su nombre en inglés *Combined DNA Index System*) recomienda la utilización de 20 STRs para identificación humana: CSF1PO, FGA, TH01, TPOX, VWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, D1S1656, D2S441, D2S1338, D10S1248, D12S391, D19S433 y D22S1045 [126, 127].

En el mercado existen diversos kits comerciales para amplificación de STRs para electroforesis capilar con buen rendimiento y eficiencia pero altos costos [128-132]. Para reducir los costos, en el Laboratorio Genia se diseñó un *pool* de cebadores propio para la amplificación múltiple de los marcadores: D3S1358, TH01, D21S11, D18S51, PENTA E, D5S818, D13S317, D7S820, D16S539, CSF1PO, PENTA D, VWA, D8S1179, TPOX, FGA, D19S433, D1S1656, D12S391 y D2S1338. De todos estos marcadores, se seleccionaron 14 (no incluidos: D3S1358, D19S433, D1S1656, D12S391, D2S1338) para evaluar por NGS. Esta selección se debió a que los cebadores para dichos marcadores se pudieron diseñar correctamente por el software *Ampliseq Designer* [87] con modificaciones específicas para ser compatibles con el sistema de construcción de bibliotecas para NGS.

Una vez diseñados los cebadores para la amplificación de los 14 STRs, se hizo un *pool* con todos los cebadores y se amplificaron muestras control (sangre) para probar el funcionamiento de los mismos. De los 14 STRs evaluados, 6 de ellos (TPOX, D16S539, D13S818, D5S818, TH01 y CSF1PO) funcionaron correctamente, teniendo una cobertura mayor a 250 lecturas por amplicón e identificando el genotipo correcto mediante el *plug in* HID STR Genotyper [133], cuando se comparó con el resultado de la electroforesis capilar. Los STRs D7S820, D8S1179, VWA, D18S51, FGA, Penta E, PentaD y D21S11 no tuvieron cobertura adecuada o no se pudo realizar la identificación correcta de alelos mediante el *plug in* utilizado. En algunos casos las

lecturas terminaban sobre el repetido y no se podía determinar el genotipo, o se cortaban las lecturas de un amplicón por ser una región complicada, no pudiendo determinar el número de repetidos exacto. A continuación, se procedió a probar la amplificación de los 6 STRs en muestras FFPE que se caracterizan por tener ADN fragmentado. Asimismo en las pruebas se consideró utilizar una concentración de cebadores para los STRs menor a la utilizada en la amplificación de exones dado que se quería que los STRs consumieran como máximo un 10% de las lecturas totales (STRs más los amplicones de *KRAS*, *NRAS* y *EGFR*). Para ello, se hicieron dos ensayos, uno incorporando los STRs a la mezcla de amplificación utilizando una concentración de cebadores de 33nM (cada cebador) y otro con 11nM de cada cebador de STRs. La primera concentración representa 1/6 de la concentración de los cebadores utilizados para la amplificación de los genes de interés (de aquí en adelante 1/6x) y la segunda concentración representa 1/18 de los mismos (de aquí en adelante 1/18x). Se analizaron cinco muestras provenientes de tejido FFPE y una muestra control de sangre (ADN ctrl). De la prueba realizada se observó que el conjunto de 6 STRs amplificados con un 1/18x de cebadores presentó un gran desequilibrio entre los amplicones, teniendo algunos de ellos muy pocas lecturas. Para el conjunto de 6 STRs amplificados con un 1/6x de cebadores se logró amplificar correctamente 5 de los 6 STRs en todas las muestras analizadas. El marcador D16S539 se logró amplificar en la mayoría de los casos pero con pocas lecturas. El porcentaje de lecturas de STRs amplificados con un 1/6x de cebadores se encontró entre un 8 y un 10% del total de lecturas (Tabla 7).

Tabla 7: Número de lecturas para cada STR en las 6 muestras evaluadas con 1/6x de cebadores y cálculo de porcentaje de lecturas de STRs en relación a las lecturas totales de la secuenciación para cada muestra.

STR	FFPE_1	FFPE_2	FFPE_3	FFPE_4	FFPE_5	ADN ctrl
TPOX	247	154	209	172	238	173
D5S818	234	327	123	173	211	302
CSF1PO	199	166	336	143	182	243
TH01	261	286	246	217	154	158
D13S317	171	245	210	164	193	220
D16S539	29	37	1	47	36	38
Nº de reads STRs	1141	1215	1125	916	1014	1134
Nº de reads totales	11657	13857	11441	9731	12367	13168
%Reads STRs/Reads totales	9,8	8,8	9,8	9,4	8,2	8,6

En la Figura 14 se observa la comparación de los 6 STRs en estudio, CSF1PO, D16S539, D13S317, D5S818, TH01, TPOX para la muestra control de sangre (ADN ctrl) utilizando 1/6x de cebadores. Si se comparan los resultados obtenidos mediante electroforesis capilar (Figura 14 izquierda) con los alelos secuenciados por NGS (Figura 14 derecha), se observa una concordancia del 100%, funcionando correctamente todos los STRs.



Figura 14: Resultado de 6 marcadores STRs (CSF1PO, D16S539, D13S317, D5S818, TH01, TPOX) para una muestra control (ADN ctrl) con la condición 1/6x. Resultado de la electroforesis capilar (izquierda) y de NGS (derecha).

En la Figura 15, se presenta un ejemplo del resultado de la secuenciación por NGS de los STRs de una muestra FFPE (FFPE_1) utilizando una concentración de cebadores de STRs 1/6x. En este caso, se puede observar que todos los STRs amplificaron en las dos técnicas. Sin embargo, el marcador D16S539 presentó una diferencia entre las dos metodologías ya que no se detectó por NGS el alelo 11 del par 8/11 observado en la EC. Esto puede deberse a que para este marcador la confirmación por NGS no es óptima, como se mencionó anteriormente. En este caso, no se tomó en cuenta el marcador D16S539 para la comparación. La segunda condición evaluada fue la concentración de cebadores de STRs 1/18x. Un ejemplo de esta condición con muestras FFPE se presenta en la Figura 16 (FFPE_2). En este caso, dos de los STRs no funcionaron por NGS: D5S818 y D16S539. A su vez, el marcador TH01 amplificó uno de los alelos (alelo 9.3) en menor proporción que el otro (alelo 7), provocando que quedara por debajo del umbral de detección.

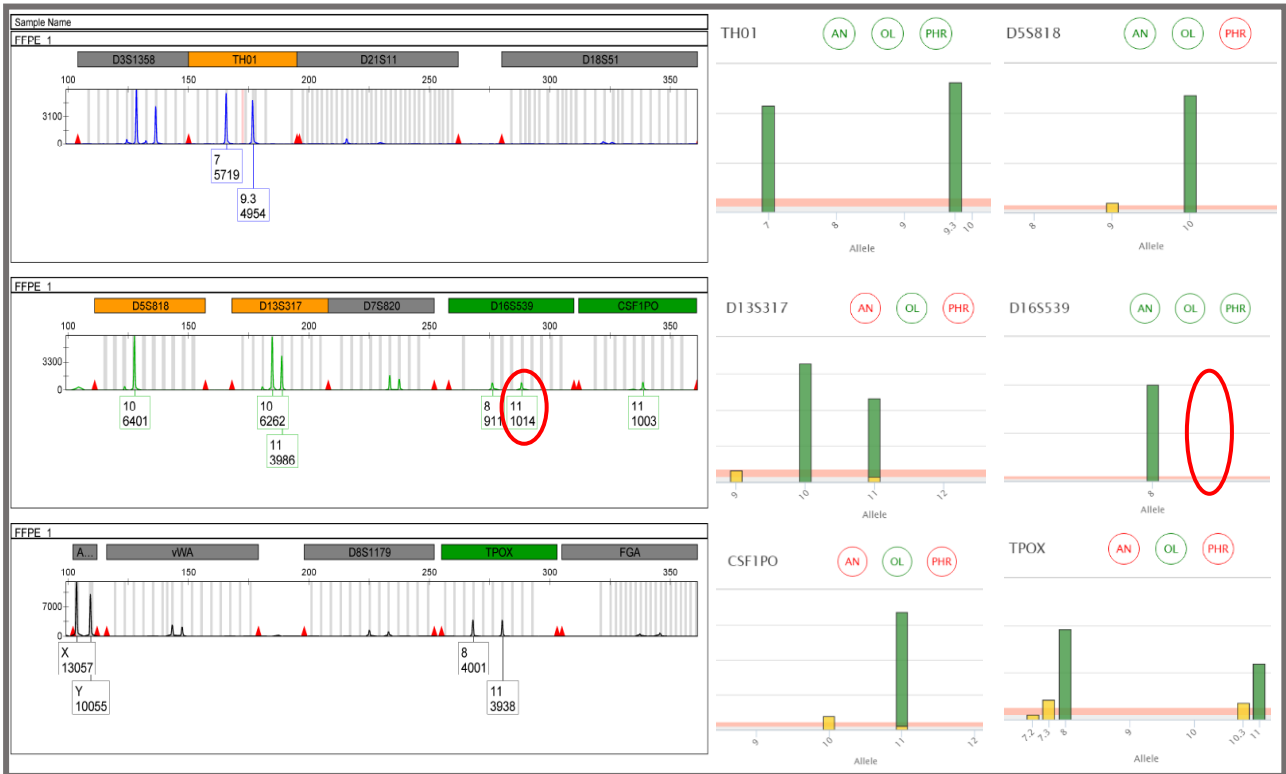


Figura 15: Resultado de 6 marcadores STRs (CSF1PO, D16S539, D13S317, D5S818, TH01, TPOX) para una muestra proveniente de tejido FFPE (FFPE_1) con la condición 1/6x. Resultado de la electroforesis capilar (izquierda) y de NGS (derecha).



Figura 16: Resultado de 6 marcadores STRs (CSF1PO, D16S539, D13S317, D5S818, TH01, TPOX) para una muestra proveniente de tejido FFPE (FFPE_2) con la condición 1/18x. Resultado de la electroforesis capilar (izquierda) y de NGS (derecha).

Ante estos resultados, se decidió estudiar el poder de discriminación que tenían los 6 STRs analizados para determinar si era suficiente para diferenciar entre muestras analizadas en el proceso de NGS. El poder de discriminación (PD) de un conjunto de marcadores genéticos se define como la probabilidad *a priori* de que dos muestras o individuos, seleccionados al azar de una población, puedan distinguirse en uno o más *loci*, los cuales son estadísticamente independientes [134-136]. El PD depende del número de *loci* analizados, del grado de polimorfismo de cada uno y de la frecuencia alélica de cada polimorfismo. El valor de PD para un *locus* se obtiene con la siguiente fórmula: $PD = 1 - \sum_i^n x_i^2$ donde X_i es igual a la frecuencia de cada uno de los n genotipos posibles. Para este cálculo se utilizaron los datos de frecuencias en la población argentina en 2017 reportados por la Sociedad Argentina de Genética Forense Asociación Civil (SAGF) (ver Anexo 8.2) [137]. El resultado obtenido del PD para los seis STRs (TPOX, D16S539, D13S818, D5S818, TH01 y CSF1PO) fue de 1 en 1.180.078 (Tabla 8), lo que indica que utilizando sólo dichos STRs se logra un poder de discriminación suficiente para distinguir entre dos muestras que puedan llegar a ser analizadas en paralelos por NGS. En aquellos casos donde el marcador D16S539 no se pudo utilizar (por bajo número de lecturas) el poder de discriminación descendió a 1 en 89.545, valor que sigue siendo suficiente para distinguir muestras.

Tabla 8: Poder de discriminación de cada STR.

STR	Repetido	Chr	Nº de alelos posibles	PD
CSF1PO	[TATC] <i>n</i>	5	23	1:08
D16S539	[GATA] <i>n</i>	16	22	1:13
D13S317	[TATC] <i>n</i>	13	18	1:20
D5S818	[TCTA] <i>n</i>	5	20	1:09
TH01	[AATG] <i>n</i>	11	23	1:11
TPOX	[AATG] <i>n</i>	2	23	1:06
				1:1.180.078

A la luz de estos resultados se adoptó el uso de los STRs, incorporándolos en las amplificaciones para la construcción de bibliotecas con los paneles KN y EGFR, realizando en paralelo, a partir del tubo original de la muestra, una PCR con cebadores marcados evaluada por electroforesis capilar. Esto permitió verificar que la identidad de una muestra secuenciada fuera la misma que la identidad de la muestra de ADN extraída inicialmente, corroborando que no hubiera un intercambio tras la amplificación por PCR, en la asignación de *barcodes* o en alguno de los diferentes pasos de armado de las bibliotecas, así como controlar contaminación entre muestras. También ayudó a diferenciar dos muestras de un mismo lote y con las mismas variantes, verificando que fueran distintas y no se debiera a una repetición de la misma muestra. Este método de trazabilidad deja abierta la posibilidad de ante una duda sobre la identidad de muestras poder recurrir a una nueva extracción de ADN de la muestra original y cotejar la identidad únicamente mediante STRs.

4.2. Evaluación y puesta a punto de alternativas al protocolo del proveedor para la detección de mutaciones somáticas utilizando la plataforma *Ion Torrent PGM™*

La búsqueda de una alternativa al protocolo de NGS (utilizando la plataforma *Ion Torrent PGM™*) planteado por el proveedor (*Thermo Fisher Scientific*) se realizó tanto para las etapas de laboratorio como para las etapas de análisis bioinformático. A nivel de laboratorio se comenzó por identificar los puntos de mayor costo con el fin de encontrar alternativas que fueran igual de eficientes pero con menor costo que las originales. Por otro lado se evaluó el desempeño del programa de análisis de secuencias del proveedor frente a otros programas de análisis disponibles en el mercado. Las alternativas encontradas se ensayaron sobre ADN de muestras FFPE para la detección de mutaciones somáticas en los genes *KRAS*, *NRAS* y *EGFR*. Para buscar opciones alternativas fue importante entender cada paso de la técnica de NGS utilizando la plataforma *Ion Torrent PGM™* y conocer sus costos. En la Figura 17, se indican los pasos desde la amplificación por PCR hasta el análisis bioinformático. A continuación se detalla cada paso en función de los kit utilizados.

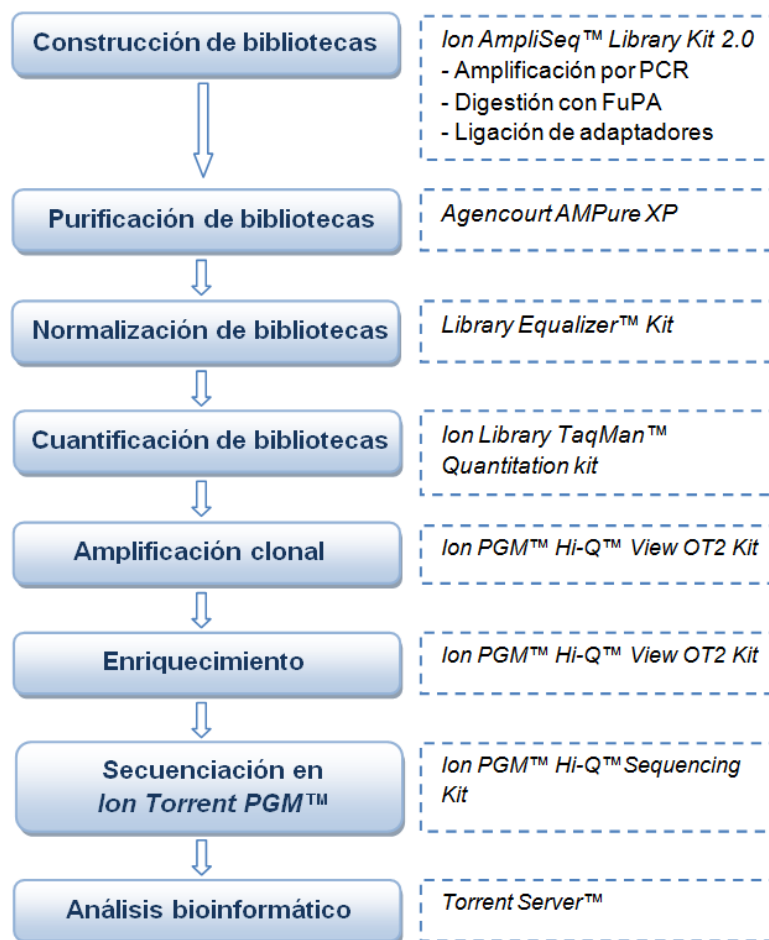


Figura 17: Esquema que resume los pasos de la técnica de NGS (del proveedor *Thermo Fisher Scientific*) utilizando la plataforma *Ion Torrent PGM™*, desde la construcción de bibliotecas hasta el análisis bioinformático de las secuencias. Se detallan los kit utilizados en cada etapa.

Construcción de bibliotecas con *Ion AmpliSeq™ Library Kit 2.0* [92]: este kit está diseñado para la preparación rápida de bibliotecas de amplicones utilizando paneles *Ion AmpliSeq™*. Incluye las etapas de amplificación, digestión con FuPA y ligación de adaptadores. Este kit tiene un costo elevado (ver apartado 4.2.4). La primera etapa "Amplificación por PCR" utiliza la enzima HiFi del kit *Ion AmpliSeq™ Library Kit 2.0*, la cual permite la amplificación de múltiples amplicones en simultáneo, utilizando cebadores con modificaciones necesarias para la etapa siguiente. La "Digestión con FuPA" es un proceso que se encuentra bajo protección de patente y es fundamental en el sistema Ampliseq [138]. La enzima FuPA digiere los cebadores modificados creando terminaciones que facilitan la ligación de los adaptadores en la siguiente etapa. La digestión de los cebadores evita que éstos se secuencien innecesariamente y además mejora la eficiencia de la ligación de los adaptadores a ambos extremos. Por último, la "Ligación de adaptadores" utiliza adaptadores con barcode (código de barras) que son secuencias cortas de ADN que se unen a cada fragmento de ADN amplificado y permiten estudiar múltiples reacciones de PCR en una única reacción de secuenciación (múltiplex), permitiendo identificar bioinformáticamente a cada muestra mediante el barcode (secuencia conocida de 8 nucleótidos). La enzima necesaria para la ligación de los adaptadores es un componente del kit *Ion AmpliSeq™ Library Kit 2.0* [92]. Sin embargo, los adaptadores con barcode se venden de forma independiente al kit y tienen bajo costo [139]. Al producto de la ligación de los adaptadores se la denomina biblioteca.

Purificación de bibliotecas con el kit *Agencourt AMPure XP* [140]: esta etapa utiliza el kit *Agencourt AMPure XP* de Beckman Coulter. Se basa en un sistema de purificación de ADN por perlas magnéticas que permite eliminar contaminantes como dNTPs, sales, cebadores y dímeros de cebadores, conservando los fragmentos mayores a 100pb y eliminando a los pequeños fragmentos. Los reactivos que se utilizan en esta etapa no suponen un costo elevado dentro del proceso (Figura 18).

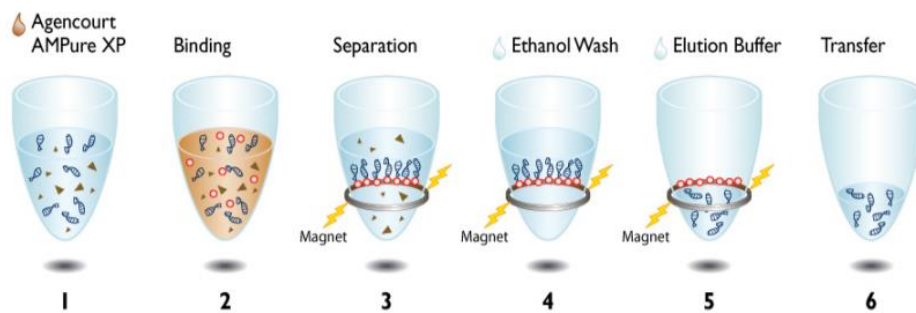


Figura 18: Pasos de la purificación con *Agencourt AMPure XP* [140].

Normalización con *Library Equalizer™ Kit* [141]: el objetivo de esta etapa es normalizar las bibliotecas previamente purificadas, llevándolas a todas a la misma concentración (100pM). Es una solución simple basada en perlas que reemplaza la necesidad de la cuantificación de las bibliotecas (etapa posterior) y las diluciones de las mismas para su normalización. Ayuda a reducir la mano de obra, los costos de reactivos y permite la optimización del espacio del chip en la etapa de secuenciación. Sin embargo, es un paso opcional por lo cual se puede prescindir del mismo dado su costo.

Cuantificación de bibliotecas con sondas *TaqMan®* utilizando el *Ion Library TaqMan™ Quantitation kit* [85]: este paso es crucial en la técnica de NGS. Se realiza para cuantificar las bibliotecas y llevarlas todas a una concentración deseada y luego generar un *pool* de bibliotecas con una concentración final entre 10-12pM para entrar en la etapa de amplificación clonal (etapa de alto costo). Para ello el kit posee sondas *TaqMan®* [142] y un estándar de bibliotecas que permiten detectar y cuantificar cantidades femtomolares de bibliotecas. Este kit presenta un costo relativamente alto dentro del proceso.

Amplificación clonal / Enriquecimiento / Secuenciación por *Ion Torrent PGM™* [95]: en estas etapas se produce la reacción de amplificación clonal de las bibliotecas unidas a perlas y la posterior eliminación de las perlas vacías (sin biblioteca unida) mediante el enriquecimiento. La amplificación clonal y el enriquecimiento se evalúan mediante un control de calidad, el cual indica el funcionamiento de las etapas anteriores y permite decidir si continuar o no con la secuenciación. Finalmente se realiza la secuenciación en el equipo *Ion Torrent PGM™* (ver detalles en el apartado 1.3.1.1). Estas últimas etapas del proceso son complejas y utilizan equipamiento específico, por lo tanto no se pueden variar ni sustituir aunque se puede apuntar a buscar un mejor rendimiento, optimizando el número de muestras estudiadas en simultáneo de forma tal que el costo relativo de esta etapa por muestra disminuya.

Análisis bioinformático con *Torrent Suite™*: esta es la última etapa del proceso en la cual se analizan los resultados de la secuenciación en el *Ion Torrent PGM™*. El análisis de datos se puede dividir en cuatro operaciones principales: asignación de base (del inglés *base calling*), alineación de lecturas contra el genoma de referencia, identificación de variantes y anotación de variantes. Estas operaciones se llevan a cabo mediante diferentes algoritmos y *software* que corren en el *Torrent Suite™* provisto por *Thermo Fisher Scientific*.

Tras el análisis previo se identificaron dos etapas con alto costo sobre las que se podía centrar la búsqueda de alternativas que permitieran reducir los costos del proceso. Las mismas fueron: "construcción de bibliotecas" (independizándose del kit *Ion Ampliseq Library Kit 2.0*) y "cuantificación de bibliotecas" (independizándose del kit *Ion Library TaqMan™ Quantitation Kit*).

4.2.1. Construcción de bibliotecas

La construcción de bibliotecas comprende las siguientes etapas: amplificación por PCR con HiFi, digestión con FuPa y ligación de los adaptadores con *barcode*. Para realizar estas etapas, el proveedor de *Ion Torrent* utiliza el kit *Ion Ampliseq Library kit 2.0* [92] (contiene las enzimas para las tres etapas) y el *Ion Xpress™ Barcode adapter* [139] (contiene los adaptadores con *barcode*).

Para tener otra opción al kit *Ion Ampliseq Library Kit 2.0*, se buscó un protocolo alternativo para la construcción de bibliotecas compatible con la metodología *Ion Torrent™*. En el mercado existe un kit comercial para este propósito: *NEBNext® Fast DNA Library Prep Set for Ion Torrent™* de *New England BioLabs® Inc* [143], mucho más económico que el propuesto por el proveedor (ver apartado 4.2.4). El sistema de NEB está ideado para ser utilizado sobre ADN genómico. Puesto que en este proyecto se trabaja con un conjunto de amplicones obtenidos por PCR, se planteó realizar una adaptación de dicho kit que permitiera trabajar con este tipo de material de partida, lo que implicó buscar una alternativa a la enzima HiFi para la amplificación del ADN por PCR.

4.2.1.1. Amplificación por PCR

La amplificación por PCR de los ADNs extraídos de tejido FFPE es la primera etapa en la construcción de bibliotecas y es un punto crítico en la técnica. La enzima recomendada por el proveedor de *Ion Torrent* se encuentra contenida en el reactivo *5X Ion Ampliseq™ HiFi Mix*. Los cebadores que se utilizan están modificados de forma que sean sustratos de la enzima en la etapa posterior a la amplificación (digestión con FuPa). Debido a que la información sobre el tipo de modificaciones que poseen no está disponible, los cebadores se deben comprar al proveedor (*Thermo Fisher Scientific*), el cual ofrece un servicio para el diseño y la síntesis de cebadores compatibles con el *Ion Ampliseq Library kit 2.0* [87]. En el caso particular de genes de interés para el estudio de cáncer colorrectal y de pulmón, existe un panel de cebadores prediseñado que ha sido creado por expertos en el tema y utilizado por muchos laboratorios de diagnóstico e investigación [144-146]. Dicho panel es el *Ion AmpliSeq Colon and Lung Research Panel v2* [147], el cual permite amplificar regiones de interés para el diagnóstico en 22 genes incluyendo a los genes que se estudian en este proyecto: *KRAS*, *NRAS* y *EGFR*.

Con el fin de disminuir los costos y/o independizarse del kit *Ion Ampliseq Library kit 2.0*, se plantearon los siguientes ajustes y alternativas: (a) hacer uso del *Ion AmpliSeq Colon and Lung Research Panel v2*, prescindiendo de los amplicones que no son necesarios para este trabajo y generar un panel adaptado sólo con los genes de interés; (b) buscar una enzima alternativa a

la enzima HiFi que sea de menor costo y con el mismo poder de amplificación múltiple de esta última.

Con respecto al **punto (a)**, se evaluó la utilización del panel comercial *Ion AmpliSeq Colon and Lung Research Panel v2* [147]. Este panel amplifica *hotspots* mutacionales y regiones seleccionadas (un total de 14,6 kb) de 22 genes asociados con tumores de colorrecto y pulmón. Los genes son: **KRAS**, **EGFR**, *BRAF*, *PIK3CA*, *AKT1*, *ERBB2*, *PTEN*, **NRAS**, *STK11*, *MAP2K1*, *ALK*, *DDR2*, *CTNNB1*, *MET*, *TP53*, *SMAD4*, *FBX7*, *FGFR3*, *NOTCH1*, *ERBB4*, *FGFR1* y *FGFR2*. El diseño consiste en un único *pool* con 92 amplicones con un tamaño medio de 160pb. Para este panel el proveedor recomienda utilizar un *chip* 314, *chip* 316 o *chip* 318 para el análisis de 2, 8 o 16 muestras, respectivamente, para obtener una cobertura de 500x en promedio para cada amplicón.

Las necesidades actuales de los estudios solicitados al Laboratorio Genia no requieren de la evaluación de todos los genes que incluye el panel antes mencionado. Esto se debe a que al Laboratorio se le solicitan dos tipos de estudios, uno para evaluar mutaciones en el gen *EGFR* y otro para evaluar los genes *KRAS* y *NRAS* (para más detalles ver apartado 1.2). En base a esto se planteó realizar dos sub paneles, uno de ellos para evaluar a los genes *KRAS* y *NRAS* (**panel KN**) y el otro para evaluar únicamente a *EGFR* (**panel EGFR**). Para el panel KN se planteó estudiar los exones 2, 3 y 4 de *KRAS* y *NRAS*, y para el panel EGFR se planteó estudiar los exones 18, 19, 20 y 21. En comparación con el uso del panel *Ion AmpliSeq Colon and Lung Research Panel v2*, la utilización de los dos sub paneles mejora notablemente el rendimiento del uso de los *chip*, permitiendo estudiar una mayor cantidad de muestras en paralelo, ya que en lugar de 92 amplicones se pasa a analizar 6 u 4 amplicones dependiendo el sub panel.

Con respecto al **punto (b)** se buscó una enzima alternativa que tuviera un rendimiento similar pero con menor costo que la enzima HiFi. Como se mencionó anteriormente, la enzima HiFi es parte del reactivo *5X Ion AmpliSeq™ HiFi Mix* del kit *AmpliSeq Library 2.0* [92] y permite la amplificación múltiple de 12 a 24000 amplicones en una única reacción usando 10ng de ADN de partida. Además, logra una cobertura uniforme y alta especificidad. La necesidad de sustituir la enzima HiFi para reducir costos se debió a que para paneles de bajo número de genes como es el caso del panel KN y EGFR (personalizados en este trabajo), el costo de utilizar esta enzima es muy alto cuando se le compara con la del método tradicional de evaluación de los genes *KRAS*, *NRAS* y *EGFR* por secuenciación Sanger. Además la capacidad de amplificación múltiple de la enzima HiFi no fue necesaria en este caso debido a la decisión de utilizar paneles de menor número de genes (panel KN y panel EGFR).

En el Laboratorio Genia se utiliza la enzima *Platinum® Multiplex PCR Master Mix* [148] tanto en la PCR de *KRAS*, *NRAS* y *EGFR* para secuenciación Sanger como en la amplificación de microsatélites para la realización de perfiles genéticos. Esta enzima está diseñada para realizar PCRs múltiples de hasta 20 amplicones de tamaños entre 50pb y 2,5 kb. Para poder determinar la eficiencia de la enzima *Platinum® Multiplex PCR Master Mix* y compararla con la enzima *5X Ion Ampliseq™ HiFi Mix* del kit *Ampliseq Library 2.0*, se realizó la PCR sobre cuatro muestras, 2 muestras de tejido FFPE (FFPE_1 y FFPE_2), 1 control de ADN de sangre (CP) y 1 control negativo (CN), utilizando los cebadores del panel KN con cada una de las enzimas. Se analizaron los productos de las reacciones mediante un gel de agarosa al 3%. El resultado de esta prueba (Figura 19) mostró un rendimiento similar para ambas enzimas, tanto para las muestras provenientes de tejido FFPE (FFPE_1 y FFPE_2) como para el control de sangre (CP). Según las bandas que se observan, se puede determinar que se amplificaron de igual forma los amplicones del panel KN en ambas PCRs (los amplicones poseen tamaños similares, entre 130 y 200pb). No se detectó amplificación de productos inespecíficos ya que no se observan bandas extras a la de los amplicones de interés. Con los resultados obtenidos se decidió utilizar la enzima *Platinum® Multiplex PCR Master Mix* para las amplificaciones subsiguientes del proceso de NGS.

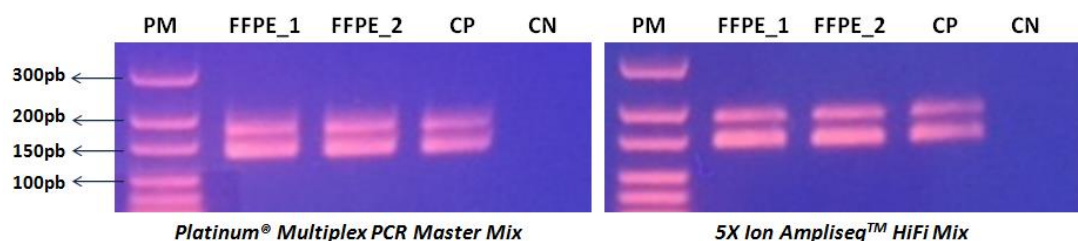


Figura 19: Comparación de la actividad de dos enzimas de amplificación para dos muestras de tejido FFPE (FFPE_1 y FFPE_2), 1 control de ADN de sangre (CP) y 1 control negativo (CN), utilizando los cebadores del panel KN. Izquierda: *Platinum® Multiplex PCR Master Mix* para panel KN; derecha: *5X Ion Ampliseq™ HiFi Mix* para panel KN.

Una vez obtenidos los productos de la PCR se los debió evaluar. Uno de los métodos utilizados fue la electroforesis en gel de agarosa al 3%. Este método no es práctico a la hora de trabajar con muchas muestras en paralelo, además de que consume mucho tiempo del proceso de diagnóstico. Por lo tanto, se probó comprobar la amplificación cuantificando el producto de PCR por *Qubit* utilizando el kit *Qubit™ dsDNA HS Assay* [86] (ver apartado 3.4). Para determinar a partir de qué valores de cuantificación se detectaba una amplificación adecuada para continuar con el proceso, se analizaron 10 muestras FFPE (5 con el panel KN y 5 con el panel EGFR) con 0,2ng de ADN (límite de cantidad de ADN establecido para el diagnóstico). Al analizar los valores de cuantificación de las muestras se observó un promedio de concentración de 21,6ng/μl (Tabla 9), con un valor mínimo de 11ng/μl. A la luz de estos resultados de cuantificación por *Qubit*, se pudo establecer que a partir de 11ng/μl se podía

continuar con el proceso de NGS. Aquellas muestras cuyo valor de cuantificación no superara los 11ng/μl aprox., debió volver a ser amplificada para poder proseguir con el siguiente paso de la construcción de bibliotecas. De esta forma se estableció un punto de chequeo en el proceso, mas rápido que la electroforesis en gel de agarosa, que evitó un fracaso en la detección de mutaciones somáticas por una ineficiente amplificación del ADN.

Por otro lado, para determinar el nivel de concentración debido al ruido en la PCR se evaluaron las amplificaciones de 10 controles negativos (CN, 5 CN con el panel KN y 5 con el panel EGFR). Los resultados de cuantificación se resumen en la Tabla 9. Se observa que el promedio de cuantificación de los CN es de 2,4ng/μl con un valor máximo de 3,0ng/μl. Al analizar en un gel de agarosa los productos de los controles negativos, no se observó las bandas específicas de *KRAS*, *NRAS* o *EGFR*, solamente se observó una banda tenue menor a 50pb que podría deberse a dímeros de cebadores (datos no mostrados).

Tabla 9: Cuantificación de ADN de la PCR realizada sobre controles negativos y muestras FFPE a 33 ciclos.

CN (sin ADN)	[PCR] ng/μl	FFPE (0,2ng)	[PCR] ng/μl
CN_KN_1	3,0	FFPE_KN_1	16,1
CN_KN_2	2,1	FFPE_KN_2	19,2
CN_KN_3	2,8	FFPE_KN_3	26,0
CN_KN_4	1,3	FFPE_KN_4	21,0
CN_KN_5	2,8	FFPE_KN_5	32,5
CN_EGFR_1	2,7	FFPE_EGFR_1	11,1
CN_EGFR_2	2,4	FFPE_EGFR_2	17,4
CN_EGFR_3	2,5	FFPE_EGFR_3	21,0
CN_EGFR_4	2,0	FFPE_EGFR_4	30,0
CN_EGFR_5	2,7	FFPE_EGFR_5	22,0
Promedio	2,4	Promedio	21,6

De aquí en más se incorporó la cuantificación de los productos de PCR por *Qubit* en el proceso de NGS.

4.2.1.2. NEBNext® Fast DNA Library Prep Set for Ion Torrent™

La etapa siguiente en la construcción de bibliotecas es la digestión de los productos amplificados con la enzima FuPa y luego la ligación de los adaptadores con *barcode* (ver Figura 18). Ambas etapas utilizan el kit *Ion Ampliseq Library Kit 2.0*. Para sustituir el proceso de construcción de bibliotecas se probó utilizar el kit *NEBNext® Fast DNA Library Prep Set for Ion Torrent™* de *New England BioLabs® Inc* [143], cuyo costo es menor costo que el kit ofrecido por el proveedor de la tecnología *Ion Torrent™*. El sistema de NEB para la construcción de bibliotecas está ideado para ser utilizado sobre ADN genómico y requiere un paso inicial de fragmentación del ADN. La construcción de bibliotecas con ADN fragmentado requiere de la

reparación de los extremos 3' y 5' antes de la ligación de los adaptadores. Una vez que se ligan los adaptadores, las bibliotecas son purificadas y seleccionadas por tamaño, y luego enriquecidas por amplificación mediante PCR (opcional) (Figura 20).

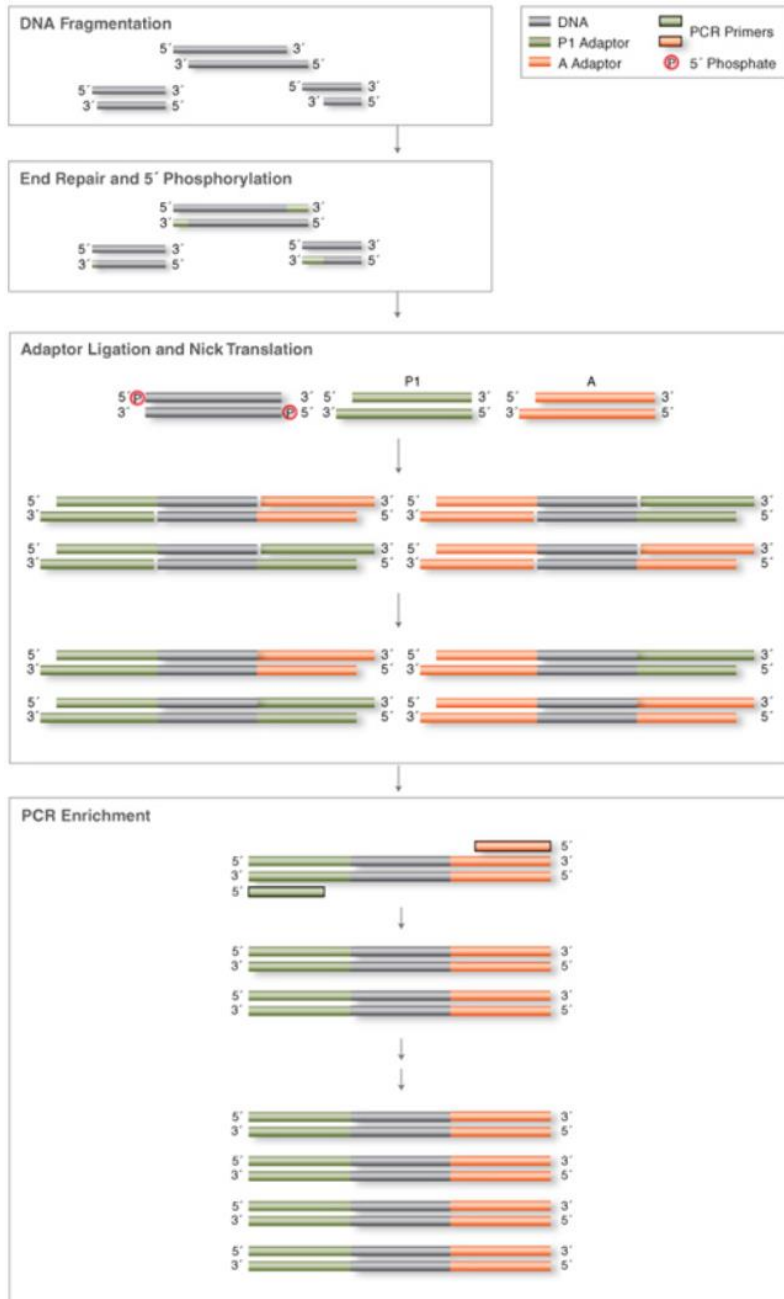


Figura 20: Sistema NEBNext® Fast DNA Library Prep Set for Ion Torrent™ de New England BioLabs® Inc. Pasos para la construcción de bibliotecas a partir de ADN genómico: 1. fragmentación del ADN, 2. reparación de los extremos 3' y 5', 3. ligación de adaptadores, 4. enriquecimiento por amplificación mediante PCR (opcional).

Como se mencionó anteriormente, en este trabajo, no se secuenció ADN genómico sino un conjunto de amplicones obtenidos por PCR (ver apartado 4.2.1.1). Por lo tanto, se probó realizar una adaptación del kit *NEBNext® Fast DNA Library Prep Set for Ion Torrent™* para poder trabajar con este tipo de material de partida [143]. Para ello, se comenzó realizando una prueba con una muestra de ADN control (ADN ctrl, de 0,2ng de ADN), una muestra proveniente de tejido FFPE (FFPE, de 0,6ng de ADN) y un control negativo (CN, con agua ultra pura), comprobando que el sistema funcionara a partir de productos de PCR. Se amplificó el ADN de las muestras antes mencionadas utilizando el panel KN y la enzima *Platinum® Multiplex PCR Master Mix* según las condiciones descritas en el apartado 3.5.3 a 27 ciclos, como recomiendan en el protocolo de *Ampliseq* para muestras de baja calidad de ADN (como son los FFPE). Con el propósito de eliminar los restos de cebadores tras la PCR que pueden interferir en el proceso de secuenciación se utilizó el sistema de purificación de ADN *Agencourt AMPure XP* de *Beckman Coulter* [140] que permite eliminar fragmentos menores a 100pb (este paso se consideró necesario debido a la ausencia de la enzima FuPA). Los productos de PCR purificados con *AMPure* se cuantificaron por *Qubit*. En la Tabla 10 se muestra el resultado de la cuantificación de las PCRs purificadas para las dos muestras evaluadas (FFPE y ADN ctrl) y el control negativo (CN).

Tabla 10: Cuantificación por *Qubit* de ADN en la PCR realizada para el panel KN sobre las muestras FFPE, ADN ctrl y CN, después de la purificación con *AMPures*.

Muestra	ng de ADN	ng totales en PCR purificada
FFPE	0,6	16
ADN ctrl	0,2	22
CN	-	<1

Según las recomendaciones del kit de NEB, se necesita entre 10ng y 1µg de ADN fragmentado para comenzar con la etapa de "reparación de los extremos y fosforilación 5'" [94]. Al observar los resultados de la cuantificación de las PCRs purificadas (Tabla 10) se verificó que las muestras correspondientes a FFPE y ADN ctrl tenían una cantidad mayor a 10ng, suficiente para continuar con el proceso de **"reparación de los extremos y fosforilación 5'"** (condiciones especificadas en apartado 3.9.2), etapa que se mantuvo para mantener las condiciones de los *buffer* indicados por el kit de NEB. El control negativo no se continuó. Luego de este paso, el kit de NEB indica la **"ligación de adaptadores con barcode"** (condiciones especificadas en apartado 3.9.2), lo cual se hizo para obtener las bibliotecas. Los *barcode* no ligados se eliminaron mediante una purificación con *AMPure* tomando como referencia al paso de purificación de bibliotecas del protocolo de *Ampliseq* (relación de *AMPure*/ biblioteca de 1,5x) [149]. Posteriormente se cuantificaron las bibliotecas purificadas por *Qubit* (en el apartado 4.2.2 se explica en detalle la cuantificación de bibliotecas). Para la secuenciación de las bibliotecas se armó un *pool* de las mismas en una relación equimolar, el cual se utilizó en la etapa de amplificación clonal (OT), para luego continuar con las etapas de enriquecimiento

(ES) y secuenciación en el *Ion Torrent PGM™* (según lo indicado por el proveedor, *Thermo Fisher Scientific*).

4.2.1.2.1. Resultado de la prueba de adaptación del kit de NEB

En una primera instancia de análisis de los resultados de la prueba de adaptación del kit de NEB, se evaluaron los indicadores de calidad de la secuenciación, la cobertura de los amplicones y el llamado de variantes, con el propósito de realizar los ajustes necesarios para optimizar el protocolo.

En relación a los indicadores de calidad de la secuenciación, llamaron la atención el porcentaje de policlonalidad y el porcentaje de dímeros de adaptadores por encontrarse fuera de los rangos indicados en el manual del equipo *Ion Torrent PGM™* [150]. Concretamente, el porcentaje de perlas policlonales (con más de un molécula de ADN de origen diferente por perla) fue de 41%, valor que sobrepasa el límite superior aceptable (40%), y el valor de dímeros de adaptadores fue de 28%, cuando un valor esperable se encuentra entre 0 y 10%. A pesar de que los valores de calidad permitieron continuar analizando los resultados de la secuenciación, se tuvo que dedicar esfuerzos para mejorar estos parámetros que se detallan más adelante (apartado 4.2.1.2.2).

A continuación, se evaluó la cobertura de los amplicones del panel KN para las dos muestras secuenciadas (FFPE y ADN ctrl). Para todos ello se obtuvo más de 1000 lecturas, que es el número mínimo recomendado para detectar mutaciones somáticas (Tabla 11) [71, 151]. A partir de estos datos se detectó un desequilibrio en la proporción de los amplicones, de 14 veces (entre el amplicón con menos lecturas y el que tuvo mayor número de lecturas) en la muestra FFPE y de 19 veces en la muestra ADN ctrl. Sin embargo, al observar la cobertura de la muestra ADN ctrl realizada con el protocolo *Ampliseq*, se vio que el desequilibrio de amplicones era mucho menor, 2,4 veces. El desbalance con el protocolo NEB utilizando la enzima Platinum se buscó disminuir (ver más adelante) ya que conlleva un alto consumo del espacio en el *chip*, si se considera que todos los amplicones deben de tener una cobertura mínima de 1000 lecturas.

Tabla 11: Cobertura de los amplicones del panel KN para la muestra FFPE y ADN ctrl. Comparación con muestra control utilizando el protocolo Ampliseq (Thermo Fisher Scientific).

Amplicón	Protocolo NEB		Protocolo Ampliseq
	FFPE	ADN ctrl	ADN ctrl
KRAS ex2	13960	6945	12483
KRAS ex3	30120	18089	29510
KRAS ex4	2089	1065	12813
NRAS ex2	7237	2941	20897
NRAS ex3	21231	19932	19433
NRAS ex4	18540	17367	14433

En cuanto al análisis de las variantes mediante el software *Torrent Suite™*, se logró detectar dos variantes en la muestra FFPE (las mismas que fueron detectadas por secuenciación Sanger), ambas en el gen *KRAS* (c.33T>C y c.34G>T) y ningún llamado de variantes en el control de sangre (ADN ctrl) (también corroborado por secuenciación Sanger). En la Figura 21, se presentan los cambios detectados en la muestra FFPE utilizando el programa *Integrative Genome Viewer (IGV)*.

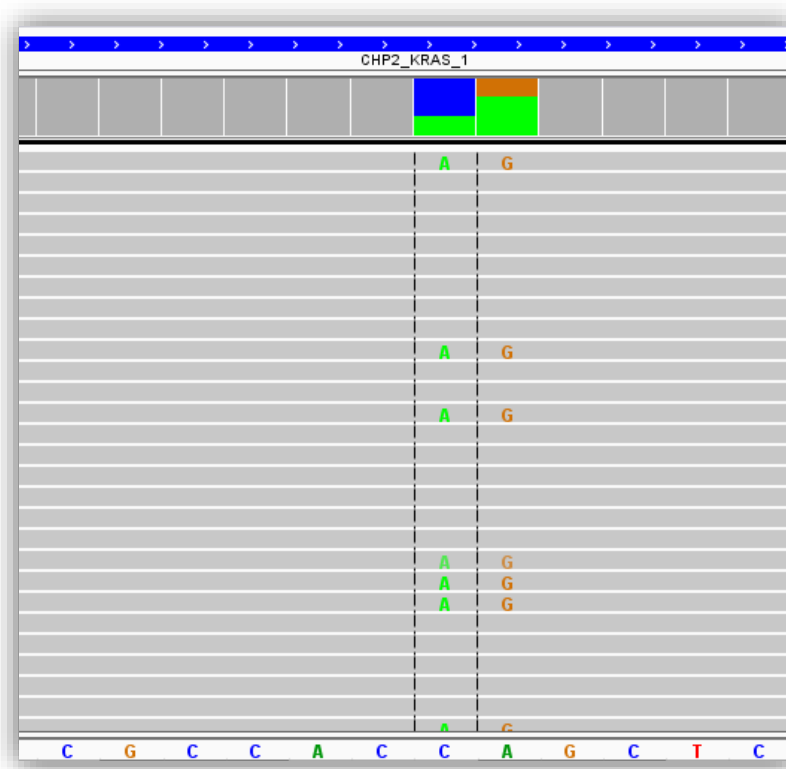


Figura 21: IGV de muestra FFPE, KRAS exón 2, mutación en c.33T>C y c.34G>T. Nótese que el gen *KRAS* se transcribe a partir de la hebra complementaria a la representada (por convención hebra +).

De esta prueba se concluyó que el kit de NEB funcionó con productos de PCR como material de partida, logrando amplificar los amplicones del panel KN y detectar correctamente las variantes presentes en la muestra FFPE. En lo que corresponde a los problemas con los

indicadores de calidad de la secuenciación (porcentaje de policlonalidad y porcentaje de dímeros de adaptadores) y el desequilibrio de amplicones, se trabajó en la solución de los mismos en etapas posteriores, que se presentan a continuación.

4.2.1.2.2. Optimización de los indicadores de la calidad de secuenciación

Los resultados de la secuenciación en el *Ion Torrent PGM™* se visualizan utilizando el *software Torrent Suite*, el cual genera un informe de la secuenciación que contiene indicadores de calidad, como lo son la densidad de perlas o ISP (*Ion Sphere™ Particle*), el porcentaje de ISP policlonales (ISP que llevan clones de dos o más *template*), porcentaje de baja calidad (porcentaje de ISPs con un señal baja o irreconocible), el porcentaje de dímeros de adaptadores (ISP que llevan unido un inserto de ADN de una longitud inferior a 8 pb entre ambos adaptadores) y el porcentaje de lecturas útiles (el porcentaje de ISP con biblioteca que pasan los filtros policlonales, de baja calidad y dímeros de cebadores) [74].

En las pruebas realizadas hasta el momento los dos indicadores de calidad de secuenciación que no funcionaron de forma óptima fueron el porcentaje de ISP policlonales y el porcentaje de dímeros de adaptadores. El porcentaje de perlas policlonales depende de la concentración del *pool* de bibliotecas que ingresa a la etapa de amplificación clonal (OT). Cuando el molde (bibliotecas) es bajo en relación al número de perlas, queda un exceso de perlas vacías (que se eliminan en la etapa ES) y pocas perlas clonales para la secuenciación posterior, teniendo como consecuencia un bajo número de lecturas. Si hay un exceso de molde en el OT, algunas perlas van a tener más de una molécula de ADN de orígenes distintos, generando una señal múltiple (policlonal). Esta última situación fue la observada en la prueba de adaptación del kit de NEB. Para solucionar este problema se fueron probando diferentes concentraciones de bibliotecas para seleccionar aquella que permitiera un equilibrio entre el número de lecturas útiles y la presencia de ISPs policlonales. Al ensayar 0,5ng/μl de concentración del *pool* de bibliotecas, el porcentaje de perlas policlonales superó el máximo establecido (40%) en las dos pruebas a esa concentración (Tabla 12). Con este resultado se decidió disminuir la concentración del *pool* a 0,3ng/μl. Con esta concentración, 2 de las 3 pruebas tuvieron valores de policlonalidad adecuados pero la tercera superó el valor máximo. Algo similar sucedió con la condición a 0,1ng/μl, donde una de las pruebas no cumplió con los parámetros de policlonalidad adecuados. Debido a esto se decidió continuar disminuyendo la concentración del *pool* hasta 0,05ng/μl que fue la concentración en la cual las tres pruebas tuvieron valores de policlonalidad por debajo de 40% (límite superior aceptable). A partir de este resultado, se decidió continuar con 0,05ng/μl de concentración del *pool* de bibliotecas para ingresar al OT.

Tabla 12: Porcentaje de ISPs policlonales dependiendo de la concentración del pool de bibliotecas.

[pool bibliotecas] ng/ μ l	% de ISPs policlonales
0,5	41
0,5	56
0,3	34
0,3	27
0,3	46
0,1	56
0,1	38
0,05	31
0,05	25
0,05	26

El siguiente parámetro que se evaluó fue el porcentaje de dímeros de adaptadores. El valor de dímeros depende de la cantidad de adaptadores que se utilizan en la reacción de “ligación de adaptadores con *barcode*”. Los valores deseables para este indicador se encuentran entre 0 y 10%. Debido a la presencia de un alto porcentaje de dímeros en la prueba de adaptación del protocolo NEB (28%), se realizaron pruebas para poder disminuir este valor (Tabla 13). El primer cambio fue reducir a la mitad la cantidad de adaptadores utilizados en la ligación. Con esta reducción, disminuyó notablemente el porcentaje de dímeros aunque el valor (10%, Tabla 13) se encontró en el límite de lo aceptable. Por este motivo, se decidió disminuir aún más la cantidad de adaptadores en el proceso de ligación realizando una reducción al quinto. Con este último cambio, el valor quedó dentro del rango aceptable en las 8 pruebas evaluadas. Tras estos resultados, se decidió utilizar un 1/5 de de la cantidad original de los adaptadores con *barcode* en el proceso de ligación de NEB.

Tabla 13: Variaciones en el porcentaje de dímeros de adaptadores en 10 secuenciaciones según cambios en el volumen de adaptadores utilizados.

Cantidad de adaptadores en la PCR	% de dímeros de adaptadores
Original para 100ng de PCR	28%
1/2 del original	10%
1/5 del original	3%
1/5 del original	2%
1/5 del original	4%
1/5 del original	4%
1/5 del original	5%
1/5 del original	3%
1/5 del original	3%
1/5 del original	4%

4.2.1.2.3. Optimización del equilibrio de amplicones

Para mejorar el desbalance de amplicones del panel KN observado en la prueba de adaptación del kit de NEB, se probaron tres condiciones de ciclado: 27, 30 y 33 ciclos. Las muestras evaluadas fueron FFPE y ADN ctrl (0,6 y 0,2ng respectivamente). La cuantificación de las PCRs purificadas, para las dos muestras evaluadas (FFPE y ADN ctrl) con los tres ciclados de amplificación diferentes se muestra en la Tabla 14.

Tabla 14: Cuantificación por Qubit de las PCR purificadas de los tres ciclados estudiados para las muestras FFPE y ADN ctrl.

Muestra	ng totales en PCR purificada
FFPE_27x	16
ADN ctrl_27x	22
FFPE_30x	104
ADN ctrl_30x	140
FFPE_33x	268
ADN ctrl_33x	254

Tomando en cuenta los valores de la Tabla 14 y que el kit de NEB menciona que la eficiencia de la construcción de bibliotecas se puede favorecer utilizando 100ng de material de partida cuando se utilizan los *barcode Ion Xpress™ Barcode adapter* [93], se decidió utilizar 100ng de cada producto de PCR purificado. En el caso de las muestras amplificadas a 27 ciclos no se llegó a esta cantidad y se tuvieron que utilizar puras. Por otro lado, las amplificaciones con 30 y 33 ciclos alcanzaron correctamente la cantidad de ADN necesario para continuar con el proceso. Para determinar si los cambios en el número de ciclos incidieron en el balance de lecturas por amplicón, se analizó el resultado de la cobertura (Tabla 15). Se volvió a observar un gran desbalance entre los amplicones obtenidos a 27 ciclos en comparación con los otros dos ciclados. A su vez, hubo amplicones cercanos al límite inferior de lecturas, como es el caso del exón 2 de *NRAS*, que en casos donde la cantidad de biblioteca que se utilice para el OT (amplificación clonal) sea más baja, sería un problema ya que podría no ser secuenciado con la cobertura mínima deseada. Con 30 ciclos todos los amplicones superaron el mínimo de 1000 lecturas y el desbalance fue de 3,5 veces para la muestra FFPE y de 4 veces para el ADN ctrl. Con 33 ciclos se vio un buen balance de amplicones con una diferencia entre el amplicón con menos lecturas y el de más lecturas de tan solo 2,9 veces para la muestra FFPE y 3,7 veces para el ADN ctrl. De esta prueba se concluyó que para lograr una cobertura uniforme de los amplicones, sería mejor trabajar con 33 ciclos en la PCR inicial.

Tabla 15: Comparación de los resultados de secuenciación de la prueba realizada con reactivos de NEB y PCR con distinto número de ciclos (27, 30 y 33) para las muestras FFPE y ADN ctrl.

Amplicones	27x_FFPE	27x_ADN ctrl	30x_FFPE	30x_ADN ctrl	33x_FFPE	33x_ADN ctrl
KRAS EX2	9338	12476	26662	29307	23744	25549
KRAS EX3	10888	7937	42594	57530	43496	51480
KRAS EX4	7984	8311	18009	22640	20544	26889
NRAS EX2	2118	1061	21110	35026	41327	52985
NRAS EX3	37547	14179	15176	24292	18416	19881
NRAS EX4	47420	62283	53112	89783	54004	73213

4.2.1.2.4. Reducción del volumen de reacción del protocolo de NEB

Una vez que se corroboró el correcto funcionamiento del kit de NEB para productos de PCR y se realizaron las modificaciones necesarias para mejorar los indicadores de secuenciación y el equilibrio de amplicones, se probó reducir el volumen de reacción de las dos etapas del kit de NEB ("reparación de los extremos y fosforilación 5'" y "ligación de adaptadores con *barcode*") con el fin de disminuir los costos del proceso. Para ello, se tomaron PCRs purificadas obtenidas a partir de dos muestras (FFPE y ADN ctrl) y se procesaron en paralelo utilizando el **volumen completo, mitad de volumen y cuarto del volumen** en las etapas del kit de NEB.

Los resultados de esta prueba mostraron que para los tres volúmenes de reacción (completo, $\frac{1}{2}$ y $\frac{1}{4}$) la secuenciación funcionó de forma similar. La cobertura fue homogénea entre amplicones (una relación de 3 y 4 veces entre el amplicón con más cobertura y el de menos, para las dos muestras a $\frac{1}{2}$ y $\frac{1}{4}$) y se obtuvo una concentración de bibliotecas superior a la requerida. En los tres casos se detectaron correctamente las dos variantes en el gen *KRAS* (c.33T>C y c.34G>T) presentes en la muestra FFPE, y ninguna variante en ADN ctrl. Por lo que se puede concluir que la disminución del volumen de reacción a $\frac{1}{2}$ y $\frac{1}{4}$ del volumen original permitió la correcta secuenciación de todos los amplicones de interés sin interferir en la detección de las variantes. A partir de esta prueba se seleccionó la reducción del volumen al cuarto para continuar trabajando con el protocolo de NEB y de esa forma reducir costos.

4.2.1.2.5. Prueba del protocolo de NEB sobre muestras variadas

Luego de realizar varias modificaciones al protocolo de NEB, se planteó probar el sistema con muestras de baja concentración de ADN y con mutaciones en baja proporción respecto al ADN normal, es decir exigiendo a la técnica. Para ello se analizaron 3 ADNs de referencia "Horizon" con mutaciones y concentración conocida (ver apartado 3.1.3), 3 muestras provenientes de tejido FFPE de baja concentración con mutaciones previamente analizadas por secuenciación

Sanger y un control de sangre (ADN ctrl, con la cantidad de ADN límite: 0,2ng) con el panel KN. Se trabajó con las muestras "Horizon" con tres cantidades diferentes: 1ng (para usar como referencia en el proceso), 0,1ng y 0,02ng. En la Tabla 16 se resumen los resultados de la prueba, detallando la cantidad de ADN utilizada como molde en la PCR, la concentración de amplicones tras la PCR, la cantidad de producto de PCR purificado obtenido y la concentración de biblioteca, así como el resultado de la secuenciación en el *Ion Torrent PGM™*.

Tabla 16: Detección de variantes patogénicas en KRAS y NRAS sobre muestras con diferentes de ADN inicial.

Muestra	Cantidad de ADN en la PCR (ng)	[PCR] (ng/ul)	PCR purif. (ng)	[Biblioteca ca 1/10] (ng/ul)	Cobertura por amplicón	Variante esperada vs. detectada
Horizon_1	1,0	34	226	0,27	>1000	6,25% vs 4,5% (KRAS_c.436G>A)
Horizon_2	1,0	41	280	0,25	>1000	6,25% vs 4,9% (KRAS_c.183A>C)
Horizon_3	1,0	39	272	0,24	>1000	6,25% vs 8,1% (KRAS_c.35G>T)
Horizon_1	0,1	13	76	0,24	>1000	6,25% vs 2,4% (KRAS_c.436G>A)
Horizon_2	0,1	12	88	0,22	>1000	6,25% vs 3,5% (KRAS_c.183A>C)
Horizon_3	0,1	12	72	0,23	>1000	6,25% vs 10,2% (KRAS_c.35G>T)
Horizon_1	0,02	3,1	6	<0,05	<500	NO detectada
Horizon_2	0,02	3,0	2	<0,05	<500	NO detectada
Horizon_3	0,02	2,7	4	<0,05	<500	NO detectada
FFPE_1	0,07	41	220	0,19	>1000	Sanger=NGS (KRAS_c.35G>A)
FFPE_2	0,02	23	120	0,12	>1000	Sanger=NGS (WT)
FFPE_3	0,04	30	186	0,18	>1000	Sanger=NGS (KRAS_c.183A>C)
ADN ctrl	0,2	12	44	0,29	>1000	Sanger=NGS (WT)

A partir de esta prueba en la cual se exige al sistema de NEB, se pudieron observar diferentes resultados. Lo primero que se observa es un efecto estocástico al comparar las muestras Horizon con distintas cantidades de ADN. Cuando la cantidad de ADN inicial es mayor al límite estipulado, 0,2ng, se logra detectar la variante es un porcentaje cercano al esperado, 6,25% (las pequeñas desviaciones se pueden deber a la generación de esa muestra, la cual se debe diluir o mezclar con ADN normal para llegar a dicho porcentaje). Sin embargo, cuando disminuye la cantidad de ADN a una décima parte, 0,1ng, el porcentaje de detección de la variante se aleja más del valor esperado (6,25%). Un caso más extremo fueron las mismas muestras Horizon con 0,02ng de ADN donde no se detectó ninguna de las variantes analizadas, mostrando el efecto que provoca la cantidad de ADN de partida cuando se quieren detectar variantes de muy bajo porcentaje. Este efecto de la cantidad de ADN se puede observar también en la cuantificación de la PCR ya que todas las muestras que superaron el valor mínimo definido de 11ng/μl (ver apartado 4.2.1.1), lograron secuenciarse obteniendo el

resultado esperado, mientras que aquellas muestras que presentaron valores inferiores, no lograron llegar a secuenciarse correctamente. Sin embargo, la concentración de ADN inicial no siempre se correspondió con el valor de cuantificación de la PCR. Hubo casos en que las muestras tuvieron muy baja cantidad de ADN inicial (FFPE1, FFPE2 y FFPE3), aproximadamente diez veces menos que el límite de 0,2ng (ver apartado 4.1.1) y sin embargo, los valores de concentración en la PCR fueron superiores al valor mínimo definido (11ng/μl) y lograron secuenciarse obteniendo el resultado esperado. Esto pudo deberse a que las muestras presentaban inhibidores de la PCR que inhibieron la PCR de la cuantificación con *SYBR Green®* subestimando la cantidad de ADN contenida en la muestra FFPE. Dicha inhibición no afectó de igual manera la PCR con la enzima Platinum.

Con estos resultados se demuestra que el sistema de NGS permite secuenciar, en algunos casos, muestras de baja cantidad de ADN inicial (menor al límite establecido de 0,2ng), siempre que la concentración de la PCR sea la adecuada (mayor a 11ng/μl). Sin embargo, no se puede perder de vista que podría haber un riesgo de no detectar variantes patogénicas por un efecto estocástico en la toma de muestra, cuando la cantidad de ADN inicial es baja. El análisis de un mayor número de réplicas fue realizado en la sección donde se evalúa la sensibilidad del método según la cantidad de ADN de partida (ver apartado 4.3. Validación).

4.2.1.2.6. Utilización del protocolo adaptado de NEB para evaluar *EGFR*

El protocolo de NEB adaptado para el panel KN se aplicó al estudio de *EGFR* (panel EGFR). Para ello, se analizaron 9 muestras de controles de calidad (control de calidad EMQN) (ver apartado 3.1.2) y 2 muestras provenientes de tejido FFPE (con resultado conocido por secuenciación Sanger), las cuales se amplificaron utilizando el panel EGFR como se indica en el apartado 3.5. Todas las muestras partieron de una cantidad de ADN superior a 0,2ng para la PCR.

Los resultados de esta prueba fueron alentadores ya que el protocolo de NEB adaptado utilizando el panel EGFR funcionó adecuadamente. En todas las muestras se logró obtener más de 1000 lecturas y se detectaron las variantes esperadas en los 9 casos de controles de calidad (comparados con el resultado provisto por *EMQN*) y de las 2 muestras de FFPE (comparadas con el resultado de la secuenciación Sanger). En la Tabla 17 se detallan los resultados de la detección de variantes patogénicas detectadas en las muestras analizadas, que incluyeron cambios puntuales y deleciones de 15 nucleótidos.

Tabla 17: Variantes patogénicas detectadas en las muestras analizadas con el panel EGFR utilizando el protocolo de NEB adaptado. En todos los casos los resultados coincidieron con el valor esperado.

EGFR_CC_1	c.2573T>G (10%) c.2369C>T (11%)
EGFR_CC_2	c.2155G>A p.(Gly719Ser) (15%)
EGFR_CC_3	WT
EGFR_CC_4	WT
EGFR_CC_5	WT
EGFR_CC_6	WT
EGFR_CC_7	WT
EGFR_CC_8	WT
EGFR_CC_9	WT
EGFR_FFPE_1	c.2236_2250 del15 (13,6%)
EGFR_FFPE_2	c.2235_2249 del15 (6,8%)

Al estudiar el balance de amplicones se vio cierta variabilidad entre muestras, observando una diferencia de 2 a 11 veces entre el exón 18 (siempre con menor cobertura que los demás) y el exón de mayor cobertura (Tabla 18). Por lo tanto se necesitarían al menos 33 mil lecturas para cada muestra realizada con este panel para lograr cubrir todos los amplicones con un mínimo de 1000 lecturas.

Tabla 18: Planilla de cobertura de las muestras realizadas con el panel EGFR utilizando el protocolo de NEB adaptado.

	EGFR CC_1	EGFR CC_2	EGFR CC_3	EGFR CC_4	EGFR CC_5	EGFR CC_6	EGFR CC_7	EGFR CC_8	EGFR CC_9	EGFR FFPE1	EGFR FFPE2
EGFR Ex 18	7185	12223	9919	17018	11333	12320	9470	38462	9959	8481	7912
EGFR Ex 19	69982	76696	61710	54124	75705	70377	68761	49970	16095	21809	31228
EGFR Ex 20	75021	101836	83811	135973	87464	73284	78674	79210	28390	23961	50889
EGFR Ex 21	79083	100398	97732	151100	63809	90327	87243	63891	27568	24431	58704

Los resultados de esta prueba permitieron concluir que el sistema de NEB permitió amplificar y secuenciar los amplicones del panel EGFR, detectando correctamente los cambios esperados.

4.2.1.2.7. Evaluación de STRs

El último tema que se evaluó en la puesta a punto del protocolo de NEB fue la detección de los microsatélites (STRs) realizados por NGS para corroborar la trazabilidad de las muestras desde la extracción hasta el resultado final (este punto fue detallado en el apartado 4.1.4). Por lo tanto, se realizó la amplificación de los 6 STRs, definidos para este trabajo, tanto por NGS (ver apartado 3.5.2) como por electroforesis capilar (ver apartado 3.8) en 10 muestras FFPE (5

utilizando el panel KN y 5 el panel EGFR). Como se mencionó anteriormente los STRs para NGS se amplificaron junto con los amplicones para *KRAS/NRAS* o *EGFR*.

En la Tabla 19, se observa la comparación de los STRs de las 10 muestras analizadas en esta prueba. En un 70% de los casos los resultados obtenidos con las dos metodologías de análisis (NGS y electroforesis capilar) coincidieron. La falta de coincidencia se debió principalmente al resultado obtenido con el marcador D16S539, lo que era de esperar según lo evaluado en el apartado 4.1.4, donde se vio que dicho marcador no tuvo un funcionamiento óptimo. Por otro lado, hubo un caso para el marcador THO1 en donde uno de los alelos (alelo 5) analizado por NGS tuvo un bajo número de lecturas (3 lecturas), por lo cual no se pudo detectar correctamente y no se pudo chequear con el resultado de la electroforesis capilar.

Tabla 19: Comparación de los 6 STRs evaluados por NGS versus electroforesis capilar (EC) para 10 muestras. Con * se indica los resultados de los STRs por NGS que no coincidieron con EC.

Muestra	CSF1PO		D13S317		D16S539		D5S818		THO1		TPOX	
	NGS	EC	NGS	EC	NGS	EC	NGS	EC	NGS	EC	NGS	EC
KN_1	10/12	10/12	12	12	11/12	11/12	12	12	7/9	7/9	8/11	8/11
KN_2	12	12	11/12	11/12	9/10	9/10	9/11	9/11	7/9,3	7/9,3	8/11	8/11
KN_3	10/12	10/12	9/12	9/12	0*	10	11/12	11/12	6/7	6/7	8	8
KN_4	10/11	10/11	11/12	11/12	11	11	11/12	11/12	7/8	7/8	8/11	8/11
KN_5	11/12	11/12	12/13	12/13	11	11	11	11	7*	5/7	8	8
EGFR_1	10/12	10/12	11	11	11	11	11	11	6/7	6/7	8	8
EGFR_2	10	10	11	11	11	11	11/12	11/12	9,3	9,3	8/9	8/9
EGFR_3	10/11	10/11	11	11	0	0	7/12	7/12	7/9,3	7/9,3	8/12	8/12
EGFR_4	11/12	11/12	8/12	8/12	9/11	9/11	11/13	11/13	6/9,3	6/9,3	11	11
EGFR_5	11/12	11/12	11/12	11/12	11*	11/12	9/11	9/11	9,3	9,3	8	8

Con los datos obtenidos en esta prueba se logró cumplir el objetivo de la utilización de STRs para chequear la identidad de muestras ya que en todos los casos hubo, por lo menos, 5 STRs que se pudieron comparar permitiendo un poder de discriminación de 1 en 89.545, suficiente *a priori* para poder discriminar e identificar las muestras que se procesan en paralelo en el laboratorio. Utilizando STRs se constató que no hubo cambio de muestras durante el proceso de NGS y permitió tener trazabilidad de las muestras que se procesaron.

4.2.2. Cuantificación de bibliotecas

El siguiente punto evaluado, en esta tesis, fue la cuantificación de las bibliotecas. Esta etapa es de suma importancia ya que permite normalizar las bibliotecas, llevándolas todas a la misma concentración para poder generar un *pool* de concentración adecuada para ingresar en la etapa de amplificación clonal (OT). Según el protocolo del proveedor (*Thermo Fisher Scientific*),

para la cuantificación de las bibliotecas se utiliza el *Ion Library TaqMan™ Quantitation Kit* [85], que trabaja con sondas Taqman® [142]. Debido a que esta etapa presenta un costo relativamente alto dentro del proceso (ver apartado 4.2.4), se buscaron alternativas de cuantificación que permitieran disminuir los costos.

4.2.2.1. Cuantificación de bibliotecas por PCR en tiempo real con *SYBR Green®*

La primera alternativa que se evaluó fue la cuantificación de las bibliotecas mediante PCR en tiempo real utilizando el reactivo *Power SYBR® Green PCR Master Mix* [82, 84]. Para realizar esta prueba, se evaluaron tres juegos de cebadores distintos (con el propósito de establecer si era factible utilizarlos en la cuantificación de bibliotecas): 1. cebadores universales que utiliza el *Library Equalizer™ Kit* para amplificar bibliotecas (desde ahora "cebadores Equalizer"); 2. cebadores que trae el *Ion AmpliSeq™ Library Kit 2.0* para re-amplificar bibliotecas (desde ahora "cebadores Ampliseq"); 3. cebadores que trae el kit *NEBNext® Fast DNA Library Prep Set for Ion Torrent™* para re-amplificar bibliotecas (desde ahora "cebadores NEB"). Se evaluaron dos bibliotecas realizadas con la metodología de *Ampliseq* (protocolo del proveedor *Thermo Fisher Scientific*) (desde ahora "bibliotecas Ampliseq") y dos bibliotecas realizadas con la alternativa de NEB (desde ahora "bibliotecas NEB") a distintas concentraciones: **puras, diluídas 1/10 y diluídas 1/100**. La mezcla de reacción, el ciclado y la curva estándar se realizó según las condiciones especificadas en el apartado 3.3.

Se analizaron las curvas estándar para corroborar que los parámetros de amplificación se encontraran dentro de los valores adecuados. En la Figura 22 se aprecian las tres curvas estándar en paralelo, resultado de la amplificación con los cebadores Ampliseq (curva estándar 1), cebadores Equalizer (curva estándar 2) y cebadores NEB (curva estándar 3). Tras el análisis de las tres curvas estándar (CT vs. Log(concentración)) se observó que todas presentaron una pendiente cercana al óptimo (-3,32, ver valor "slope"), un R² perfecto (ver valor "R²") y una eficiencia dentro del rango aceptable (mayor a 90%, ver valor "Eff%").

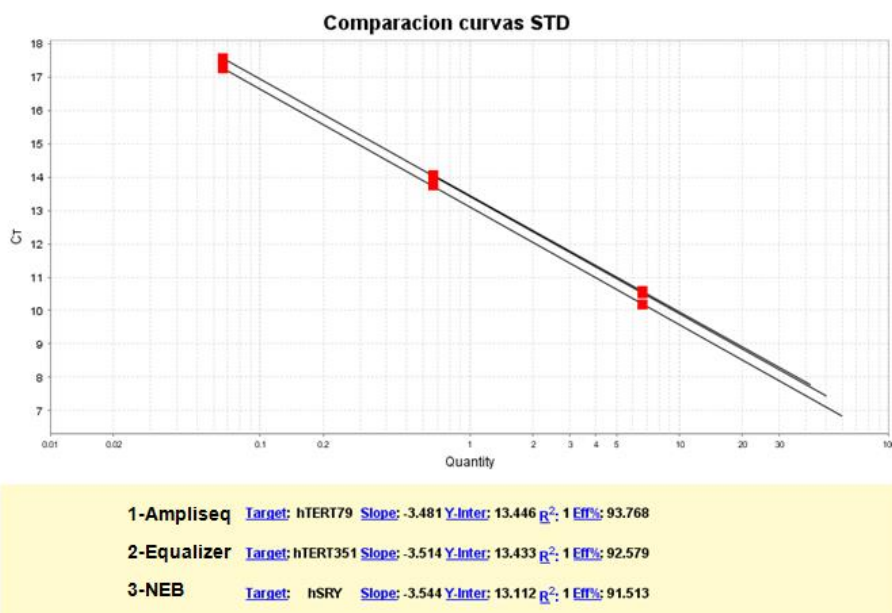


Figura 22: Comparación de las tres curvas estándar tras la amplificación con los cebadores Ampliseq (curva estándar 1), cebadores Equalizer (curva estándar 2) y cebadores NEB (curva estándar 3).

Posteriormente, se evaluaron las curvas de amplificación de las bibliotecas en estudio. Se observó que las bibliotecas cuantificadas puras se inhibieron por el exceso de ADN, tanto las bibliotecas NEB como las Ampliseq, ya que la relación de la concentración de las bibliotecas puras con la concentración de las diluciones 1/10 fue bastante menor a 10 veces (Tabla 20). Lo mismo sucedió entre la dilución 1/10 y 1/100 de las bibliotecas NEB. Por lo tanto, los valores confiables de cuantificación serían aquellos de las diluciones 1/100 para NEB y 1/10 o 1/100 para Ampliseq. Por otro lado se observó que el rango de concentración para la dilución 1/100 de las bibliotecas NEB fue similar al rango de concentración de la dilución 1/10 de las bibliotecas Ampliseq. Es importante recordar que las bibliotecas NEB que fueron analizadas (A y B) no fueron las mismas que las bibliotecas Ampliseq (C y D).

Tabla 20: Resultados de la cuantificación con SYBR Green® para dos muestras realizadas con el protocolo NEB (bibliotecas NEB) y dos muestras realizadas con el protocolo Ampliseq (bibliotecas Ampliseq) utilizando tres pares de cebadores diferentes: NEB, Ampliseq y Equalizer.

Cebadores	Bibliotecas NEB				Bibliotecas Ampliseq			
	Muestra	Pura (pM)	1/10 (pM)	1/100 (pM)	Muestra	Pura (pM)	1/10 (pM)	1/100 (pM)
NEB	A	51,3	69,9	11,9	C	183,9	10,7	1,4
	B	63,3	47,8	6,1	D	166,1	10,8	1,4
Ampliseq	A	17,7	28,6	10,1	C	29,3	9,9	1,8
	B	14,8	25,2	6,1	D	33,9	8,9	1,6
Equalizer	A	19,1	38,1	9,4	C	47,9	9,7	1,3
	B	12,9	31,8	5,0	D	50,3	8,6	1,1

Debido a la similitud detectada entre la dilución 1/100 de las bibliotecas NEB y la dilución 1/10 de las bibliotecas Ampliseq, se decidió medir la concentración de estas mismas bibliotecas utilizando el *Ion Library TaqMan™ Quantitation Kit* [85], método propuesto por el proveedor (*Thermo Fisher Scientific*). Al evaluar los resultados de la cuantificación con sondas *Taqman®*,

se apreció que la curva estándar utilizada cumplía con los parámetros de amplificación (Figura 23), un R^2 cercano a 1 (ver valor “ R^2 ”) y una eficiencia dentro del rango aceptable (mayor a 90%, ver valor “Eff%”). En lo que refiere a las amplificaciones de las bibliotecas evaluadas (diluciones 1/100 de las bibliotecas NEB y diluciones 1/10 de las bibliotecas Ampliseq), lo que se observó fue que las curvas de amplificación de las bibliotecas NEB se detectaron un ciclo más tarde que las curvas de las bibliotecas Ampliseq, a diferencia de lo que se observó en la cuantificación con *SYBR Green*® donde las cuatro muestras presentaron una amplificación similar. Esto se muestra en la Tabla 21 con los valores de cuantificación con cada técnica donde las bibliotecas NEB cuantificadas por Taqman tuvieron una concentración 10 veces menor que las bibliotecas Ampliseq. Esto podría deberse a que para la cuantificación por *Taqman*® la biblioteca NEB no posee lo necesario, ya sea a nivel de secuencia en los adaptadores o algún otro tipo de modificación, para que se pueda realizar de forma correcta la cuantificación. Por lo tanto, con estos resultados, se pudo determinar que la cuantificación por *Taqman*® no sirvió para bibliotecas NEB, con lo que se reafirmó la necesidad de buscar una alternativa de cuantificación. Sin embargo, la cuantificación de las bibliotecas de Ampliseq con *SYBR Green*® y *Taqman*® fue muy similar. Esta comparación corroboró que la cuantificación con *SYBR Green*® permitió cuantificar bibliotecas correctamente.

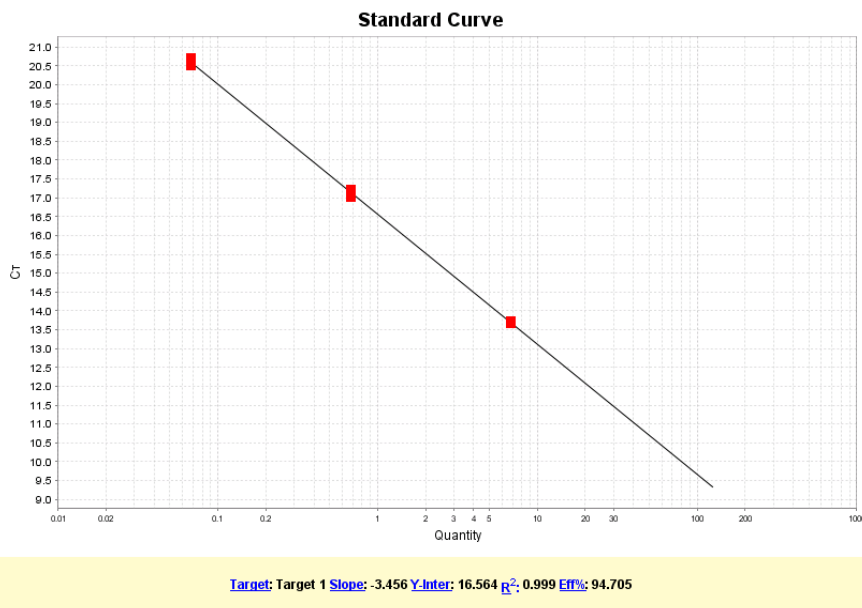


Figura 23: Curva estándar utilizando el Ion Library TaqMan™ Quantitation Kit.

Tabla 21: Comparación de resultados de la cuantificación con *Taqman*® vs. *SYBR Green* (valores promedio de la cuantificación con los diferentes cebadores evaluados) para dos bibliotecas NEB y dos bibliotecas Ampliseq.

Bibliotecas NEB			Bibliotecas Ampliseq		
	Taqman	SYBR Green		Taqman	SYBR Green
Muestra	1/100 (pM)	1/100 (pM)	Muestra	1/10 (pM)	1/10 (pM)
A	1,3	10,5 (promedio)	C	6,8	10,1 (promedio)
B	0,4	5,7 (promedio)	D	8,4	9,4 (promedio)

A partir de los resultados de las dos pruebas realizadas (cuantificación con *SYBR Green*® y *Taqman*®), se pudo concluir que la cuantificación de las bibliotecas por *SYBR Green* podría ser una opción para cuantificar las bibliotecas NEB diluídas 1/100 utilizando cualquiera de los tres juegos de cebadores evaluados ya que los mismos arrojaron resultados similares.

4.2.2.2. Cuantificación por *Qubit*® utilizando el kit *Qubit*® *dsDNA HS*

La segunda alternativa planteada para sustituir la cuantificación por *Taqman*®, fue medir las bibliotecas generadas con el protocolo de *NEB* mediante *Qubit*. Para ello se utilizó el kit *Qubit*™ *dsDNA HS Assay*, el cual trabaja con una sonda que se une específicamente al ADN doble hebra y al hacerlo fluoresce [86]. Las concentraciones que permite medir van desde 10pg/μl a 100ng/μl. La prueba consistió en medir diez bibliotecas NEB puras y diluídas 1/10. El resultado de la cuantificación se detalla en la Tabla 22.

Tabla 22: Resultado de la cuantificación por *Qubit* de 10 bibliotecas puras y diluídas 1/10.

Biblioteca	[Pura] (ng/μl)	[Dil 1/10] (ng/μl)
NEB_1	4,0	0,64
NEB_2	2,4	0,21
NEB_3	3,3	0,27
NEB_4	1,9	0,10
NEB_5	2,5	0,23
NEB_6	1,3	0,10
NEB_7	3,1	0,25
NEB_8	3,9	0,43
NEB_9	1,9	0,15
NEB_10	1,7	0,13

Para poder llevar las bibliotecas a 12pM (que es lo recomendado para realizar la amplificación clonal), se tomó en consideración los valores de la cuantificación de bibliotecas NEB por *SYBR Green*® (Tabla 20, promedio de los resultados con los 3 juegos de cebadores) y se comparó

con los valores de la cuantificación de las mismas bibliotecas puras medidas por *Qubit*. La biblioteca A (Tabla 20) diluída 1/100 tuvo una concentración, en promedio, de 10,5pM, y por *Qubit* midió 3,2ng/μl la pura. La biblioteca B (Tabla 20) diluída 1/100 tuvo una concentración, en promedio, de 5,7pM, y por *Qubit* midió 2,4ng/μl la pura. Al hacer la correlación de las dos cuantificaciones, se llega a que para diluір las bibliotecas a 12pM, se las debe llevar a 0,04 y 0,05 ng/μl, respectivamente. Estos valores se relacionan con las pruebas realizadas en el apartado 4.2.1.2.2 (Tabla 12) donde el caso óptimo para lograr obtener una secuenciación con suficientes perlas cargadas con biblioteca y pocas perlas policlonales se presentó con una concentración de 0,05ng/μl. Con estos datos se estableció un valor de 0,05ng/μl para la concentración de las bibliotecas NEB.

Una vez evaluados los dos métodos alternativos para sustituir el método tradicional de *Taqman®*, se hizo una evaluación de costos de cada uno ellos. Al compararlos se vio que las dos alternativas presentaban una disminución de costos de un 70% respecto al método proporcionado por el proveedor (*Thermo Fisher Scientific*) siendo ambos métodos muy similares en costos (para más detalles leer el apartado 4.2.4). Dada esta similitud, el método elegido fue *Qubit* por su facilidad de manipulación y la rapidez en la obtención de los resultados (menos de 5 minutos). En un laboratorio de diagnóstico cualquier mejora en el proceso que permita ahorrar tiempo es valioso.

4.2.3. Análisis de secuencias y comparación de *software* alternativos

En esta etapa se realizó el análisis de las secuencias generadas por el *Ion Torrent PGM™* utilizando el sistema de análisis de *Thermo Fisher Scientific* para la detección de variantes somáticas. Para ello, se tuvo en cuenta las consecuencias de trabajar con muestras FFPE y los cebadores del panel de genes KN. Por un lado se evaluó la presencia de desaminaciones (cambios C:G>T:A) que generan ruidos inespecíficos de forma aleatoria. Por otro lado, la interferencia por el pseudogen de *KRAS* (*KRAS1*). Luego, se realizó la comparación del análisis de secuencias con dos *software* alternativos al propuesto por *Thermo Fisher Scientific*, y se evaluó si estos presentan mejoras en la sensibilidad de la detección de mutaciones somáticas en bajo porcentaje alélico y la presencia de INDELS.

4.2.3.1. Desaminaciones y pseudogen *KRAS1*

Las desaminaciones son sustituciones C>T o G>A, colectivamente C:G>T:A, causadas por desaminación de citosinas son artefactos comunes al trabajar con tejido FFPE. Los cambios en los nucleótidos inducidos por la desaminación pueden ser causados por la fijación de los tejidos FFPE u otros factores externos, como la radiación UV, y se sabe que están afectados por el

pH, la hipoxia y el calor. En varios estudios de NGS se han informado cambios C:G>T:A inducidos por desaminación en muestras frescas y FFPE. Debido a que la mayoría de estos cambios ocurren con baja frecuencia y se han observado incluso en tejidos libres de tumores, se consideran falsos positivos [152, 153]. Esto genera un problema al interpretar los datos ya que los cambios patogénicos reales de baja frecuencia deben distinguirse de los falsos positivos o el ruido de fondo.

Durante esta tesis, se detectó un gran número de ruidos en muestras de resultado conocido. Para determinar si se trataba de desaminaciones se procesaron por triplicado 5 de estas muestras con alto número de variantes en bajo porcentaje (la cantidad de ADN utilizado para cada una de estas muestra fue: 0,27, 0,45, 0,32, 0,25 y 0,53 ng). De las variantes que se encontraron en estas muestras, un 88,5% eran cambios C:G>T:A. Al realizar la prueba se vio que las variantes que se detectaron en cada una de las muestras no coincidieron entre las réplicas y tampoco coincidieron, en ningún caso, entre las distintas muestras analizadas. Esto permitió deducir que se trataba de ruido y no de variantes reales, ya que el ADN de partida de cada muestra era adecuado (mayor a 0,2ng) por lo que la estocástica no era una variable que afectara. Además se pudo observar que las desaminaciones eran variadas dentro de una misma muestra. Los genomas que la conforman pueden presentar, cada uno de ellos, distintos cambios por desaminaciones, y va a depender de la alícuota que se tome de la muestra, las variantes C:G>T:A (por desaminación) que se detecten.

En la Tabla 23 se detallan las variantes detectadas en las tres réplicas de una de las muestras analizadas en la cual se identificó previamente, por Secuenciación Sanger, la variante patogénica *KRAS* c.35G>T (Figura 24). Dicha variante se logró detectar correctamente mediante NGS en las tres réplicas (chr12: 25398284, C>A) en porcentajes entre 39% y 64% (Tabla 23). Además de la variante patogénica, se observaron otros cambios tanto en el gen *NRAS* (cromosoma 1) como en el gen *KRAS* (cromosoma 12), muchos de las cuales correspondieron a cambios del tipo C:G>T:A. Dentro de ellos, la mayoría no correspondía a sitios *hotspot* de interés en este estudio, por lo cual se pudieron descartar tempranamente. Sin embargo, en la primera réplica, tres variantes pertenecían a posiciones *hotspot* (chr1:115252289, chr1:115256536 y chr12:25380283), pero dado que en las repeticiones de la muestra dichas variantes no volvieron a aparecer, se descartaron como reales. Con este análisis se logró determinar que la única variante real en la muestra era *KRAS* c.35G>T, siendo el resto ruido, en su mayoría probablemente por el efecto de desaminaciones. Además permitió observar que las desaminaciones son aleatorias entre los distintos genomas de una misma muestra, y por tanto en baja proporción, por lo que si se realizan réplicas de la muestra, se pueden descartar aquellas variantes procedentes de desaminaciones por la falta de repetitividad. También se pueden descartar variantes con un porcentaje inferior al límite de detección establecido en este trabajo (ver más adelante en el apartado 4.3. Validación).

Tabla 23: Variantes detectadas en tres réplicas de una muestra analizada desde la amplificación hasta la secuenciación en el Ion Torrent PGM™. Las variantes marcadas en rojo son cambios del tipo C:G>T:A. El cambio marcado en azul es la variante patogénica detectada en la muestra.

Réplica	Chr	Posición	Ref	Cambio	Frecuencia (%)	Hotspot?
1	1	115252289	C	T	4,6	SI
	1	115256519	G	A	3,2	NO
	1	115256536	C	T	3,5	SI
	1	115258751	C	T	2,4	NO
	1	115258766	G	A	2,5	NO
	1	115252221	G	A	4,1	NO
	12	25378585	C	T	4,9	NO
	12	25380283	C	T	3,7	SI
	12	25380290	G	A	4,3	NO
	12	25380316	C	T	2,5	NO
	12	25398187	G	A	4,6	NO
	12	25398213	T	C	2,6	NO
	12	25398219	G	A	2,5	NO
	12	25398223	A	T	2,7	NO
		12	25398284	C	A	39,1
2	1	115252246	C	T	6,8	NO
	1	115256514	G	A	6,4	NO
	12	25380313	C	T	3,8	NO
	12	25380341	G	A	3,8	NO
	12	25398195	A	G	2,8	NO
	12	25398280	G	A	9,8	NO
	12	25398284	C	A	64,1	SI
	12	25398297	C	T	11,5	NO
3	1	115256574	A	G	9,1	NO
	1	115256582	T	C	4,1	NO
	1	115258739	C	T	4,4	NO
	12	25378604	C	T	7,3	NO
	12	25380346	C	T	6,0	NO
	12	25398284	C	A	57,6	SI
12	25398288	C	T	10,1	NO	

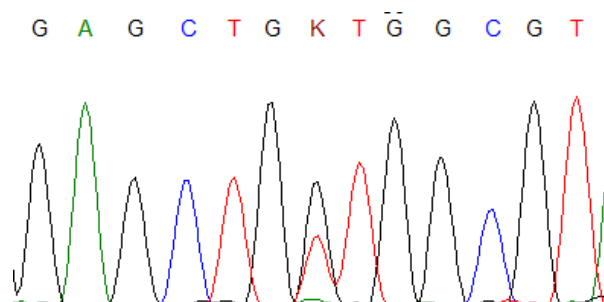


Figura 24: Variante patogénica KRAS c.35G>T detectada por Secuenciación Sanger en la muestra analizada.

Otro aspecto que se evaluó en la etapa de análisis de secuencias fue la presencia del pseudogen de *KRAS* (*KRAS1*) (ver apartado 1.2.1.1). La ubicación de *KRAS1* es en el cromosoma 6 [154-156]. En este trabajo, se identificó la presencia de lecturas correspondientes al pseudogen *KRAS1* tras la secuenciación de los exones 2, 3 y 4 de *KRAS* y *NRAS* con el panel KN. Analizando los resultados de archivos BAM sin procesar (formato binario de alineamiento), es decir sin considerar las regiones *target*, se detectó que la amplificación del pseudogen se generó con los cebadores *KRAS_Ex2_R* y *KRAS_Ex3_F* (en el pseudogen regiones homólogas al exón 2 y exón 3 de *KRAS* se encuentran muy próximas en el ADN sin que exista intrón de por medio). El cebador *KRAS_Ex2_R* no es 100% complementario a la región del pseudogen a la que se une ya que difiere en una base en el extremo 5' del cebador pero igualmente esta diferencia le permitiría hibridar. En cuanto al cebador *KRAS_Ex3_F*, este tiene una base más cerca del extremo 3' que la región complementaria del pseudogen. El amplicón generado con estos cebadores y a partir del pseudogen es de 217pb (Figura 25), un tamaño de amplicón mayor al generado tanto para el amplicón del exón 2 de *KRAS* como para el amplicón del exón 3 de *KRAS*.

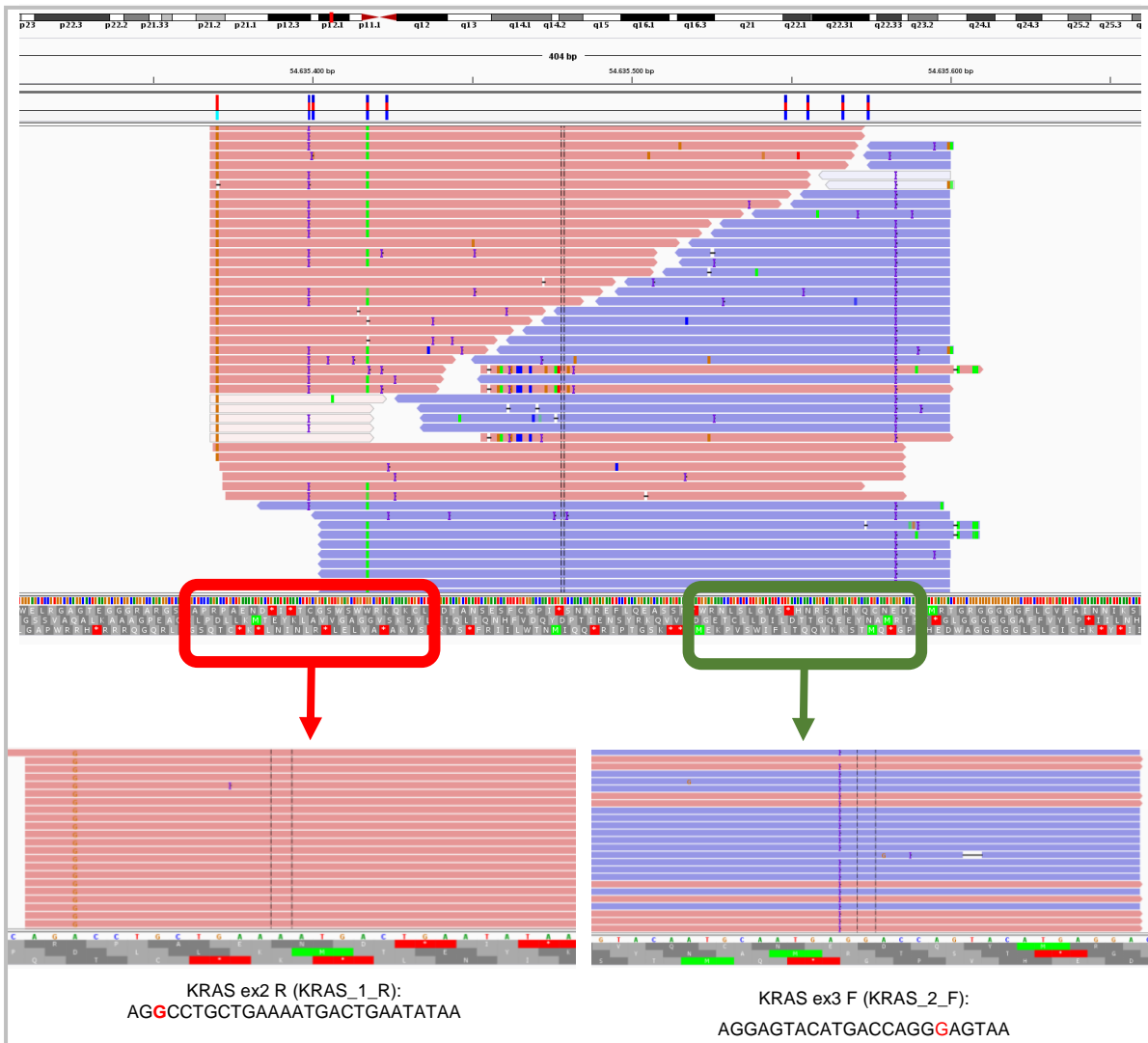


Figura 25: Imagen de IGV en donde se señalan los cebadores que amplifican el pseudogen de *KRAS* (*KRAS1*).

La amplificación del pseudogen no interfiere en el llamado de variantes de los genes *KRAS* y *NRAS*, ya que los fragmentos del amplicón del pseudogen secuenciados alinean con otra región del genoma que no se encuentra en el archivo de "targetregion.bed" con el que se generan los BAM procesados. Sin embargo, durante el proceso de validación (explicado más adelante en el apartado 4.3) se observó que aparecían variantes en *cis* en el exón 2 de *KRAS* y otras en *cis* en el exón 3 de *KRAS*, y todas al mismo porcentaje alélico (un porcentaje siempre menor al 4%). Tras un análisis detallado de estas secuencias y varias hipótesis de por medio, se llegó a la teoría de que podían ser secuencias provenientes del pseudogen. El análisis detallado de las mismas llevó a la conclusión de que las 4 variantes en *cis* encontradas en el exón 2 de *KRAS* y las 4 variantes en *cis* encontradas en el exón 3 de *KRAS* se daban debido a que la secuencia en la que se encontraban las variantes era 100% concordante con la región homóloga en el pseudogen y eran del tamaño de esos amplicones. Una hipótesis fue que en ciertas PCRs, en determinado ciclo, luego de haberse amplificado el pseudogen de *KRAS* y los amplicones correspondientes al exón 2 y 3 de *KRAS*, hubo entrecruzamientos entre amplicones en donde fragmentos incompletos de los amplicones del exón 2 y exón 3 hibridaron sobre el pseudogen generando amplicones quiméricos. De esa forma se habrían generado productos secundarios no deseados provenientes del pseudogen, con el tamaño y secuencia similar a los amplicones del exón 2 y exón 3 de *KRAS*, que alinearían con estos exones y serían analizados posteriormente. Con esta suposición, en el alineamiento de secuencias, las lecturas procedentes del amplicón quimérico se alinearon con los exones de *KRAS* respectivos, dando lugar a variantes en *cis*. En la Figura 26 se observa una lectura con cuatro variantes en *cis* alineada con el exón 2 de *KRAS*. Lo mismo sucede con el exón 3 de *KRAS*, Figura 27.

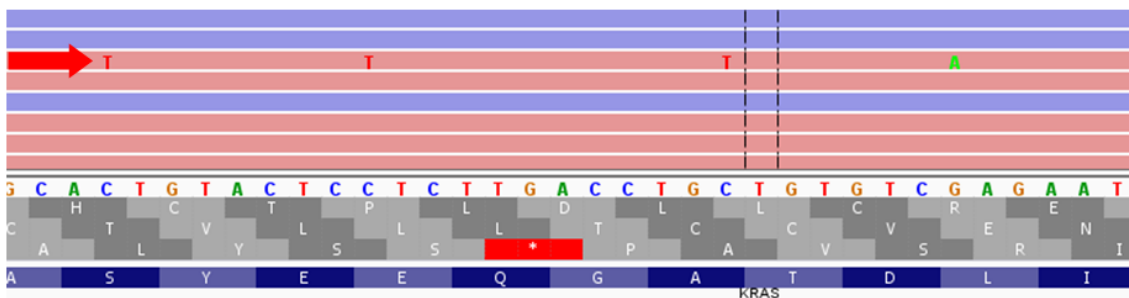


Figura 26: Imagen de IGV de alineamiento de secuencias contra el exón 2 de *KRAS*. Se observa una lectura con cuatro variantes en *cis*.

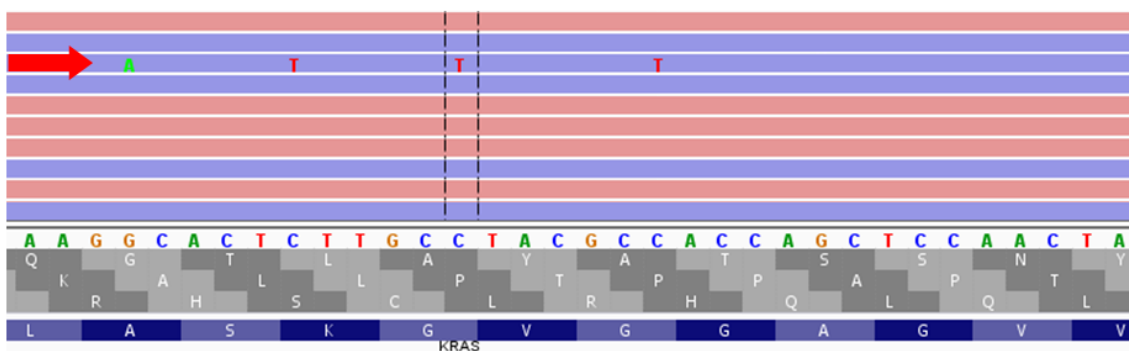


Figura 27: Imagen de IGV de alineamiento de secuencias contra el exón 3 de *KRAS*. Se observa una lectura con cuatro variantes en *cis*.

Con el fin de resolver la amplificación inespecífica dada con el panel KN, se realizó una prueba con cambios en las condiciones de amplificación. Se aumentó 2°C la temperatura de hibridación de los cebadores y se probó reducir a la mitad la cantidad de cebadores utilizados en la reacción de PCR. Para ello, se utilizó una muestra en la que se habían visto lecturas con las variantes en *cis*, descritas anteriormente, en un 4%. En la Tabla 24 se observa la comparación de la muestra realizada con diferentes condiciones de amplificación y las variantes detectadas en cada una de las pruebas. La primera condición: A, condiciones de amplificación con las que se venía trabajando: temperatura de hibridación de cebadores 60°C y mezcla de cebadores 0,2µM (cada uno); B. mezcla de cebadores al medio de la concentración utilizada en A, 0,1µM; C. temperatura de hibridación de cebadores 62°C; y D. mezcla de cebadores al medio de la cantidad utilizada en A y temperatura de hibridación de cebadores 62°C. Los resultados de esta prueba fueron muy claros y alentadores. En la condición A, se observaron las 4 variantes en el exón 3 de *KRAS* y 3 variantes en el exón 2, con porcentajes entre 3 y 4%. Todas estas variantes estaban en *cis* según se vieron en el IGV. Expresamente se procesó una muestra que poseía una variante sobre el exón 2 de *KRAS* y en una de las posiciones *hotspot* donde se encuentra una de las variantes provenientes del pseudogen. Para este caso del cambio en la posición chr12: 25398281 (*KRAS* c.38G>A), en la condición de amplificación A (condición estándar) se vio un porcentaje cercano al 4% de lecturas con este cambio en *cis* con las demás variantes provenientes del pseudogen y también se vio un porcentaje de lecturas con este cambio de forma única en ciertas lecturas. La observación de estas dos poblaciones de lecturas hizo pensar que la variante encontrada en las lecturas de forma única era real, proveniente de la región de interés en el exón 2 de *KRAS*. En la condición B, se observó una situación similar a la anterior, con lo cual se dedujo que la disminución de concentración en los cebadores no generó un cambio significativo. En la condición C, se detectaron menos variantes en *cis*, haciendo pensar que el aumento de la temperatura en la PCR fue un cambio beneficioso. Se observó una mejora en la última condición, donde se probó reducir los cebadores y aumentar la temperatura de hibridación de los mismos. En esta situación, únicamente se detectó la variante chr12: 25398281 al 6,7%, lo que reafirma que esta variante (*KRAS* c.38G>A) es un cambio real en la muestra y no se observaron los cambios de variantes en *cis* en una misma lectura provenientes del pseudogen.

Tabla 24: Detección de variantes en una muestra con diferentes condiciones de amplificación.

Prueba	Chr	Posición	Exón	Ref	Variante	Frecuencia (%)
Condición A 60°C temp. de hibridación de cebadores Mezcla de cebadores panel KN 0,2µM	chr12	25380264	3	C	T	4.6
	chr12	25380272	3	C	T	4.0
	chr12	25380283	3 (hotspot)	C	T	4.1
	chr12	25380290	3	G	A	4.2
	chr12	25398266	2	G	A	3.7
	chr12	25398276	2	C	T	3.2
	chr12	25398281	2 (hotspot)	C	T	7.5
chr12	25398299	2	A	G	4.2	
Condición B 60°C temp. de hibridación de cebadores Mezcla de cebadores panel KN 0,1µM	chr12	25380264	3	C	T	2.8
	chr12	25380272	3	C	T	2.9
	chr12	25380290	3	G	A	2.5
	chr12	25398266	2	G	A	4.5
	chr12	25398276	2	C	T	3.5
	chr12	25398281	2 (hotspot)	C	T	6.4
	chr12	25398299	2	A	G	6.0
Condición C 62°C temp. de hibridación de cebadores Mezcla de cebadores panel KN 0,2µM	chr12	25380264	3	C	T	3.4
	chr12	25380272	3	C	T	3.1
	chr12	25380290	3	G	A	2.6
	chr12	25398281	2 (hotspot)	C	T	6.3
	chr12	25398299	2	A	G	3.0
Condición D 62°C temp. de hibridación de cebadores Mezcla de cebadores panel KN 0,1µM	chr12	25398281	2 (hotspot)	C	T	6.7

Con la prueba realizada se consiguió una condición de PCR que evitó la generación de amplicones procedentes del pseudogen que daban lugar a un alineamiento contra los exones 2 y 3 de *KRAS*, y ello llevaba al llamado de variantes de nucleótidos en los que difiere el pseudogen con el gen funcional. Con ninguna de las condiciones evaluadas se logró disminuir la amplificación del pseudogen de *KRAS* pero sí ese producto quimera. Debido a que este problema se observó al final del trabajo de esta tesis, durante la validación, los cambios en las condiciones de PCR no se llegaron a aplicar en la misma. Por lo tanto, lo que se decidió hacer en las muestras que se realizaron con el protocolo de amplificación original (y aparecieron variantes en *cis* inespecíficas) fue mirar directamente en el IGV dicha posición y analizar si aparecía la variante de forma única en la lectura (en *trans*), si era así se consideraba real. Queda como perspectiva a futuro una nueva validación del sistema modificado. De cualquier forma se chequeó el equilibrio entre amplicones al utilizar las nuevas condiciones de PCR y el mismo no fue afectado, por lo que el sistema prácticamente no difiere del validado, sí es un sistema mejorado respecto a esta característica del producto no deseado que en algunos casos llegó a ser del 4% sobre el producto específico.

4.2.3.2. Comparación de *software* alternativos

Con el propósito de contar con alternativas informáticas para el análisis de las secuencias generadas en el *Ion Torrent PGM™* se seleccionaron dos *software*: *Sophia DDM®* (*Sophia Genetics*) [98] y *CLC Genomics Workbench* (*Qiagen*) [96]. *Sophia DDM®* es una plataforma para la detección de variantes que garantiza un rendimiento analítico de grado clínico y ayuda a los expertos a identificar de manera rápida y precisa mutaciones accionables. Detecta, anota y preclasifica todos los tipos de variantes genómicas (SNV, INDELS, amplificaciones y fusiones) en un experimento único. Para ello, el alineamiento debe realizarse previamente con otro programa. En lo que refiere al *CLC Genomics Workbench*, este *software* está desarrollado para realizar una amplia gama de estudios bioinformáticos de NGS. Los flujos de trabajo pueden combinar pasos de control de calidad, recorte de adaptadores, mapeo de lecturas, detección de variantes y múltiples pasos de filtrado y anotación. Los dos programas evaluados son *software* pagos, sin embargo, para este trabajo, se utilizaron de manera gratuita en modo de prueba, con el objetivo de evaluar y comparar con el ofrecido por el proveedor.

Para comparar los *software* de interés (*Torrent Suite™*, *Sophia DDM®* y *CLC*), se seleccionaron 9 muestras, las cuales presentaron diferentes dificultades de análisis. Algunas de ellas (de resultado conocido) tenían variantes que no fueron detectadas correctamente por el *software* *Torrent Suite™*, otras presentaban deleciones, inserciones o cambios complejos (normalmente cambios difíciles de analizar), y otras muestras a las cuales se les detectó variantes por NGS que no habían sido detectadas previamente por Secuenciación Sanger. En el caso de *Sophia DDM®* se utilizaron las secuencias alineadas (contra el genoma humano) por el *software* *TMAP* (archivos BAM, generado con el programa de *Thermo Fisher Scientific*). En el caso de *CLC* se realizó nuevamente el alineamiento contra el genoma humano (hg19) ingresando los archivos FASTQ obtenidos luego de la secuenciación (el formato FASTQ es un formato basado en texto para almacenar tanto una secuencia biológica, generalmente una secuencia de nucleótidos, como sus correspondientes puntuaciones de calidad). En la Tabla 25 se presentan los resultados de la comparación de las 9 muestras evaluadas. La muestra n^o1 tuvo un resultado *WT* por secuenciación Sanger (sin variantes patogénicas), sin embargo, por NGS se detectó una variante al 7,5% aproximadamente, que se observó por igual con los tres *software* utilizados. Deleciones de varios nucleótidos, como en el caso n^o2, fueron detectadas por los tres *software*. Sin embargo, *CLC* detectó la deleción en menor porcentaje que los otros dos programas, lo que posiblemente se haya debido a una diferencia en el alineamiento generado por el programa. Este tipo de deleciones es muy común en el gen *EGFR*, por lo que es de gran importancia que puedan ser detectadas correctamente. Cambios complejos, como el caso n^o3 y n^o4, e inserciones/duplicaciones de varios nucleótidos, como en el caso n^o5, fueron detectadas en porcentajes similares en todos los casos por los tres *software*. Los casos 6 a 9 se trataron de muestras de resultado conocido (muestras "Horizon", ver apartado 3.1.3) con cuatro variantes patogénicas en una proporción conocida. En estos casos, el *Torrent*

Suite[™] no hizo el llamado de muchas de estas variantes aunque las mismas se encontraban presentes luego del alineamiento (visualizadas mediante IGV). El motivo por el cual no se realizaron los llamados se desconoce. Se hicieron cambios en los parámetros del *plug in* del *Variant Caller* que se encarga del llamado de variantes en el *Torrent Suite*[™] pero ninguno permitió que las reportara correctamente. Sin embargo, *Sophia DDM*[®] y *CLC* realizaron el llamado de las variantes esperadas correctamente. En cuanto a los porcentajes en la detección de las variantes, se observó que la variante en *NRAS* se detectó a un porcentaje bastante menor al esperado, lo que puede deberse a un problema en la realización de la mezcla de ADN con dicha mutación y el ADN normal, y no a un problema bioinformático. Por otro lado, se observó en los cuatro casos (n°6 a 9) la detección de la variante *KRAS c.175G>A* en un porcentaje entre 2,5 y 4%. Esta variante proviene de lecturas del pseudogen que se alinearon incorrectamente con el exón 3 de *KRAS*. Debido a que es un tema en la amplificación (como se vio en el apartado anterior), los tres *software* detectaron las variantes provenientes del pseudogen.

Tabla 25: Comparación de tres software (*Torrent Suite*[™], *Sophia DDM*[®] y *CLC Genomics Workbench*) para el análisis de 9 secuencias generadas en el *Ion Torrent PGM*[™]. Con * se marcan los resultados conocidos a partir de secuenciación Sanger.

Nº	Resultado conocido	<i>Torrent Suite</i> [™]	<i>Sophia DDM</i> [®]	<i>CLC Genomics Workbench</i>
1	*WT	KRAS c.38G>A (7,5%)	KRAS c.38G>A (7,5%)	KRAS c.38G>A (7,7%)
2	*EGFR c.2235_2249del15	EGFR c.2235_2249del15 (13,7%)	EGFR c.2235_2249del15 (12,6%)	EGFR c.2235_2249del15 (2,7%)
3	*EGFR c.2236_2248 CAAC (Complex)	EGFR c.2236_2248 CAAC (Complex) (18,1%)	EGFR c.2236_2248 CAAC (Complex) (16,9%)	EGFR c.2236_2248 CAAC (Complex) (15%)
4	*EGFR c.2573TG>GT	EGFR c.2573TG>GT (75,8%)	EGFR c.2573TG>GT (75%)	EGFR c.2573TG>GT (74,5%)
5	*EGFR c.2300_2308dup CCAGCGTGG	EGFR c.2300_2308dup CCAGCGTGG (21%)	EGFR c.2300_2308dup CCAGCGTGG (22,6%)	EGFR c.2300_2308dup CCAGCGTGG (15,9%)
6	NRAS c.182A>G (6,25%)	No hace el llamado (en IGV se ve el cambio al 1%)	NRAS c.182A>G (1,5%)	NRAS c.182A>G (1,4%)
	KRAS c.436G>A (6,25%)	KRAS c.436G>A (5,5%)	KRAS c.436G>A (5,2%)	KRAS c.436G>A (5,3%)
	KRAS c.183A>C (6,25%)	KRAS c.183A>C (6,1%)	KRAS c.183A>C (6,0%)	KRAS c.183A>C (5,9%)
	-	KRAS c.175G>A (4,7%)	KRAS c.175G>A (4,4%)	KRAS c.175G>A (4,1%)
7	KRAS c.34G>T (6,25%)	KRAS c.34G>T (7,7%)	KRAS c.34G>T (6,9%)	KRAS c.34G>T 7,1%)
	NRAS c.182A>G (3,125%)	No hace el llamado (en IGV se ve el cambio al 2%)	NRAS c.182A>G (1,7%)	NRAS c.182A>G (1,6%)
	KRAS c.436G>A (3,125%)	No hace el llamado (en IGV se ve el cambio al 2%)	KRAS c.436G>A (1,6%)	KRAS c.436G>A (1,6%)
	KRAS c.183A>C (3,125%)	No hace el llamado (en IGV se ve el cambio al 2%)	KRAS c.183A>C (2,5%)	KRAS c.183A>C (2,4%)
8	-	-	KRAS c.175G>A (2,6%)	KRAS c.175G>A (2,5%)
	KRAS c.34G>T (3,125%)	KRAS c.34G>T (3,8%)	KRAS c.34G>T (4,1%)	KRAS c.34G>T (4,0%)
	NRAS c.182A>G (3,125%)	No hace el llamado (en IGV se ve el cambio al 2%)	NRAS c.182A>G (1,8%)	NRAS c.182A>G (1,9%)
	KRAS c.436G>A (3,125%)	KRAS c.436G>A (4,2%)	KRAS c.436G>A (3,6%)	KRAS c.436G>A (4,1%)
9	KRAS c.183A>C (3,125%)	No hace el llamado (en IGV se ve el cambio al 2%)	KRAS c.183A>C (1,9%)	KRAS c.183A>C (4,0%)
	-	-	KRAS c.175G>A (3,4%)	-
	KRAS c.34G>T (3,125%)	KRAS c.34G>T (3,2%)	KRAS c.34G>T (3,6%)	KRAS c.34G>T (4,8%)
	NRAS c.182A>G (3,125%)	No hace el llamado (en IGV se ve el cambio al 3%)	NRAS c.182A>G (2,6%)	NRAS c.182A>G (2,6%)
9	KRAS c.436G>A (3,125%)	No hace el llamado (en IGV se ve el cambio al 3%)	KRAS c.436G>A (3,2%)	KRAS c.436G>A (3,1%)
	KRAS c.183A>C (3,125%)	KRAS c.183A>C (3,6%)	KRAS c.183A>C (3,5%)	KRAS c.183A>C (3,5%)
	-	KRAS c.175G>A (4,2%)	KRAS c.175G>A (4,5%)	KRAS c.175G>A (4,2%)
	KRAS c.34G>T (3,125%)	No hace el llamado (en IGV se ve el cambio al 3%)	KRAS c.34G>T (3,0%)	KRAS c.34G>T (3,0%)

A la luz de estos resultados, se pudo observar que el análisis realizado por el *Torrent Suite*[™] no fue óptimo ya que no logró hacer el llamado de las variantes en el 100% de los casos, específicamente de las variantes en bajo porcentaje (menores a 3%). En cambio el software *Sophia DDM*[®] mostró un rendimiento bueno respecto a lo esperado ya que hizo el llamado correcto de todas las variantes. En cuanto al *CLC*, este software permitió realizar un nuevo alineamiento de las secuencias independiente del paquete de *Thermo Fisher Scientific* y fue capaz de reportar las variantes presentes en las muestras que el *Torrent Suite*[™] no pudo. Por

lo tanto, CLC y Sophia mostraron tener un mejor desempeño que el *Torrent Suite*[™] en la detección de mutaciones cuando el alelo mutado se encuentra en un porcentaje menor a un 3%. Puesto que la tesis establece, más adelante en la etapa de validación (ver apartado 4.3.3), que la sensibilidad es de un 12% para muestras con 0,2 ng de ADN y puede llegar a un 6% con el doble de cantidad de ADN inicial, entonces en estas situaciones se puede analizar con los *plug in* presentes en el *Torrent Suite*[™].

4.2.4. Análisis de costos

Como punto final del proceso de evaluación y puesta a punto de alternativas al protocolo del proveedor *Thermo Fisher Scientific* para la detección de mutaciones somáticas en *KRAS*, *NRAS* y *EGFR*, se realizó un análisis de costos del proceso de NGS desde la extracción de ADN de tejido FFPE hasta el análisis bioinformático. Hubo puntos del proceso donde se hicieron mejoras que no provocaron un cambio en los precios (pero sí en las condiciones de trabajo o en los tiempos del proceso), otros puntos donde la disminución de costos fue notoria y otras etapas que no pudieron ser modificadas porque requerían de los equipos y procesos tal cual se planteaban en el protocolo original. Para este análisis se tomó como base el costo de cada etapa del proceso realizado según las recomendaciones del proveedor y sobre estos costos se realizó la comparación utilizando el sistema puesto a punto en este trabajo. Se incluyó el costo de reactivos, los plásticos y los gastos de importación. El costo en mano de obra no se analizó porque se consideró que es prácticamente el mismo en los dos procesos evaluados. El costo de realizar el panel KN o el panel EGFR no difiere significativamente, por lo que los costos son independientes del panel que se utilice.

El primer punto que se trabajó en esta tesis fue la extracción de ADN de tejidos parafinados (FFPE). El cambio sustancial en esta etapa, en cuanto a reactivos utilizados, fue la sustitución del Xylol por aceite mineral. El precio de ambos reactivos es muy bajo, por lo que no generan un cambio en el costo final de la extracción. El mayor costo de la etapa proviene del kit de extracción *QIAamp DNA FFPE Tissue Kit* de *Qiagen*. Por lo tanto, en esta etapa no se redujo el precio de la extracción con la alternativa planteada, ya que sólo el 0,8% del precio total corresponde al costo del aceite mineral.

El siguiente punto evaluado fue la construcción de bibliotecas. Según lo recomendado por *Thermo Fisher Scientific* cuando se realiza la secuenciación de paneles de genes (diseñados por el proveedor) se debe utilizar el kit *Ion AmpliSeq*[™] *Library Kit 2.0* para las etapas de amplificación, digestión con FuPa y ligación de adaptadores. El valor de este kit implica un costo fijo por muestra y es relativamente alto. Sin embargo, al sustituir dicho kit por la alternativa, amplificación con *Platinum*[®] *Multiplex PCR Master Mix* más el kit *NEBNext*[®] *Fast*

DNA Library Prep Set for Ion Torrent™, el costo final de la construcción de bibliotecas con el sistema NEB (incluida la amplificación con Platinum) permite una **reducción de los costos en un 70%**, es decir que el valor final de una biblioteca por NEB queda igual a un 30% del valor original de Ampliseq.

En cuanto a la cuantificación de bibliotecas (otro de los puntos en el cual se realizó una disminución de costos) la metodología utilizada en el sistema Ampliseq es la cuantificación con sondas *Taqman®* mediante el *Ion Library TaqMan™ Quantitation Kit*. El costo de la cuantificación de una muestra (dentro de un lote de 16 muestras) incluyendo los tres puntos de la curva estándar es relativamente elevado. En cambio, cuando se utilizó la cuantificación por *Qubit®* (con el kit *Qubit® dsDNA HS*) para medir la concentración de la biblioteca pura y diluida al décimo, el costo incluyendo los dos puntos de la curva estándar **disminuyó un 65%**. La otra opción para la cuantificación de bibliotecas utilizando PCR en tiempo real con *Power SYBR® Green PCR Master Mix* tuvo un costo muy similar a *Qubit®* (incluyendo los tres puntos de la curva estándar), no existiendo una diferencia significativa entre estas dos técnicas alternativas y puestas a punto en este trabajo.

Posterior a la cuantificación de las bibliotecas, se realiza la amplificación clonal del *pool* de bibliotecas (OT), el enriquecimiento (ES) y la secuenciación en el equipo Ion Torrent PGM™. Estas tres etapas del proceso se debieron realizar utilizando los equipos y reactivos propuestos por el proveedor de *Ion Torrent™*, por lo que no se planteó una alternativa en el protocolo. Estas etapas producen un costo variable por muestra, dependiendo de la cantidad de muestras que se procesen en un mismo lote, el tipo de *chip* utilizado (*chip 314*, *chip 316* o *chip 318*) y el número de *chip* dentro de una misma inicialización del equipo (*Ion Torrent™ PGM*). Tomando como base un *chip 314* en donde se secuencian 8 muestras (único *chip* por inicialización), se realizó una comparación del costo por muestra con otras condiciones en cuanto al tipo de *chip*, número de muestras y número de *chip* por inicialización (Tabla 26). Se eligieron 8 muestras como referencia dado que es el número mínimo a partir del cual comienza a ser rentable utilizar esta técnica en el Laboratorio Genia.

Tabla 26: Comparación de las etapas OT, ES y secuenciación con diferentes opciones de n° de muestras, tipo de chip y n° de chip por inicialización tomando como base (1) un chip 314 con 8 muestras, único en la inicialización.

OT+ES+ SEC	1 chip/inicialización	2 chip/inicialización
Chip 314 (8 muestras)	1	0,75
Chip 314 (16 muestras)	0,46	0,33

Como se observa en la Tabla 26, el mayor costo se da cuando se procesan 8 muestras en un *chip 314*, único en una inicialización. Cuando aumenta el número de muestras dentro de un lote, el costo por muestra disminuye a la mitad aproximadamente. Es importante destacar que en cada *chip* es recomendable utilizar el mayor número de muestras posibles para reducir los

costos, en el caso de un *chip* 314, el máximo número de muestras realizadas con el panel KN o EGFR (con 1000 lecturas por amplicón) es de 16 (verificado empíricamente). Otro punto importante es poder realizar dos procesos de secuenciación en el equipo *Ion Torrent PGM™* utilizando una misma inicialización del equipo de forma de amortizar el costo de la inicialización. Esto no es siempre posible ya que depende de los tipos de *chip* que se quieran correr y del número de flujos de dNTPs que se quieran utilizar en cada uno y si en el laboratorio es necesario utilizar 2 *chip*. Es decir, en el caso de que sea posible realizar dos corridas en una misma inicialización el costo de la misma se reduce a la mitad por *chip*, incidiendo en el costo por muestra entre un 25% y un 28% (según número de muestras por *chip*), lo que implica una reducción significativa.

Con el propósito de comparar costos con el panel *Ion AmpliSeq Colon and Lung Research Panel v2* (ver apartado 4.2.1.1), se tomó como base el costo por muestra secuenciada en un *chip* 314 con 8 muestras dentro del lote (único *chip* por inicialización) realizada con el panel KN o EGFR (1000x) *versus* el costo de secuenciar una muestra en un *chip* 318 en un lote de 8 muestras (1 *chip* por inicialización, 1000x) realizada con el panel de 22 genes. El costo de realizar una muestra con el panel completo es 4,6 veces mayor que si se realizan los subpaneles KN o EGFR. Con esto se desprende la necesidad de reducir el panel completo cuando lo único que se quiere estudiar es *KRAS/NRAS* o *EGFR*, para poder hacer más accesible estos tipos de estudios.

Para evaluar el costo total de procesar una muestra desde la extracción de ADN hasta la secuenciación y comparar los dos protocolos evaluados (Ampliseq y NEB) se presentan dos ejemplos donde varía el número de muestras dentro del *chip* (Tabla 27). Se toma como referencia un *chip* 314 con 8 muestras (2 *chip* por inicialización) realizados con el sistema Ampliseq. Lo que se observa con los ejemplos planteados es que la diferencia entre procesar una muestra con la alternativa planteada en este trabajo (sistema NEB) es en promedio entre **40% y 50% más barato** que procesar la misma muestra con el sistema planteado por el proveedor de *Ion Torrent™* (*Thermo Fisher Scientific*).

Tabla 27: Comparación de los costos totales de procesar una muestra desde la extracción hasta la secuenciación con los dos sistemas evaluados, Ampliseq y NEB. Varía el nº de muestras por *chip*. Se considera el costo de correr dos *chip* en una misma inicialización.

	Sistema Ampliseq	Sistema NEB
Chip 314 (8 muestras)	1	0,62
Chip 314 (16 muestras)	0,77	0,38

En lo que refiere al análisis bioinformático, última etapa del protocolo, el *Torrent Suite™* software recomendado por el proveedor (*Thermo Fisher Scientific*) está incluido dentro del costo de utilizar el equipo *Ion Torrent PGM™*, por lo que no implica un gasto extra dentro del proceso. En cuanto a los dos software alternativos evaluados, *Sophia DDM®* y *CLC Genomics*

Workbench, ambos fueron utilizados en modo prueba, por lo que no tuvieron costo a la hora de realizar los análisis. En el caso de querer incluirlos en el proceso implicarían un costo extra pero el objetivo en este caso sería mejorar y complementar el análisis de las secuencias si se quiere informar con una sensibilidad mayor al 3%.

Por último, un punto importante que se valoró fue la comparación en costos entre NGS y secuenciación Sanger. Se conoce que NGS tiene muchas ventajas para la detección de mutaciones somáticas con respecto a la técnica de referencia, secuenciación Sanger (ver apartado 1.3), tanto en tiempo de procesamiento, facilidad de manipulación de varias muestras en paralelo y mejor análisis de los resultados puesto que son automatizados los generados por NGS. Sin embargo, fue importante evaluar la diferencia en costos entre estas técnicas para determinar si es competitivo económicamente pasar de Sanger a NGS. Realizando los cálculos de costos de evaluar las mismas regiones de los genes *KRAS/NRAS* o *EGFR*, se observó que la secuenciación Sanger es 12% más costosa que realizar NGS con el nuevo sistema NEB puesto a punto en esta tesis, y realizando 8 muestras por *chip*. Por lo tanto, la elección de NGS frente a Sanger también se vio favorecida por la disminución de costos. Cabe aclarar y se deduce que la técnica de NGS original utilizando el sistema Ampliseq es 70% más cara que la secuenciación Sanger, por ello el cambio de tecnología no era viable desde el punto de vista económico. Debido a los cambios introducidos en la técnica gracias a esta tesis es que el Laboratorio Genia se encuentra en posición de evaluar dicho cambio de tecnología. Además del aspecto económico es importante también la mejora de NGS en la posibilidad de realizar más muestras en paralelo por semana y producir resultados automatizados. Una desventaja de NGS es que es necesario procesar al menos 8 muestras en cada *chip* para que sea viable económicamente, lo que puede significar un factor negativo si solo se tienen 8 muestras por semana pues no habría tiempo para repetir una muestra en el caso de ser necesario. Por ello se plantea un mínimo de 2 *chip* por semana separados en el tiempo por 3 días, para poder incorporar el protocolo de NGS en el laboratorio y mantener un tiempo de entrega de 7 días a nivel de laboratorio de biología molecular.

4.3. Validación

La validación de un procedimiento analítico se realiza para demostrar que un procedimiento es adecuado para su propósito. Implica medir el rendimiento analítico de una prueba sobre un conjunto de métricas predefinidas para demostrar que cumple con las indicaciones de uso y especificaciones predefinidas. Se evalúa si la prueba identifica o mide con éxito, dentro de los límites estadísticos definidos, la presencia o ausencia de una variante que proporcionará información sobre una enfermedad u otra condición en un paciente [71, 157, 158].

Según la norma ISO 9001:2015 [159], la validación se define como “la confirmación, a través de la presentación de pruebas objetivas, de que los requisitos para un uso o aplicación específica prevista se han cumplido”. La normativa internacional solicita que se realice la validación para los productos o servicios de diagnóstico *in vitro* con el propósito de asegurar sus resultados y, por lo tanto, las decisiones terapéuticas que se tomen en base a ellos. Más específicamente la norma ISO 17025:2017 [120] (para laboratorios de ensayo y calibración) recomienda que "El laboratorio debe validar los métodos no normalizados, los métodos desarrollados por el laboratorio y los métodos normalizados utilizados fuera de su alcance previsto o modificados de otra forma. La validación debe ser tan amplia como sea necesario para satisfacer las necesidades de la aplicación o del campo de aplicación dados".

Por las recomendaciones antes mencionadas, se decidió validar el procedimiento puesto a punto en este trabajo para la detección de mutaciones somáticas en los genes *KRAS*, *NRAS* y *EGFR*, desde la etapa de extracción de ADN de tejido FFPE hasta el análisis de las secuencias generadas. Los parámetros evaluados para el panel KN (genes *KRAS* y *NRAS*) fueron el rango de trabajo informable, la especificidad analítica (sustancias interferentes), la sensibilidad analítica (límite de detección), la precisión (repetibilidad) y la exactitud. Para el panel EGFR (gen *EGFR*) se evaluó el rango de trabajo informable, la precisión (repetibilidad) y la exactitud [71].

4.3.1. Rango de trabajo informable

El rango de trabajo informable debe incluir las regiones seleccionadas que cumplan con los requisitos mínimos de calidad, los tipos de variantes que se han validado y los límites de detección para estos. Dicha información debe incluirse en el informe para que se comprenda claramente qué regiones, variantes y porcentajes de mutaciones se detectarán, y cuales no [71, 157].

Para el panel KN, la cantidad de ADN debe encontrarse entre 0,2 y 2ng, los genes evaluados deben ser *KRAS* (exón 2, 3 y 4) y *NRAS* (exón 2, 3 y 4) y la cobertura de la secuenciación mínima admitida debe ser de 1000 lecturas por amplicón. Las variantes evaluadas con este panel se especifican en la Tabla 28 (*KRAS*) y en la Tabla 29 (*NRAS*). Estas variantes en los genes *KRAS* y *NRAS* se asocian a resistencia a los anticuerpos monoclonales cetuximab (Erbix[®]) y panitumumab (Vectibix[®]). Cuando estos genes se encuentran mutados se ha demostrado una reducción de la respuesta en los pacientes con CCRm [37, 39, 46].

Para el panel EGFR, la cantidad de ADN debe encontrarse entre 0,2 y 2ng, el gen evaluado debe ser *EGFR* (exón 18, 19, 20 y 21) y la cobertura de la secuenciación mínima admitida debe ser de 1000 lecturas por amplicón. Las variantes analizadas en el gen *EGFR* (Tabla 30), puntualmente en los exones 18, 19 y 21 en pacientes con NSCLC, se asocian a sensibilidad al tratamiento con inhibidores de tirosina kinasa (TKI) de EGFR. Las variantes en el exón 20 se asocian a resistencia al tratamiento con TKI de EGFR. La ausencia de mutaciones en este gen, se asocia a resistencia a la terapia con TKI [160].

Tabla 28: Mutaciones analizadas en los exones 2, 3 y 4 del gen de *KRAS*. Las variantes se clasifican de acuerdo con las guías del Colegio Americano de Genética Médica y Genómica (ACMG).

Exón	Cambio aminoácido	Cambio de base
2	p.G12C	c.34G>T
	p.G12S	c.34G>A
	p.G12R	c.34G>C
	p.G12V	c.35G>T
	p.G12D	c.35G>A
	p.G12A	c.35G>C
	p.G13C	c.37G>T
	p.G13S	c.37G>A
	p.G13R	c.37G>C
	p.G13V	c.38G>T
	p.G13D	c.38G>A
	p.G13A	c.38G>C
	3	p.A59T
p.A59S		c.175G>T
p.A59E		c.176C>A
p.A59G		c.176C>G
p.Q61K		c.181C>A
p.Q61E		c.181C>G
p.Q61P		c.182A>C
p.Q61L		c.182A>T
p.Q61R		c.182A>G
p.Q61H	c.183A>C	
4	p.K117N	c.351A>C
	p.K117N	c.351A>T
	p.A146T	c.436G>A
	p.A146P	c.436G>C
	p.A146G	c.437C>G
	p.A146V	c.437C>T

Tabla 29: Mutaciones analizadas en los exones 2, 3 y 4 del gen de NRAS. Las variantes se clasifican de acuerdo con las guías del Colegio Americano de Genética Médica y Genómica (ACMG) [161].

Exón	Cambio aminoácido	Cambio de base
2	p.G12C	c.34G>T
	p.G12S	c.34G>A
	p.G12R	c.34G>C
	p.G12V	c.35G>T
	p.G12D	c.35G>A
	p.G12A	c.35G>C
	p.G13C	c.37G>T
	p.G13S	c.37G>A
	p.G13R	c.37G>C
	p.G13V	c.38G>T
	p.G13D	c.38G>A
	p.G13A	c.38G>C
	3	p.A59T
p.A59D		c.176C>A
p.Q61K		c.181C>A
p.Q61E		c.181C>G
p.Q61P		c.182A>C
p.Q61L		c.182A>T
p.Q61R		c.182A>G
p.Q61H		c.183A>C
p.Q61H		c.183A>T
4	p.K117N	c.351G>C
	p.K117N	c.351G>T
	p.A146T	c.436G>A
	p.A146P	c.436G>C
	p.A146G	c.437C>G
	p.A146V	c.437C>T

Tabla 30: Mutaciones analizadas en los exones 18, 19, 20 y 21 del gen de EGFR. Las variantes se clasifican de acuerdo con las guías del Colegio Americano de Genética Médica y Genómica (ACMG) [161].

Exón	Cambio de Aminoácido	Cambio de Base
18	p.G719A	c.2156 G>C
	p.G719S	c.2155 G>A
	p.G719C	c.2155 G>T
19	p.E746_A750del	c.2235_2249 del 15 c.2236_2250 del 15
	p.E746_T751 > I	c.2235_2252 AAT (complex)
	p.E746_T751del	c.2236_2253 del 18
	p.E746_T751Z>A	c.2237_2251 del 15
	p.E746_S752>A	c.2237_2254 del 18
	p.E746_S752>V	c.2237_2255 >T (complex)
	p.E746_S752>D	c.2238_2255 del 18
	p.L747_A750>P	c.2238_2248 >GC (complex)
	p.L747_T751>Q	c.2238_2252 >GCA (complex)
	p.L747_E749del	c.2239_2247 del 9
	p.L747_T751del	c.2239_2253 del 15 c.2240_2254 del 15
	p.L747_S752del	c.2239_2256 del 18
	p.L747_T751>P	c.2239_2248 TTAAGAGAAG>C
	p.L747_P753>Q	c.2239_2258 >CA (complex)
	p.L747_T751>S	c.2240_2251 del 12
	p.L747_P753>S	c.2240_2257 del 18
	p.L747_T751>P	c.2239_2251 >C (complex)

	p.T790M	c.2369 C>T
	p.S768I	c.2303 G>T
	p.D770_N771 ins SVD	c.2311_2312 ins9 (GCGTGGACA)
	p.V769_D770 ins ASV	c.2308_2309 ins9 (CCAGCGTGG)
	p.A767_S768 ins TLA	c.2302_2303 ins9 (CGCTGGCCA)
	p.D770N	c.2308 G>A
	p.H773_V774 ins NPH	c.2319_2320 ins9 (AACCCCCAC)
	p.H773_V774 ins NPH	c.2319_2320 ins9 (AACCCCCAC)
20	p.V774_C775 ins HV	c.2321_2322 ins6 (CCACGT)
	p.H773_V774 ins PH	c.2319_2320 ins6 (GCCCCAC)
	p.H773_V774 ins AH	c.2320_2321 ins6 (CCCACG)
	p.D770_N771 ins GT	c.2310_2311 ins6 (GGCACA)
	p.D770_N771 ins GF	c.2310_2311 ins6 (GGTTT)
	p.D770>GY	c.2308_2309 ins3 (GTT)
	p.D770_N771 ins Y	c.2310_2311 ins3 (TAC)
	p.H773_V774 ins H	c.2319_2320 ins3 (CAC)
	p.N771_p772 ins H	c.2314_2315 ins3 (ACC)
21	p.L858R	c.2573 T>G
	p.L861Q	c.2582 T>A

4.3.2. Especificidad analítica (sustancias interferentes)

La especificidad se define como la capacidad de evaluar solamente el analito blanco en presencia de componentes que se espera que se encuentren en la muestra (impurezas, degradantes, matriz, etc.). Para laboratorios de genética molecular, se define como la capacidad de un ensayo de distinguir la secuencia blanco, alelo o mutación, de otras secuencias o alelos presentes en la muestra analizada [157, 162].

Las sustancias interferentes que se sabe afectan las pruebas moleculares, particularmente los métodos basados en amplificación, deben abordarse durante la validación. Las etapas de extracción y purificación de ácidos nucleicos normalmente eliminan posibles contaminantes. Sin embargo, se debe tener en cuenta los tipos de muestras que se utilizan en la validación así como la presencia de secuencias repetitivas y pseudogenes [71].

Para determinar la especificidad de la técnica de NGS modificada en este trabajo, se realizaron dos pruebas. El estudio se realizó únicamente con el panel KN y sobre algunas posiciones relevantes. La primera prueba consistió en analizar 12 posiciones del gen *KRAS* y 12 posiciones del gen *NRAS* (Tabla 31) (mínimo 37 veces) en muestras control que se conoce que no presentan ninguna variante en esas posiciones (muestras Horizon y ADN ctrl, ver apartado 3.1.3 y 3.1.4).

Tabla 31: Análisis de 12 posiciones del gen *KRAS* y 12 posiciones del gen *NRAS* en muestras control WT para esas posiciones, para estudiar especificidad analítica.

Gen	Exón	Posición en transcripción	Posición en genoma	Nº veces analizada la pos.	Nº de muestras con llamados	
<i>KRAS</i> (chr12)	2	c.34	25398285	37	0	
	2	c.35	25398284	66	0	
	2	c.37	25398282	81	0	
	2	c.38	25398281	81	0	
	3	c.175	25380283	81	8 (c.175G>A)	
	3	c.176	25380282	81	0	
	3	c.181	25380277	81	0	
	3	c.182	25380276	81	0	
	3	c.183	25380275	40	0	
	4	c.351	25378146	81	0	
	4	c.436	25378562	40	0	
	4	c.437	25378561	81	0	
	<i>NRAS</i> (chr1)	2	c.34	115258748	81	0
		2	c.35	115258747	79	0
2		c.37	115258745	81	0	
2		c.38	115258744	81	0	
3		c.175	115256536	81	0	
3		c.176	115256535	81	0	
3		c.181	115256530	79	0	
3		c.182	115256529	47	0	
3		c.183	115256528	79	0	
4		c.351	115252289	81	0	
4		c.436	115252204	81	0	
4		c.437	115252203	81	0	

Los resultados de esta prueba muestran que la técnica no detecta variantes inespecíficas en 23 de las 24 posiciones evaluadas. El único caso en donde se detectó una variante inespecífica fue en la posición chr12:25380283 donde en el 10% de los casos estudiados se detectó la variante c.175G>A (entre 2-4% de frecuencia alélica) en el gen *KRAS*. La presencia de esta variante inespecífica concuerda con la interferencia del pseudogen de *KRAS* planteado en el apartado 4.2.3.1, problema que fue solucionado posterior a la validación, cambiando las condiciones de PCR. Este punto de la validación permitió detectar el problema en esta posición y fue tratado en la sección mencionada anteriormente como un problema bioinformático, aunque tiene un componente técnico también.

Para la segunda prueba de especificidad, se evaluaron 9 ADNs de referencia para *KRAS* y 5 para *NRAS*, adquiridos en la empresa Horizon Diagnostic (ver apartado 3.1.3). Dado que cada uno de dichos ADNs contiene una mutación conocida, se pudo evaluar la especificidad de la técnica verificando que la misma lograra detectar cada una de las mutaciones adecuadamente.

Para cada uno de los ADNs de referencia, la técnica de NGS detectó específicamente la mutación presente sin ninguna otra variante inespecífica.

A partir de las dos pruebas realizadas, se definió que la **especificidad para la técnica de NGS fue del 99,5%**. Este valor no es 100% debido a la detección inespecífica de la variante c.175G>A en el gen *KRAS* en 8 eventos de 1763 analizados.

4.3.3. Sensibilidad analítica (límite de detección)

La sensibilidad analítica, también conocida como límite de detección (LD), permite determinar la cantidad mínima de analito que se puede detectar en una muestra [71, 123, 157, 163, 164]. En este caso el analito que se quiere identificar es el alelo mutado en un fondo de ADN normal (*WT*). El LD se calcula como la concentración más baja de analito en la cual al menos el 95% de las llamadas positivas se dan en forma correcta entre las réplicas analizadas para esa concentración [123]. El estudio se realizó únicamente para el panel KN y sobre algunas posiciones relevantes.

Para definir el LD del panel KN en este trabajo (equivalente al porcentaje mínimo de una mutación en una muestra) se estudiaron ADNs mutados de referencia para *KRAS* y *NRAS*, adquiridos en la empresa *Horizon Diagnostic* (ver apartado 3.1.3) con diferentes porcentajes de mutación. Se realizó un análisis diferencial del LD utilizando 50 copias de genoma y 500 copias como cantidad de ADN molde, de esa forma se estudió el LD a diferentes cantidades de molde, puesto que se sabía que la cantidad de molde influye en la estocástica en ciertas condiciones límite de LD. El LD para una cantidad de ADN equivalente a 50 copias de genoma se determinó analizando por quintuplicado cuatro ADNs de referencia (3 con mutaciones en *KRAS* y 1 con mutación en *NRAS*) con porcentajes variables de ADN mutado: 12%, 6% y 3% (porcentaje de alelo mutado con respecto a alelo *WT*). Para la generación de los mismos se mezcló ADN mutado de frecuencia alélica 50% con ADN normal también de referencia comprado a la empresa *Horizon Diagnostic* [72], hasta llegar a los porcentajes teóricos deseados. Se determinó el LD para cada una de las mutaciones por separado. El valor de LD asignado para cada mutación correspondió al mínimo porcentaje de mutación detectada en todas las réplicas analizadas, es decir que en el siguiente punto de dilución de la mutación se falló en detectar la mutación a la frecuencia esperada en todas las réplicas, estableciendo por ello como LD el punto anterior de dilución. Para evaluar el LD de una cantidad de ADN equivalente a 500 copias de genoma (2ng de ADN), se analizaron por quintuplicado los mismos cuatro ADNs de referencia que en el caso anterior, con porcentajes variables de ADN mutado: 6%, 3% y 1,5% (porcentaje de alelo mutado con respecto a alelo *WT*). Se determinó el LD para cada una de las mutaciones por separado. Los resultados de LD para 50 y 500 copias de

genoma se resumen en la Tabla 32 y se presentan los datos en detalle para cada réplica en el Anexo 8.3.

Tabla 32: LD para distintas mutaciones en KRAS y NRAS con cantidades de ADN equivalentes a 50 y 500 copias de genoma. Se realizó el ensayo por quintuplicado y se estableció el LD para cada mutación como el mínimo porcentaje de mutación para el cual, en todas las réplicas analizadas, se detectó la mutación.

Nº catálogo	Gen	Chr.	Posición	Referencia	Variante	LD 50 copias	LD 500 copias
HD411	KRAS	12	25378562	C	T	12%	1,5%
HD632	KRAS	12	25380275	T	G	6%	1,5%
HD378	KRAS	12	25398285	C	A	12%	1,5%
HD534	NRAS	1	115256529	T	C	6%	1,5%

Según los resultados de esta prueba, el LD para muestras con 50 copias de genoma se encontró entre 6% y 12%, dependiendo de la posición estudiada, puesto que para 2 mutaciones hubo 1 y 2 réplicas en 6% en la que no se detectó variantes. Extrapolando los valores obtenidos a otras mutaciones, se estableció como LD en muestras con 50 copias de genoma un valor de 12% (correspondiente al LD aceptable en 2 de las mutaciones, por ende se tomó como valor para todas las mutaciones). Por otro lado, el LD en ensayos con 500 copias de genoma utilizado como molde, para las cuatro variantes estudiadas, fue de 1,5%. Extrapolando este resultado a todas las mutaciones estudiadas con el panel KN, se estableció como LD en muestras con 500 copias de genoma un valor de 2%. Se constató que a mayor molde menor efecto estocástico y por ende es posible llegar a un menor LD, en 50 genomas el equivalente a 12% es 6 genomas y en 500 genomas el equivalente a un 1,5% es 6 genomas mutados, la siguiente dilución equivalente a 3 genomas no fue posible detectarla, por efecto estocástico en la toma de muestra (no es posible pipetear 3 copias de un blanco de forma reproducible). Debido a que el LD debe ser único para el diagnóstico, se estableció que para la detección de mutaciones somáticas en los genes *KRAS* y *NRAS* en muestras con más de 50 copias de genoma o más de 0,2 ng de ADN (menores a este valor, se rechazan) el **LD a informar de la técnica es de 12%** en caso de muestras WT o sin mutaciones detectadas.

4.3.4. Precisión (repetibilidad)

La precisión de un procedimiento analítico expresa la proximidad (grado de dispersión) o consistencia entre los resultados obtenidos tras una serie de medidas (muestreo múltiple) de una misma muestra homogénea [157]. Los procesos complejos y de varios pasos, como lo es la técnica de NGS, pueden introducir errores aleatorios en cada paso debido a la variación en la instrumentación, los reactivos y la técnica. Para minimizar esta variación, los instrumentos, los reactivos y el personal deben estar calificados para el propósito previsto. Sin embargo, la variación puede ocurrir y esto debe cuantificarse a través de la validación del método. Dado el número de posibles fuentes de variación, no es práctico evaluar exhaustivamente todas las

fuentes de forma independiente. Más bien, se recomienda evaluar un mínimo de tres muestras en todos los pasos y durante un período prolongado para incluir todos los instrumentos, el personal de pruebas y varios lotes de reactivos [71, 72].

Dentro de la precisión se encuentra la **repetibilidad**, que es la concordancia entre los resultados de ensayos independientes realizados bajo las mismas condiciones operativas (la misma muestra, analizadas por el mismo analista, en el mismo laboratorio, con los mismos equipos y reactivos y en diferentes días). Para evaluar la repetibilidad se deben realizar pruebas de repetición (entre diferentes ejecuciones) y replicación (dentro de la misma ejecución) [157, 162, 164, 165].

En este trabajo, se evaluó la repetibilidad (repetición y replicación) del panel KN y la repetibilidad (repetición) del panel EGFR. Para evaluar la repetibilidad (entre ejecuciones) del panel KN se analizaron 17 ADNs de distintos orígenes. Dichos ADNs fueron evaluados en tres días diferentes por el mismo analista y en el mismo laboratorio para poder compararlos y determinar la concordancia entre los resultados. En la Tabla 33 se detallan los resultados obtenidos para las 17 muestras.

Tabla 33: Repetibilidad (entre ejecuciones) de 17 muestras analizadas con el panel KN en tres días diferentes por el mismo analista y en el mismo laboratorio.

Muestra	Resultado día 1	%	Resultado día 2	%	Resultado día 3	%
1	WT	-	WT	-	WT	-
2	NRAS c.35G>A	37,4	NRAS c.35G>A	34,4	NRAS c.35G>A	37,4
3	NRAS c.37G>C	33,9	NRAS c.37G>C	31,9	NRAS c.37G>C	40,5
4	KRAS c.35G>C	19,2	KRAS c.35G>C	15,9	KRAS c.35G>C	14,0
5	KRAS c.38G>A	7,5	KRAS c.38G>A	7,0	KRAS c.38G>A	8,1
6	WT	-	WT	-	WT	-
7	KRAS c.34G>T	38,2	KRAS c.34G>T	37,5	KRAS c.34G>T	37,9
	KRAS c.35G>T	38,2	KRAS c.35G>T	37,5	KRAS c.35G>T	37,9
	NRAS c.35G>A	43,7	NRAS c.35G>A	48,1	NRAS c.35G>A	55,6
8	KRAS c.35G>T	39,1	KRAS c.35G>T	64,1	KRAS c.35G>T	57,6
9	WT	-	WT	-	WT	-
10	WT	-	WT	-	WT	-
11	KRAS c.33T>C	30,6	KRAS c.33T>C	29,2	KRAS c.33T>C	28,3
	KRAS c.34G>T	30,6	KRAS c.34G>T	29,2	KRAS c.34G>T	28,3
12	KRAS c.34G>T	53,9	KRAS c.34G>T	43,9	KRAS c.34G>T	40,2
13	KRAS c.35G>T	34,8	KRAS c.35G>T	40,9	KRAS c.35G>T	38,9
14	KRAS c.183A>C	40,9	KRAS c.183A>C	46,2	KRAS c.183A>C	43,5
15	KRAS c.436G>A	61,9	KRAS c.436G>A	48,4	KRAS c.436G>A	54,9
16	NRAS c.183A>T	47,1	NRAS c.183A>T	27,5	NRAS c.183A>T	35,4
17	WT	-	WT	-	WT	-

Al observar la Tabla 33 se aprecia que en todos los casos estudiados los resultados obtenidos en las tres réplicas fueron iguales, con pequeñas variaciones en el porcentaje de detección de las variantes. De las 17 muestras analizadas, 5 muestras fueron WT, 9 muestras tuvieron 1 o 2 mutaciones en el gen KRAS y 4 muestras se encontraban mutadas en el gen NRAS.

Para evaluar la repetibilidad (dentro de la misma ejecución) del panel KN con muestras dentro de una misma corrida, se analizaron 5 ADNs por triplicado. En los datos de la Tabla 34 se destaca que las 5 muestras procesadas en la misma ejecución arrojaron el mismo resultado y con porcentajes de mutación cercanos.

Tabla 34: Repetibilidad (en la misma ejecución) de 5 muestras analizadas con el panel KN en la misma ejecución.

Muestra	Resultado réplica 1	%	Resultado réplica 2	%	Resultado réplica 3	%
1	WT	-	WT	-	WT	-
2	NRAS c.35G>A	37,4	NRAS c.35G>A	30,3	NRAS c.35G>A	47,4
3	NRAS c.37G>C	33,9	NRAS c.37G>C	33,8	NRAS c.37G>C	41,4
4	KRAS c.35G>C	19,2	KRAS c.35G>C	15,2	KRAS c.35G>C	17,3
5	KRAS c.38G>A	7,5	KRAS c.38G>A	7,2	KRAS c.38G>A	7,9

Dada la concordancia entre los resultados de repetición y replicación de las muestras analizadas con el panel KN, se consideró que la **repetibilidad fue del 100%** tanto para la detección de mutaciones en el gen *KRAS* como para la detección de mutaciones en el gen *NRAS*.

Por último, para evaluar la repetibilidad (entre ejecuciones) del **panel EGFR** se analizaron 11 muestras, 10 de ellas provenientes de controles de calidad y 1 ADN de sangre (ADN ctrl). Se evaluaron en tres días diferentes en las mismas condiciones de analista, equipos y reactivos. En la Tabla 35 se detallan los resultados obtenidos para las 11 muestras.

Tabla 35: Repetibilidad (entre ejecuciones) de 11 muestras analizadas con el panel EGFR en tres días diferentes por el mismo analista y en el mismo laboratorio.

Muestras	Resultado corrida 1	%	Resultado corrida 2	%	Resultado corrida 3	%
1	EGFR c.2573T>G	12,3	EGFR c.2573T>G	16,4	EGFR c.2573T>G	14,3
	EGFR c.2369C>T	9,9	EGFR c.2369C>T	13	EGFR c.2369C>T	10,5
2	WT	-	WT	-	WT	-
3	EGFR c.2155G>A	9,8	EGFR c.2155G>A	15,3	EGFR c.2155G>A	12,8
4	WT	-	WT	-	WT	-
5	WT	-	WT	-	WT	-
6	WT	-	WT	-	WT	-
7	WT	-	WT	-	WT	-
8	WT	-	WT	-	WT	-
9	WT	-	WT	-	WT	-
10	WT	-	WT	-	WT	-
11	WT	-	WT	-	WT	-

Los resultados de las tres réplicas de las muestras analizadas con el panel EGFR fueron idénticas en todas las ocasiones, generando que la **repetibilidad sea del 100%** para la detección de mutaciones en el gen *EGFR*.

4.3.5. Exactitud

La exactitud de un procedimiento analítico expresa el grado de concordancia entre el valor aceptado como verdadero (o el valor de una referencia) y el valor hallado al realizar el ensayo. Para las pruebas basadas en NGS, la exactitud representa el grado de concordancia de los resultados entre una secuencia obtenida por NGS y la misma secuencia determinada por un método de comparación válido identificado como apropiado por la FDA (*Food and Drug Administration*), o entre una muestra de referencia ejecutada por NGS y el resultado de alta confianza de la referencia [72, 157, 163, 164]. Para establecer el grado de exactitud previamente se deben calcular tres parámetros: PAP (porcentaje de acuerdo positivo), PAN (porcentaje de acuerdo negativo) y VPTP (valor predictivo técnico positivo). El PAP se calcula dividiendo el número de variantes conocidas detectadas por NGS ("verdaderos positivos" (VP)) sobre el número total de variantes testeadas (VP más "falsos negativos" (FN)). El PAN se calcula dividiendo el número de "verdaderos negativos" (VN) sobre el número de resultados wild type (WT) (VN más "falsos positivos" (FP)). El VPTP se calcula dividiendo el número de VP de NGS sobre el total de resultados positivos (VP más FP) [123]. En este trabajo, se evaluó la exactitud del panel KN y del panel EGFR.

Para evaluar la exactitud del sistema puesto a punto se realizaron dos comparaciones, una de ellas contra los resultados de muestras de referencia (controles de calidad y ADN de referencia adquiridos en la empresa *Horizon Diagnostic*) y otra contra el método de referencia, secuenciación Sanger (técnica "Gold Standard") aplicado a muestras clínicas. Con respecto al panel KN, se compararon los resultados de 26 muestras analizadas con el sistema NEB contra los resultados de muestras de referencia, 12 de ellas comparadas con los resultados del control de calidad *UK NEQAS* y las restantes 14 contra el resultado conocido de las muestras de ADN Horizon. La Tabla 36 presenta los resultados de la comparación y se observa que en todos los casos, los resultados fueron coincidentes. Con estos resultados, se puede decir que la **exactitud de la técnica es del 100%**.

Tabla 36: Exactitud contra muestras de referencia. Comparación de 26 muestras analizadas con el panel KN con la técnica en estudio (NGS con sistema NEB) y los resultados de dos tipos de referencias, control de calidad UK NEQAS y ADN Horizon.

Muestras	Resultado NGS	Resultado referencia	Tipo de referencia
1	WT	WT	Control de calidad
2	WT	WT	Control de calidad
3	KRAS c.35G>A	KRAS c.35G>A	Control de calidad
4	KRAS c.35G>T	KRAS c.35G>T	Control de calidad
5	KRAS c.436 G>A	KRAS c.436 G>A	Control de calidad
6	WT	WT	Control de calidad
7	KRAS c.34G>A	KRAS c.34G>A	Control de calidad
8	NRAS c.35G>A	NRAS c.35G>A	Control de calidad
9	KRAS c.437C>T	KRAS c.437C>T	Control de calidad
10	WT	WT	Control de calidad
11	KRAS c.34G>T	KRAS c.34G>T	Control de calidad
12	NRAS c.37G>C	NRAS c.37G>C	Control de calidad
13	KRAS c.35G>C	KRAS c.35G>C	ADN de referencia "Horizon"
14	KRAS c.34G>T	KRAS c.34G>T	ADN de referencia "Horizon"
15	KRAS c.35G>A	KRAS c.35G>A	ADN de referencia "Horizon"
16	KRAS c.34G>C	KRAS c.34G>C	ADN de referencia "Horizon"
17	KRAS c.34G>A	KRAS c.34G>A	ADN de referencia "Horizon"
18	KRAS c.35G>T	KRAS c.35G>T	ADN de referencia "Horizon"
19	KRAS c.38G>A	KRAS c.38G>A	ADN de referencia "Horizon"
20	KRAS c.183A>C	KRAS c.183A>C	ADN de referencia "Horizon"
21	KRAS c.436G>A	KRAS c.436G>A	ADN de referencia "Horizon"
22	NRAS c.35G>T	NRAS c.35G>T	ADN de referencia "Horizon"
23	NRAS c.183A>T	NRAS c.183A>T	ADN de referencia "Horizon"
24	NRAS c.181C>A	NRAS c.181C>A	ADN de referencia "Horizon"
25	NRAS c.182A>T	NRAS c.182A>T	ADN de referencia "Horizon"
26	NRAS c.182A>G	NRAS c.182A>G	ADN de referencia "Horizon"

Por otro lado, se realizó la comparación del panel KN con la técnica de referencia (secuenciación Sanger) analizando 64 muestras clínicas (a partir de FFPE). Los resultados se presentan en la Tabla 37.

Tabla 37: Exactitud contra método de referencia. Comparación de 64 muestras clínicas provenientes de tejido FFPE analizadas con el panel KN con la técnica en estudio (NGS con sistema NEB) y los resultados de las mismas muestras realizadas por secuenciación Sanger.

Muestras	Resultado NGS	Resultado referencia	Tipo de referencia
1	KRAS c.34G>T	KRAS c.34G>T	Sec. Sanger
2	KRAS c.35G>T	KRAS c.35G>T	Sec. Sanger
3	KRAS c.437C>T	KRAS c.437C>T	Sec. Sanger
4	KRAS c.35G>A	KRAS c.35G>A	Sec. Sanger
5	KRAS c.34G>T	KRAS c.34G>T	Sec. Sanger
6	WT	WT	Sec. Sanger
7	KRAS c.34G>T KRAS c.35G>T NRAS c.35G>A	KRAS c.34G>T KRAS c.35G>T NRAS c.35G>A	Sec. Sanger
8	KRAS c.183A>C	KRAS c.183A>C	Sec. Sanger
9	NRAS p.179G>A	NRAS p.179G>A	Sec. Sanger
10	WT	WT	Sec. Sanger
11	KRAS c.35G>T	KRAS c.35G>T	Sec. Sanger
12	WT	WT	Sec. Sanger
13	WT	WT	Sec. Sanger
14	WT	WT	Sec. Sanger
15	WT	WT	Sec. Sanger
16	WT	WT	Sec. Sanger
17	KRAS c.35G>A	KRAS c.35G>A	Sec. Sanger
18	WT	WT	Sec. Sanger
19	WT	WT	Sec. Sanger
20	KRAS c.34G>A	KRAS c.34G>A	Sec. Sanger
21	WT	WT	Sec. Sanger
22	KRAS c.35G>T	KRAS c.35G>T	Sec. Sanger
23	WT	WT	Sec. Sanger
24	KRAS c.34G>A	KRAS c.34G>A	Sec. Sanger
25	KRAS c.34G>T	KRAS c.34G>T	Sec. Sanger
26	KRAS c.35G>A	KRAS c.35G>A	Sec. Sanger
27	WT	WT	Sec. Sanger
28	WT	WT	Sec. Sanger
29	WT	WT	Sec. Sanger
30	WT	WT	Sec. Sanger
31	WT	WT	Sec. Sanger
32	WT	WT	Sec. Sanger
33	KRAS c.35G>T	KRAS c.35G>T	Sec. Sanger
34	KRAS c.35G>A	KRAS c.35G>A	Sec. Sanger
35	NRAS c.179G>A	NRAS c.179G>A	Sec. Sanger
36	KRAS c.34G>A	KRAS c.34G>A	Sec. Sanger
37	WT	WT	Sec. Sanger
38	WT	WT	Sec. Sanger
39	KRAS c.183A>C	KRAS c.183A>C	Sec. Sanger
40	KRAS c.38G>A	KRAS c.38G>A	Sec. Sanger
41	WT	WT	Sec. Sanger
42	WT	WT	Sec. Sanger
43	WT	WT	Sec. Sanger
44	WT	WT	Sec. Sanger
45	WT	WT	Sec. Sanger
46	WT	WT	Sec. Sanger
47	WT	WT	Sec. Sanger
48	WT	WT	Sec. Sanger
49	WT	WT	Sec. Sanger
50	WT	WT	Sec. Sanger
51	WT	WT	Sec. Sanger
52	WT	WT	Sec. Sanger
53	KRAS c.35G>C	WT	Sec. Sanger
54	KRAS c.38G>A	WT	Sec. Sanger
55	WT	WT	Sec. Sanger
56	WT	WT	Sec. Sanger
57	WT	WT	Sec. Sanger
58	WT	WT	Sec. Sanger
59	WT	WT	Sec. Sanger
60	WT	WT	Sec. Sanger
61	WT	WT	Sec. Sanger
62	WT	WT	Sec. Sanger
63	WT	WT	Sec. Sanger
64	WT	WT	Sec. Sanger

Como se observa en la Tabla 37, hubo dos casos en donde los resultados no fueron coincidentes (muestra 53 y 54). Por lo tanto, se realizaron los cálculos necesarios para definir el valor de exactitud. Primero se calculó el PAP como $21/(21+0)$, dando como resultado un 100% (lo que indica que no existen FN). El PAN fue de 95,3% ($41/(41+2)$), generando un valor de FP de 4,7% ($FP=1-PAN$). Por último se calculó el valor de VPTP que fue 91% ($21/(21+2)$). Los valores necesarios para estos cálculos se detallan en la Tabla 38. Según los cálculos realizados, al comparar la técnica de NGS (con panel KN) contra el resultado de la secuenciación Sanger dio como resultado un 4,3% de FP. Sin embargo, no se trató de un FP ya que ambas variantes se confirmaron realizando la técnica de NGS por triplicado. La variante de la muestra 53 tuvo un porcentaje de mutación promedio de 16%, mientras que la muestra 54 presentó la variante en un 7,5% en promedio. Ambos porcentajes de mutación son bajos para poder ser detectados por la técnica de Sanger, cuyo límite de detección establecido en el Laboratorio Genia es de 20% [2, 3], seguramente por ello los valores entre el resultado de NGS y la técnica de referencia no coincidieron, siendo reales los valores detectados por NGS en estas muestras clínicas.

Tabla 38: Valores positivos y negativos detectados con la técnica de NGS (panel KN) versus Sanger, para todas las variantes analizadas.

		Referencia		
		+	-	Total
NGS	+	21	2	23
	-	0	41	41
	Total	21	43	64

Para evaluar la exactitud del panel EGFR, se compararon los resultados de 10 muestras contra los resultados de muestras de referencia (control de calidad *EMQM*), y luego se realizó la comparación de 35 muestras clínicas (a partir de FFPE) contra los resultados de la secuenciación Sanger. En la Tabla 39 se presentan los resultados de la comparación contra las muestras de referencia y se observa que en todos los casos, los resultados fueron coincidentes. Con estos resultados, se puede decir que la **exactitud de la técnica es del 100%**.

Tabla 39: Exactitud contra muestras de referencia. Comparación de 10 muestras analizadas con el panel EGFR entre la técnica en estudio (NGS con sistema NEB) versus los resultados del control de calidad EMQN.

Muestra	Resultado NGS	Resultado referencia	Tipo de referencia
1	EGFR c.2573T>G EGFR c.2369C>T	EGFR c.2573T>G EGFR c.2369C>T	Control de calidad
2	WT	WT	Control de calidad
3	EGFR c.2155G>A	EGFR c.2155G>A	Control de calidad
4	WT	WT	Control de calidad
5	WT	WT	Control de calidad
6	WT	WT	Control de calidad
7	WT	WT	Control de calidad
8	WT	WT	Control de calidad
9	WT	WT	Control de calidad
10	WT	WT	Control de calidad

Por otro lado, cuando se compararon las 35 muestras con la técnica de referencia (secuenciación Sanger), los resultados fueron totalmente coincidentes (Tabla 40), y no existieron FP ni FN.

Tabla 40: Exactitud contra método de referencia. Comparación de 35 muestras clínicas analizadas con el panel EGFR entre la técnica en estudio (NGS con sistema NEB) y los resultados de las mismas muestras realizadas por secuenciación Sanger.

Muestra	Resultado NGS	Resultado referencia	Tipo de referencia
1	WT	WT	Sec. Sanger
2	WT	WT	Sec. Sanger
3	WT	WT	Sec. Sanger
4	WT	WT	Sec. Sanger
5	EGFR c.2235_2249 del15	EGFR c.2235_2249 del15	Sec. Sanger
6	WT	WT	Sec. Sanger
7	EGFR c.2236_2250 del15	EGFR c.2236_2250 del15	Sec. Sanger
8	EGFR c.2235_2249 del15	EGFR c.2235_2249 del15	Sec. Sanger
9	EGFR c.2235_2249 del15	EGFR c.2235_2249 del15	Sec. Sanger
10	EGFR c.2236_2250 del15	EGFR c.2236_2250 del15	Sec. Sanger
11	WT	WT	Sec. Sanger
12	WT	WT	Sec. Sanger
13	WT	WT	Sec. Sanger
14	EGFR c.2573TG>GT	EGFR c.2573TG>GT	Sec. Sanger
15	WT	WT	Sec. Sanger
16	WT	WT	Sec. Sanger
17	WT	WT	Sec. Sanger
18	WT	WT	Sec. Sanger
19	WT	WT	Sec. Sanger
20	WT	WT	Sec. Sanger
21	EGFR c.2573T>G	EGFR c.2573T>G	Sec. Sanger
22	WT	WT	Sec. Sanger
23	WT	WT	Sec. Sanger
24	WT	WT	Sec. Sanger
25	WT	WT	Sec. Sanger
26	WT	WT	Sec. Sanger
27	WT	WT	Sec. Sanger
28	WT	WT	Sec. Sanger
29	WT	WT	Sec. Sanger
30	WT	WT	Sec. Sanger
31	WT	WT	Sec. Sanger
32	EGFR c.2573T>G	EGFR c.2573T>G	Sec. Sanger
33	EGFR c.2236_2248 CAAC (Complex)	EGFR c.2236_2248 CAAC (Complex)	Sec. Sanger
34	WT	WT	Sec. Sanger
35	WT	WT	Sec. Sanger

La conclusión final para el parámetro de exactitud fue que la técnica de NGS en estudio logró resultados coincidentes con los valores de referencia en el 100% de los casos (**exactitud del 100%**), tanto con el panel KN como con el panel EGFR, logrando detectar 2 variantes de bajo porcentaje en el gen *KRAS* que no se vieron con la técnica de referencia.

4.3.6. Resumen validación

En la Tabla 41 se presenta un resumen de la validación de la técnica de NGS adaptada en este trabajo.

Tabla 41: Resumen de los resultados de la validación de la técnica de NGS utilizando los paneles KN y EGFR con el sistema NEB.

Parámetro	Actividad	N° de muestras	Tipo de muestras	N° de ensayos	Resultado panel KN	Resultado panel EGFR
Especificidad analítica	Análisis de muestras <i>WT</i> para posiciones determinadas en los genes <i>KRAS</i> y <i>NRAS</i>	17	ADNs de referencia (Horizon y ADN ctrl)	Entre 37 y 81	99,5%*	No determinado
	Análisis de muestras con mutaciones conocidas	14	ADNs de referencia (Horizon)	14		
LD 50 copias de genoma	Análisis de muestras con porcentaje conocido de mutación	4	ADNs de referencia (Horizon)	20	12%	No determinado
LD 500 copias de genoma	Análisis de muestras con porcentaje conocido de mutación	4	ADNs de referencia (Horizon)	20	2%	No determinado
Repetibilidad (entre ejecuciones)	Análisis de un mismo grupo de muestras, realizado 3 días diferentes, por el mismo operador	17 (panel KN)	ADNs provenientes de distintos orígenes	51	100%	100%
		11 (panel EGFR)		33		
Repetibilidad (dentro de la misma ejecución)	Análisis de un grupo de muestras realizadas por triplicado en la misma ejecución	5	ADNs provenientes de distintos orígenes	15	100%	No determinado
Exactitud	Comparación de métodos	90 (panel KN)	ADNs provenientes de distintos orígenes	90	100%	100%
		45 (panel EGFR)		45		

* Con los cambios en el protocolo de amplificación (ver apartado 4.2.3.1), se elimina a futuro la interferencia asociada a la amplificación de la quimera con el pseudogen de *KRAS*.

4.3.7. Análisis global de los datos obtenidos tras la validación

Con el propósito de hacer un estudio global de los resultados obtenidos tras la validación de la técnica de NGS y determinar en qué porcentaje se encontraban las variantes patogénicas encontradas en los genes *KRAS*, *NRAS* y *EGFR*, así como la cantidad de muestras *WT* (sin mutación), se analizó un grupo de muestras provenientes de tejido FFPE. Se tomaron los resultados de 110 muestras procesadas con el panel KN y 46 muestras con el panel EGFR. En la Figura 28, se observan los resultados de las 110 muestras analizadas con el panel KN. De las muestras evaluadas, 28% fueron *WT* por NGS, y las variantes patogénicas detectadas se encontraron en diferentes rangos de porcentaje de mutación. El 56% de las muestras presentaron variantes patogénicas al 30% o más, y sólo un caso presentó una variante al 6,7% (rango 5-9%). Esto puede deberse a que mutaciones en los genes *RAS* se presentan en forma temprana (según modelos de tumorigénesis), generalmente en los adenomas tardíos que todavía no se convirtieron en carcinomas. Esta etapa se encuentra relativamente al comienzo de la tumorigénesis, por lo que es probable que cuando se detecta la mutación en el paciente (normalmente en CRC metastásico), un porcentaje alto de células presenten la variante patogénica [166].

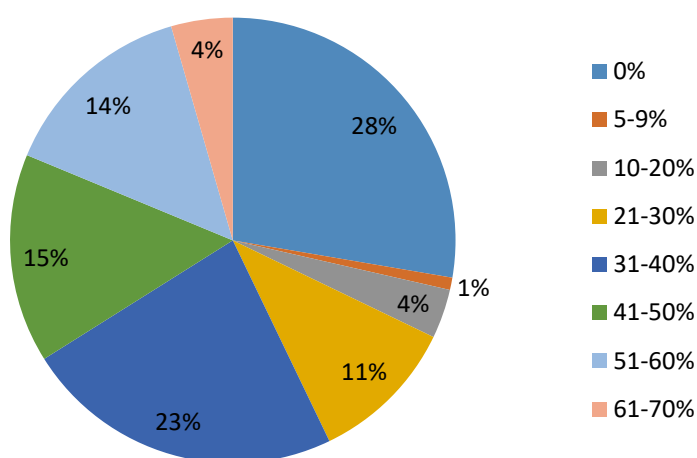


Figura 28: Gráfico circular que representa los rangos de porcentaje de variantes patogénicas encontradas en 110 muestras de tejido FFPE analizadas con el panel KN.

En cuanto a la distribución de los porcentajes de mutación en *EGFR* (Figura 29), en las 46 muestras analizadas, 35 muestras fueron *WT* (75%) y 11 muestras tenían variantes patogénicas (25%). En este caso el porcentaje de las variantes patogénicas encontradas fue menor que para el caso de *KRAS* y *NRAS*. La mayoría de las variantes se presentó con un porcentaje entre 10 y 20%. Sin embargo, según bibliografía las mutaciones en *EGFR*

comienzan temprano durante la patogénesis y pueden detectarse en el epitelio respiratorio histológicamente normal cerca de los tumores [167]. Esto podría deberse a que se tomó mayor tejido normal en la extracción o por la presencia de células infiltrantes en estos tumores.

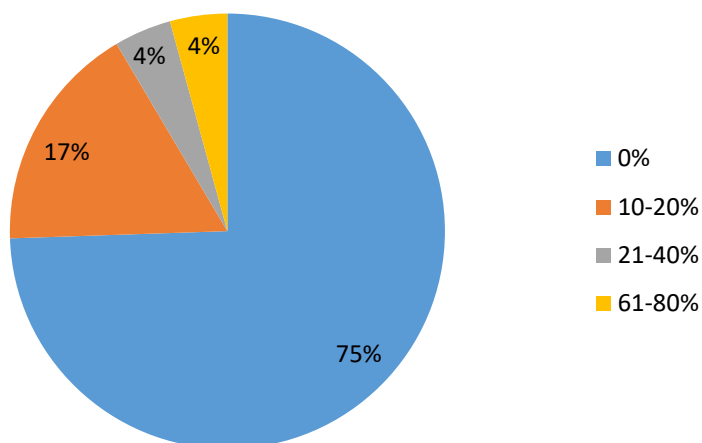


Figura 29: Gráfico circular que representa los rangos de porcentaje de las variantes patogénicas encontradas en 46 muestras de tejido FFPE analizadas con el panel EGFR.

Por otro lado, se realizó el estudio del tipo de mutaciones encontradas en *KRAS* y *NRAS* (panel KN) en las muestras clínicas analizadas en el parámetro de exactitud en la validación (ver apartado 4.3.5). De las 64 muestras estudiadas con el panel KN (66 resultados totales ya que algunas muestras tienen más de una variante), 41 muestras fueron *WT* (sin mutaciones patogénicas), lo que corresponde al 62% de los casos. De las restantes 23 muestras (25 variantes), 33% eran variantes en *KRAS*, dentro de las cuales el 86,5% se encontraba en el exón 2 (89% en el codón 12 y 11% en el codón 13), el 9% en el exón 3 y el 4,5% en el exón 4; y 3 variantes (5%) eran en el gen *NRAS*, 33% en el exón 2 y 67% en el exón 3. La Figura 30 representa mediante un gráfico circular estos datos. Los resultados obtenidos de la validación se asemejan a lo reportado en bibliografía. El porcentaje de muestras mutadas en *KRAS* (33%) se encuentra levemente por debajo de lo indicado en bibliografía que es un 40% de los casos de CCRm [48], en cambio *NRAS* (5%) se encuentra muy cercano a lo reportado (3-5% de los pacientes con CCRm).

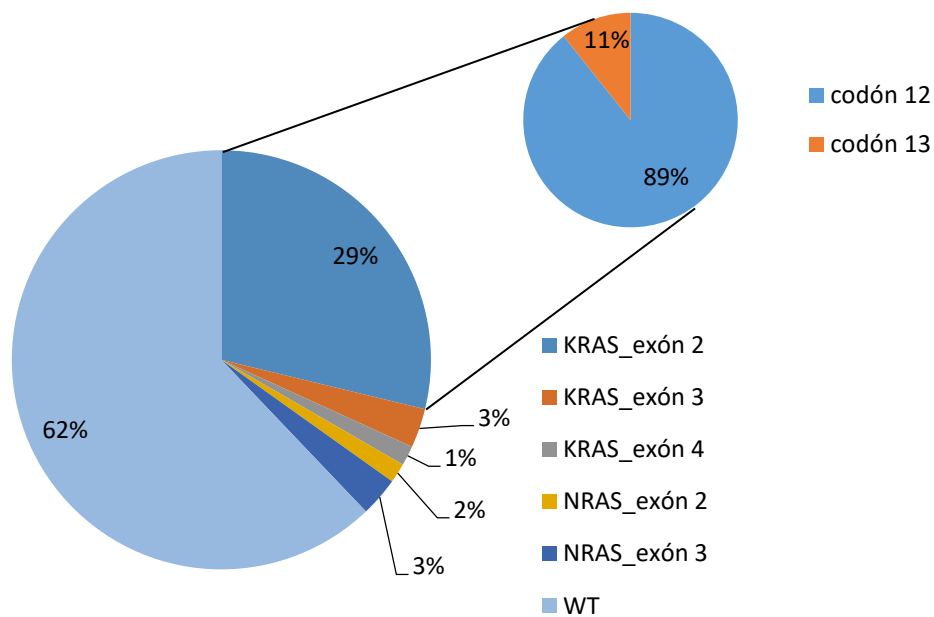


Figura 30: Gráfico circular que representa el porcentaje de muestras WT o mutadas en los genes KRAS y/o NRAS en 64 muestras analizadas con el panel KN (un total de 66 resultados).

No se pudo realizar un análisis similar del tipo de mutaciones encontradas en *EGFR* ya que las muestras analizadas durante el trabajo fueron seleccionadas específicamente, utilizando un número alto de muestras con mutaciones en *EGFR* (resultado de la secuenciación Sanger). Del total de las muestras estudiadas un 74% fueron muestras WT y 26% con mutaciones en *EGFR*. Estos valores se alejan de lo reportado en bibliografía (8-9%) [168].

5. DISCUSIÓN GENERAL

El objetivo general de este trabajo fue poner a punto un protocolo de NGS de bajo costo que fuera igual de eficiente que el utilizado hasta el momento en el Laboratorio Genia (protocolo Ampliseq para paneles de *Thermo Fisher Scientific*) y que permitiera sustituir la técnica de Sanger actualmente utilizada en el laboratorio para la identificación de mutaciones somáticas en *KRAS*, *NRAS* y *EGFR*. Teniendo en consideración este propósito, se analizaron opciones y mejoras al proceso para generar un protocolo alternativo. Debido a que el trabajo se realizó dentro del contexto de un laboratorio clínico, cuyas guías se basan en la ISO9001 e ISO17025, se realizó la validación del sistema para lograr confiabilidad y robustez en el mismo [120, 159].

La secuenciación Sanger se ha considerado por años como la técnica de referencia para la detección de mutaciones somáticas, sin embargo, sólo puede detectar variantes que estén por encima de 15 a 25% (en el Laboratorio Genia la detección se hace por encima de 20%), además es una técnica que consume mucho tiempo y mano de obra cuando se quiere realizar a gran escala [1-4]. Por las debilidades que presenta la secuenciación Sanger, en este trabajo, se propuso poder realizar un cambio de tecnología de Sanger a secuenciación de nueva generación (NGS). El laboratorio Genia cuenta con la tecnología de secuenciación masiva de *Thermo Fisher Scientific*, conocida como *Ion Torrent™*, para la cual existe un panel prediseñado, *Ion AmpliSeq Colon and Lung Research Panel v2*, para la detección de mutaciones somáticas en 22 genes implicados en cáncer de pulmón y colorrecto, entre ellos, los tres genes de interés (*KRAS*, *NRAS* y *EGFR*) y cuenta con un sistema de construcción de bibliotecas denominado *Ampliseq Library kit 2.0*. Utilizar el panel de forma completa implica amplificar 92 amplicones y cubrir cada amplicón con al menos 500x, lo que quiere decir que el amplicón que tenga menos lecturas tiene que tener 500x para cumplir con los requerimientos para determinar mutaciones somáticas con un porcentaje adecuado [88]. Esto da como resultado que sólo 2 muestras se puedan cubrir adecuadamente en un *chip* 314. Puesto que las muestras que recibe el laboratorio vienen con indicación ya sea para determinar mutaciones en los genes *KRAS/NRAS* o en *EGFR*, se armaron dos *pool* de cebadores, uno para determinar las mutaciones en *KRAS/NRAS* de 6 amplicones y otro para *EGFR* de 4 amplicones. Para CCR se solicita el estudio de los genes *KRAS* y *NRAS* ya que determinadas variantes patogénicas se asocian a resistencia a los anticuerpos monoclonales cetuximab (Erbix®) y panitumumab (Vectibix®). Cuando estos genes se encuentran mutados se ha demostrado una reducción de la respuesta en los pacientes con CCR metastásico. Por otro lado, se solicita el estudio del gen *EGFR* para determinar sensibilidad o resistencia a los inhibidores de tirosina kinasa (TKI) de *EGFR*, principalmente en pacientes con cáncer de pulmón (NSCLC). Al utilizar los subpaneles, el lugar que ocupa en el *chip* una muestra pasó a ser menor comparado con el panel completo, llegando a poder correr 16 muestras en un *chip* 314. Con esta modificación, el costo de secuenciación se ve distribuido en 16 muestras y es 8

veces menos costoso secuenciar este tipo de paneles que el panel original vendido por la empresa. De esta forma se acercó el precio de NGS al precio de Sanger, sin embargo, el costo de realizar NGS con el sistema Ampliseq sigue siendo alto, 70% mayor que el costo de la secuenciación Sanger. Por ello esta tesis hizo énfasis en buscar un método alternativo desde la extracción de ADN de tejido FFPE hasta el análisis de las secuencias, para poder llevar el método de NGS a un nivel competitivo, desde el punto de vista económico, con la secuenciación Sanger.

Las muestras con las que se trabajó, en su mayoría, fueron tejidos tumorales fijados en parafina (FFPE). Este tipo de muestras presentan grandes desafíos a la hora de extraer el ADN dando como resultado un bajo rendimiento de la extracción. Este problema es debido a que el ADN muchas veces se encuentra degradado y otras veces conteniendo inhibidores que afectan la PCR posterior. Por este motivo se utilizó un kit comercial para la extracción de ADN de tejido FFPE, *QIAamp DNA FFPE Tissue Kit de Qiagen*, que tiene un alto rendimiento de extracción y una pureza relativa del ADN extraído de buena calidad. Para adaptar el kit a las condiciones del laboratorio Genia y a las muestras con las que se contaba, se realizaron algunas modificaciones. La más importante fue la sustitución de un reactivo tóxico que utiliza el protocolo de Qiagen, Xylol, por uno totalmente inocuo, aceite mineral, ya que no se contaba con una campana de extracción de gases para evitar los efectos del Xylol. En cuanto a los inhibidores presentes en este tipo de muestras, no se pudieron identificar ni eliminar, por lo tanto, se evaluó su presencia mediante PCR en tiempo real, utilizando diluciones seriadas de la muestra, y seleccionando aquella que presentara menor inhibición sin perder mucha cantidad de ADN. Este último criterio fue muy importante ya que se planteó trabajar con cantidades de ADN superiores a 0,2ng en la PCR realizada para el diagnóstico, lo que equivale a 54 genomas humanos. Este límite es esencial para evitar tener resultados falsos negativos ya que las muestras de tejido FFPE están compuestas por células tumorales y no tumorales, dentro de las tumorales hay células mutadas y otras *WT*. Por lo tanto, si se parte de baja cantidad de células, la probabilidad de tomar las células tumorales mutadas se reduce (sobre todo cuando el porcentaje de células mutadas es bajo). Es por ello que es importante partir de un alto número de células y llegar a una buena cantidad de ADN como molde. Gracias al estudio realizado en la optimización de la extracción de ADN, se puede concluir que se necesita partir de 2 cm² como mínimo para llegar a una concentración de ADN adecuada. Si el ADN que se utiliza como molde equivale a menos de 54 genomas, se constató que existe un efecto estocástico en la toma de muestra si se quiere determinar una frecuencia alélica mutada tan baja como 6%. Con 50 genomas un 6% de alelos mutados equivale a decir que se espera que sólo 3 moléculas sean mutadas en la toma de muestra. Esto es el límite de detección para muchas técnicas de cuantificación que implican PCR, aunque en ellas se puede establecer la presencia de 3 moléculas de ADN solo si se utilizan 3 réplicas técnicas y 2 de ellas son positivas. Puesto que en NGS no es posible realizar réplicas técnicas por un tema de costos, se establece la sensibilidad como el porcentaje en el cual es posible informar todas las réplicas técnicas que se

hayan hecho en la validación (en este caso se realizaron 5 réplicas técnicas y por ello se exige que todas sean positivas para informar con esa sensibilidad). Si la intención es pipetear una muestra con 6% de alelo mutado y la cantidad de ADN es de 50 genomas entonces, por probabilidad, es posible que a veces se pipeteen 5 moléculas mutadas, otras veces 3, 1 o incluso ninguna molécula mutada, en cuyo caso el diagnóstico final sería un falso negativo (algo que se busca evitar). Entonces, para este trabajo se concluye que si uno quiere informar con un 12% de sensibilidad es necesario usar una cantidad de ADN de molde mayor al equivalente de 50 genomas, es decir mayor a 0,2 ng de ADN. Para cuantificar de forma correcta la cantidad de equivalentes de genoma presentes en la muestra (o concentración de ADN) y amplificables por PCR, lo que se realizó fue una cuantificación por PCR en tiempo real utilizando una PCR de un amplicón de tamaño similar al que luego se utilizó en la aplicación deseada (amplicones de *KRAS*, *NRAS* y *EGFR* que miden cerca de 150pb). Con esto se evitó cuantificar el ADN muy fragmentado (habitual en este tipo de muestras) que no alcanzaba un tamaño de 163pb (tamaño del fragmento hTERT que se utiliza en la cuantificación), que puede hacer que la cuantificación de ADN amplificable se sobreestime cuando se hace por otros métodos, por ejemplo métodos ópticos. Esta comprobación de la importancia en la cuantificación y la cantidad de ADN de partida sólo es posible realizarla en una técnica como NGS donde se puede cuantificar el número de moléculas mutadas y deducir la influencia de la cantidad de ADN de partida. Esta comprobación es extrapolable a la técnica de Sanger, por ello la relevancia de los estudios de estocástica realizados en el trabajo, los cuales se vieron afectados por la cantidad de ADN utilizado como molde. Los estudios realizados sirvieron para validar la técnica y para demostrar la importancia de usar más de 50 genomas como molde si se quiere informar correctamente con un 12% de sensibilidad, tanto en Sanger como NGS.

Con el propósito de disminuir los costos de la metodología de NGS, se planteó independizar, en mayor medida, de los productos provistos por el proveedor de *Ion Torrent™* buscando alternativas que tuvieran una eficiencia igual o mayor que la original, y un menor costo. Hubo dos etapas de alto costo que se pudieron reemplazar, la construcción y la cuantificación de bibliotecas. Para la construcción de bibliotecas *Thermo Fisher Scientific* recomienda el *Ion Ampliseq Library kit 2.0*, el cual incluye los reactivos necesarios para la amplificación por PCR de los genes de interés, la digestión con FuPA y la ligación de adaptadores. La alternativa para sustituir por completo el kit recomendado para esta etapa fue un kit comercial compatible con *Ion Torrent™*, *NEBNext® Fast DNA Library Prep Set for Ion Torrent™*, que permite la construcción de bibliotecas a partir de ADN genómico fragmentado [143]. Como en este trabajo se parte de la amplificación por PCR de los genes de interés, se tuvo que adaptar dicho kit para tal propósito. En primera instancia se sustituyó la etapa de amplificación, buscando una alternativa a la enzima del *Ion Ampliseq Library kit 2.0* que lograra amplificar correctamente los genes de interés con buen rendimiento. La enzima seleccionada fue la *Platinum® Multiplex PCR Master Mix* que logró amplificar los genes *KRAS*, *NRAS* y *EGFR* con buena eficiencia. La siguiente etapa en la construcción de bibliotecas es la ligación de adaptadores, la cual se

realizó con el kit de NEB. Como se dijo anteriormente, este kit se tuvo que adaptar para que aceptara productos de PCR como material de partida y además, se tuvieron que realizar ajustes para optimizar la secuenciación. Los cambios más relevantes fueron la optimización del equilibrio de amplicones que se corrigió aumentando los ciclos de la PCR a 33 y la mejora de los indicadores de calidad de la secuenciación, entre ellos, el porcentaje de policlonal y el porcentaje de dímeros de adaptadores, que se corrigieron disminuyendo la concentración del *pool* de bibliotecas a 0,05ng/μl y bajando la cantidad de adaptadores utilizados en el proceso de ligación de adaptadores, respectivamente. Además se redujo el volumen de reacción del kit de NEB a 1/4 del volumen original, sin perder la eficiencia de secuenciación. Con todos estos cambios se puso a punto el kit de NEB para la detección de mutaciones en *KRAS* y *NRAS* (panel KN) y luego se trasladó al panel EGFR, obteniéndose buenos resultados. Se logró amplificar y secuenciar los amplicones del panel EGFR, detectando correctamente los cambios esperados. El análisis de costos de la utilización del sistema NEB contra el sistema Ampliseq dio como resultado una disminución de 70% en la etapa de construcción de bibliotecas. Esto se debió principalmente a la sustitución del kit *Ion Ampliseq Library kit 2.0* por el kit *NEBNext® Fast DNA Library Prep Set for Ion Torrent™*. En lo que refiere a la cuantificación de bibliotecas, que fue otro de los grandes puntos que se evaluó, se encontró una alternativa al método de cuantificación propuesto por *Thermo Fisher Scientific*, cuantificación por *Qubit*, que generó una disminución de costos de esta etapa de un 65%. La gran ventaja de este método es la rapidez en la obtención de los resultados, tiempo que se ahorra en el proceso y es muy valioso en un laboratorio de diagnóstico.

Posteriormente a la construcción y cuantificación de bibliotecas, están los pasos de amplificación clonal, enriquecimiento y secuenciación, pasos costosos en la técnica pero que, por ahora, no se pueden reemplazar ni en reactivos ni en equipos. Por lo tanto, se utilizaron tal cual aconseja el proveedor *Thermo Fisher Scientific*. En relación a los costos de estas etapas, se observó que los mismos implican un costo variable según el número de muestras que se procesan en paralelo y es independiente del sistema que se utilice (Ampliseq o NEB). Sin embargo, debido a la gran disminución de costos que se realizó con el sistema NEB en la construcción de bibliotecas, las etapas de amplificación clonal, enriquecimiento y secuenciación pasaron a representar un alto porcentaje del costo total. En cambio para el sistema Ampliseq el gran costo proviene de la construcción de las bibliotecas. En lo que refiere al Laboratorio Genia, el número de muestras que se procesan para el análisis de *KRAS*, *NRAS* o *EGFR* ronda las 8 a 10 muestras por semana, las cuales se podrían secuenciar en un *chip* 314 sin estar utilizándolo en su máxima capacidad (16 muestras máximo). Este tipo de muestras cuenta con un plazo de entrega de resultados de una semana, por lo que se hace necesario poder correr como mínimo dos *chip* en días distintos de la semana en caso de requerir realizar la repetición de alguna muestra y poder cumplir con los plazos de entrega de informe. Esto implicaría el ingreso de 16 muestras aprox. en el correr de una semana, algo con lo que Genia no cuenta actualmente y sería un requerimiento previo para utilizar la tecnología de NGS.

Como se dijo anteriormente, el costo de la etapa de secuenciación es dependiente del número de muestras a procesar en una sola tanda por lo que a mayor número de muestras habría un descenso drástico en el precio del procesamiento por muestra, lo que es beneficioso para el sistema NEB ya que es la etapa de mayor influencia en el precio final.

Otro aspecto que se abarcó fue la trazabilidad o chequeo de identidad en el análisis de muestras que tienen procesos largos y complejos como NGS (recomendo por distintas guías [122, 123]). Para ello se implementó por primera vez en NGS, aplicado para determinar mutaciones somáticas, un método de trazabilidad utilizando STRs germinales. Esta diferencia entre el objetivo del estudio que es detectar mutaciones somáticas y determinar STRs que son germinales (y por ende se encuentran en todas las células), es importante cuando se tiene en cuenta la profundidad suficiente con la que se debe secuenciar unos y otros. Para STRs es suficiente con 100x de profundidad. Si se hubiera optado por una concentración de cebadores mayor a la seleccionada (33nM, 1/6x) la profundidad quedaba equivalente a la de los amplicones de *KRAS*, *NRAS* y *EGFR*, es decir de 1000x, y eso hubiera llevado a ocupar el doble de lugar en el *chip* (considerando que son 6 STRs y 6 los amplicones de *KRAS* y *NRAS*) y por ende encarecer el estudio, algo que iba en contra de los objetivos de este trabajo. Por lo tanto, se tuvo que poner a punto, entre otras cosas, la concentración de cebadores de forma tal que los STRs no amplificaran al mismo nivel que los amplicones de *KRAS*, *NRAS* y *EGFR*. Gracias a este método se logró obtener los resultados para STRs junto con el resultado de secuenciación de *KRAS*, *NRAS* y *EGFR*, es decir los resultados emitidos tienen incorporado el estudio de identidad biológica. Por un método alternativo y utilizando una técnica menos compleja (PCR-electroforesis capilar), se realiza el estudio de identidad de la muestra que ingresa al laboratorio. De esa forma se llega a contrastar el resultado por NGS con el resultado de identidad biológica, para chequear así que no haya habido algún problema en todos los pasos que requiere la técnica de NGS (diluciones de muestra, amplificación, purificaciones varias, ligación de adaptadores y asignación de *barcodes*, diluciones de bibliotecas, y todas las rotulaciones de tubos que hay, así como el rótulo final que lleva la muestra en el plan de análisis de NGS asignando muestras a *barcodes*). Esta identificación sabemos que es inequívoca pues con los 6 STRs funcionando, el poder de discriminación de una muestra es de 1 en 1.180.078.

Una vez finalizado el proceso en mesada, se realizó el análisis de las secuencias generadas en el secuenciador *Ion Torrent™*. Para realizar dicho análisis se utilizó el *software* propio de *Ion Torrent™*, *Torrent Suite™*, y luego se realizó una comparación con otros *software* para comparar el desempeño de los mismos. Durante el análisis de las secuencias generadas en el trabajo, se detectaron dos problemas, la presencia de desaminaciones en las muestras FFPE que generaban ruidos inespecíficos y la amplificación de un producto secundario del pseudogen de *KRAS* que alineaba con los exones 2 y 3 de *KRAS*, generando variantes en *cis* que alteraban el llamado de variantes reales. La desaminación es un proceso que está

presente en muchas de las muestras FFPE debido al método de fijación de estos tejidos. Como no se pudo interferir en ese paso, lo que se hizo fue evaluar algunas muestras con un exceso de variantes inespecíficas del tipo C:G>T:A, y realizar triplicados. Con los resultados de estas pruebas se confirmó que se trataba de cambios aleatorios propios de las desaminaciones en FFPE y no de variantes patogénicas del tumor. Por lo tanto, en aquellos casos donde aparezcan variantes de este tipo y coincidan con los sitios *hotspot*, se puede repetir la muestra para confirmar o descartar estas variantes. En cuanto al pseudogen, se observó que en la amplificación con el panel KN se generaba siempre un producto inespecífico del pseudogen que alineaba con los exones 2 y 3 de *KRAS*. En la mayoría de los análisis, la presencia del pseudogen no interfirió en el correcto alineamiento de las secuencias ni en la detección de variantes. Sin embargo, fue durante la validación (donde se procesó el mayor número de muestras) que se observó que este fenómeno afectaba la detección de variantes patogénicas. Para eliminar en mayor medida este problema, se realizaron cambios en el ciclado de la PCR aumentando la temperatura de hibridación de los cebadores a 62°C y disminuyendo la cantidad de cebadores en la mezcla de reacción. Con estos dos cambios en conjunto, se logró eliminar por completo la amplificación inespecífica (sin dejar de amplificar el pseudogen) en una muestra que presentaba claramente este problema. Se pretende seguir ahondando en este tema a futuro para lograr erradicarlo con seguridad en todas las muestras analizadas con el panel KN para luego poder realizar una nueva validación con las mejoras en este punto. Otra de las actividades que se realizó en el análisis de secuencias fue la comparación de tres *software* utilizando muestras con distintas características. En lo que refiere a la detección de inserciones, deleciones o cambios complejos, los tres *software* evaluados (*Torrent Suite*, *Sophia Genetics* y CLC) realizaron el llamado correcto de las variantes. El punto débil lo presentó el *software Torrent Suite™* ya que no realizó el llamado de algunas variantes (cambios puntuales) en bajo porcentaje en muestras de resultado conocido. Este programa es brindado por el proveedor de *Ion Torrent* por utilizar dicho sistema, por lo tanto no tiene costo extra. Sin embargo, para realizar una detección completa de todas las variantes se hace necesario complementarlo con otro *software* como por ejemplo *Sophia Genetics*. El gran inconveniente es que los dos programas alternativos son pagos. Por lo que queda abierta la posibilidad de utilizar *software* libres, los cuales se deberían evaluar para ser utilizados. De cualquier forma puesto que las ventajas de otros *software* respecto al utilizado fueron puesta de manifiesto en porcentajes muy bajos (cerca de 3%), rango en el no se va a informar el diagnóstico ya que está muy por debajo del límite de sensibilidad de la técnica (12%), el *software* utilizado responde bien dentro de los parámetros establecidos, para informar con sensibilidad por arriba del 12% (y hasta un 6% no habría problema pues el comportamiento del *software* es adecuado). Sin embargo, es importante conocer las limitantes del programa utilizado actualmente si se quiere llegar a informar porcentajes menores al 6%.

La validación de la técnica fue lo que concluyó el trabajo y permitió contar con un protocolo robusto y eficiente. La especificidad de la técnica fue del 99,5%. Este valor se debió a la

amplificación inespecífica del pseudogen de *KRAS* ya que una de las variantes del pseudogen cuando alinea con el exón 3 de *KRAS* cae en una posición *hotspot*. Como se dijo anteriormente, a este problema ya se le encontró una solución que se deberá validar y así mejorar este parámetro. En lo que respecta al límite de detección (LD), se logró un valor muy bueno en muestras con cantidad de ADN inicial alta (500 copias de genoma), permitiendo detectar variantes presentes al 2%. Sin embargo, el valor del LD aumentó a 12% cuando el material de partida fue menor (50 copias de genoma). La diferencia en los valores del LD dependiendo de la cantidad de ADN de partida es grande y eso se debe a que a menor cantidad de ADN inicial juega un papel más importante la estocástica y por lo tanto, si bien en algunos casos se lograba detectar las variantes en menores porcentajes con 50 copias, no se logró en todas las réplicas realizadas. Los otros dos parámetros, repetibilidad y exactitud fueron óptimos, 100% para ambos. Además, en la prueba de exactitud del panel KN se lograron detectar dos variantes en *KRAS* en dos muestras clínicas que habían sido *WT* por Sanger, en bajo porcentaje. Una de ellas se detectó en un 7% aprox. Dicha detección se pudo realizar dado que se partió de una cantidad de genomas mayor a 500.

Existen otros genes importantes en la vía de señalización mediada por EGFR que están implicados en el cáncer de colorrecto y pulmón, y que podrían condicionar la eficiencia de las drogas anti-EGFR y TKi, como por ejemplo *BRAF* y *PIK3CA*. Dichos genes podrían ser diagnosticados si se necesitaran, utilizando la técnica de NGS con el protocolo adaptado en este trabajo con una implementación relativamente fácil después de la experiencia generada y no se alterarían en gran medida los costos.

6. CONCLUSIÓN

En este trabajo se logró adaptar el método de extracción de muestras de tejido FFPE a las necesidades del laboratorio Genia. Se consiguió poner en marcha un método de trazabilidad útil para verificar la identidad de las muestras mediante STRs. Se encontró un proceso alternativo para la construcción y cuantificación de bibliotecas, el cual se puso a punto y validó, permitiendo comprobar que la técnica de NGS es una técnica robusta y sensible para el diagnóstico de mutaciones somáticas en los genes *KRAS*, *NRAS* y *EGFR* con una sensibilidad del 12% (partiendo de 0,2ng de ADN). Con todos los cambios y mejoras realizados en el proceso de NGS utilizando NEB se logró disminuir los costos de la construcción de bibliotecas en un 70% y en un 65% en la cuantificación de bibliotecas, llegando a una reducción final de 50% con respecto al sistema Ampliseq y quedando justo por debajo del costo de Sanger. Por lo tanto, se puede decir que se cumplieron con éxito los objetivos planteados en el origen del trabajo.

7. BIBLIOGRAFÍA

1. Normanno N, *et al.* KRAS mutations testing in colorectal carcinoma patients in Italy: from guidelines to external quality assessment. *PLoS One*. 2011. 6(12).
2. Gerstner A, *et al.* Low level somatic variant detection by Sanger sequencing of formalin-fixed paraffin-embedded (FFPE) samples. 2019.
3. Athanasios C, *et al.* Comparison of Sanger Sequencing, Pyrosequencing, and Melting Curve Analysis for the Detection of KRAS Mutations. *J Mol Diagn*. 2010. 12(4): p.425–432.
4. Xu X, *et al.* Assessment of the clinical application of detecting EGFR, KRAS, PIK3CA and BRAF mutations in patients with non-small cell lung cancer using next-generation sequencing. *Scandinavian Journal of Clinical and Laboratory Investigation*. 2016. 76(5): p.6.
5. Serrati S, *et al.* Next-generation sequencing: advances and applications in cancer diagnosis. *OncoTargets and Therapy*. 2016. 9: p.7355–7365.
6. Jie Gao HW, *et al.* Validation of targeted next-generation sequencing for RAS mutation detection in FFPE colorectal cancer tissues: comparison with Sanger sequencing and ARMS-Scorpion real-time PCR. *BMJ Open*. 2016. 6.
7. Barrios E, Garau M. Cáncer: magnitud del problema en el mundo y en Uruguay, aspectos epidemiológicos. *Facultad Med (Univ Repúb Urug)*. 2017. 4(1): p.9-47.
8. Cáncer [internet]. Organización Mundial de la Salud. 2018. Disponible desde: <https://www.who.int/es/news-room/fact-sheets/detail/cancer>.
9. 10 gráficos para entender el grave impacto del cáncer en el mundo [internet]. Perasso V. BBC. 2016. Disponible desde: https://www.bbc.com/mundo/noticias/2016/02/160203_cancer_graficos_impacto_men.
10. Estadísticas del cáncer [internet]. Instituto Nacional del Cáncer. 2018. Disponible desde: <https://www.cancer.gov/espanol/cancer/naturaleza/estadisticas>.
11. Tarver T. Cancer Facts and Figures. American Cancer Society (ACS); 2012. Vol. 16: p 366–367.
12. Resumen de la situación epidemiológica del Uruguay en relación al cáncer. Registro Nacional de Cáncer: Comisión honoraria de lucha contra el cáncer. 2018.
13. Cancer Today [internet]. Organización Mundial de la Salud. 2018. Disponible desde: <http://qco.iarc.fr/today/home>.
14. Cáncer en Uruguay: principales tipos de cáncer [internet]. Comisión Honoraria de Lucha contra el Cáncer. 2017. Disponible desde: <http://www.comisioncancer.org.uy/>
15. Tratamiento del cáncer de colon (PDQ®) [internet]. Instituto Nacional del Cáncer. 2019. Disponible desde: <https://www.cancer.gov/espanol/tipos/colorrectal/paciente/tratamiento-colorrectal-pdq#section/all?redirect=true>.
16. Cáncer de colon [internet]. MD Anderson Cancer Center. 2019. Disponible desde: <https://www.mdanderson.es/elcancer/elcancerdelaaalaz/cancer-de-colon>.
17. Hernández J, *et al.* Recomendaciones para la determinación de mutaciones de K-RAS en cáncer de colon. *Rev Esp Patol*. 2012. 2 (45).
18. Barrios E, *et al.* IV atlas de incidencia del cáncer en el Uruguay (2007-2011). Comisión Honoraria de Lucha Contra el Cáncer. 2014.
19. Cáncer de pulmón [internet]. Medline Plus. 2018. Disponible desde: <https://medlineplus.gov/spanish/lungcancer.html>.
20. Lung Cancer [internet]. MD Anderson Cancer Center. 2018. Disponible desde: <https://www.mdanderson.org/cancer-types/lung-cancer.html>.
21. Barrios E, *et al.* V Atlas de mortalidad por cáncer en el Uruguay 2009-2013. Comisión Honoraria de Lucha Contra el Cáncer. 2015.

22. Medicina de Precisión y Terapia Dirigida en el tratamiento del cáncer [internet]. HC Marbella. 2018. Disponible desde: <https://www.hcmarbella.com/es/medicina-de-precision-y-terapia-dirigida-en-el-tratamiento-del-cancer/>.
23. Medicina de Precisión [internet]. Cancer Center. 2019. Disponible desde: <http://www.cancercenter.mx/mx/AdvancedTreatment/PrecisionMedicine/>.
24. Jordan E, *et al.* Prospective Comprehensive Molecular Characterization of Lung Adenocarcinomas for Efficient Patient Matching to Approved and Emerging Therapies. American Association for Cancer Research. 2017. 7(6): p.596-609.
25. Campos A, Cruz-Rico G, Arrieta O. Genotyping Non-Small Cell Lung Cancer (NSCLC). GAMO. 2012. 11(1).
26. Nicholson R, Gee JM, Harper M. EGFR and cancer prognosis. European Journal of Cancer. 2001. 37: p.9-15.
27. Vincenzi B, *et al.* The biological properties of cetuximab. Crit Rev Oncol Hematol. 2008. 68(2): p.93-106.
28. Wong SF. Cetuximab: an epidermal growth factor receptor monoclonal antibody for the treatment of colorectal cancer. Clin Ther. 2005. 27(6): p.684-694.
29. Markman B, *et al.* EGFR and KRAS in Colorectal Cancer. Elsevier: San Francisco. 2010. p. 71-114.
30. EGFR Gene (Protein Coding) [internet]. GeneCards. 2019. Disponible desde: <https://www.genecards.org/cgi-bin/carddisp.pl?gene=EGFR&keywords=egfr>.
31. Gómez J, *et al.* Recomendaciones para la determinación de biomarcadores en el carcinoma de pulmón no microcítico avanzado. Consenso nacional de la Sociedad Española de Anatomía Patológica y de la Sociedad Española de Oncología Médica. Rev Esp Patol. 2012. 45: p.14.
32. Benvenuti S, *et al.* Oncogenic Activation of the RAS/RAF Signaling Pathway Impairs the Response of Metastatic Colorectal Cancers to Anti-Epidermal Growth Factor Receptor Antibody Therapies. Cancer Res. 2007. 67: p.6.
33. Bethune G, *et al.* Epidermal growth factor receptor (EGFR) in lung cancer: an overview and update. J Thorac Dis. 2010. 2(1): p.48-51.
34. Genes del Cáncer [internet]. Cancer Quest. 2016. Disponible desde: <https://www.cancerquest.org/es/biologia-del-cancer/genes-de-cancer#ras>.
35. Paz Ares L, *et al.* Afatinib versus gefitinib in patients with EGFR mutation-positive advanced non-small-cell lung cancer: overall survival data from the phase IIb LUX-Lung 7 trial. Ann Oncol. 2017. 28(2): p.270-277.
36. Wu Y L, *et al.* Dacomitinib versus gefitinib as first-line treatment for patients with EGFR-mutation-positive non-small-cell lung cancer (ARCHER 1050): a randomised, open-label, phase 3 trial. Lancet Oncol. 2017. 18(11): p.1454-1466.
37. Jancik S, *et al.* Clinical relevance of KRAS in human cancers. J Biomed Biotechnol. 2010. 1: p.1-13.
38. Troiani T, *et al.* Predictive Biomarkers to Anti-EGFR Inhibitors Treatment in the Management of Metastatic Colorectal Cancer. Carcinog & Mutagen. 2014. 9.
39. Misale S, *et al.* Resistance to anti-EGFR therapy in colorectal cancer: from heterogeneity to convergent evolution. Cancer Discov. 2014. 4(11): p.1269-1280.
40. NRAS Gene(Protein Coding) [internet]. GeneCards. 2019. Disponible desde: <https://www.genecards.org/Search/Keyword?queryString=nras>.
41. KRAS Gene(Protein Coding) [internet]. GeneCards. 2019. Disponible desde: <https://www.genecards.org/cgi-bin/carddisp.pl?gene=KRAS&keywords=kras>.
42. HRAS Gene(Protein Coding) [internet]. GeneCards. 2019. Disponible desde: <https://www.genecards.org/cgi-bin/carddisp.pl?gene=HRAS&keywords=hras>.
43. McGuire W. Oncogenes. Kluwer Academic Publishers; 1989. p.
44. Jančík S, *et al.* Clinical Relevance of KRAS in Human Cancers. J Biomed Biotechnol. 2010.

45. Siddiqui AD, Piperdi B. KRAS mutation in colon cancer: a marker of resistance to EGFR-I therapy. *Ann Surg Oncol*. 2010. 17(4): p.1168-1176.
46. Demes M, *et al*. Signature of microsatellite instability, KRAS and BRAF gene mutations in German patients with locally advanced rectal adenocarcinoma before and after neoadjuvant 5-FU radiochemotherapy. *J Gastrointest Oncol*. 2013. 4(2): p.182-192.
47. Hope on the Horizon for KRAS-positive Lung Cancer Patients [internet]. Upal Basu R. Lungevity. 2018. Disponible desde: <https://lungevity.org/for-patients-caregivers/blogs/experts-blog/hope-on-horizon-for-kras-positive-lung-cancer-patients>.
48. Liu X, Jakubowski M, Hunt JL. KRAS gene mutation in colorectal cancer is correlated with increased proliferation and spontaneous apoptosis. *Am J Clin Pathol*. 2011. 135(2): p.245-252.
49. De Roock W, *et al*. Effects of KRAS, BRAF, NRAS, and PIK3CA mutations on the efficacy of cetuximab plus chemotherapy in chemotherapy-refractory metastatic colorectal cancer: a retrospective consortium analysis. *Lancet Oncol*. 2010. 11(8): p.753-762.
50. Brohawn P, *et al*. Translating Biomarker Discovery into Companion Diagnostics through Validation and Regulatory Consideration. Elsevier. 2014. p. 157-181.
51. Guidance for Industry and Food and Drug Administration Staff [internet]. Food and Drug Administration. 2014. Disponible desde: <https://www.fda.gov/ucm/groups/fdagov-public/@fdagov-meddev-gen/documents/document/ucm262327.pdf>.
52. Reacción en cadena de la polimerasa [internet]. Wikipedia. 2018. Disponible desde: https://es.wikipedia.org/wiki/Reacci%C3%B3n_en_cadena_de_la_polimerasa.
53. Chang KC, *et al*. Development and validation of a clinical trial patient stratification assay that interrogates 27 mutation sites in MAPK pathway genes. *PLoS One*. 2013. 8(8).
54. Morlan J, Baker J, Sinicropi D. Mutation Detection by Real-Time PCR: A Simple, Robust and Highly Selective Method. *Plos One*. 2009. 4(2): p.11.
55. cobas® EGFR Mutation Test v2. Cobas. Nº de catálogo: 07248563190.
56. Rodríguez-Santiago B, Armengol L. Tecnologías de secuenciación de nueva generación en diagnóstico genético pre- y postnatal. *Diagn Prenat*. 2012. 23(2): p.11.
57. Metzker M. Sequencing technologies: the next generation. *Nature Review*. 2010. 11.
58. Mardis E. Next-Generation DNA Sequencing Methods. *Annu. Rev. Genomics Hum. Genet.* . 2008. 9: p.387–402.
59. Ansorge W. Next-generation DNA sequencing techniques. *New Biotechnology*. 2009. 25(4).
60. NGS: Secuenciación de Segunda Generación [internet]. Garrigues F. *Genética Médica*. 2017. Disponible desde: <https://revistageneticamedica.com/blog/ngs-secuenciacion/>.
61. Introduction to SBS Technology [internet]. Illumina. 2019. Disponible desde: <https://www.illumina.com/science/technology/next-generation-sequencing/sequencing-technology.html>.
62. Quail MA, *et al*. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics*. 2012. 13: p.341.
63. Ross MG, *et al*. Characterizing and measuring bias in sequence data. *Genome Biol*. 2013. 14(5).
64. Ion Torrent (ThermoFisher). T.S. Marketplace.
65. The Ion Torrent S5™ and S5™XL Next Generation Sequencing Systems [internet]. BioSciences. Disponible desde: <http://www.biosciences.ie/ion-torrent-s5-and-s5xl>.
66. How AmpliSeq Works [internet]. FLORAGENEX. 2018. Disponible desde: <https://www.floragenex.com/ampliseq>.
67. Wonyong K. Application of Metagenomic Techniques: Understanding the Unrevealed Human Microbiota and Explaining the in Clinical Infectious Diseases. *Journal of Bacteriology and Virology*. 2012. 42(263).
68. Ion Torrent Next-Generation Sequencing Technology. T.F. Scientific.

69. Ion Torrent: Proton / PGM sequencing [internet]. EMBL-EBI. Disponible desde: <https://www.ebi.ac.uk/training/online/course/ebi-next-generation-sequencing-practical-course/what-next-generation-dna-sequencing/ion-torre>.
70. Ion semiconductor sequencing [internet]. Wikipedia. 2018. Disponible desde: https://en.wikipedia.org/wiki/Ion_semiconductor_sequencing.
71. Jennings L, *et al.* Guidelines for Validation of Next-Generation Sequencing Based Oncology Panels. A Joint Consensus Recommendation of the Association for Molecular Pathology and College of American Pathologists. *The Journal of Molecular Diagnostics*. 2017. 19(3).
72. Singh R, *et al.* Clinical Validation of a Next-Generation Sequencing Screen for Mutational Hotspots in 46 Cancer-Related Genes. *The Journal of Molecular Diagnostics*. 2013. 15(5).
73. de Leng W, *et al.* Targeted Next Generation Sequencing as a Reliable Diagnostic Assay for the Detection of Somatic Mutations in Tumours Using Minimal DNA Amounts from Formalin Fixed Paraffin Embedded Material. *PLoS ONE* 2016. 11(2).
74. Torrent Suite™ Software 5.8. T.F. Scientific. N° de catálogo: MAN0017530.
75. Ion Reporter Software [internet]. Thermo Fisher Scientific. Disponible desde: <https://ionreporter.thermofisher.com/ir/>.
76. Van Nimwegen KJ, *et al.* Is the \$1000 Genome as Near as We Think? A Cost Analysis of Next-Generation Sequencing. *Clin Chem*. 2016. 62(11): p.1458-1464.
77. Marino P. Cost of cancer diagnosis using next-generation sequencing targeted gene panels in routine practice: a nationwide French study. *European Journal of Human Genetics*. 2018. 26: p.314–323
78. Weymann D, *et al.* The cost and cost trajectory of whole-genome analysis guiding treatment of patients with advanced cancers. *Molecular Genetics & Genomic Medicine*. 2017.
79. Muir, *et al.* The real cost of sequencing: scaling computation to keep pace with data generation *Genome Biology*. 2016. 17(53).
80. EMQN [internet]. Manchester Centre for Genomic Medicine. 2019. Disponible desde: <https://www.emqn.org/about-emqn/>.
81. UK NEQAS: International Quality Expertise [internet]. UK NEQAS. 2019. Disponible desde: <https://ukneqas.org.uk/>.
82. Power SYBR™ Green PCR Master Mix. T.F. Scientific. N° de catálogo: 4367659.
83. SYBR® Green based qPCR [internet]. Merck KGaA. 2019. Disponible desde: <https://www.sigmaaldrich.com/life-science/molecular-biology/pcr/quantitative-pcr/sybr-green-based-qpcr.html>.
84. SYBR Green I [internet]. Wikipedia. 2018. Disponible desde: https://en.wikipedia.org/wiki/SYBR_Green_I.
85. Ion Library TaqMan™ Quantitation Kit. T.F. Scientific. N° de catálogo: 4468802.2018.
86. Qubit® dsDNA HS Assay Kit. T.F. Scientific. N° de catálogo: Q32854.
87. Ion AmpliSeq Designer [internet]. Thermo Fisher Scientific. Disponible desde: <https://www.ampliseq.com/login/login.action>.
88. Ion AmpliSeq™ Colon and Lung Cancer Research Panel v2. T.F. Scientific. N° de catálogo: CP1004.
89. PCR Cycling Parameters—Six Key Considerations for Success [internet]. Thermo Fisher Scientific. 2019. Disponible desde: <https://www.thermofisher.com/uy/en/home/life-science/cloning/cloning-learning-center/invitrogen-school-of-molecular-biology/pcr-education/pcr-reagents-enzymes/pcr-cycling-considerations.html>.
90. López-Barragán M, *et al.* Effect of PCR extension temperature on high-throughput sequencing. *Mol Biochem Parasitol*. 2011. 176(1): p.64-67.
91. Molecular Evolutionary Genetics Analysis [internet]. Megasoftware. 2019. Disponible desde: <https://www.megasoftware.net/>.

92. Ion AmpliSeq™ Library Kit 2.0. T.F. Scientific. N° de catálogo: 4475345.
93. LIBRARY PREPARATION : NEBNext® Fast DNA Library Prep Set for Ion Torrent™ Instruction Manual. 2015.
94. NEBNext® for Ion Torrent™: LIBRARY PREPARATION KITS [internet]. New England Biolabs. 2014. Disponible desde: <https://international.neb.com/products/e6270-nebnext-fast-dna-library-prep-set-for-ion-torrent#Protocols,%20Manuals%20&%20Usage>.
95. Ion PGM™ Hi Q™ View OT2 Kit USER GUIDE. T.F. Scientific. N° de catálogo: A29900
96. CLC Genomics Workbench [internet]. Qiagen. Qiagen. 2017. Disponible desde: <https://www.qiagenbioinformatics.com/products/clc-genomics-workbench/>.
97. Torrent Suite Software and Torrent Server [internet]. Thermo Fisher Scientific. 2019. Disponible desde: <https://www.thermofisher.com/uy/en/home/life-science/sequencing/next-generation-sequencing/ion-torrent-next-generation-sequencing-workflow/ion-torrent-next-generation-sequencing-data-analysis-workflow/ion-torrent-suite-software.html>.
98. Sophia Genetics [internet]. Sophia Genetics. 2017. Disponible desde: <http://www.sophiagenetics.com/home.html>.
99. HID STR Genotyper Plugin USER GUIDE v5.2.2. T.F. Scientific. N° de catálogo: A30939.
100. Integrative genome viewer (IGV) [internet]. Broad Institute. 2019. Disponible desde: <http://www.broadinstitute.org/igv/>.
101. Ingenuity Pathway Analysis [internet]. Qiagen. Disponible desde: <https://www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis/>.
102. COSMIC (Sanger Institute Catalogue of Somatic Mutations in Cancer) database [internet]. Sanger Institute. Disponible desde: <http://www.sanger.ac.uk/cosmic>.
103. Human Genome Variation Society [internet]. HGVS. 2018. Disponible desde: <http://www.hgvs.org/>.
104. Muestras FFPE [internet]. Sygnis. 2015. Disponible desde: <http://www.sygnis.com/muestras-ffpe/?lang=es>.
105. Atanesyan L, *et al.* Optimal Fixation Conditions and DNA Extraction Methods for MLPA Analysis on FFPE Tissue-Derived DNA. *Am J Clin Pathol.* 2017. 174(1): p.60-68.
106. QIAamp® DNA FFPE Tissue Handbook. Qiagen.2012.
107. Muñoz-Cadavid C, *et al.* Improving Molecular Detection of Fungal DNA in Formalin-Fixed Paraffin-Embedded Tissues: Comparison of Five Tissue DNA Extraction Methods Using Panfungal PCR. *J Clin Microbiol.* 2010. 48(6): p.2147–2153.
108. Janecka A, Adamczyk A, Gasin´ ska A. Comparison of eight commercially available kits for DNA extraction from formalin-fixed paraffin-embedded tissues. *Analytical Biochemistry* 2015. p.8–10.
109. Reseña Toxicológica de los Xilenos [internet]. Agencia para Sustancias Tóxicas y el Registro de Enfermedades. 1995. Disponible desde: <http://www.ladep.es/ficheros/documentos/DTO%20DE%20SEGURIDAD%20Y%20SALUD%20DE%20EEUU.pdf>.
110. Cardoso F. Desparafinado muestras FFPE por calor. Comunicación personal (e-mail). 2016.
111. Shan-Rong S, *et al.* DNA Extraction from Archival Formalin-fixed, Paraffin-embedded Tissue Sections Based on the Antigen Retrieval Principle: Heating Under the Influence of pH. *The Journal of Histochemistry & Cytochemistry.* 2002. 50(8): p.1005–1011.
112. Mansour A, *et al.* A novel xylene-free deparaffinization method for the extraction of proteins from human derived formalin-fixed paraffin embedded (FFPE) archival tissue blocks. *MethodsX.* 2014. 1: p.90-95.
113. Heikal N, Nussenzveig R, Archana M. Deparaffinization with mineral oil: a simple procedure for extraction of high-quality DNA from archival formalin-fixed paraffin-embedded samples. *Appl Immunohistochem Mol Morphol.* 2013.

114. Jianghai L. High-quality genomic DNA extraction from formalin-fixed and paraffin-embedded samples deparaffinized using mineral oil. *Anal Biochem.* 2009. 395(2): p.265–267.
115. Dolez J, Greilhuber J. Nuclear Genome Size: Are We Getting Closer? *International Society for Advancement of Cytometry.* 2010. 77: p.635-642.
116. Schade C. The challenges of FFPE sample materials – where does variation in quantity of purified DNA come from? *Qiagen.* 2016.
117. Primer Design [internet]. *Macrogen.* Disponible desde: https://dna.macrogen.com/eng/support/ces/guide/universal_primer.jsp.
118. Granato A. DNA and RNA isolation from canine oncologic formalin-fixed, paraffin-embedded tissues for downstream “-omic” analyses: possible or not? *Journal of Veterinary Diagnostic Investigation.* 2014. 26(1): p.117–124.
119. Chung M, *et al.* Tissue Requirements and DNA Quality Control for Clinical Targeted Next- Generation Sequencing of Formalin-Fixed, Paraffin-Embedded Samples: A Mini-Review of Practical Issues. *J Mol Genet Med* 2017. 11: p.262.
120. UNIT-ISO / IEC 17025:2017: Requisitos generales para la competencia de los laboratorios de ensayo y calibración. *ISO 17025:2017. UNIT.* 2017.
121. UNIT-ISO / IEC 15189:2012: Clinical Laboratory Testing and In Vitro Diagnostic Systems. *ISO 15189:2012 UNIT.* 2012.
122. Palmirotta R, *et al.* DNA fingerprinting for sample authentication in biobanking: recent perspectives. *Biorepository Science for Applied Medicine* 2015. (3): p.35–45.
123. Considerations for Design, Development, and Analytical Validation of Next Generation Sequencing (NGS) – Based In Vitro Diagnostics (IVDs) Intended to Aid in the Diagnosis of Suspected Germline Diseases. *Guidance for Stakeholders and Food and Drug Administration Staff.* 2018.
124. Coleman W, Tsongalis G. The role of genomic instability in human carcinogenesis. *Anticancer Res.* 1999. 19: p.4645-4664.
125. GeneMapper™ Software 5, full installation. T.F. Scientific. N° de catálogo: 4475073.
126. Combined DNA Index System (CODIS) [internet]. *FBI.* 2019. Disponible desde: <https://www.fbi.gov/services/laboratory/biometric-analysis/codis>.
127. Hares DR. Selection and implementation of expanded CODIS core loci in the United States. *Forensic Sci. Int. Genetics.* 2015. (17): p.33-4.
128. PowerPlex® 18D System. *Promega.* N° de catálogo: DC1802.
129. PowerPlex® 21 System. *Promega.* N° de catálogo: DC8902.
130. AmpFLSTR™ Identifiler™ PCR Amplification Kit. T.F. Scientific. N° de catálogo: 4322288.
131. GlobalFiler™ PCR Amplification Kit. T.F. Scientific. N° de catálogo: 4482815.
132. AmpFLSTR™ NGM SElect™ Express Kit. T.F. Scientific. N° de catálogo: 4472193.
133. HID STR Genotyper Plugin. T.F. Scientific. N° de catálogo: A30939.
134. Ibarra A. Estudio de marcadores genéticos bialélicos para aplicaciones forenses: SNPs autosómicos e Indels específicos de cromosoma X. 2015.
135. Fernández F. Análisis de 15 loci tipo short tandem repeats (STR) en la población de Paraguay para su uso en identificación forense. 2008.
136. Hernández A, Trejo FM. Estudio Genético Poblacional de Frecuencias Alélicas para 15 marcadores STR presentes en la Población del Estado de Zacatecas Aplicado a la Práctica Forense. *iMedPub Journals.* 2014. 10(1).
137. Sociedad Argentina de Genética Forense Asociación Civil (SAGF) [internet]. *SAGF.* 2018. Disponible desde: <https://www.sagf.org.ar/>.
138. DNA database establishing method for high-throughput sequencing. *L. Technology.* CN105568393A.
139. Ion Xpress™ Barcode Adapters 1-16 Kit. T.F. Scientific. N° de catálogo: 4471250.

140. PCR Purification: Agencourt AMPure XP [internet]. Beckman Coulter. Disponible desde: <http://www.beckman.com/nucleic-acid-sample-prep/purification-clean-up/pcr-purification>.
141. Ion Library Equalizer™ Kit. T.F. Scientific. N° de catálogo: 4482298.
142. Sondas Taqman [internet]. Wikipedia. 2018. Disponible desde: <https://es.wikipedia.org/wiki/TaqMan>.
143. NEBNext® Fast DNA Library Prep Set for Ion Torrent™. N.E.B. Inc. N° de catálogo: E6270S/L.
144. Scarpa A, *et al.* Molecular Typing of Lung Adenocarcinoma on Cytological Samples Using a Multigene Next Generation Sequencing Panel. Plos One. 2013. 8(11).
145. Hirsch B, *et al.* Multicenter validation of cancer gene panel-based next-generation sequencing for translational research and molecular diagnostics. Virchows Archiv 2017. (472): p.557–565.
146. Takeda M, *et al.* Clinical application of amplicon-based next-generation sequencing to therapeutic decision making in lung cancer. Annals of Oncology. 2015. 26(12): p.2477–2482.
147. Ion AmpliSeq Colon and Lung Cancer Research Panel v2 and Ion AmpliSeq RNA Fusion Lung Cancer Research Panel [internet]. Thermo Fisher Scientific. 2016.
148. Platinum® Multiplex PCR Master Mix. T.F. Scientific. N° de catálogo: 4463722.
149. Ion AmpliSeq™ Library Kit 2.0 USER GUIDE. T.F. Scientific. 2019.
150. Ion AmpliSeq™ DNA and RNA Library Preparation. T.F. Scientific. N° de catálogo: MAN0006735.
151. Ankala A, Hegde M. Genomic technologies and the new era of genomic medicine. J Mol Diagn. 2014. 16(1): p.7-10.
152. Kim S, *et al.* Deamination Effects in Formalin-Fixed, Paraffin-Embedded Tissue Samples in the Era of Precision Medicine. The Journal of Molecular Diagnostics. 2017. 19(1).
153. Do H, Dobrovic A. Dramatic reduction of sequence artefacts from DNA isolated from formalin-fixed cancer biopsies by treatment with uracil- DNA glycosylase. OncoTargets and Therapy. 2012. 3(5).
154. KRASP1 [internet]. Incrnadb. 2015. Disponible desde: <http://Incrnadb.com/KRASP1/>.
155. KRASP1 [internet]. NCBI. 2019. Disponible desde: <https://www.ncbi.nlm.nih.gov/gene/3844>.
156. V-KI-RAS2 [internet]. OMIM. 2019. Disponible desde: <https://www.omim.org/entry/190070>.
157. Validation of analytical procedures: Text and methodology Q2 (R1). ICH. 1994. 4: p.1-16.
158. Validation Resources [internet]. Association for Molecular Pathology. 2019. Disponible desde: <https://www.amp.org/resources/validation-resources/>.
159. UNIT-ISO / IEC 9001:2015: International Organization for Standardization. Quality Management Systems, Fundamentals and Vocabulary. ISO 9001:2015. UNIT. 2015.
160. J Y Douillard, *et al.* UK NEQAS for Molecular Pathology-Molecular genetic analysis of non small cell lung cancer. J Thorac Oncol. 2014. 9.
161. American College of Medical Genetics and Genomics. 2018.
162. Castillo B, González R. Protocolo de validación de métodos analíticos para la cuantificación de fármacos. Rev Cubana Farm. 1997. 30(1).
163. Tholen D, *et al.* Protocols for determination of limits of detection and limits of quantitation. NCCLS. 2010.
164. Jennings L, Van Deerlin VM, Gulley ML. Recommended principles and practices for validating clinical molecular pathology tests. Arch Pathol Lab Med. 2009. 133(5): p.743-755.
165. Principles and methods of validation of diagnostic assays for infectious diseases. NCCLS. 2013. p.17.

166. Armaghany T, *et al.* Genetic Alterations in Colorectal Cancer. *Gastrointest Cancer Res.* 2012. 5(1): p.19-27.
167. Gazdar A, Minna J. Deregulated EGFR Signaling during Lung Cancer Progression: Mutations, Amplicons, and Autocrine Loops. *Cancer Prevention Research.* 2008.
168. Graham R, *et al.* Worldwide Frequency of Commonly Detected EGFR Mutations. *Archives of Pathology & Laboratory Medicine* 2018. 142(2).

8. ANEXOS

8.1. Ejemplo de consentimiento informado para estudio de los genes *KRAS* y *NRAS*

Consentimiento Informado

Genotipificación de los genes *KRAS* y *NRAS*

A través del presente manifiesto en forma libre y voluntaria que he sido ampliamente informado por el / la Dr. / Dra., sobre las características de la afección que padezco (cáncer de colon) y las distintas alternativas terapéuticas para su tratamiento..

Me ha explicado y he comprendido que en relación con mi enfermedad, es posible mediante la tipificación del:

gen *KRAS* (Exón 2 - codones 12, 13 // Exón 3 - codones 59 y 61 // Exón 4 – codones 117 y 146) y

gen *NRAS* (Exón 2 - codones 12 y 13 // Exón 3 - codones 59 y 61 // Exón 4 – codones 117 y 146)

y de acuerdo a los resultados obtenidos, se podría elegir el tratamiento más apropiado para mi, disponible a la fecha.

Luego de haber sido informado y habiendo realizado todas las preguntas necesarias sobre las distintas alternativas terapéuticas disponibles para mi tratamiento, he comprendido y he decidido solicitar la realización de la tipificación del:

gen *KRAS* (Exón 2 - codones 12, 13 // Exón 3 - codones 59 y 61 // Exón 4 – codones 117 y 146) y

gen *NRAS* (Exón 2 - codones 12 y 13 // Exón 3 - codones 59 y 61 // Exón 4 – codones 117 y 146)

También se me informó que el destino de la muestra está vinculado exclusivamente al estudio de tipificación de los genes de *KRAS* y *NRAS*. El laboratorio de biología molecular que efectúe el estudio diagnóstico devolverá al médico tratante cuanto no hubiese sido utilizado de la muestra del material de biopsia, sin perjuicio de que podrá conservar parte de dicha muestra con la finalidad de respaldar los resultados informados. Estoy de acuerdo de que una vez que el laboratorio de biología molecular evalúe que no es necesario conservar la parte de la muestra conservada con fines de respaldo, la misma podrá ser utilizada por dicho laboratorio para validaciones metodológicas u otros fines científicos estadísticos respetando siempre y en todo la normativa aplicable.

Además, se me informó que en el caso de que se efectúen publicaciones científicas, las mismas únicamente podrán consistir en datos estadísticos o que no vinculen la muestra con mi identidad, ya que se realizarán bajo la modalidad de disociación irreversible, es decir que nadie podrá relacionar la publicación con mi persona.

También se me explico que en cualquier momento puedo solicitar información y acceso sobre el tratamiento de mis datos personales y de salud, como así también requerir la modificación y/o supresión de los datos, sin que ello afecte mi derecho a la atención de mi salud.

Declaro que he comprendido la información que se me ha brindado y que he tenido todas las oportunidades de preguntar respecto a la decisión de realizar la tipificación de los genes *KRAS* y *NRAS* y que las mismas han sido debidamente respondidas y comprendidas.

FIRMA PACIENTE y/o REPRESENTANTE LEGAL:

NOMBRE Y APELLIDO:

C.I.:

FECHA:

HORA:

Figura 32: Imagen del consentimiento informado del Laboratorio Genia para pacientes a los cuales se le va a analizar los genes *KRAS* y *NRAS*. Recuadrado en rojo, frase que da permiso de utilizar las muestras con fines científicos.

8.2. Frecuencias alélicas de la Sociedad Argentina de Genética Forense Asociación Civil (SAGF) 2017

CSF1PO		TPOX		D13S317	
Alelo	Frecuencia	Alelo	Frecuencia	Alelo	Frecuencia
6	0,0004	5	0,00016	6	0,00055
7	0,00167	6	0,00261	7	0,00027
8	0,00448	7	0,00142	8	0,09117
8,3	0,00006	8	0,48349	9	0,15994
9	0,02185	9	0,0779	10	0,07562
10	0,26929	10	0,04761	11	0,22469
10,3	0,00006	11	0,28963	12	0,24182
11	0,28222	12	0,09435	13	0,12787
12	0,34753	13	0,00256	14	0,07606
13	0,06209	14	0,00012	15	0,00196
14	0,00868	16	0,00005	16	0,00005
15	0,00167	21	0,00005		
		21,2	0,00005		
TH01		D16S539		D5S818	
Alelo	Frecuencia	Alelo	Frecuencia	Alelo	Frecuencia
4	0,00023	5	0,00005	7	0,0693
5	0,00064	7	0,00011	8	0,00635
6	0,29564	8	0,01758	9	0,0421
7	0,26031	9	0,16177	10	0,05326
8	0,08001	10	0,10915	11	0,42216
9	0,12342	11	0,27673	12	0,27108
9,3	0,23091	12	0,27623	13	0,12503
10	0,00808	13	0,1379	14	0,00843
11	0,0007	14	0,01917	15	0,00208
13,3	0,00006	15	0,00131	16	0,00021

Figura 33: Datos de frecuencia alélica para 6 marcadores aportados por la Sociedad Argentina de Genética Forense Asociación Civil (SAGF) del año 2017.

8.3. Resultados del límite de detección del panel KN realizado para 50 y 500 copias

Tabla 42: Frecuencias detectadas en cuatro posiciones evaluadas con el panel KN en 5 réplicas para el estudio de LD con 50 copias de genoma. * Variantes no detectadas.

Condiciones	Gen	Chr	Posición	Ref	Cambio	Réplica 1 (%)	Réplica 2 (%)	Réplica 3 (%)	Réplica 4 (%)	Réplica 5 (%)
50 copias de genoma y 12,5% de frec. alélica	KRAS	12	25378562	C	T	12.1	16.2	18.0	19.2	25.5
	KRAS	12	25380275	T	G	12.3	18.0	12.6	9.2	8.8
	KRAS	12	25398285	C	A	15.6	17.0	16.7	5.4	20.9
	NRAS	1	115256529	T	C	15.3	3.4	3.7	5.4	5.1
50 copias de genoma y 6,25% de frec. alélica	KRAS	12	25378562	C	T	0*	5.2	0*	16.3	5.0
	KRAS	12	25380275	T	G	5.2	7.9	8.5	4.2	9.5
	KRAS	12	25398285	C	A	0*	9.2	14.7	13.5	11.8
	NRAS	1	115256529	T	C	4.3	4.7	3.7	4.5	5.3
50 copias de genoma y 3,125% de frec. alélica	KRAS	12	25378562	C	T	4.0	3	0*	3.0	4.8
	KRAS	12	25380275	T	G	4.3	3	0*	2.0	7.3
	KRAS	12	25398285	C	A	5.3	0*	0*	4.9	0*
	NRAS	1	115256529	T	C	0*	0*	0*	6.0	3.0

Tabla 43: Frecuencias detectadas en cuatro posiciones evaluadas con el panel KN en 5 réplicas para el estudio de LD con 500 copias de genoma. * Estos porcentajes fueron calculados a partir del IGV.

Condiciones	Gen	Chr	Posición	Ref	Cambio	Réplica 1 (%)	Réplica 2 (%)	Réplica 3 (%)	Réplica 4 (%)	Réplica 5 (%)
500 copias de genoma y 6,25% de frec. alélica	KRAS	12	25378562	C	T	6.2	3.8	6.8	5.5	6.7
	KRAS	12	25380275	T	G	4.8	7.7	4.9	6.1	7.3
	KRAS	12	25398285	C	A	11.9	7.0	12.0	7.7	7.8
	NRAS	1	115256529	T	C	4.0	3.9	3.3	1.0*	3.5
500 copias de genoma y 3,125% de frec. alélica	KRAS	12	25378562	C	T	2.0*	3.0*	4.2	3.0*	3.8
	KRAS	12	25380275	T	G	2.0*	3.6	2.0*	3.6	3.0*
	KRAS	12	25398285	C	A	3.8	3.7	3.2	3.0*	4.3
	NRAS	1	115256529	T	C	2.0*	1.0	2.0*	3.0*	1.0*
500 copias de genoma y 1,56% de frec. alélica	KRAS	12	25378562	C	T	1.0*	1.0*	2.0*	2.0*	2.0*
	KRAS	12	25380275	T	G	2.0*	1.0*	1.0*	1.0*	2.0*
	KRAS	12	25398285	C	A	3.0*	3.0*	2.0*	2.0*	2.0*
	NRAS	1	115256529	T	C	2.0*	1.0*	2.0*	2.0*	2.0*