

UNIVERSIDAD DE VALENCIA
DOCTORADO EN BIOMEDICINA Y FARMACIA
Departamento de Farmacología

FUNDACIÓN INVESTIGACIÓN CLÍNICO DE VALENCIA-INCLIVA
Unidad de Genómica y Diagnóstico Genético

**DESARROLLO DE UN SISTEMA DE DETECCIÓN DE MUTACIONES
SOMÁTICAS DE INTERÉS CLÍNICO MEDIANTE SECUENCIACIÓN
MASIVA EN MUESTRAS ONCOLÓGICAS**

Tesis Doctoral Presentada por:
ENRIQUE SEDA GARCÍA

Dirigida por:
Dr. Sebastián Blesa Luján
Dra. M^a Rosario Abellán Sánchez
Dr. Benjamín Sarriá Chust

Valencia, MARZO 2018



VNIVERSITAT
DE VALÈNCIA

INCLIVA | VLC
Instituto de Investigación Sanitaria

D. SEBASTIÁN BLESA LUJÁN, Doctor en Ciencias Biológicas y Colaborador de Investigación en la Unidad de Genómica y Diagnóstico Genético-INCLIVA.

Dña. M^a ROSARIO ABELLÁN SÁNCHEZ, Doctora en Bioquímica y Colaboradora de Investigación en el Servicio de Bioquímica y Patología Molecular-INCLIVA.

D. BENJAMÍN SARRIÁ CHUST, Doctor en Medicina y Profesor Titular de Medicina, Departamento de Farmacología, Universidad de Valencia.

INFORMAN:

Que D. **ENRIQUE SEDA GARCÍA**, Licenciado en Ciencias Biológicas por la Universidad de Valencia, ha realizado bajo nuestra dirección en la Fundación de Investigación Clínico de Valencia-INCLIVA, en la Unidad de Genómica y Diagnóstico Genético, el trabajo experimental conducente a la elaboración de la presente memoria de investigación destinada a Tesis Doctoral, que se presenta bajo el título de: “**DESARROLLO DE UN SISTEMA DE DETECCIÓN DE MUTACIONES SOMÁTICAS DE INTERÉS CLÍNICO MEDIANTE SECUENCIACIÓN MASIVA EN MUESTRAS ONCOLÓGICAS**”.

Y para que así conste, expido el presente en Valencia, a 13 de marzo de 2018.

Dr. Sebastián
Blesa Luján

Dra. M^a Rosario
Abellán Sánchez

Dr. Benjamín
Sarriá Chust

“Lo importante no es lo que sucede, sino cómo lo interpretamos”

Epicteto

A mi hermana M^a Cristina

AGRADECIMIENTOS

Me gustaría que estas líneas sirvieran para expresar mi más profundo y sincero agradecimiento a todas aquellas personas que con su ayuda han colaborado en la realización del presente trabajo.

En especial al Dr. Sebastián Blesa Luján, a la Dra. M^a Rosario Abellán Sánchez y al Dr. Benjamín Sarriá Chust, directores de esta Tesis Doctoral, por la orientación, el seguimiento y la supervisión continua de la misma, pero sobre todo por la motivación y el apoyo recibido a lo largo de estos años.

Especial reconocimiento merece el interés mostrado por mi trabajo y las sugerencias recibidas del Dr. Felipe Javier Chaves Martínez, con el que me encuentro en deuda por el ánimo infundido y la confianza en mí depositada.

También me gustaría agradecer:

Al Departamento de Farmacología, por aceptarme en el Programa de Doctorado “Biomedicina y Farmacia” donde he podido presentar este trabajo de investigación, en especial a Mamen, por toda la ayuda para formalizar dicho trabajo.

Al Servicio de Anatomía Patológica del Hospital Clínico de Valencia, a Cristina y Elisa por su dedicación y aporte material.

Quiero expresar mi más sincero agradecimiento:

A Charo por su supervisión y dirección, siempre apoyándome y guiándome durante la realización de esta Tesis Doctoral.

A Pilar, por su amistad y por la inestimable ayuda en la orientación para la presentación de esta Tesis y en general por su colaboración desinteresada en este trabajo de investigación.

Quisiera hacer extensiva mi gratitud a mis compañeros de la Unidad de Genómica y Diagnóstico Genético:

A Cristina y Victoria “mis queridísimas Mamis”, para mi han sido el gran ejemplo a seguir en el trabajo, por toda la formación que recibí de ellas al entrar en la Unidad, con sus arduas explicaciones de las zonas “PRE y POST”.

A Ana Bárbara, Ana Ochando, Azahara, Carolina, Griselda, Inma, Javi, Laura, Loles, Raquel, Vanessa, Verónica González, Verónica Lendínez y Verónica Navarro por su amistad, apoyo y colaboración. Aunque la mayoría de ellas ya no estén trabajando en esta Unidad, siempre serán UGDG.

No quisiera olvidarme de mis otras compañeras Nerea y Carmen, agradecerles su amistad y apoyo.

Al Servicio de Bioinformática Alba, Alex, Dani, Miguel y Pablo, y por extensión a todo el personal del INCLIVA por toda la ayuda obtenida.

Me gustaría agradecer, ya en el terreno más familiar:

A mi otra familia Tani, Fernan, David, Sergio y Nuria, por haberme hecho participe de vuestra vida. En especial a Tani, que me ha dirigido y aconsejado en todas las fases de esta Tesis Doctoral, ha estado siempre “ahí”, en los momentos felices y no tan felices de mi vida, de todo corazón gracias “brother”.

A mis padres por haberme dado una educación basada en el respeto y valores, gracias a vosotros he llegado hasta aquí, especialmente a ti Mamá, que siempre has estado apoyándome en todas las facetas de mi vida.

A mis hermanos Paco, Irene, M^a Cristina, Samuel y Abraham, cuñados y sobrinos por la comprensión y el cariño recibidos.

A mi compañero de vida Luis, por todo su apoyo, paciencia y confianza depositada en mí.

En resumen, a todas las personas que de una u otra forma permitieron que este trabajo de investigación pudiera desarrollarse.

A todos, muchas gracias...

ÍNDICE

ABREVIATURAS.....	1
RESUMEN.....	7
INTRODUCCIÓN.....	11
1. El cáncer de colon	14
2. Patología molecular del cáncer de colon	17
3. Principales genes implicados en el cáncer de colon.....	19
4. Subtipos de Cáncer Colorrectal	27
5. Datos clínicos en la terapia contra el cáncer de colon.....	28
6. Técnicas de detección de mutaciones en oncología clínica	33
6.1 Técnicas de genotipado	34
6.2 Técnicas de secuenciación	41
7. Validación	53
HIPÓTESIS Y OBJETIVOS.....	57
METODOLOGÍA.....	61
1. Diseño del método de análisis	63
2. Obtención y preparación de las muestras	65
3. Metodología de la secuenciación.....	67
3.1 Descripción de los adaptadores universales 454.	67
3.2 Obtención de los fragmentos por PCR.....	68
3.3 Purificación de la librería	69
3.4 Cuantificación de la librería	70
3.5 Amplificación de la librería por PCR en emulsión o emPCR.....	70
3.6 Enriquecimiento de la librería.....	72
3.7 Secuenciación.....	72
3.8 Calidad de la secuenciación y procesado de datos	73
4. Rendimientos de la técnica.....	76

5. Pruebas de Validación.....	77
5.1 Exactitud del método de análisis.....	77
5.2 Precisión del método de análisis.....	78
5.3 Sensibilidad y especificidad analítica.....	80
5.4 Concordancia con otros métodos.....	81
6. Controles de calidad.....	87
RESULTADOS.....	89
1. Diseño del panel con adaptadores 454 universales.....	91
1.1 Obtención de los fragmentos por PCR.....	92
1.2 Estandarización de la librería.....	93
2. Calidad de la secuenciación.....	96
2.1 Rendimientos de la técnica.....	97
3. Pruebas de validación.....	101
3.1 Exactitud del método de análisis.....	101
3.2 Precisión del método de análisis.....	102
3.3 Sensibilidad y especificidad analítica.....	105
3.4 Concordancia con otros métodos.....	109
4. Resultados en la participación de los controles de calidad.....	115
5. Resultados mutacionales de la población analizada.....	117
DISCUSIÓN.....	119
CONCLUSIONES.....	139
BIBLIOGRAFÍA.....	143

ABREVIATURAS

- **ADN:** Ácido desoxirribonucleico.
- **AKT:** Gen que codifica la proteína AKT también llamada PKB, proteína cinasa B.
- **APC:** (*Adenomatous Polyposis Coli*). Gen *APC*, Poliposis Adenomatosa Colorrectal
- **APS:** Adenosín 5'-fosfosulfato.
- **ASCO:** (*American Society of Clinical Oncology*). Sociedad Americana de Oncología Clínica.
- **ATP:** Adenosín trifosfato.
- **AVA:** (*Amplicon Variant Analyzer*). Programa desarrollado por Roche para el procesado de datos.
- **BOND:** Estudio aleatorizado en el que se observó la combinación de cetuximab + irinotecan
- **BRAF:** Gen que codifica la proteína BRAF, proteína de la familia de cinasas serina-treonina.
- **CCD:** (*Charge-coupled device*). Dispositivo de carga acoplada.
- **CCR:** Cáncer colorrectal.
- **CCRM:** Cáncer colorrectal metastásico.
- **cdNA:** ADN complementario. Es una hebra de ADN de doble cadena, una de las cuales constituye una secuencia totalmente complementaria del ARN mensajero a partir del cual se ha sintetizado.
- **CELIM:** Estudio donde se ha analizado la tasa de respuestas y de reseccabilidad de metástasis hepáticas en pacientes tratados con cetuximab más FOLFOX6 o FOLFIRI.
- **CLIA:** (*Clinical Laboratory Improvement Amendments*). Enmiendas de Mejora de Laboratorio Clínico.
- **CMS:** (*Consensus Molecular Subtypes*). Consenso Molecular de Subtipos de Cáncer Colorrectal.
- **CPNCP:** Cáncer de pulmón de célula no pequeña.
- **CRAF:** Gen que codifica la proteína CRAF, proteína de la familia de cinasas serina-treonina.
- **CRYSTAL:** (cetuximab combined with irinotecan in first-line therapy for metastatic colorectal cancer). Estudio que comparaba FOLFIRI con FOLFIRI más cetuximab en pacientes con CCRM en primera línea de tratamiento.

- **CV:** Coeficiente de variación, mide la dispersión o variación de nuestras medidas, frente a la media de éstas.
- **DCC:** (*Detected in Colorectal Carcinoma*). Gen *DCC*, Detectado en Carcinoma Colorrectal
- **ddNTPs:** didesoxinucleótidos.
- **DMSO:** Dimetil sulfóxido.
- **EGF:** (*Epidermal Growth Factor*). Factor de crecimiento epidérmico
- **EGFR:** (*Epidermal Growth Factor Receptor*). Receptor del factor de crecimiento epidérmico.
- **EMEA:** Agencia Europea de Medicamentos.
- **EPIC:** Estudio que compara irinotecan en monoterapia vs irinotecan con cetuximab.
- **EPMA:** Asociación Europea de la Medicina Predictiva, Preventiva y Personalizada.
- **ErbB:** Familia de cuatro receptores: HER-1 (EGFR), HER-2 (neu/erb B2), HER-3 y HER-4.
- **ERK:** (*Extracelular signal-regulated kinase*). Cinasa regulada por señales extracelulares.
- **ESMO:** (*European Society for Medical Oncology*). Sociedad Europea de Oncología Médica.
- **FDA:** (*Food and Drug Administration*). Administración de alimentos y medicamentos.
- **FFPE:** (*Formalin Fixed Paraffin Embedded*). Muestras embebidas en parafina fijadas en formol.
- **FOLFIRI:** (**FOL:** Folinato, **F:** Fluoruracilo, **IRI:** Irinotecán) Tipo de quimioterapia combinada que se usa en el tratamiento de cáncer colorrectal.
- **FOLFOX:** (**FOL:** Folinato, **F:** Fluoruracilo, **OX:** oxaliplatino). Tipo de quimioterapia combinada que se usa en el tratamiento de cáncer colorrectal. Hay varios regímenes FOLFOX diferentes.
- **Gb:** Gigabase.
- **GDP:** Guanosín difostato.
- **GEF:** Factor intercambiador de nucleótidos de guanina.
- **GTP:** Guanosín trifosfato.

- **ISFET:** (*Ion Sensitive Field Effect Transistor*). Transistor de efecto campo sensible a iones.
- **KRAS:** (*Kirsten rat sarcoma viral oncogene homolog*). Gen *V-Ki-ras2*. Homólogo del oncogén viral del sarcoma de Kirsten.
- **LOD:** (*lower limit of detection*). Límite inferior de detección.
- **MAPK:** (*Mitogen activated protein kinase*). Vía de las proteínas cinasas activadas por mitógenos.
- **MEK:** Enzima codificada en humanos por el gen *MAP3K1*.
- **MID:** (*Multiplex identifier*). Secuencia de 10 nucleótidos que es identificadora específica de cada individuo, para la plataforma JR 454/Roche.
- **MITF:** Factor de transcripción asociado a microftalmia.
- **mTOR:** (*mammalian Target of Rapamycin*). Gen que codifica una proteína llamada la diana de rapamicina en células de mamífero.
- **MYC:** Familia de protooncogenes compuesta por varios miembros (*L-myc*, *N-myc* y *c-myc*). Se encuentran en las células normales y codifican proteínas del núcleo de la célula que se unen al ADN y facilitando su transcripción.
- **NaOH:** Hidróxido sódico.
- **NGS:** (*Next Generation Sequencing*). Secuenciación de Nueva Generación.
- **NRAS:** Homólogo del oncogén del virus del neuroblastoma.
- **OPUS:** Estudio en fase II aleatorizado que comparaba FOLFOX4 + cetuximab frente a FOLFOX4.
- **P53:** Gen *P53* o gen *TP53*, gen que codifica una proteína de 53KDa, llamado también guardian del genoma.
- **Pb:** Pares de bases.
- **PCR:** (*Polymerase Chain Reaction*). Reacción en cadena de la polimerasa, es una técnica de biología molecular desarrollada en 1983 por Kary Mullis. Su objetivo es obtener un gran número de copias de un fragmento de ADN.
- **PIK3CA:** Gen que codifica la fosfatidilinositol-4,5-bisfosfato 3-cinasa o fosfoinositida-3-cinasa (PI3K).
- **PRIME:** (*Panitumumab Randomized trial in combination with chemotherapy for Metastatic colorectal cancer to determine Efficacy*). El ensayo clínico PRIME es un estudio global, multicéntrico y aleatorizado de Fase 3 diseñado para evaluar Vectibix

más FOLFOX en comparación con FOLFOX en pacientes con CCRM metastásico y *KRAS wt*.

- **PTEN**: Gen que codifica la enzima fosfatidilinositol-3,4,5-trisfosfato 3-fosfatasa.
- **RAF**: Proto-oncogén serina / treonina-proteína cinasa
- **RAS**: Gen que codifica para la proteína RAS que es una proteína G monomérica, GTPasa con actividad reguladora GTP-hidrolasa.
- **SBS**: (*Sequencing By Synthesis*). Secuenciación por síntesis.
- **SD**: Desviación estándar.
- **SEAP**: Sociedad Española de Anatomía patológica.
- **SEOM**: Sociedad Española de Oncología Médica.
- **SFF**: Archivos de datos estándar de formato de diagrama de flujo.
- **SMRT**: (*Single Molecule Real Time Sequencing*). Secuenciación a tiempo real de una única molécula de ADN.
- **Tm**: (*Temperature Melting*). Temperatura de fusión del ADN, en la cual las dobles hebras de ADN se separan en solución en dos hebras completamente independientes. En nuestro caso es la temperatura a la cual el ADN molde está separado, pero permite la unión de los cebadores.
- **TGFβ**: Factor de crecimiento transformante beta. Ruta que afecta a varias etapas celulares como el crecimiento, diferenciación celular, el proceso de apoptosis y la homeostasis celular.
- **WNT**: Vía de transducción de señales al exterior de la célula.
- **WT**: (*Wild-Type*). Tipo salvaje. Se considera como alelo salvaje o silvestre a la primera variante alélica descrita de un gen.
- **ZMW**: (*Zero Mode Waveguide*). Guía de onda de modo cero.

RESUMEN

DESARROLLO DE UN SISTEMA DE DETECCIÓN DE MUTACIONES SOMÁTICAS DE INTERÉS CLÍNICO MEDIANTE SECUENCIACIÓN MASIVA EN MUESTRAS ONCOLÓGICAS

INTRODUCCIÓN. La descripción de los mecanismos moleculares de algunos de los tumores más frecuentes ha permitido el desarrollo de nuevos fármacos capaces de antagonizar moléculas propias de los tumores. Esto implica la necesidad de determinar la existencia de determinadas alteraciones génicas cuya presencia va asociada a la decisión terapéutica en el uso de estos fármacos. Por lo que ha sido necesario el desarrollo de técnicas de análisis genético que sean lo suficientemente sensibles para detectar la presencia de mutaciones somáticas de interés clínico, a la vez que lo suficientemente óptimas en términos de consumo económico, de muestra y que proporcionen una respuesta rápida.

OBJETIVOS. Desarrollar y validar un sistema de detección rápido, sensible, reproducible y fácilmente integrable en la rutina clínica, para detectar mutaciones de interés clínico. Este sistema debe ser escalable y modulable, y el diseño del método de validación debe ser extrapolable a otras plataformas de secuenciación.

METODOLOGÍA. A partir de biopsias incluidas en parafina de pacientes con cáncer colorrectal metastásico (CCRM) se analizaron muestras portadoras de mutaciones en los genes *BRAF/KRAS/NRAS* y no portadoras. Se seleccionaron las condiciones de amplificación por técnicas de PCR para obtención de los amplicones, y se secuenciaron mediante el sistema 454 GS Junior de Roche. Se validaron los resultados obtenidos por medio de la exactitud, precisión, sensibilidad y concordancia con otros métodos estandarizados para asegurar la fiabilidad de los resultados.

RESULTADOS. Se diseñaron y optimizaron las condiciones para la amplificación de las zonas de interés de los genes descritos. Los resultados de las carreras analizadas nos permitieron establecer una calidad del proceso con un promedio de *Passed Filter* del 45.94% y de *Dot+Mixed* del 17.73%. Así como unos rendimientos de la técnica expresados en coberturas de 135.000 lecturas por carrera, 8.300

lecturas por muestra y 1.000 lecturas por amplicón. La exactitud del método de análisis mostró un error relativo inferior al 9% en muestras comerciales, portadoras de mutación al 50%, aumentando cuando se analizan en muestras con varias mutaciones y a porcentajes más bajos. De la precisión en términos de repetibilidad (intra-run) se obtuvieron coeficientes de variación (CV) del 8-23%, y de repetibilidad (inter-run) del 8-15%. Al estudiar la sensibilidad y especificidad analíticas se observó que al 2% de presencia de la mutación y a 1000 lecturas de cobertura, se obtuvieron valores del 100% de sensibilidad y 97% de especificidad analítica. Se obtuvieron unas concordancias con nuestro sistema de secuenciación NGS del 100% para el *kit BRAF Mutation analysis*, del 81% con el *test cobas KRAS Mutation* y del 63-85% para NRAS con el método de Sanger, todos los casos con un p-valor < 0.05. La participación en los controles de calidad hizo visible que los resultados de nuestro método eran fiables. Los resultados mutacionales de la población analizada coincidieron con datos reportados en poblaciones equiparables.

CONCLUSIONES. El método desarrollado en este trabajo permite una detección rápida y fácilmente integrable en la rutina clínica para la detección de mutaciones de interés oncológico. Mediante el uso de este panel es posible detectar estas mutaciones con una sensibilidad del 2% en el exón 15 del gen *BRAF* y en los exones 2, 3 y 4 de los genes *KRAS* y *NRAS*. El diseño de la validación es aplicable a otras plataformas de NGS.

INTRODUCCIÓN

En las últimas décadas, se ha producido una gran cantidad de conocimientos relacionados con los mecanismos patogénicos del cáncer, sobre todo gracias a la descripción de los procesos moleculares de algunos de los tumores más frecuentes. Esto ha permitido el desarrollo de nuevos fármacos capaces de antagonizar moléculas propias de los tumores, que se comportan como “dianas terapéuticas”. Este tipo de planteamiento, implica la necesidad de determinar algunas alteraciones génicas en la neoplasia. Estudios donde se han realizado este tipo de detecciones en muestras poblacionales, comparando las mutaciones presentes en las biopsias analizadas con el efecto de fármacos en desarrollo, han permitido el establecimiento de una serie de biomarcadores genéticos cuya presencia va asociada a la decisión terapéutica en el uso de estos fármacos (LiVolsi V *et al.*, 2012).

Las directrices de la Asociación Europea de la Medicina Predictiva, Preventiva y Personalizada (EPMA) de 2012 recomiendan introducir aproximaciones integradoras para diagnósticos avanzados, dirigidos a la prevención y tratamientos personalizados, con un máximo aprovechamiento del coste para los sistemas de salud. La existencia de biomarcadores validados es crucial para la aplicación en la práctica clínica de estos tratamientos.

De la misma manera, una vez establecidos con claridad estos biomarcadores, se hace necesaria, en la rutina clínica, su detección con seguridad y rapidez en las biopsias cuyo análisis se encarga desde los servicios de Oncología. Esta necesidad, junto al avance de la tecnología de secuenciación de última generación, ha permitido el desarrollo de sistemas rápidos, fiables y cada vez más económicos basados en las técnicas de secuenciación masiva (Golubnitschaja O *et al.*, 2012).

1. El cáncer de colon

El Cáncer Colorrectal (CCR) es la patología tumoral con mayor incidencia en los países occidentales, junto a los tumores de mama y pulmón, siendo los que conllevan mayor mortalidad mundial. El CCR es la cuarta causa de muerte por cáncer en el mundo, aproximadamente 1.4 millones de nuevos casos y casi 700.000 muertes en 2012 (**Fig. 1**). Basándonos en las características temporales de incidencia y mortalidad en los últimos 10 años, se han identificado tres grupos de países (Arnold M *et al.*, 2017):

- Grupo 1: grupo de países donde ha aumentado la incidencia y la mortalidad de casos por CCR. Dentro de este grupo se encuentra España junto con Filipinas, China, Colombia, Bulgaria, Costa Rica, Brasil, Rusia, Bielorrusia, Estonia, Lituania, Croacia, Letonia y Polonia.
- Grupo 2: grupo de países donde la incidencia ha aumentado pero la mortalidad ha disminuido. Como Canadá, Dinamarca, Suiza, Irlanda, Suecia, Singapur, Finlandia, Noruega, Eslovaquia, Reino Unido, Holanda, Italia, Malta y Eslovenia.
- Grupo 3: grupo de países donde tanto la incidencia como la mortalidad han disminuido. Países como Estados Unidos, Austria, Nueva Zelanda, República Checa, Islandia, Francia, Japón, Australia e Israel.

Alrededor del 25-30% de los pacientes diagnosticados de CCR presentan al inicio tumores metastásicos (estadio IV) y el resto, principalmente diagnosticados en estadios II o III, más del 20-50% van a progresar a CCRM. La supervivencia global a 5 años es de aproximadamente el 50%, llegando a ser el 10% en pacientes con enfermedad en estadio IV (Cunningham D *et al.*, 2010).

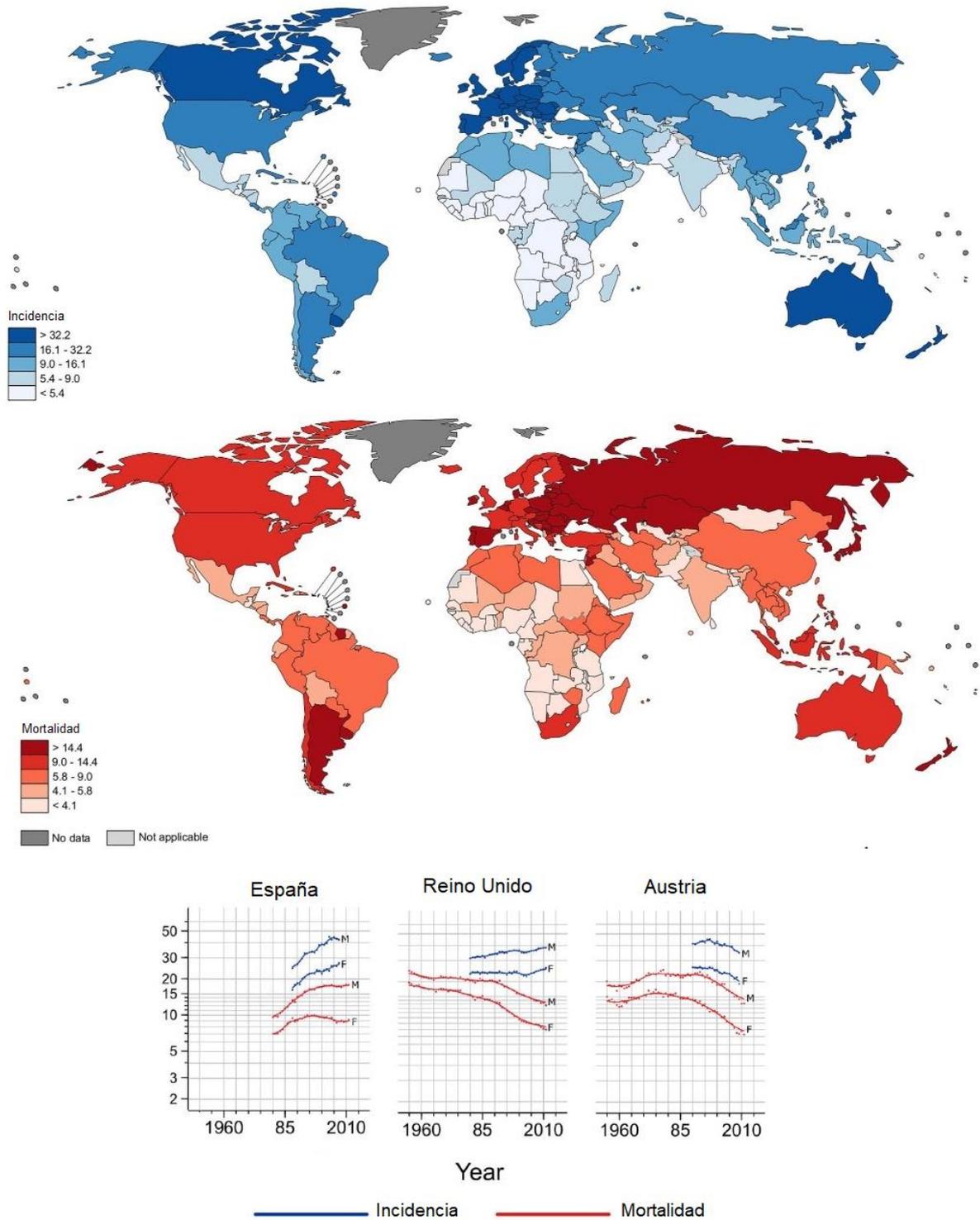


Figura 1: Incidencia y tasas de mortalidad mundial por cáncer colorrectal por cada 100000 habitantes en 2012. M: sexo masculino y F: sexo femenino.

Dada la gran incidencia y las altas tasas de mortalidad, se han ido aplicando diversos protocolos terapéuticos basados en agentes quimioterápicos generales y, más recientemente, se han implementado nuevos protocolos que prevén el uso combinado de dichos quimioterápicos con anticuerpos monoclonales y/o moléculas frente a dianas oncogénicas específicas como el receptor del factor de crecimiento epidérmico (EGFR). Los protocolos más efectivos para pacientes con CCRM utilizados actualmente combinan el uso de anticuerpos monoclonales (cetuximab, panitumumab) con quimioterápicos clásicos, como el irinotecan, (Cripps C *et al.*, 2010; Ochendusko SL *et al.*, 2010).

2. Patología molecular del cáncer de colon

Las características que dotan a las células de un fenotipo maligno son causadas por mutaciones en las células somáticas de los llamados oncogenes, con ganancia de función, y de los genes supresores de tumor, con pérdida de la función inhibitoria.

Un protooncogén es aquel gen, normalmente implicado en la regulación de la proliferación, que puede convertirse en un oncogén promotor de cáncer mediante una serie de mecanismos, que fundamentalmente pueden ser: mutaciones puntuales, sobreexpresión y reordenamiento cromosómico. Este oncogén puede actuar de manera dominante para ayudar a que una célula se transforme en maligna. Los productos de los oncogenes incluyen factores de transcripción, remodeladores de la cromatina, factores de crecimiento, receptores de factores de crecimiento, moléculas transductoras de señal y reguladores de la apoptosis.

Por contra, un gen supresor de tumor es aquel gen que evita la formación de un cáncer, por tanto, las mutaciones que producen una pérdida de su función aumentan la susceptibilidad al cáncer. Están involucrados en el mantenimiento de la integridad genómica y en el control del ciclo celular. Los genes supresores de tumor suelen requerir la alteración de ambos alelos para que haya un efecto oncogénico. La mutación en los genes supresores tiene como consecuencia la adquisición del denominado fenotipo mutador, que conlleva una mayor tasa de mutación y la acumulación del repertorio de mutaciones necesarias para desencadenar un cáncer (Croce CM *et al.*, 2008; Papke B *et al.*, 2017).

En el CCR, la acumulación de alteraciones oncogénicas se produce de manera progresiva. Desde lesiones iniciales hiperplásicas (aumento anormal de tamaño de un tejido orgánico debido al incremento del número de células normales que lo forman), con displasia (anomalía en el desarrollo de un tejido de una parte anatómica del organismo), hasta la formación de adenomas, de carcinomas intraepiteliales, infiltrantes y finalmente metástasis. Todo este proceso de cambios histopatológicos requiere la acumulación de numerosas alteraciones moleculares que pueden conllevar más de 10 años (Grady WM *et al.*, 2008).

La alteración más frecuente es la pérdida de función del gen *APC*, gen supresor tumoral que codifica una proteína que juega un importante papel en el desarrollo del cáncer. Su pérdida facilita la aparición del CCR y las alteraciones en la ruta de WNT, vía de transducción de señales extracelulares, incluso previas a las alteraciones morfológicas. Posteriormente, y en fases de adenomas de pequeño tamaño, es frecuente detectar mutaciones de los oncogenes *RAS* y *BRAF*. En adenomas grandes se pueden observar alteraciones en los genes *SMAD4* y *DCC*, implicados en el crecimiento celular. En los carcinomas pueden coexistir mutaciones del gen *TP53* con otras alteraciones genéticas (Fearhead NS *et al.*, 2001; Logan CW *et al.*, 2004; Popal S *et al.*, 2005; Pritchard CC *et al.*, 2011; Papke B *et al.*, 2017)

Entre la gran cantidad de alteraciones moleculares, hasta varias decenas por tumor, se pueden distinguir dos tipos fundamentales de alteraciones genéticas:

a) Centrales para el desarrollo del tumor: promueven la proliferación celular, se denominan desencadenantes o *drivers*.

b) Acompañantes: se cree que juegan un papel secundario en el desarrollo del tumor, se denominan pasajeras o *passengers*.

Gracias a los trabajos realizados por *Hanahan* en relación a la biología tumoral (Hanahan D *et al.*, 2000), hoy en día se conocen con más detalle los mecanismos implicados en la transformación y la progresión tumoral. Se ha descrito que las células tumorales presentan alteraciones en las grandes rutas de señalización celular como son: el ciclo celular, la apoptosis, la invasividad y la angiogénesis. En ese sentido, en la mayoría de los tumores la activación de las vías de señalización, independientemente de las alteraciones oncogénicas presentes, da lugar a una activación constitutiva de las mismas. Concretamente, en el CCR la señalización mediada por el EGFR desencadena un aumento de la proliferación tumoral (Hernández-Losa J *et al.*, 2012).

3. Principales genes implicados en el cáncer de colon

Las vías de señalización celular activadas por el EGFR son uno de los mecanismos de mayor importancia en la oncogénesis del CCR, por este motivo, el EGFR y las proteínas que continúan la vía de transducción de señal son dianas moleculares en el tratamiento del CCRM.

El EGFR pertenece a la familia de receptores de membrana de tipo I de ErbB, constituido por cuatro miembros. Estas proteínas están implicadas en una gran variedad de vías de transducción de señal que promueven la proliferación celular, la angiogénesis, la invasión de las células tumorales y metástasis.

El gen *EGFR* se encuentra en el brazo corto del cromosoma 7 (7p12), formado por 28 exones. Este gen codifica una glucoproteína transmembrana de 170 kDa compuesta por un dominio extracelular de unión a ligando, un dominio transmembrana y una porción citoplasmática, dividida en un dominio tirosina-quinasa y una región carboxi-terminal, que presentan residuos de tirosina y elementos reguladores del receptor (**Fig. 2**) (Nair P, 2005).



Figura 2: Localización del gen *EGFR* y estructura de su proteína.

La unión del ligando al dominio extracelular del EGFR produce su dimerización (homo- o hetero-dimerización) con otro receptor de la misma familia ErbB, lo cual provoca su autofosforilación y la consecuente activación de la región tirosina-cinasa, que, a su vez, activa una cascada de señalización mitogénica a través de vías de transducción de señal. Subyacente al EGFR y tras la activación del mismo, la señal puede dirigirse por dos grandes vías de señalización hacia el interior de la célula: a través de la activación de RAS-BRAF-MAPK y tras la activación de PIK3CA-AKT1-PTEN-mTOR (**Fig. 3**). Por tanto, como genes *drivers* de la señal proliferativa se podrían encontrar tanto *EGFR* como *RAS*, *BRAF*, *PIK3CA*, *AKT*, etc... (Hernández-Losa J *et al.*, 2012).

EGFR regula las funciones esenciales y tiene un papel oncogénico en diferentes tumores mediante la activación de mecanismos que incluyen la sobreexpresión, la presencia de alteraciones activadoras de la proteína en la secuencia del gen y la producción autocrina de ligando. Las alteraciones del gen *EGFR* se asocian con la proliferación tumoral, crecimiento, invasión, capacidad metastásica, inhibición de la apoptosis y angiogénesis tumoral.

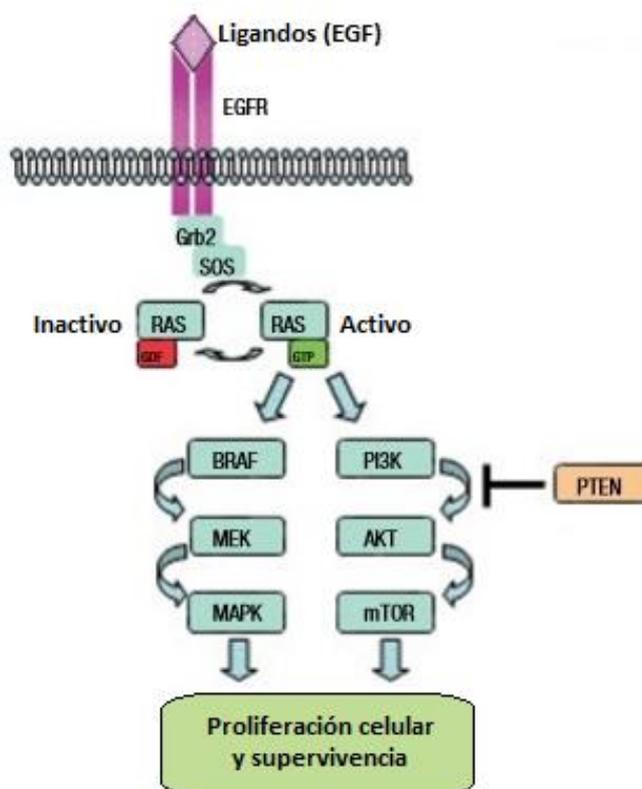


Figura 3: Esquema de las vías de señalización del EGFR.

En el CCR, el gen *EGFR* puede estar activado de forma aberrante, por lo que es una potencial diana terapéutica. Mientras que la proteína EGFR se encuentra sobreexpresada en, aproximadamente, un 80% del CCR, el aumento en el número de copias del gen se detecta entre un 5-50%. Las mutaciones activadoras en el dominio cinasa del gen son muy poco frecuentes (Papke B *et al.*, 2017).

Dentro de la familia de los genes *RAS* se conocen 3 miembros *HRAS* (homólogo del oncogén del sarcoma de la rata de Harvey), *NRAS* (homólogo del oncogén del virus *RAS* viral (v-ras)) y *KRAS* (homólogo del oncogén viral del sarcoma de Kirsten). Este último presenta dos isoformas: *KRAS 4A* y *KRAS 4B*. La isoforma B es cinco veces más abundante que la A, por lo que se encuentra con mayor frecuencia de mutación en los carcinomas de colon (McCormick F, 2015).

El gen *KRAS* se encuentra en el brazo corto del cromosoma 12 (12p12.1). Está formado por 6 exones que codifican, en condiciones normales, una proteína de 189 aminoácidos (**Fig. 4**). La proteína KRAS pertenece a la superfamilia de proteínas G, las cuales transmiten la señalización producida mediante la activación de receptores de membrana. KRAS inactiva está unida a guanosín difosfato (GDP) y, al estimularse, un factor intercambiador de nucleótidos de guanina (GEF) favorece la formación de GTP-RAS, que es la forma activa. El guanosín trifosfato (GTP) es hidrolizado a GDP por la actividad GTPasa intrínseca de las proteínas RAS, inactivándose (McCormick F, 2015).

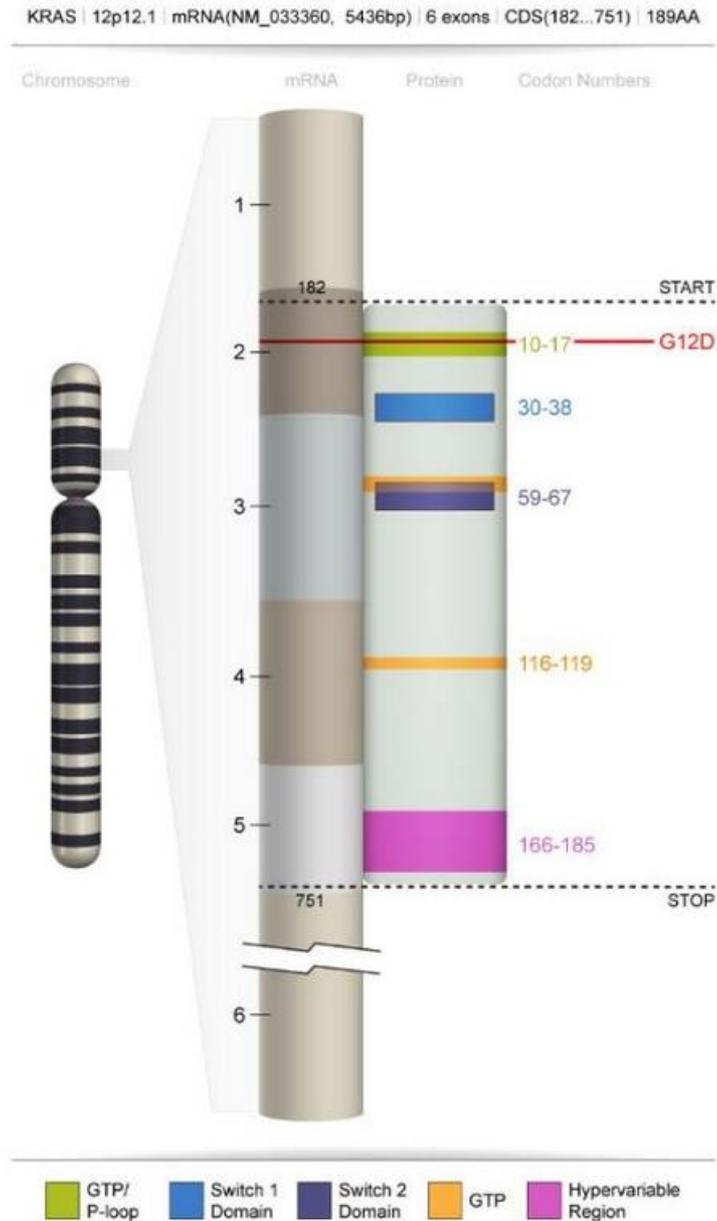


Figura 4: Localización y estructura del gen *KRAS* y de su proteína.

Históricamente, la mayoría de los estudios bioquímicos y estructurales de *RAS* se han centrado en *HRAS*. Sin embargo, el gen *HRAS* es la isoforma *RAS* con menor tasa de mutación en los cánceres humanos (<4%), mientras que el gen *KRAS* es la isoforma predominantemente mutada (85%), seguida por la forma mutada del gen *NRAS* (11%). Cuando hay mutaciones del gen *KRAS*, la actividad GTPasa queda bloqueada y la proteína *KRAS* permanece constitutivamente activada y unida a GTP. Dichas mutaciones suelen ocurrir en los codones 12 y 13 (exón 2 de del gen

KRAS), y con menor frecuencia en los codones 61 y 146 (exones 3 y 4 del gen *KRAS*) (**Fig. 7**) (Malumbres M *et al.*, 2003; Rizzo S *et al.*, 2010).

En cuanto al gen *NRAS*, este se localiza en el brazo corto del cromosoma 1 (1p13.2), y está formado por 7 exones que codifican una proteína de 189 aminoácidos (**Fig. 5**). Como en el caso de *KRAS*, pertenece a la superfamilia de proteínas G. La mayoría de las mutaciones del gen *NRAS* se detectan en los codones 60 y 61 del exón 3 (80%), y en los codones 12 y 13 del exón 2 (20%) (**Fig. 7**) (McCormick F, 2015).

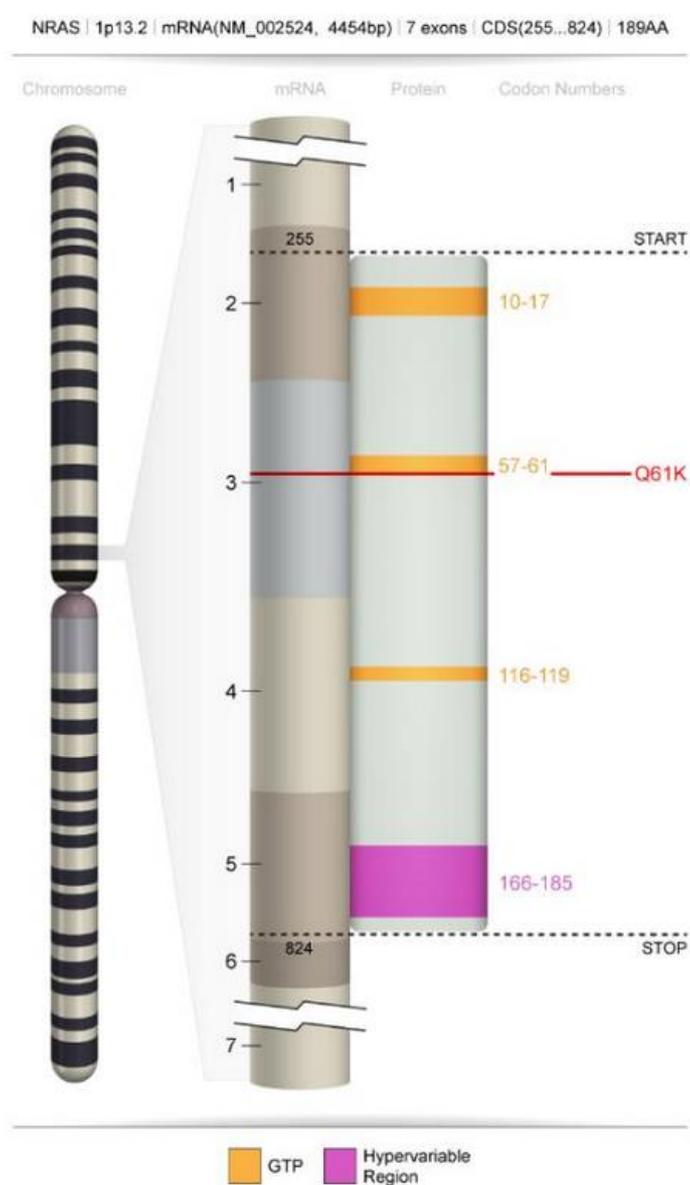


Figura 5: Estructura y localización del gen *NRAS* y su proteína.

El gen *BRAF* se encuentra en el brazo largo del cromosoma 7 (7q34), está formado por 18 exones que codifican una proteína de 766 aminoácidos (**Fig. 6**). Esta proteína es una serin-treonina cinasa, sustrato de la actividad de las proteínas RAS. Las mutaciones del gen *BRAF* se detectan entre el 6-10% de los CCR. La activación de la proteína BRAF, en su forma de homo o heterodímero, inicia la fosforilación de la MAPK cinasa MEK, enzima codificada en humanos por el gen *MAP3K1*, que induce a su vez la fosforilación de ERK (cinasa regulada por señales extracelulares), que es su único sustrato. La activación de la cinasa ERK promueve el crecimiento y la transformación de señales a través de su interacción con un número de moléculas críticas en la patogénesis tumoral (Beeram M *et al.*, 2005; Garnett MJ *et al.*, 2005).

Las mutaciones más frecuentemente observadas en el gen *BRAF* que se encuentran en el exón 15, son la V600E (80%) y la V600K (18%), siendo mucho menos frecuentes las V600D, V600R, V600A, V600G y K601E (**Fig. 7**). Las mutaciones de *BRAF* tienden a asociarse con un período libre de enfermedad más corto y una disminución de la supervivencia global, independientemente del estatus mutacional de los genes *RAS* (Edlundh-Rose E *et al.*, 2006; Tran NH *et al.*, 2015; Loupakis F *et al.*, 2016; Johnson B *et al.*, 2017).

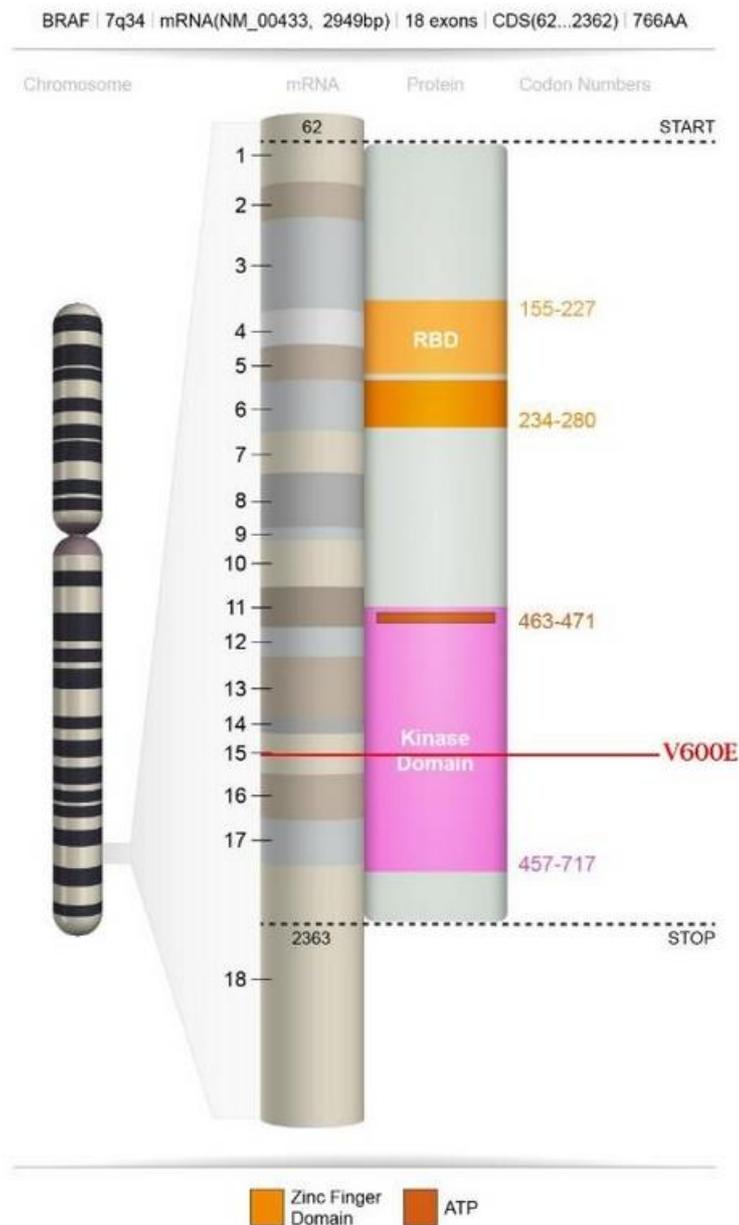


Figura 6: Localización genómica del gen *BRAF* y estructura de su proteína.

Otros genes subyacentes al *EGFR* pueden estar activados por mutaciones hasta en el 9-10% de los casos, como el gen *PIK3CA*; o por pérdida de función, como en el gen *PTEN*, hasta en el 40% de los casos (Van Cutsem E *et al.*, 2010).

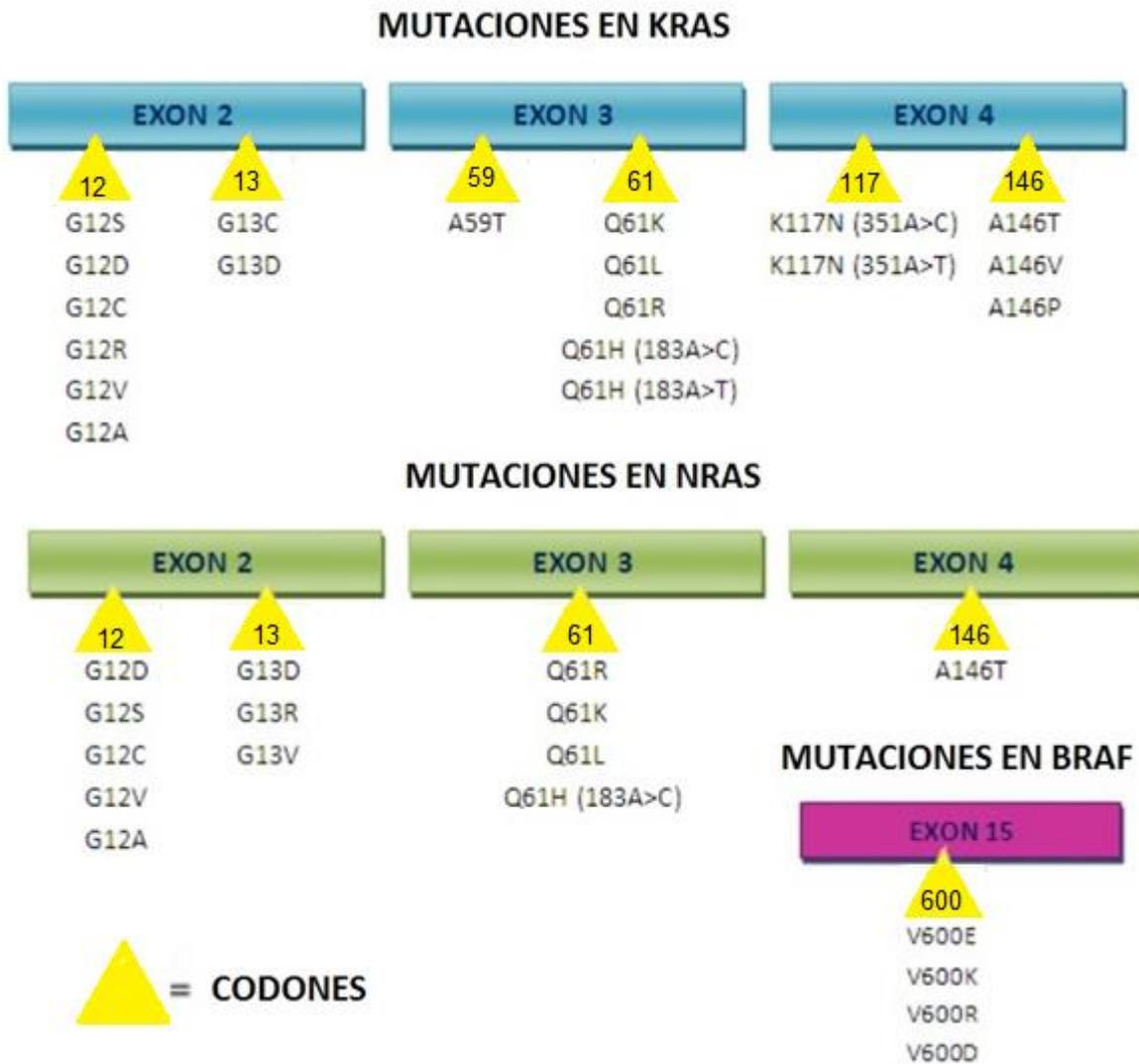


Figura 7: Esquema de las mutaciones más frecuentes encontradas en los genes *KRAS*, *NRAS* y *BRAF*.

4. Subtipos de Cáncer Colorrectal

Un trabajo colaborativo realizado gracias al “Consenso Molecular de Subtipos de Cáncer Colorrectal” (CMS) a nivel internacional ha establecido una nueva clasificación para CCR en base a datos genómicos (Guinney J *et al.*, 2015).

El estudio reunió 4.000 muestras de pacientes y definió un patrón de subtipos genómicos que además de identificar las características biológicas intrínsecas de los tumores, también correlaciona estos datos con el comportamiento clínico y el pronóstico de los pacientes.

Los grupos que se han propuesto por este CMS fueron cuatro:

- CMS1 (inestabilidad microsatélite inmune, 15% de los tumores). Subtipo que se caracteriza por una potente activación inmunológica y por tener desregulación de las vías metabólicas asociadas a mutaciones del gen *BRAF*. El grupo de pacientes afectados por este grupo es el que más beneficio obtendría con la inmunoterapia.
- CMS2 (epitelial canónico, 40% de los tumores), Grupo con más inestabilidad cromosómica, lo que sigue la carcinogénesis clásica del CCR con activación de vías de transducción: de señales al exterior celular (WNT), protooncogenes *MYC* y *EGFR*.
- CMS3 (epitelial metabólico, 15% de los tumores), se caracteriza por tener desregulación de las vías metabólicas asociadas a mutaciones del gen *RAS*.
- CMS4 (mesenquimal, 30% de los tumores), presenta un pronóstico muy desfavorable con el tratamiento estándar de CCR debido a la activación de las vías de señalización del factor de crecimiento transformante beta ($TGF\beta$), angiogénesis e invasión estromal.

5. Datos clínicos en la terapia contra el cáncer de colon

El estado mutacional de *RAS* es el marcador más usado en la práctica clínica diaria en el tratamiento del CCRM; la **Figura 8** nos muestra la evolución cronológica de *RAS* como biomarcador. Su determinación como marcador de resistencia es imprescindible para la toma de decisiones terapéuticas con un fármaco anti-EGFR. La evidencia viene dada por los análisis de *RAS* en los datos de los estudios de fase III pivotaes de cetuximab y panitumumab, tanto en monoterapia como en combinación con quimioterapia en las primeras líneas de tratamiento. Según los resultados de estos estudios, las autoridades sanitarias tanto americanas como europeas recomiendan la detección de estado mutacional de *RAS* en el tumor de los pacientes con CCRM candidatos a anti-EGFR, y si está mutado, debe administrarse terapias alternativas. (Rizzo S *et al.*, 2010; Kassouf E *et al.*, 2016)

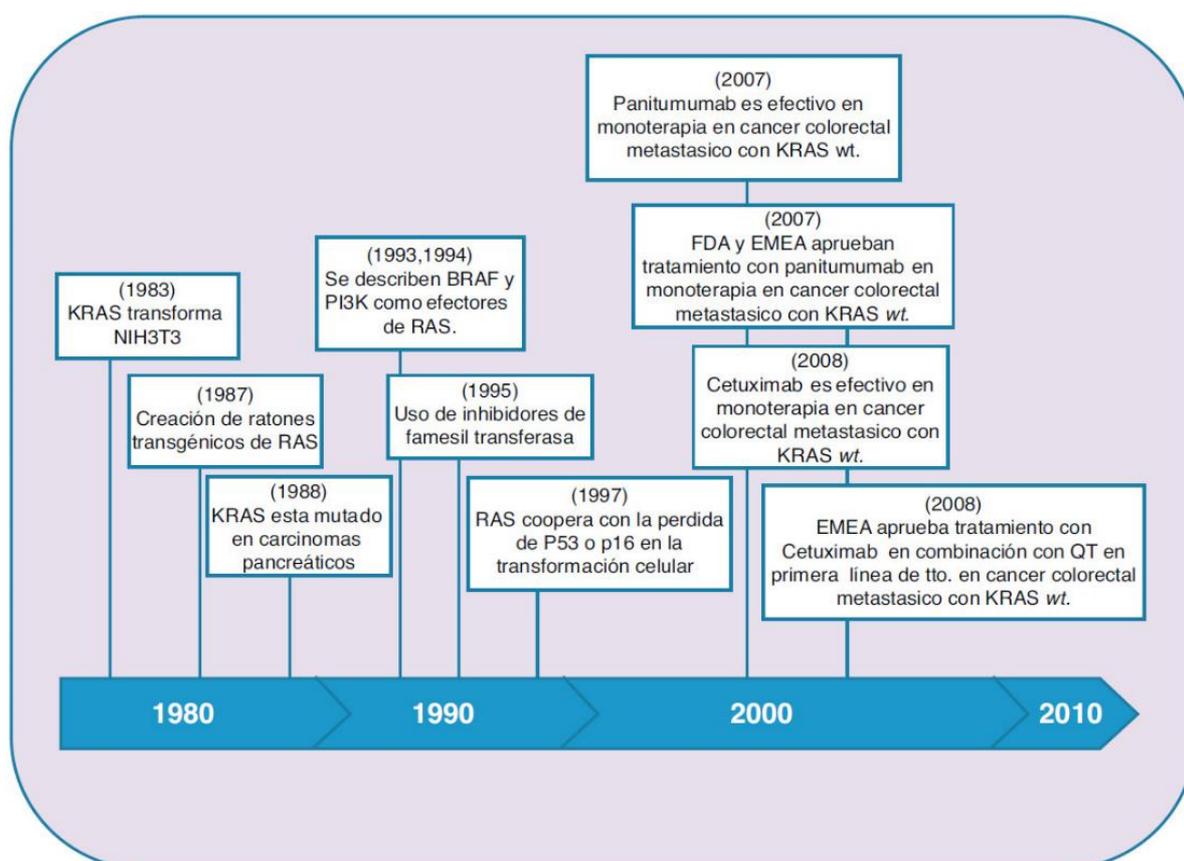


Figura 8: Cronología de la evolución de *KRAS* como biomarcador.

El cetuximab es un anticuerpo monoclonal quimérico que reconoce el dominio extracelular del EGFR. La administración de cetuximab fue inicialmente ensayada en monoterapia en pacientes no seleccionados con CCRM. En este contexto, el cetuximab inducía una tasa de respuestas modesta (8-11%) en pacientes previamente tratados con irinotecan, tasa que se incrementaba cuando se combinaba con irinotecan en esta misma población de pacientes (Cunningham D et al., 2004; Saltz LB et al., 2004).

El estudio BOND fue un estudio aleatorizado en el que se observó que la combinación de cetuximab + irinotecan inducía el doble de respuestas (22,9 vs 10,8%) que el cetuximab en monoterapia y una supervivencia libre de progresión más prolongada (4,1 vs 1,5 meses), demostrando que la adición de dicho fármaco era capaz de revertir la resistencia al irinotecan en un número significativo de pacientes (Cunningham D *et al.*, 2004; Saltz LB *et al.*, 2004; Saltz LB *et al.*, 2007).

En el mismo subgrupo de pacientes refractarios al irinotecan el estudio BOND-2 demostró que el uso de irinotecan en combinación con bevacizumab y cetuximab inducía una mejoría en cuanto a la tasa de respuesta (38%) y al tiempo hasta la progresión (8,5 meses) en esta población de pacientes pretratados (Ochenduazko SL *et al.*, 2010)

En 2008 se demostró que la presencia de mutaciones en el gen *KRAS* era un factor predictivo negativo relevante, ya que seleccionaba aproximadamente al 40% de pacientes que no se benefician de los tratamientos dirigidos frente al EGFR. Así, se ha observado que en pacientes con *KRAS WT* (nativo, *wild-type*), el empleo de cetuximab en monoterapia, comparado con el mejor tratamiento de soporte, prolongaba significativamente la supervivencia libre de progresión (3,7 vs 1,9 meses) y la supervivencia global (9,5 vs 4,8 meses) con tasas de respuesta del 12,8 frente al 1,2% (Karapetis CS *et al.*, 2008).

En segunda línea de tratamiento los resultados del estudio EPIC, comparando irinotecan en monoterapia frente a irinotecan con cetuximab, tras progresión a una primera línea basada en oxaliplatino, han mostrado un beneficio en la tasa de repuestas (16,4 vs 4,2%) y supervivencia libre de progresión (4,0 vs 2,6 meses), aunque el estudio no consiguió demostrar el objetivo primario de incremento de

supervivencia global (10,7 vs 10 meses). No obstante, este estudio se llevó a cabo en una población no seleccionada en función de *KRAS* (Amado RG *et al.*, 2008).

El cetuximab también se ha ensayado en primera línea de tratamiento tanto en combinación con esquemas basados en oxaliplatino como en combinación con FOLFIRI. De nuevo estos estudios se llevaron a cabo antes de que se implantara el uso de la determinación de estatus mutacional de *KRAS*, si bien retrospectivamente se ha logrado analizarlo en la mayoría de la población incluida. El estudio OPUS es un estudio fase II aleatorizado que comparaba la terapia con FOLFOX4 + cetuximab frente a FOLFOX4 (Pritchard CC *et al.*, 2011). La subpoblación con *KRAS* WT se benefició claramente del tratamiento con cetuximab, con tasas de respuesta (57 vs 34%) y supervivencia libre de progresión (8,3 vs 7,2 meses) significativamente superiores que las de los pacientes que solo recibieron FOLFOX. (Bokemeyer C *et al.*, 2009; Bokemeyer C *et al.*, 2010).

De manera similar, los resultados globales del estudio CRYSTAL (Van Cutsem E *et al.*, 2010), que comparaba FOLFIRI con FOLFIRI más cetuximab en pacientes con CCRM avanzado en primera línea de tratamiento, mostraron un beneficio a favor del cetuximab en cuanto a tasa de respuestas (46,9 vs 38,7%) y supervivencia libre de progresión (8,9 vs 8 meses). No obstante, tras reanalizar de manera retrospectiva la eficacia en función del estado mutacional de *KRAS*, se volvió a confirmar que los beneficios asociados a la adición de cetuximab a FOLFIRI se observaban en la población con *KRAS* WT, siendo la magnitud de estos beneficios mucho mayor que los comunicados inicialmente en término de la tasa de respuesta (57,3 vs 39,7%), supervivencia libre de progresión (9,9 vs 8,4 meses) y supervivencia global (23,5 vs 20 meses) (Van Cutsem E *et al.*, 2009).

En el estudio CELIM se ha analizado la tasa de respuestas en metástasis hepáticas en pacientes tratados con cetuximab más FOLFOX 6 o FOLFIRI, confirmándose los resultados de los estudios anteriores; en los pacientes con *KRAS* WT la tasa de respuestas fue del 70%, y en los pacientes con *KRAS* mutado, del 41% (Kohne C *et al.*, 2010; Folprecht G *et al.*, 2010).

El Panitumumab es un anticuerpo monoclonal totalmente humanizado que reconoce el dominio extracelular del EGFR. En 2008 se pudo observar que, en

pacientes con un *KRAS* WT, el empleo de panitumumab en monoterapia, comparado con el mejor tratamiento de soporte, prolongaba significativamente la supervivencia libre de progresión (12,3 vs 7,3 semanas) y la supervivencia global (8,1 vs 7,6 meses), con tasas de respuesta del 17 frente al 0% (Peeters M *et al.*, 2010).

El panitumumab, por su lado, ha sido comparado en segunda línea en combinación con FOLFIRI, pudiendo determinarse en esta población el estado mutacional de *KRAS* donde se han mostrado un beneficio en la tasa de repuestas en los pacientes con *KRAS* WT (35 vs 10%) y supervivencia libre de progresión (5,9 vs 3,9 meses), aunque el estudio no consiguió demostrar un incremento de supervivencia global (14,5 vs 12,5 meses) (Peeters M *et al.*, 2010).

En el estudio Prime se ha podido evaluar el papel del panitumumab en primera línea en combinación con FOLFOX, evaluándose el estado mutacional de *KRAS* de la población, mostrando en la población con *KRAS* WT un aumento significativo de supervivencia libre de progresión (9,6 vs 8 meses), no mostrando un aumento significativo en la supervivencia total (23,9 vs 19,7 meses) (Douillard JY *et al.*, 2010; Peeters M *et al.*, 2014).

De estos estudios se ha realizado un meta análisis, cuyos resultados muestran una reducción del riesgo de progresión del 34%, un aumento de la supervivencia global y de tasa de respuestas en pacientes con CCRM con *KRAS* WT tratados en primera línea con quimioterapia más cetuximab (Van Cutsem E *et al.*, 2010). En base a estos estudios, la Agencia Europea de Medicamentos (EMA) ha concedido su aprobación para el empleo del cetuximab o panitumumab en el tratamiento de primera línea del CCRM, en pacientes con *KRAS* WT.

En cuanto a *NRAS*, las mutaciones se han determinado en los exones 2 y 3 y en menor medida en el exón 4; además en estrecha relación con 10% de los pacientes que tienen *KRAS* Exón 2 no mutado (Guillermo Borja H *et al.*, 2017).

Existe controversia sobre si algunas mutaciones presentes tanto en *KRAS* como en *NRAS* (por ejemplo, la G13D) podrían no conferir resistencia al tratamiento, con lo cual los pacientes portadores de esta mutación podrían ser candidatos al tratamiento anti-EGFR (Johnson B *et al.*, 2017).

El estudio LBA-07 ha demostrado que una combinación de anticuerpos anti-EGFR e inhibidores de BRAF son capaces de bloquear la transducción de la señal en sentido desde EGFR hacia proteína cinasa activada por mitógeno (MEK) en los casos con *BRAF* mutado (Kassouf E *et al.*, 2016)

Actualmente e independientemente del subtipo genómico de consenso, si el CCR se encuentra en estadio IV (tumor metastásico) se analiza el estado mutacional siguiendo el siguiente esquema (**Fig. 9**) (Kassouf E *et al.*, 2016):

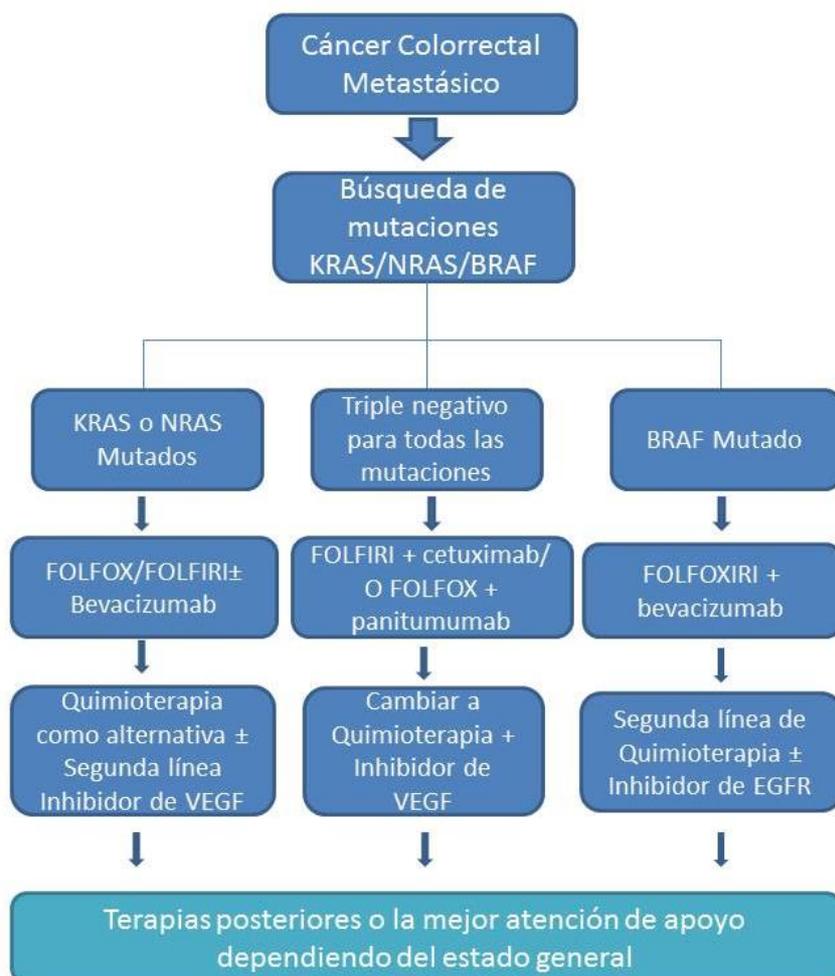


Figura 9: Esquema del enfoque inicial sugerido para los pacientes con CCRM.

6. Técnicas de detección de mutaciones en oncología clínica

La naturaleza del campo biológico relacionado con la genómica en oncología ha forzado un desarrollo tecnológico particular en cuanto a la detección de mutaciones. La conservación de las muestras, embebidas en parafina fijadas en formol (FFPE), la calidad de las muestras, junto a las pequeñas cantidades de muestra de partida de la que extraer el material genético a analizar, así como la necesidad de alcanzar sensibilidades elevadas – a diferencias de sensibilidades cercanas a un 50% de presencia del alelo a detectar, como es el caso de los análisis de mutaciones germinales – hizo preciso el desarrollo de técnicas de análisis que, o bien fuesen lo suficientemente sensibles para detectar pequeños porcentajes de presencia de mutación, o bien se pudiesen acoplar a las anteriores para enriquecer el alelo mutado en el ADN a analizar (Milbury CA *et al.*, 2009).

En este segundo caso, básicamente, se han utilizado dos técnicas: PNA-LNA PCR clamp (Skronski M *et al.*, 2009) y COLD PCR (Li J *et al.*, 2009; Yeo CD *et al.*, 2013). La primera no se ha desarrollado con éxito ningún kit comercial puesto que se necesitan sondas específicas de mutación, la segunda en realidad es una variación de las condiciones de la reacción de PCR que necesita elevados conocimientos técnicos por requerir un proceso complejo de optimización, con lo que ambas implementaciones han quedado reservadas al campo de la investigación sin saltar a la rutina analítica clínica.

Podemos dividir las técnicas más utilizadas en dos tipos:

- Técnicas de genotipado: aquellas técnicas de análisis genético en las que la detección se limita a posiciones exactas, de manera que se interrogan bases concretas situadas a lo largo del genoma.
- Técnicas de secuenciación: conjunto de métodos y técnicas bioquímicas cuya finalidad es la determinación del orden de los nucleótidos (A, C, G y T) en un fragmento de ADN estudiado.

6.1 Técnicas de genotipado

La mejora tecnológica en este tipo de técnicas, permitió alcanzar sensibilidades suficientes como para ser aceptables en los análisis oncológicos, así como alcanzar cada vez mayores números de bases (de mutaciones de interés) interrogadas. En general, las limitaciones han sido notables, ya que aun alcanzando en algún caso sensibilidades elevadas, el número de posiciones interrogadas nunca ha llegado a alcanzar valores competitivos si se comparan con las nuevas técnicas basadas en secuenciación masiva.

Entre las técnicas de genotipado más utilizadas en el campo que nos ocupa, tenemos: Amplificación de sonda dependiente de ligación por multiplex (MLPA), PCR cuantitativa a tiempo real (*Real time QPCR*), análisis de curvas de desnaturalización de alta resolución (*High Resolution Melting*), minisequenciación y secuenciación por MALDI-TOF.

6.1.1 Amplificación de sonda dependiente de ligación por multiplex (MLPA)

La técnica por MLPA fue desarrollada para la aplicación en el campo de la microbiología en un centro de Investigación de Holanda en 2002 y ahora se aplica para el análisis de grandes reordenamientos dentro de un gen, deleciones, duplicaciones, así como para la detección de mutaciones puntuales.

Básicamente, la técnica aprovecha la hibridación de dos sondas adyacentes a la posición a interrogar, de manera que solo en el caso de presencia de la base a detectar habrá hibridación de una de las dos sondas. Tras una ligación de ambas sondas, se crea una molécula artificial que es amplificada mediante PCR fluorescente (**Fig. 10**). El diseño de estas sondas permite un análisis de los diferentes productos amplificados en un equipo estándar de análisis de ácidos nucleicos por electroforesis capilar acoplada a fluorescencia (Łaczmańska I *et al.*, 2009)

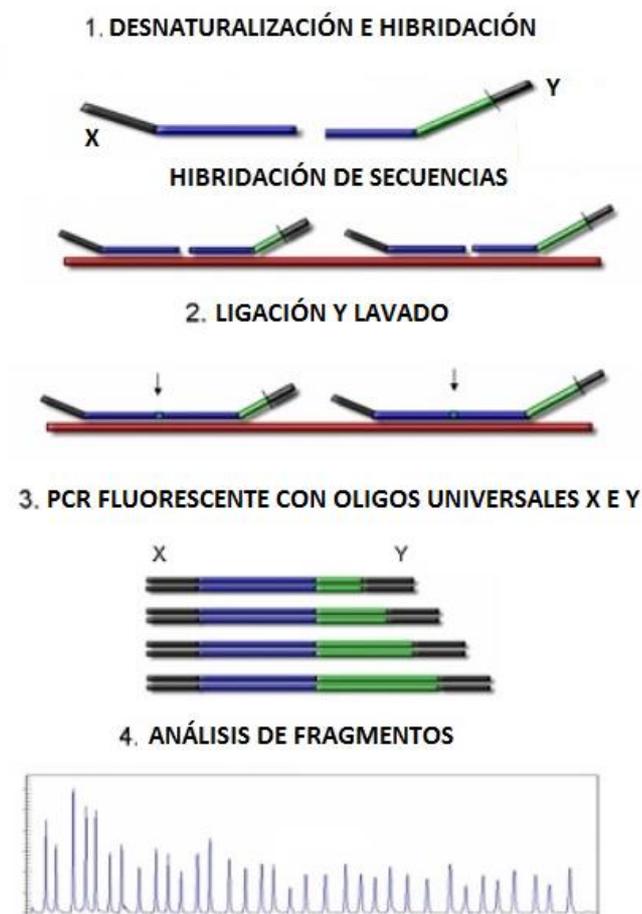


Figura 10: Esquema del proceso de genotipado por MLPA.

La técnica MLPA posee muchas aplicaciones, incluyendo la detección de mutaciones puntuales, inserciones o deleciones de unas pocas bases a exones completos, el análisis de perfiles de metilación, la caracterización cromosómica de líneas celulares y muestras de tejido, la detección de variaciones en el número de copias de genoma, la detección de duplicaciones y deleciones en genes relacionados con la predisposición al cáncer humano (Łaczmańska I *et al.*, 2009).

Un ejemplo de un kit comercial es la *Salsa MLPA probemix P520-A2 MPN mix2* de MRC-Holland MLPA, capaz de detectar la mutación puntual V617F y las deleciones N542_E543del y E543_D544del en el gen *JAK2* y otras mutaciones puntuales en los genes *MPL* y *KARL* en Neoplasias Mieloproliferativas (MRC-Holland MLPA, 2017).

6.1.2 PCR cuantitativa a tiempo real (Real time QPCR).

Esta técnica, desarrollada originalmente para la cuantificación de ADN mediante PCR, se acopló a un método de amplificación específica de alelo (ASO PCR), utilizando cebadores para la reacción de amplificación específicos para la posición que se interroga. Para cada posición se utilizan al menos dos pares de cebadores: los que hibridan completamente con la secuencia que contiene la mutación, y los que hibridan con la secuencia que contiene la forma no mutada (**Fig. 11**). La comparación de los valores de cuantificación relativa entre los dos pares de cebadores para cada individuo son indicadores de la presencia o no de la mutación problema.

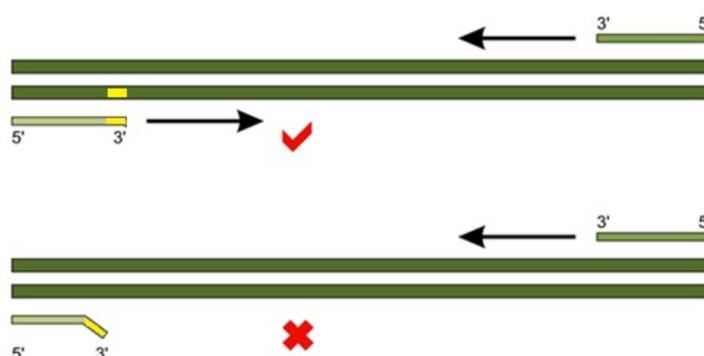


Figura 11: Esquema del proceso de una QPCR.

La realización de una curva patrón con cantidades conocidas de mutación permite también una aproximación al porcentaje presente en cada muestra. Estas técnicas tienen como principal ventaja su elevada sensibilidad, alcanzan hasta un 1% de presencia del alelo mutante con fiabilidad, y la no necesidad de un equipamiento especialmente costoso. El desarrollo comercial de este tipo de detección se ha centrado en la inclusión de un creciente número de posiciones interrogadas, siendo ésta su principal limitación (Morlan J *et al.*, 2009).

Un sistema mejorado de esta técnica sería el kit *cobas® KRAS Mutation Test* de Roche, que con el analizador *cobas z 480* puede medir, en tiempo real, el nivel de fluorescencia que generan productos de PCR específicos.

6.1.3 Análisis de curvas de desnaturalización de alta resolución (High Resolution Melting).

Esta técnica se desarrolló también a partir de técnicas de PCR cuantitativa a tiempo real. En este caso, se analiza la variación entre curvas de disociación con la temperatura en productos amplificados, con fluoróforos de intercalación al terminar estas reacciones de amplificación. En función de la presencia de una variación puntual, estas curvas son diferentes entre los distintos genotipos que se producen (**Fig. 12**). La puesta en el mercado de fluoróforos no saturantes, junto a la mejora en la óptica y análisis de equipos de detección, permitió que estas curvas, imposibles de ser detectadas individualmente por la saturación de la señal, pudiesen ser discriminadas (Lin J *et al.*, 2014).

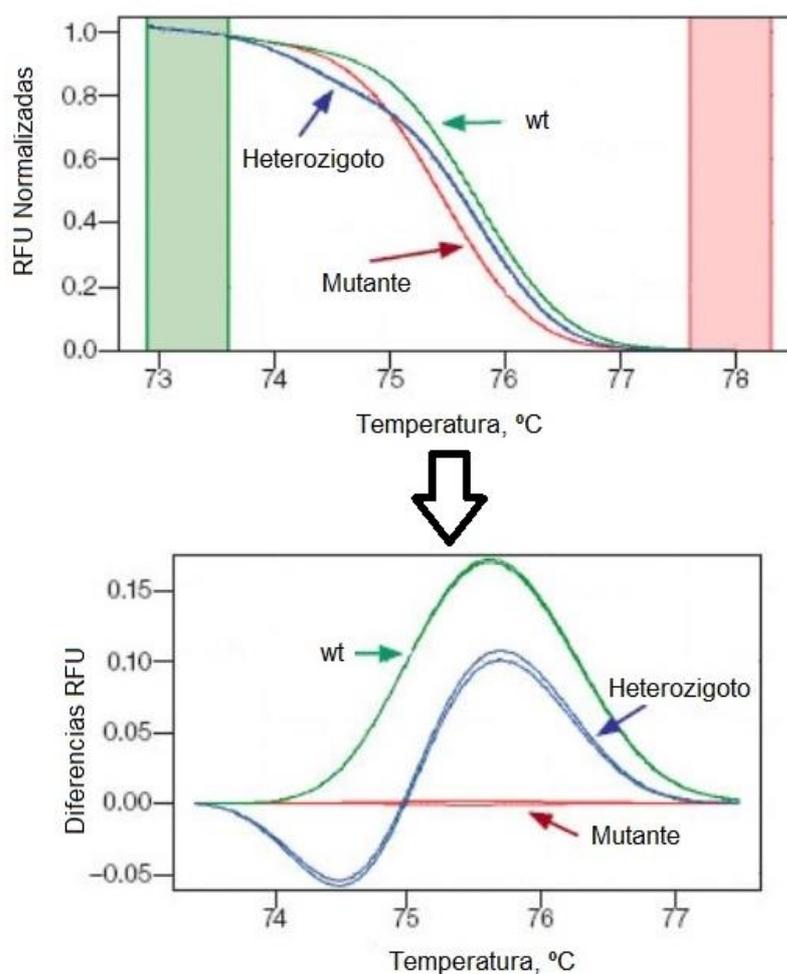


Figura 12: Análisis de curvas de desnaturalización de alta resolución (*High Resolution Melting*), WT: tipo salvaje, no mutado.

Por utilizar el mismo equipamiento (termocicladores cuantitativos), este método comparte las ventajas descritas en la qPCR: Alta sensibilidad y no precisar de equipos caros, pero en este caso, al igual que el anterior, la falta de escalabilidad, junto con problemas de interpretación en aquellas curvas correspondientes a fragmentos donde puede presentarse más de una variación. Éstas producen perfiles de curvas muy difíciles de resolver, suponen un evidente problema cuando se integran en la rutina clínica (Ross JS *et al.*, 2011).

6.1.4 Minisequenciación.

La minisequenciación se basa en la incorporación de dideoxi-dNTPs (ddNTPs): nucleótidos que carecen de un grupo 3'-hidroxilo (-OH) en la desoxirribosa. La falta de este grupo hidroxilo implica la imposibilidad de formar un enlace fosfodiéster con otros nucleótidos, el cual se produce entre el grupo 5'-fosfato de uno y el grupo 3'-hidroxilo de otro. Por tanto, durante la replicación de una molécula de ADN, la adición de un ddNTP a la cadena de nueva síntesis supone que el proceso se detenga. Estas moléculas llevan acoplados marcadores fluorescentes, de modo que cada cebador hibridado sólo extiende la base inmediatamente posterior (**Fig. 13**). La combinación de estas técnicas con una amplificación inicial por PCR y posterior carga en un equipo clásico de electroforesis fluorescente capilar, permite un análisis multiplexado, con varias posiciones interrogadas en el mismo pocillo. A diferencia de las anteriores, que requieren la separación física de cada variación estudiada. Esto hace que la minisequenciación sea mejor en este aspecto a las técnicas anteriores. Sin embargo, esta técnica no permite una aproximación cuantitativa para establecer un valor de proporción de la mutación. Siendo una técnica relativamente sensible (entre un 5 y 10% de presencia de alelos mutantes) su aplicación comercial hacia la clínica oncológica ha sido bastante limitada (Łaczmański Ł *et al.*, 2013).

Un ejemplo de un kit comercial en el cual se aplica esta técnica combinada con otras es el kit para BRAF *Mutation Analysis Reagents* de Applied Biosystems.

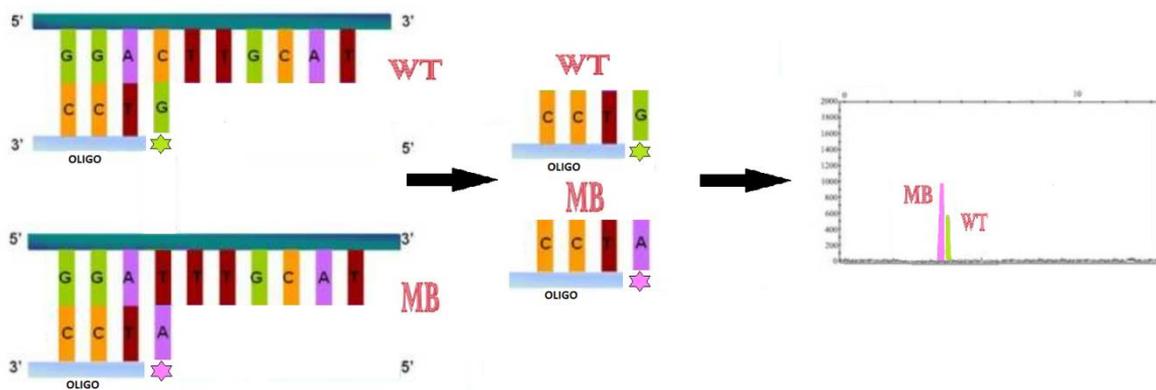


Figura 13: Esquema de una minisequenciación y posterior visión en un equipo de electroforesis fluorescente capilar. WT: tipo salvaje, MB: tipo mutante.

6.1.5 Secuenciación por MALDI-TOF

MALDI-TOF (*Matrix-Assisted Laser Desorption/Ionization Time-Of-Flight*) es una técnica de ionización utilizada en espectrometría de masas. Para llevar a cabo el análisis, sobre una placa metálica conductora se mezcla la muestra con una matriz orgánica para lograr una cocrystalización muestra-matriz. En alto vacío, es tratada con pulsos de luz láser provocando que la matriz absorba esta energía y la convierta en energía de excitación y transferencia de iones (**Fig. 14**).

En el campo del análisis de ácidos nucleicos, esta técnica se ha utilizado sobre todo en la búsqueda de polimorfismos de un único nucleótido (SNPs) y pequeñas inserciones o deleciones (Gut IG *et al.*, 2004). Esto ha permitido su introducción en el análisis de muestras oncológicas (Sherwood JL *et al.*, 2014). Tiene como principal inconveniente que produce secuencias muy cortas, tanto que en la práctica, se puede considerar más como un sistema de detección para genotipado que para secuenciación propiamente dicha. Su principal desventaja es su baja escalabilidad, junto con la necesidad de un equipo específico y costoso (Ross JS *et al.*, 2011).

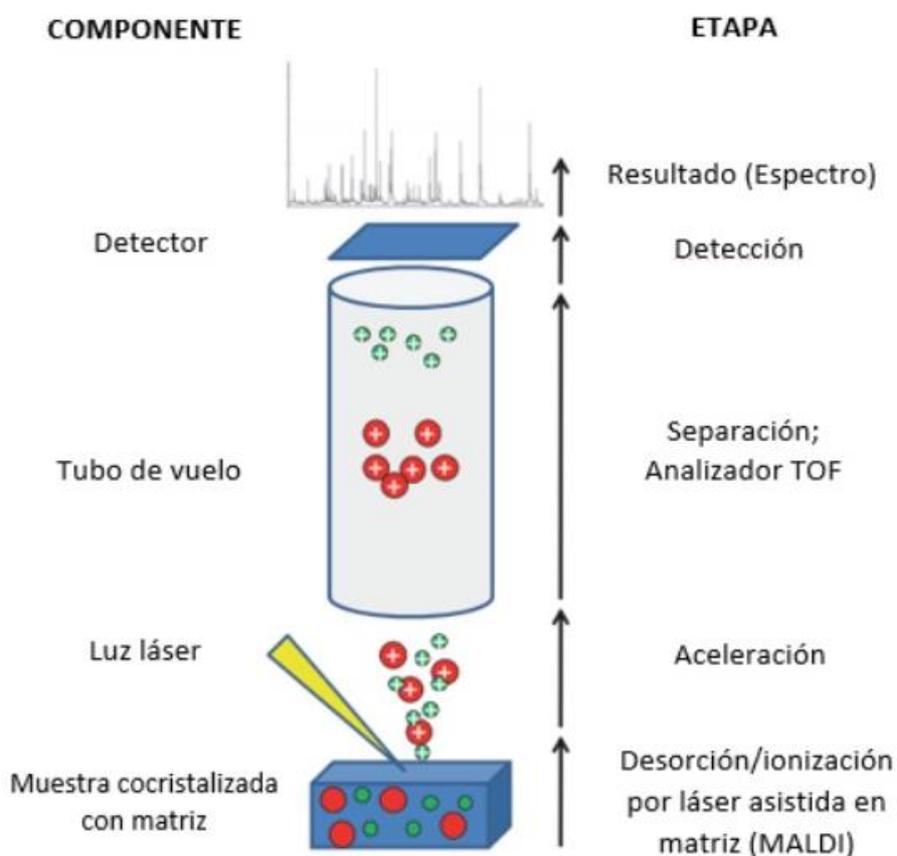


Figura 14: Esquema de los componentes y etapas para la secuenciación por la técnica de ionización en espectrometría de masas MALDI-TOF.

En resumen, estas técnicas de genotipado tienen como principales ventajas la alta sensibilidad, no precisar equipos caros, ni de un manejo muy especializado (menos MALDI-TOF). Por otro lado, a pesar de ser las técnicas más utilizadas en diagnóstico de biomarcadores, es limitante su poca capacidad de análisis en paralelo, su elevado consumo en términos de coste económico por mutación, de tiempo y de cantidad de muestra, así como una excesiva subjetividad en la interpretación de resultados.

El creciente número de mutaciones descritas con un papel activo en el tratamiento de los diferentes tipos de cáncer, con el consecuente número de mutaciones a detectar cada vez más elevado, ha ido haciendo que se vayan mejorando estas técnicas de genotipado en la rutina clínica y la implantación de técnicas de secuenciación.

6.2 Técnicas de secuenciación

Con las técnicas de secuenciación del ADN somos capaces de determinar el orden de los nucleótidos en un fragmento de interés. Esto implica, a diferencia de las técnicas descritas anteriormente, que lo que se obtiene es una secuencia más o menos larga, donde se pueden detectar directamente todas las potenciales variaciones puntuales (cambios de base y pequeñas inserciones o deleciones) presentes en la zona analizada. Inicialmente, en oncología clínica, estas técnicas se usaron para buscar variaciones en la línea germinal relacionadas con riesgo, o en farmacogenómica para estratificar poblaciones en función de su respuesta frente a tratamientos basándose en sus genotipos, especialmente en los genes relacionados con la metabolización de fármacos.

Una división histórica (y en realidad también conceptual) de este tipo de tecnologías analíticas, las separaría en dos grandes grupos: técnicas de secuenciación clásicas y técnicas de secuenciación de nueva generación (NGS, *Next Generation Sequencing*) (Liu L *et al.*, 2012; Morey M *et al.*, 2013)

6.2.1 Técnicas de secuenciación clásicas

Su aplicación al estudio de los genomas de biopsias en el ambiente clínico, aun siendo evidente, no ha podido ser generalizada dada la baja sensibilidad de estas técnicas. Sin embargo, éstas han sido las universalmente aceptadas para la detección de nuevas mutaciones, especialmente la secuenciación por el método de Sanger. Esta técnica se ha empleado para detectar mutaciones presentes en una alta proporción de las copias de ADN presentes en una muestra, como sucede en líneas celulares tumorales, clones tumorales o muestras con elevados porcentajes de células tumorales. Estas técnicas de secuenciación clásicas han sido de vital importancia para el desarrollo de las nuevas técnicas de secuenciación masiva en paralelo.

6.2.1.1 *Secuenciación de Maxam y Gilbert*

Este método fue desarrollado a principio de la década de los años 70 del siglo pasado, por Walter Gilbert y Allan Maxam (Maxam AM, Gilbert W, 1977). Consiste, en una fragmentación selectiva de la región a secuenciar gracias a una modificación química previa. Los agentes utilizados (DMSO, ácido fórmico e hidracina) rompían el ADN en cada caso sobre una base (o secuencia de dos bases) distinta. Los fragmentos eran después marcados radiactivamente en uno de sus extremos y el resultado, que dependía del orden de las bases en el ADN, se analizaba en un gel de poliacrilamida, obteniendo entonces la secuencia de tal molécula.

Este sistema fue desplazado por el método de Sanger debido a sus numerosos inconvenientes: Técnicamente era un proceso muy laborioso que además precisa el uso de agentes químicos altamente tóxicos, era muy difícil de escalar, y producía unas secuencias muy cortas (un máximo de 100 pares de bases).

6.2.1.2 *Secuenciación de Sanger*

Desarrollada por Frederick Sanger (Sanger F *et al.*, 1977), este método es también conocido como secuenciación enzimática o método de secuenciación por terminación de cadena, al utilizar ADN polimerasa y ddNTPs como terminadores de extensión para copiar la cadena a secuenciar. Originalmente, este método consistía en cuatro reacciones separadas, donde en cada caso se utiliza un ddNTP que al incorporarse a las nuevas cadenas impedía que continuase la extensión. El marcaje inicial era radiactivo, y las reacciones eran cargadas en geles de acrilamida donde la posición de cada fragmento marcado frente a los demás en las otras carreras (una carrera por cada base) permitía construir la secuencia copiada.

Gracias a su menor laboriosidad, peligrosidad (no utilizaba materiales tóxicos y la radiactividad necesaria era menor que en el método de Maxam) y capacidad de lectura de secuencias más largas, fue ésta la técnica elegida en adelante para la secuenciación de ácidos nucleicos. De hecho, fue la optimización de esta técnica la que permitió la secuenciación del genoma humano completo. Esta optimización consistió en (Smith LM *et al.*, 1986):

i) Introducción del uso de marcadores fluorescentes diferentes para cada ddNTPs de modo que, además de evitar el uso de materiales radiactivos, el método era capaz de realizar los cuatro marcajes en una sola reacción. Esto hizo que se desarrollasen sistemas de excitación por LASER y de detección de fluorescencia acoplados.

ii) Se pasó al uso de geles de policacrilamida (y posteriormente polímeros mucho más resolutivos) en la carga de los resultados en electroforesis, lo que permitían una resolución fiable, además de acortar notablemente los tiempos de carrera de la electroforesis.

En 1987, la compañía AppliedBiosystems introdujo la primera máquina de secuenciación automática, donde el paso fundamental de implementación fue la adopción de electroforesis capilar, que hizo que la secuenciación fuese mucho más rápida y precisa, al poder detectar hasta 96 muestras individuales de una vez en un periodo de carrera de poco menos de una hora, alcanzando las 500.000 bases en un día, con una longitud de lectura que podría llegar a 600 bases. Diez años más tarde, el modelo equivalente de la misma compañía, el ABI3730XL ya era capaz de registrar 900 bases de lectura, con un rendimiento de 2.88 Mb al día. Fueron estos equipos, o equipos similares (equipos de 384 capilares), los que permitieron la finalización del proyecto genoma humano con mucha más rapidez de lo inicialmente previsto (Collins FS et al., 2003).

Hasta hace unos pocos años, la miniaturización y multiplexación han dado lugar a modelos donde todo el proceso, reacción y análisis, se da en un dispositivo de unos 10 cm de diámetro, aunque este tipo de tecnología no acaba de introducirse comercialmente dada la enorme competencia que supone la tecnología NGS. En cualquier caso, para la detección de variaciones en la línea germinal, la secuenciación por el método de Sanger se sigue considerando como método de referencia o *gold standard*. Es importante destacar que es una técnica con muy baja sensibilidad para la detección de mutaciones en muestras tumorales (nunca por debajo de un 15% con fiabilidad) (Chen YJ et al., 2010).

6.2.1.3 Pirosecuenciación

La Pirosecuenciación consiste en una amplificación también enzimática, pero con dos principales diferencias: La primera, es que no hay terminación de la cadena; y la segunda, en cada paso de elongación se le acoplan dos reacciones: en la primera se forma ATP, en la segunda, éste es el sustrato de la siguiente reacción enzimática, produciendo la liberación de fotones de luz visible, que son detectados con un dispositivo acoplado de alta sensibilidad (CCD) (Ahmadian A *et al.*, 2006).

Comparada con los otros métodos de secuenciación clásica, la Pirosecuenciación es notablemente más sensible, aunque su longitud de lectura sea más corta. Esta técnica se utiliza con bastante frecuencia en entornos clínicos relacionados con los análisis oncológicos, ya que en el balance entre sensibilidad y análisis de zonas concretas del genoma, algo a tener en cuenta recurrentemente en este tipo de análisis, el primer aspecto compensa con mucho la poca longitud de lecturas (Olson MT *et al.*, 2014; de Macêdo MP *et al.*, 2015).

6.2.2 Técnicas de secuenciación de nueva generación o next-generation sequencing (NGS)

La secuenciación completa del genoma humano, en el momento de su publicación en el año 2001, requirió 12 años de trabajo a cargo de un enorme consorcio internacional formado por un gran número de laboratorios en todo el mundo y que costó más de 3.000 millones de dólares. Una vez terminado, y desde un punto de vista tecnológico, el escenario era la constatación de que la secuenciación por el método de Sanger había llegado a su máxima capacidad dada su naturaleza *singleplex*, que lo hacía un método difícilmente escalable. Además, la técnica que había permitido este hito histórico en la ciencia, no era, sin embargo, capaz de responder a necesidades fundamentales en campos como el de la oncología, al ser incapaz de detectar mutaciones en bajas proporciones. Es en este momento cuando comienzan a desarrollarse - de la mano de los cada vez mayores y más rápidos avances en las técnicas de captura de imagen (*scanners*), en nuevos sistemas de almacenamiento y procesamiento de datos informáticos, y sobre todo

en un nuevo concepto en cuanto a la naturaleza de las reacciones, que ahora pasan a poder desarrollarse sobre soportes sólidos - nuevas técnicas de secuenciación masiva en paralelo. Este desarrollo produjo un gran descenso en los costes de secuenciación por base, así como un aumento de varios órdenes de magnitud en la capacidad de secuenciación por unidad de tiempo, iniciando una nueva revolución en este campo del conocimiento.

Conceptualmente, la gran diferencia con las técnicas de secuenciación clásica es que ahora, en las químicas que utilizan estas nuevas técnicas se produce una lectura en paralelo de moléculas amplificadas de un modo clonal e identificadas con algún tipo de etiqueta molecular. De este modo lo que se obtiene es una representación masiva de multitud de eventos de amplificación, que habitualmente seguirán una distribución de Poisson, y donde aquellas moléculas con una variación en pequeñas proporciones frente al total de moléculas podrán ser leídas de la misma manera (y con la misma intensidad de señal) que las mayoritariamente representadas (Mardis ER, 2008 y 2008b).

La elevada demanda de secuenciación de bajo coste ha dado lugar a las distintas tecnologías de secuenciación de alto rendimiento (Church GM *et al.*, 2006; Hall N *et al.*, 2007) que son capaces de paralelizar muchas operaciones de secuenciación, produciendo miles o millones de lecturas a la vez, reduciendo los costes gracias a ello. Son las llamadas también secuenciaciones de nueva generación o "*next-generation sequencing*" o secuenciación masiva. Estos esfuerzos han sido financiados por instituciones públicas y privadas, así como desarrolladas y comercializadas dentro de la empresa privada por las compañías de biotecnología.

En esta metodología, no es lo suficientemente sensible para la secuenciación de una sola molécula, la mayoría de los métodos utilizan un paso con clonación *in vitro* para generar muchas copias de cada molécula individual. Uno de los métodos es la PCR de emulsión, en la que se aíslan las moléculas individuales de ADN junto con microesferas recubiertas con cebadores en burbujas acuosas dentro de una fase oleosa. Posteriormente, una reacción de PCR recubre cada microesfera con copias clonales de las moléculas aisladas. La PCR de emulsión se usa en los métodos publicados por Margulies, comercializado por 454 Life Sciences, posteriormente

adquirido por Roche (Margulies M et al., 2005); Shendure, conocido como "secuenciación polony ", término formado por polimerasa "pol" y colonia "colony" (Shendure J et al., 2005); y la secuenciación SOLiD desarrollada por la empresa Agencourt y adquirida por Applied Biosystems (Applied Biosystems SOLiD® 4 System, 2012).

Una vez que las secuencias clonales de ADN se localizan físicamente en posiciones separadas de la superficie, se pueden utilizar varios métodos de secuenciación para determinar las secuencias de ADN de todas las localizaciones en paralelo. La "secuenciación por síntesis", como en la popular secuenciación electroforética con terminador marcado, usa el proceso de síntesis de ADN por ADN polimerasa para identificar las bases presentes en la molécula complementaria de ADN. Los métodos de terminador reversible, usados por Illumina y Helicos, utilizan versiones reversibles de terminadores marcados fluorescentemente, añadiendo un nucleótido cada vez, y detectando la fluorescencia correspondiente a esa posición y eliminando posteriormente el grupo de bloqueo para permitir la polimerización de otro nucleótido. La Pirosecuenciación, utilizada por 454, también usa la polimerización del ADN para añadir nucleótidos, añadiendo cíclicamente un tipo diferente para detectar y cuantificar el número de nucleótidos añadidos en una determinada localización a través de la luz emitida por la liberación de los pirofosfatos unidos a ellos (Ambardar S et al., 2016; Kruglyak KM et al., 2016).

6.2.2.1 Sistema 454 GS FLX (Roche®)

454 Life Sciences fue una empresa de biotecnología especializada en secuenciación de ADN de alto rendimiento, adquirida por Roche en 2007. El sistema *Genome Sequencer (GS) FLX* 454 utiliza un sistema de pirosecuenciación en paralelo a gran escala capaz de secuenciar aproximadamente 450-700 megabases (Mb) de ADN por cada 10-23 horas de funcionamiento, dependiendo del equipo (Goodwin S et al., 2016), con reactivos de la serie GS Titanium. A finales de 2009, 454 Life Sciences introdujo el sistema 454 GS Junior, una versión de banco de este sistema (Voelkerding KV et al., 2009).

El sistema se basa en la PCR en emulsión. El ADN unido a un oligo de estas microesferas se amplifica entonces por PCR. En este proceso, en cada una de esas microesferas se parte de una sola molécula de ADN unida a una bola de captura o *capture bead* a través de unos oligonucleótidos anclados en ellas y complementarios a los oligos de los extremos del ADN unido. La amplificación en cada microrreactor será, por lo tanto, a partir de una sola molécula monocatenaria de ADN.

Después de la amplificación, cada microesfera se coloca en una placa con pocillos de unos 30 μm de diámetro llamada *PicoTiterPlate*. El sistema introduce los reactivos de secuenciación a través de los pocillos de la placa. Los cuatro nucleótidos de ADN se añaden secuencialmente en un orden fijo a través del dispositivo *PicoTiterPlate* durante un ciclo de secuenciación. Durante el flujo de nucleótidos, se secuencian en paralelo miles de millones de copias de ADN unidas a cada una de las microesferas. Cuando se añade un nucleótido complementario a la cadena molde en un pocillo, la polimerasa extiende la cadena de ADN existente añadiendo un nucleótido. La adición de uno o más nucleótidos genera una señal luminosa que es grabada por la cámara en el CCD del instrumento.

Esta técnica se basa en secuenciación por pirosecuenciación. La intensidad de la señal es proporcional al número de nucleótidos unidos; por ejemplo, los homopolímeros, incorporados en un único flujo de nucleótidos generan una señal mayor que los nucleótidos individuales. Los datos son almacenados en archivos estándar de formato de diagrama de flujo (SFF) para el análisis posterior. (King C *et al.*, 2008; Voelkerding KV *et al.*, 2009)

6.2.2.2 *Illumina*

Otro método para la amplificación clonal *in vitro* es la "PCR de puente", en la que los fragmentos se amplifican a partir de los cebadores unidos a una superficie sólida, desarrollados y usados por Solexa (de la que ahora es propietaria la empresa Illumina).

En este método, se usan polimerasas modificadas por la empresa y nucleótidos fluorescentes con terminadores reversibles. Las hebras de ADN y los cebadores se unen covalentemente a una superficie plana, y se lleva a cabo una amplificación por

la polimerasa, de forma que se crean colonias locales de ADN o *clusters* de ADN (**Fig. 15**). Tras esto, se usan los cuatro tipos de ddNTPs, de forma que cuando una de ellas se une a la secuencia se para la reacción, y seguidamente se lavan las demás bases que no han reaccionado. Entonces, una cámara recoge la fluorescencia y determinará qué nucleótido es el que se ha unido en esa posición. Una vez que se ha identificado se eliminará el terminal de bloqueo del extremo 3' que impedía continuar con la síntesis de la cadena y el marcaje fluorescente. Se vuelve a cargar una nueva tanda de nucleótidos y se continúa la secuenciación hasta completar toda la cadena de ADN (Ambardar S *et al.*, 2016; Kruglyak KM *et al.*, 2016).

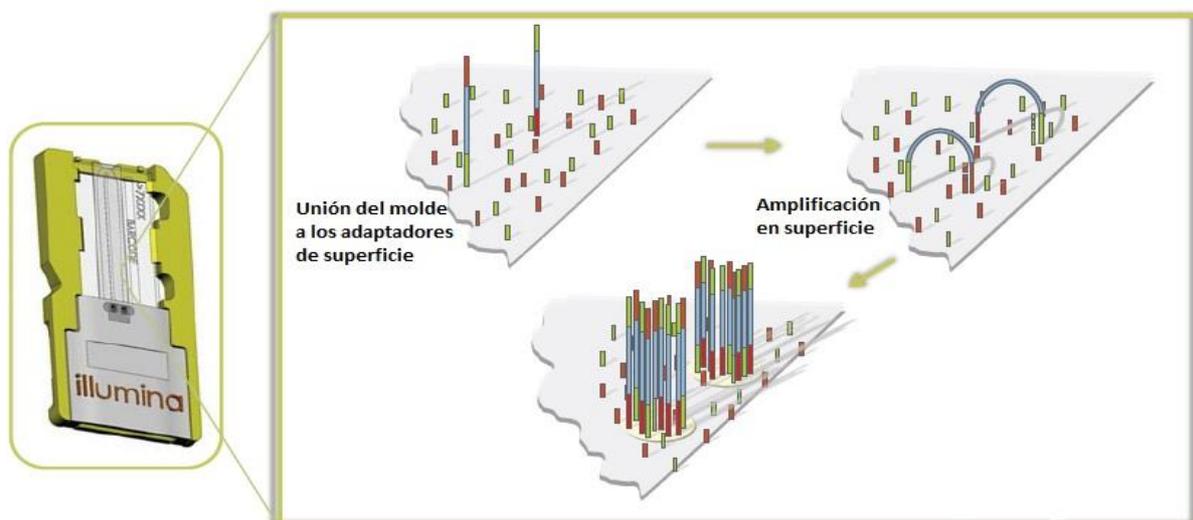


Figura 15: Esquema del proceso llevado a cabo en la celda de flujo, unión de los adaptadores, amplificación y creación de "*clusters de ADN*"

A diferencia de la pirosecuenciación, por cada ciclo se incorpora un único nucleótido, el tamaño de los fragmentos a secuenciar también es menor (150 pb). Hay equipos que hacen hasta 15.000 millones de reacciones en una carrera.

6.2.2.3 Ion Torrent semiconductor

Este método de secuenciación se basa en la detección de iones de hidrógeno que se generan durante la polimerización del ADN. Al igual que el sistema 454, hay una amplificación clonal basada en la PCR en emulsión, con un posterior paso de enriquecimiento de las microesferas unidas a una sola molécula monocatenaria de ADN (Goodwin S *et al.*, 2016).

Posee micropocillos en los que se inserta la cadena de ADN a secuenciar, y se inunda con un único tipo de nucleótido. La incorporación de un nucleótido en la polimerización de forma natural implica la formación de un enlace covalente, y la liberación de un pirofosfato y una carga positiva en forma de iones de hidrógeno. Aprovechando este hecho y la capacidad de los sensores de pH ISFET se ha desarrollado la tecnología del Ion Torrent semiconductor. Si el nucleótido añadido a la reacción no es complementario a la secuencia de ADN a secuenciar, no se incorporará y no se dará ninguna reacción. Si el nucleótido es complementario a la cadena de ADN, se incorporará, provocando la liberación de un ion hidrógeno, cuya señal será recogida por un sensor de tipo ISFET (**Fig. 16**). Tras cada ciclo, se lavan los nucleótidos para añadir un nuevo tipo, y la reacción se volverá a producir o no dependiendo de la secuencia los nucleótidos de la cadena a secuenciar. Al usar polimerasas naturales, la reacción ocurre a tiempo real. En caso de que haya dos o más bases iguales consecutivas, se incorporarán múltiples nucleótidos en un único ciclo, se liberarán más átomos de hidrógeno y la señal electrónica será proporcionalmente mayor. Las ventajas de este sistema es la tecnología de medición eléctrica, sin necesidad del uso de medición óptica mediante nucleótidos modificados, que permiten que el proceso sea más barato tanto en los costes iniciales como de operación, así como su alta velocidad gracias a las polimerizaciones a tiempo real (Perkel J *et al.*, 2011).

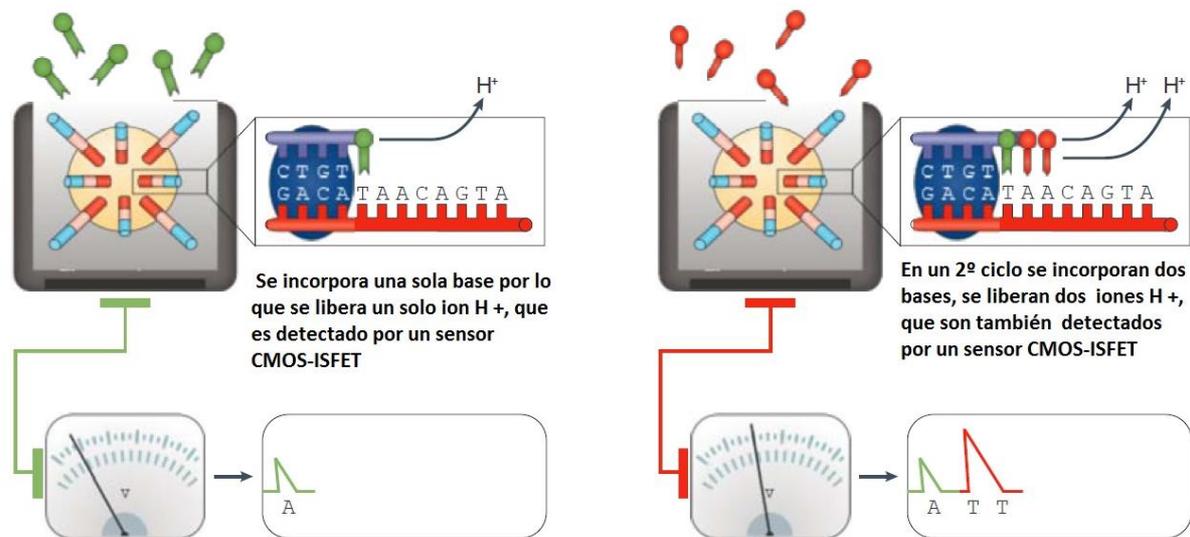


Figura 16: Esquema del proceso llevado durante la secuenciación en los micropocillos de un Ion Torrent. La incorporación de un nucleótido en la polimerización produce la liberación de un pirofosfato y una carga positiva en forma de iones de hidrógeno.

6.2.3 Secuenciación de molécula única de ADN

Las técnicas de Segunda Generación todavía son plataformas que precisan de unas necesidades específicas. Los protocolos propios de dichos métodos, incorporan una etapa de amplificación del ADN a secuenciar. Por ello, la idea de buscar una alternativa más sencilla, rápida y eficaz es una cuestión que siempre ha estado y está en la mente de los investigadores. Con este fin, nace lo que se conoce como secuenciación de tercera generación. Aunque algunas de estas plataformas, o bien se encuentran todavía en fase de experimentación, o requieren de la mejora en muchos de sus aspectos. Todas las estrategias se fundamentan en la idea de realizar la propia secuenciación a partir de una única molécula de ADN, consiguiendo lecturas más largas, y ampliando el abanico de aplicaciones, como la detección de grandes variaciones en la estructura de los cromosomas.

Existe en el mercado la tecnología *Single Molecule Real-time* (SMRT), desarrollada por la casa *Pacific Biosciences*, que trata de ofrecer una secuenciación de estas características. Es una plataforma que utiliza la fluorescencia para secuenciar una única molécula de ADN en tiempo real. El diseño

se basa en nanopocillos, en cuya base existe un orificio, denominado *zero-mode waveguide* (ZMW), por donde es posible el paso de la luz. En cada ZMW, queda inmovilizada una única molécula junto con una sola enzima de ADN polimerasa. Tras la incorporación de los nucleótidos, cada uno marcado con un fluoróforo de una fluorescencia distinta, por parte de la ADN polimerasa a la hebra en síntesis, se libera el fluoróforo que es excitado por el haz de luz ultravioleta, siendo detectable por el detector (**Fig. 17**). De esta forma, es posible saber que nucleótido se ha incorporado y, así, obtener la secuencia. SMRT permite incorporar lecturas de unos 60.000 pares de bases. Sin embargo se trata de una metodología que requiere de grandes cantidades de ADN para poder ofrecer un resultado sólido. Además, adquirir el equipamiento necesario supone una gran inversión inicial y el posterior mantenimiento resulta muy costoso (Eid J *et al.*, 2016; Goodwin S *et al.*, 2016).

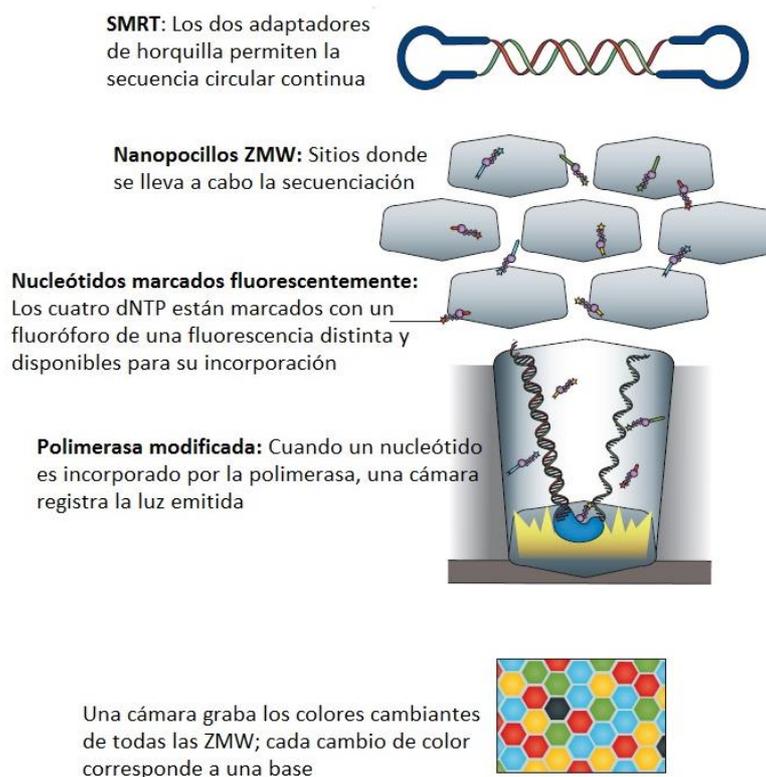


Figura 17: Esquema del proceso de secuenciación mediante la tecnología SMRT desarrollada por la casa *Pacific Biosciences*.

Recientemente ha surgido la técnica de secuenciación mediante Nanoporos, desarrollada por *Oxford Nanopore Technologies*, un método que permite secuenciar moléculas de ADN individualizadas, de forma directa, barata y en el momento. Aunque todavía se encuentra en vías de optimización, otra ventaja que supone frente al resto de tecnologías es que será capaz de secuenciar toda la longitud de la secuencia a analizar, en lugar de generar lecturas de un tamaño específico. Esto es lo que se comienza a llamar como capacidad de ofrecer lecturas “ultra largas”. Todo esto se consigue gracias a que esta tecnología está basada en el principio de leer directamente el fragmento de ADN a secuenciar mediante el paso de dicha molécula única por un nanoporo, midiendo los efectos, a nivel iónico y eléctrico, que supone el avance del ADN (**Fig. 18**). Al igual que en SMRT, la secuenciación mediante nanoporos también presenta la limitación de que se requiere de una gran cantidad de ADN, para poder secuenciar y obtener un resultado fiable. Al secuenciar una única molécula, el inconveniente es que se necesita de la secuenciación de muchas moléculas únicas para poder obtener una secuencia consenso aceptable (Schaffer A et al., 2012; Lu H et al., 2016; Goodwin S et al., 2016).

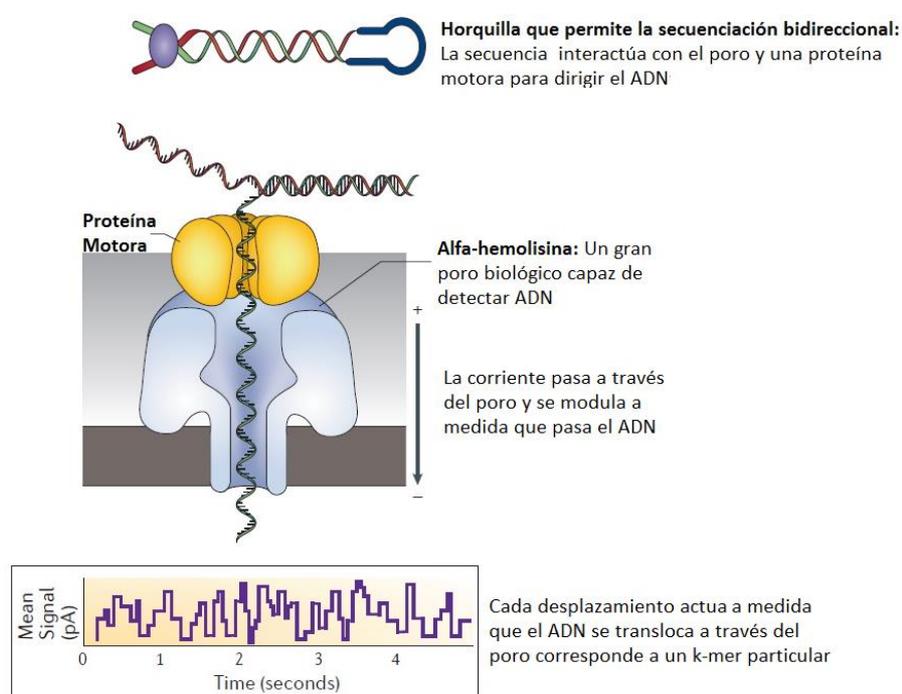


Figura 18: Esquema del proceso de secuenciación mediante Nanoporos, desarrollada por *Oxford Nanopore Technologies*

7. Validación

La validación de un método es un paso fundamental para asegurar que los resultados entregados son fiables. Cuando se realiza la validación de un método por parte del laboratorio, lo que se busca es poder determinar con fundamento estadístico que el método es adecuado para los fines previstos.

En los Estados Unidos, los reglamentos de las Enmiendas de Mejora de Laboratorio Clínico (CLIA) exigen que los laboratorios verifiquen las pruebas desarrolladas en éstos, por lo que deben realizar una validación para establecer y documentar las especificaciones de dichas pruebas para las siguientes características de rendimiento: exactitud, precisión, sensibilidad y especificidad analítica y otras características según corresponda (Gargis A *et al.*, 2015).

En Europa, la norma ISO 15189 (ISO/IEC 15189, 2012) define las instrucciones para la validación, que incluye estimaciones de las características de rendimiento analítico y diagnóstico de una prueba: rendimientos de la técnica, exactitud y precisión del método de análisis, sensibilidad y especificidad analítica y la concordancia con otros métodos.

Desde 2014, la mayoría de centros que realizan pruebas de NGS continúan diseñando, optimizando y validando sus pruebas desarrolladas en laboratorio. Esto coloca la responsabilidad de establecer y documentar las especificaciones de la validación de las pruebas de NGS en los profesionales de laboratorio. La plataforma de secuenciación, que se instala con configuración predeterminada y software, es optimizada por el laboratorio para la aplicación clínica prevista antes de la validación de la prueba. El protocolo de validación del ensayo para pruebas de NGS se realiza como un único flujo de trabajo, donde se evalúa tanto la plataforma, como la prueba y el soporte informático (Gargis A *et al.*, 2015).

Los paneles comerciales, a pesar estar validados por los propios fabricantes, deben ser vueltos a validar por los usuarios para ser introducidos en la rutina del laboratorio. Para ello, lo más habitual es la utilización de ADN de referencia, comerciales o validados por otras técnicas, con frecuencia alélica conocida para las distintas mutaciones de interés. Estos ADN son obtenidos a partir de líneas celulares tumorales o biopsias tumorales procesadas en parafina, con el objetivo

que sean lo más similares a las muestras de rutina. A partir de los resultados obtenidos de estas muestras, es posible determinar los parámetros para la validación: límite de detección, sensibilidad y especificidad analítica, precisión y exactitud (Singh RR *et al.*, 2013; Lin MT *et al.*, 2014; D'Haene N *et al.*, 2015; Ananda G *et al.*, 2014; Simen BB *et al.*, 2015).

Uno de los aspectos generales que necesita una técnica para su validación es determinar su capacidad de generar rendimientos similares en diferentes estudios o experimentos a lo largo del tiempo. Esta estabilidad, que en el contexto de secuenciación masiva se referiría a la cobertura y a la calidad de las secuenciaciones obtenidas, ya sea por carrera, por muestra o por amplicón, facilita alcanzar los parámetros de exactitud, precisión y sensibilidad analítica mínimos para ser validada (Ten Bosch JR *et al.*, 2008).

También hay que tener en cuenta que, a bajas coberturas, se incrementan los niveles de ruido del ensayo, en tanto que aumenta la probabilidad de perder sensibilidad analítica; dependiendo de la cobertura se van a ver afectados los parámetros de la validación (Simen BB *et al.*, 2015).

A partir de la replicación del análisis de muestras tumorales se obtienen los parámetros de precisión (*intra-run* o repetibilidad, e *inter-run* o reproducibilidad). Finalmente, la comparación de los resultados obtenidos con otras técnicas permite obtener la concordancia, que sirve para conocer si con un método o instrumento nuevo se obtienen resultados equivalentes. Su importancia en el área de la salud reside en que existen diversas maneras de valorar los distintos parámetros y, por lo tanto, aparecen distintas aproximaciones o métodos diagnósticos usados para medir lo mismo. Por lo tanto, la concordancia adquiere importancia cuando se desea conocer si con un método o instrumento nuevo, diferente al habitual, se obtienen resultados equivalentes de tal manera que eventualmente uno y otro puedan ser remplazados o intercambiados ya sea porque uno de ellos es más sencillo, menos costoso, y por lo tanto, más costo-efectivo, o porque uno de ellos resulta más seguro para el paciente, entre otras múltiples razones. En términos generales, la concordancia es el grado en que dos o más observadores, métodos, técnicas u

observaciones están de acuerdo sobre el mismo fenómeno observado (D'Haene N *et al.*, 2015; Ananda G *et al.*, 2014; Simen BB *et al.*, 2015).

Así, la concordancia no evalúa la validez o la certeza sobre una u otra observación con relación a un estándar de referencia dado, sino cuán acordes están entre sí observaciones de distinto origen sobre el mismo fenómeno. En estos casos se considera que los estudios evalúan la consistencia entre los métodos o instrumentos. En los estudios en los que uno de los métodos o instrumentos nuevos se compara frente al método que constituye el patrón de referencia o *gold standard*, se evalúa la conformidad del método respecto al patrón de referencia que también se denomina validez o desempeño operativo de una prueba diagnóstica (Cortés-Reyes E *et al.*, 2010).

Para el análisis de concordancia de las nuevas técnicas de NGS, es habitual que éstas sean comparadas con técnicas ya introducidas en la rutina clínica, como son las técnicas de genotipado, de minisequenciación y de secuenciación con la técnica de Sanger; además de la comparación entre las distintas técnicas NGS (D'Haene N *et al.*, 2015; Ananda G *et al.*, 2014; Simen BB *et al.*, 2015).

HIPÓTESIS Y OBJETIVOS

El desarrollo de diferentes fármacos anticancerígenos cuya eficacia aumenta o disminuye considerablemente en función de la presencia de determinadas mutaciones somáticas hace imprescindible la detección de estas mutaciones.

Nuestra hipótesis es que, en el contexto de los nuevos tratamientos oncológicos, en el CCRM se hace necesaria la implantación de sistemas de NGS para determinar la existencia de alteraciones génicas asociadas a la decisión terapéutica en el uso de estos fármacos.

Por tanto, el presente trabajo tiene como objetivos:

- Desarrollar un sistema de detección rápido, sensible, reproducible y fácilmente integrable en la rutina clínica para el análisis de mutaciones de interés oncológico, en pacientes con CCRM a partir de biopsias incluidas en parafina seleccionadas en diferentes Centros de Salud, mediante el uso de técnicas de secuenciación masiva en la plataforma de secuenciación 454 GS Junior de Roche.
- Validar dicho sistema de detección para determinar la idoneidad de nuestra técnica. Las pruebas de validación han de incluir estimaciones como son: los rendimientos de la técnica y calidad de los resultados, exactitud, precisión, sensibilidad y especificidad analítica y concordancia con otras técnicas.
- Mediante el uso de este panel, detectar las mutaciones de interés clínico descritas hasta el momento en los genes *BRAF*, *KRAS* y *NRAS*, por medio de la ampliación de las zonas de interés de estos genes, con unas condiciones tales que permitan esta detección incluso en muestras con mutaciones solo presentes a un porcentaje bajo del ADN obtenido.
- Que dicho sistema de detección de mutaciones sea modulable y escalable, y que además el proceso de diseño, desarrollo y validación sea extrapolable a otras plataformas de secuenciación.

METODOLOGÍA

1. Diseño del método de análisis

Para determinar las mutaciones de interés en la terapia del CCRM, se preparó un listado de mutaciones correspondientes a los genes *BRAF*, *KRAS* Y *NRAS* (Tabla 1). Para el diseño de los cebadores de PCR, así como para la preparación de la información necesaria para el análisis informático de los resultados, se buscó la secuencia de los transcritos y las posiciones de las alteraciones en las bases de datos públicas NCBI (www.ncbi.nlm.nih.gov/) con el genoma de referencia GRCh37 y Ensembl (www.ensembl.org/index.html). En el diseño de cebadores se utilizó el programa informático PRIMER3 (<http://bioinfo.ut.ee/primer3-0.4.0/>) (Koressaar T *et al.*, 2007; Untergrasser A *et al.*, 2012).

MUTACIONES DE INTERÉS ONCOLÓGICO*

GEN	EXON	MUTACIÓN	CÓDIGO EN COSMIC	GEN	EXON	MUTACIÓN	CÓDIGO EN COSMIC
BRAF	15	p.E586K	COSM463	KRAS	2	p.G12A	COSM522
		p.V590I	COSM51383			p.G12C	COSM516
		p.D594G	COSM467			p.G12D	COSM521
		p.D594V	COSM466			p.G12R	COSM518
		p.F595L	COSM53198			p.G12S	COSM517
		p.F595L	COSM468			p.G12V	COSM520
		p.G596R	COSM469			p.G13A	COSM533
		p.L597V	COSM470			p.G13C	COSM527
		p.T599I	COSM472			p.G13D	COSM532
		p.T599R	COSM4172020			p.G13R	COSM486
		p.V600E	COSM476			p.G13S	COSM528
		p.K601E	COSM478			p.G13V	COSM574
		p.K601Q	COSM27897			p.V14I	COSM12722
NRAS	2	p.G12A	COSM565	3	p.L19F	COSM20818	
		p.G12C	COSM562		p.L19F	COSM12703	
		p.G12D	COSM564		p.Q22K	COSM543	
		p.G12S	COSM563		p.Q61H	COSM554	
		p.G12V	COSM566		p.Q61H	COSM555	
		p.G13D	COSM573		p.Q61K	COSM549	
		p.G13R	COSM569		p.Q61L	COSM553	
		p.G13V	COSM574		p.Q61R	COSM552	
		p.T50I	COSM222543		p.T74P	COSM28672	
		p.G60E	COSM28673		p.K117N	COSM19940	
		p.Q61H	COSM586		p.K117N	COSM28519	
		p.Q61K	COSM580		p.A146P	COSM19905	
		p.Q61L	COSM583		p.A146T	COSM19404	
p.Q61R	COSM584	p.A146V	COSM19900				
4	p.A146V	COSM4170228					

Tabla 1: Mutaciones de interés oncológico implicadas en el Cáncer Colorrectal Metastásico. La referencia del transcrito para el gen BRAF es NM_004333.4, para el gen KRAS es NM_004985.4 y para el gen NRAS es NM_002524.4, respectivamente. * Bases de datos consultadas: *MY CANCER GENOME* (www.mycancergenome.org), *MD ANDERSON CANCER CENTER PERSONALIZED CANCER THERAPY* (pct.mdanderson.org) y *COSMIC* (www.cancer.sanger.ac.uk/cosmic).

- Por un lado, los cebadores debían cumplir las condiciones siguientes:
 - Tamaños entre 20 y 30 pb, siendo el tamaño óptimo en 25 pb.
 - Temperatura de unión de los cebadores (T_m) entre 58 y 62°C, siendo la temperatura óptima de 60°C y que la diferencia máxima de la T_m fuese de 2°C.
 - Número máximo de homopolímeros fuese inferior a 5.
 - Rangos del tamaño de los productos de PCR, sin adaptadores, debía estar entre 125 y 250 pares de bases.
- Por otro lado, estos diseños siguieron tres premisas:
 - Que el tamaño de los amplicones, es decir, los productos de PCR con los adaptadores universales 454, no fuera superior a las 400 pb para poder ser secuenciados en su totalidad.
 - Que estos fragmentos incluyan todas las mutaciones de interés.
 - Y que los tamaños de estos fragmentos de PCR fueran lo suficientemente diferentes en tamaño para poder ser visualizados mediante técnicas de análisis de imagen de productos de PCR, y así poder hacer un seguimiento de cada fragmento en la optimización de las condiciones de amplificación.

2. Obtención y preparación de las muestras

Este trabajo se ha realizado con diferentes tipos de muestras, entre ellas muestras biológicas de pacientes (biopsias de tumores colorrectales). Este estudio pasó por comité ético que permitía el análisis de los datos sin tener información de los pacientes, los cuales firmaron consentimiento informado. El estudio reúne las características adecuadas referentes a información a los pacientes, consentimiento informado y cumplimiento de los criterios éticos para la investigación médica y biomédica establecidos en la Declaración de Helsinki de la Asamblea Médica Mundial, y sus revisiones (General Assembly of the World Medical Association, 2014).

Las muestras analizadas a lo largo de todo el estudio fueron:

- **Muestras comerciales**, procedentes de líneas celulares, de perfil mutacional y frecuencia alélica conocidos (Horizon Diagnostics, Cambridge, United Kingdom) (n=8). Estas muestras comerciales se suministran en cortes de parafina, de forma que fueran lo más parecidas a las muestras clínicas. Éstas vienen identificadas como WT o, si son mutadas, debidamente identificadas con su mutación o mutaciones correspondientes, y la frecuencia alélica de cada una de ellas.
- **Biopsias de pacientes con CCRM**, remitidas desde diferentes Hospitales de la Comunidad Valenciana (años 2013-2016, n=875). Estas muestras están incluidas en parafina para su conservación. En el Servicio de Anatomía Patológica del Hospital Clínico de Valencia, se les practicaron varios cortes con un microtomo de precisión, obteniendo secciones de 5µm de espesor de las zonas con más porcentaje de células tumorales. Después, 5 secciones de cada muestra se introdujeron en tubos de eppendorf de 1,5 ml, debidamente identificados y etiquetados.
- **Muestras control**, del Programa de Garantía de Calidad de la Sociedad Española de Anatomía Patológica, SEAP, (Módulos de Patología Molecular 9-14, 2014-2016, n=40). Estas muestras son cortes de parafina adheridos a portaobjetos (entre 3 ó 4 por caso), que posteriormente son rascados y

depositados en tubos eppendorf de 1,5 ml, debidamente identificados y etiquetados.

Para la desparafinización y obtención del ADN se utilizó el kit de extracción Maxwell® 16 FFPE Tissue LEV DNA Purification Kit (Promega Corporation, Madison, USA), mediante un método de purificación automatizada de ADN genómico de secciones de tejidos embebidos en parafina fijado en formol (FFPE) (Fig. 19).

Tras la digestión de los cortes de las muestras en parafina con proteinasa K, este método purifica el ADN utilizando partículas paramagnéticas de sílice revestido, que proporcionan una fase sólida móvil que optimiza la captura, el lavado y la elución del ADN.



Figura 19: Sistema Maxwell® 16 para purificación automatizada de ADN

Una vez extraído el ADN de cada muestra, se procedió a su cuantificación utilizando el equipo de medición de absorbancia Nanodrop® ND-100 Spectrophotometer (ThermoScientific™) en el rango de 260-280 nm de longitud de onda del espectro. Finalmente, las muestras se almacenaron a -80°C hasta su uso. Para este trabajo se utilizaron concentraciones de ADN entre 2 y 200 ng/μL.

3. Metodología de la secuenciación

Para realizar la secuenciación masiva de los fragmentos amplificados de ADN en la búsqueda de mutaciones de interés oncológico, se seleccionó el sistema 454 GS Junior de Roche® (Fig. 20).



Figura 20: Sistema 454 GS Junior de Roche®

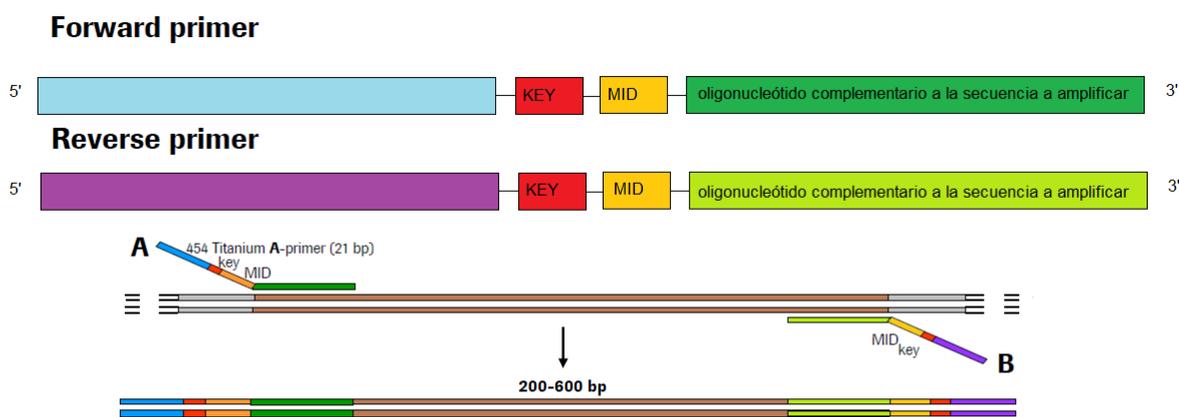
3.1 Descripción de los adaptadores universales 454.

El diseño de los oligonucleótidos en el sistema 454 requiere la adición de unos adaptadores específicos para reconocimiento de cada muestra en el proceso de la secuenciación masiva. Dentro de la secuencia del adaptador se incluyen (Fig. 21):

- El “MID”, una secuencia de 10 nucleótidos que es identificadora específica de cada individuo/muestra.
- La “KEY”, una agrupación de 4 bases (TCAG), que sirven de referencia para calibrar el detector de intensidad de luz, emitida en cada ciclo del aparato de secuenciación masiva.
- Los oligonucleótidos *Forward* o *Reverse*, secuencias de nucleótidos diferenciales, que son complementarias a la secuencia de los oligonucleótidos presentes en las microesferas A o B de emPCR, y que llevan todos los fragmentos en posición 5' ó 3', respectivamente.

De esta manera, cada fragmento quedará identificado por el MID en cada extremo, así ambos MID son secuenciados en todas las reacciones, ya sean en un sentido o en otro.

La adición de la secuencia del adaptador tras la reacción de PCR hace que se obtengan unos fragmentos con una extensión extra de 70 pb en los extremos (35 pb a cada lado).



Forward primer (Primer A-Key):

5'-CGTATCGCCTCCCTCGGCCATCAG-MID-template-specific-sequence-3'

Reverse primer (Primer B-Key):

5'-CTATGCGCCTTGCCAGCCCGCTCAG-MID-template-specific-sequence-3'

Figura 21: Proceso general de la obtención de una librería.

3.2 Obtención de los fragmentos por PCR

Para la obtención de todos los fragmentos esperados, con una proporción similar de todos ellos, se pusieron a punto unas condiciones de amplificación en cuanto a la preparación del mix de PCR con el Kit QIAGEN Multiplex PCR (Qiagen, Hilden, Germany) y programa de amplificación idóneos. Las concentraciones de cada par de cebadores, junto con el programa de amplificación idóneo conseguido tras el proceso de optimización, se describen en el apartado 1.1 de Resultados.

La visualización de los resultados se llevó a cabo con el sistema QiAxcel (modelo SK-QiAxcel-Slider, Qiagen), que permite separar ácidos nucleicos dependiendo de su tamaño mediante electroforesis capilar en gel.

La optimización se realizó mediante la amplificación de los fragmentos, a diferentes concentraciones de cebadores y condiciones de PCR, y posterior observación en dicho sistema de electroforesis capilar, modificando dichos parámetros hasta llegar a las condiciones óptimas.

3.3 Purificación de la librería

Con el objeto de eliminar productos inespecíficos, resto de cebadores no consumidos en la reacción de amplificación y dímeros de estos (**Fig. 22**), se utilizó el kit de purificación: MagSi-NGS PREP Manual Kit (MagnaMedics Diagnostics B.V., The Netherlands). Este kit permite la selección de tamaños y la limpieza de las sucesivas reacciones para aplicaciones NGS. La unión dirigida al tamaño de fragmentos de ADN de doble hebra varía modificando el volumen de reactivo específico en relación al volumen de la muestra. Al disminuir el volumen de la solución con respecto al volumen de la muestra a purificar, aumenta la eficacia de la unión a fragmentos más grandes pudiendo eliminar los más pequeños no unidos. Esto permite retener o descartar selectivamente tamaños de fragmentos no deseados.

En nuestro caso, se utilizó a una concentración 0.6X, previamente puesta a punto a diferentes concentraciones donde ésta era la idónea para la eliminación de tamaños inferiores a 100 pb.

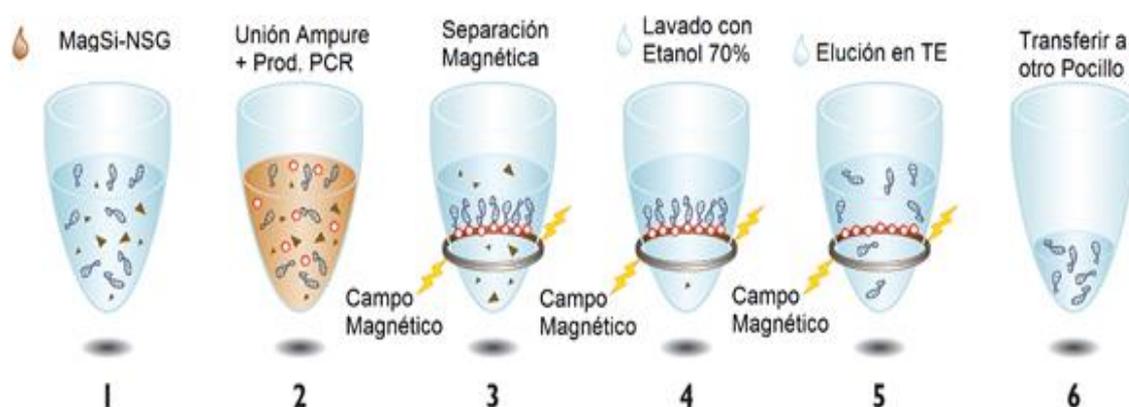


Figura 22: Etapas del protocolo de purificación

3.4 Cuantificación de la librería

Tras la purificación de las librerías de ADN, con los fragmentos de tamaño deseado, se realizó el agrupamiento de todas las muestras a secuenciar (*pool* de muestras) en la carga del equipo, de modo que el número de lecturas totales entre individuos/muestras fuera lo más homogéneo posible.

Para ello, se cuantificó la librería con el kit: QuantiFluor® dsDNA System (Promega). Midiendo la fluorescencia emitida entre 510-570nm en un espectrofluorímetro GLOMAX Multi-Detection System (Promega) según las especificaciones de la casa comercial.

Se realizó una recta patrón a partir de ADN de concentración conocida con el fin de calcular la concentración del *pool* purificado. A continuación, este *pool* se diluyó a una concentración final de $5 \cdot 10^5$ moléculas/ μL , concentración final indicada en el protocolo del sistema 454 GS Junior, para que hubiera una sola molécula de ADN unida a una bola de captura.

3.5 Amplificación de la librería por PCR en emulsión o emPCR

Este proceso consiste en una amplificación por PCR dentro de microesferas generadas a partir de una emulsión de agua y aceite (emPCR Amplification Method Manual, 2012). El objetivo de este proceso es que en cada una de esas microesferas haya miles de copias obtenidas a partir de una sola molécula de ADN unida a una bola de captura o *capture bead* a través de unos oligonucleótidos anclados en ellas y complementarios a los oligos *Forward* o *Reverse*.

El siguiente paso fue calcular el número de lecturas necesarias para llevar a cabo el experimento. El fabricante aconsejaba considerar un número inicial de lecturas de 70.000, que es el número de secuencias mínimo que se debe obtener con el sistema GS Junior por carrera. Se utilizó este valor como referencia para realizar el cálculo de los fragmentos y número de muestras que se van a incluir en cada una de ellas. La fórmula a utilizada fue la siguiente:

$$\mathbf{N^{\circ} \text{ Lecturas} = \text{Lecturas por amplicón} \cdot \mathbf{N^{\circ} \text{ de amplicones} \cdot \mathbf{N^{\circ} \text{ de individuos}}$$

A continuación, se estimó cuánto volumen había que añadir de cada *pool* a los reactivos para que al final hubiera una media de 1 molécula por microesfera. Teniendo en cuenta que hay $3 \cdot 10^6$ microesferas por reacción y que el *pool* está a una concentración de $5 \cdot 10^5$ moléculas/ μL , necesitábamos 6 μL del *pool* obtenido para cada uno de los sentidos *forward* y *reverse*.

Una vez ajustados los valores, se prosiguió con la amplificación por emPCR de la librería. Se utilizaron dos tipos de microesferas: las esferas A, que contienen oligos complementarios a los oligos *Forward*, y las B, que contienen oligos complementarios a los oligos *Reverse* de los adaptadores universales añadidos en la PCR. Con este mecanismo, cada bola unirá una única monocadena de cada fragmento de cada individuo y la amplificará, de forma que dentro de cada microesfera de emulsión habrá un solo fragmento amplificado numerosas veces (Fig. 23).

Las condiciones de la emPCR fueron: un primer paso de 4 minutos a 94°C , seguido de 50 ciclos que consistían en 94°C , 30 segundos, 58°C , 4,5 minutos y 68°C , 30 segundos. El producto tenía una estabilidad máxima de 16 horas a 10°C .

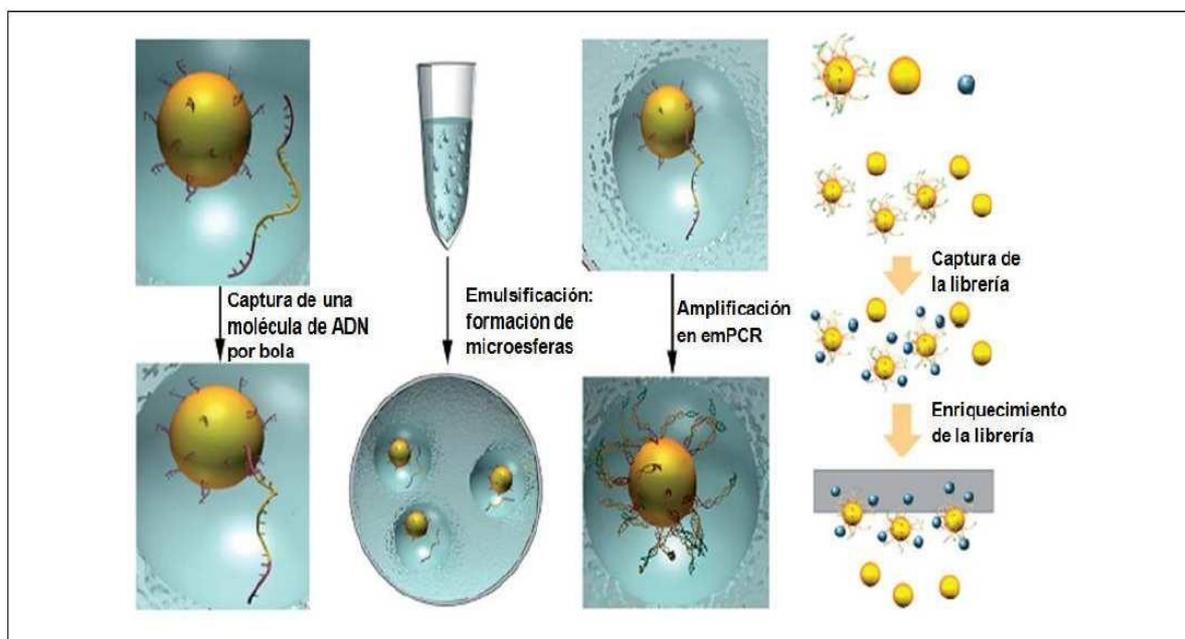


Figura 23: Esquema del proceso llevado a cabo durante la emPCR y el enriquecimiento de la librería.

3.6 Enriquecimiento de la librería

Este paso consistió en la rotura de la emulsión y enriquecimiento de la librería con microesferas de captura unidas a los fragmentos deseados, eliminando aquellas sin fragmentos, con el fin de aumentar la proporción de microesferas con ADN amplificado.

La desnaturalización del ADN con NaOH dejó únicamente moléculas monocadena unidas a la microesfera. Posteriormente, al otro extremo de la molécula de ADN, se unía el oligonucleótido de enriquecimiento unido a biotina. La muestra se mezcló con unas microesferas magnéticas cubiertas de estreptavidina, de forma que, con la ayuda de un imán y la unión biotina-estreptavidina, se separaron las microesferas de captura con muestra unidas a las microesferas magnéticas de aquellas microesferas sin muestra para eliminarlas (**Fig. 23**). Una vez enriquecida la librería con microesferas de captura con los fragmentos deseados, se añadieron los oligonucleótidos de secuenciación para pasar a la siguiente etapa.

3.7 Secuenciación

En este punto se utilizaron los oligonucleótidos de secuenciación, añadidos en el punto anterior, complementarios al extremo libre del fragmento como cebadores para que se pudiera realizar la síntesis de la nueva cadena.

La librería se cargó en una placa llamada *Pico Titer Plate Device*. Su estructura está formada por multitud de pocillos en forma de panal. En cada uno se depositó una sola bola con ADN amplificado y se recubrió de microesferas de empaquetamiento o *PackingBeads* para evitar que la señal emitida en un pocillo interfiera con la de los demás.

El sistema 454 GS Junior utiliza una metodología de secuenciación basada en la pirosecuenciación (**Fig. 24**). Este proceso ya descrito en la introducción, consiste en el acople del pirofosfato inorgánico (PPi), liberado al incorporar cada nuevo nucleótido a la cadena en síntesis, a dos reacciones en cadena llevadas a cabo mediante las enzimas sulfurilasa y luciferasa por las que, en el último paso, se libera señal de luz detectada por la cámara CCD de captación de imágenes del sistema.

Cada base es añadida en ciclos de manera secuencial y siempre en el mismo orden, mediante un proceso encadenado, por lo que se puede conocer qué base y en qué cantidad se ha unido, al emitir una señal luminiscente que recoge la cámara CCD. La intensidad de luz que se produce es proporcional al número de nucleótidos añadidos.

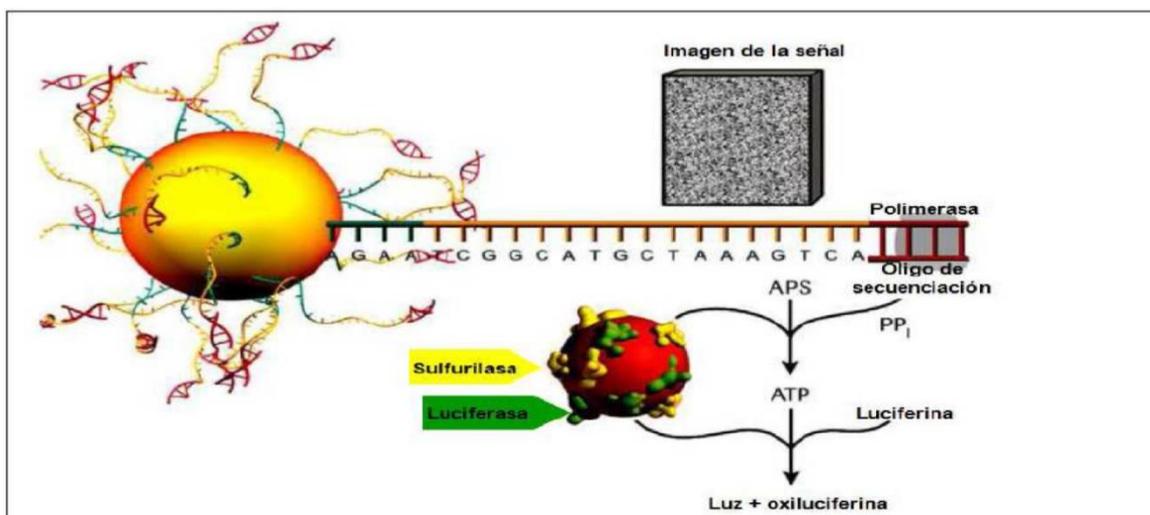


Figura 24: Esquema de la reacción de secuenciación llevada a cabo en la que se detecta una señal luminiscente en los pocillos donde se ha añadido un nuevo nucleótido.

La secuenciación realizada es bidireccional, puesto que se obtienen lecturas en sentido *forward* y sentido *reverse* en diferentes microesferas, aumentándose así la calidad y precisión de las lecturas, y minimizando el error de secuenciación. Las lecturas tienen una longitud media de 400pb de forma que cubren la secuencia amplificada completa junto con la secuencia de los MIDs en cada uno de los extremos.

3.8 Calidad de la secuenciación y procesamiento de datos

Una vez terminada la secuenciación, se revisó que todo el proceso funcionara correctamente. Para ello, se siguió el protocolo proporcionado por Roche (454 Sequencing System Software Manual, Version 2.7, 2012). Este consiste en la aplicación de una serie de filtros que nos indican la calidad de las lecturas. También nos indica si se ha calculado bien la cantidad de fragmentos de ADN en el *pool*, para

que solo hubiera como máximo una molécula de ADN amplificada por cada microesfera, y como había ido el proceso de secuenciación en general.

Estos filtros son (**Fig. 25**):

- *Dots Filter*: Filtro de lecturas con demasiados flujos negativos. Un 'Dot' es una instancia de tres flujos negativos sucesivos (no hay señal para tres flujos de base).
- *Mixed Filter*: Filtro de lecturas con demasiados flujos positivos o ambiguos. Cuando hay demasiadas incorporaciones de nucleótidos, posiblemente a partir de un pocillo que contiene más de una hebra de ADN, contaminación de señal de un pozo vecino o bien una relación señal/ruido baja.
- *Short Quality Filter*: Este filtro descarta lecturas que son demasiado cortas.
- *Passed Filter*: Todas las lecturas de buena calidad, que han pasado todos los filtros anteriores.

En carreras de muy buena calidad la suma del porcentaje de los *Dots Filter* más *Mixed Filter* es inferior al 20%.

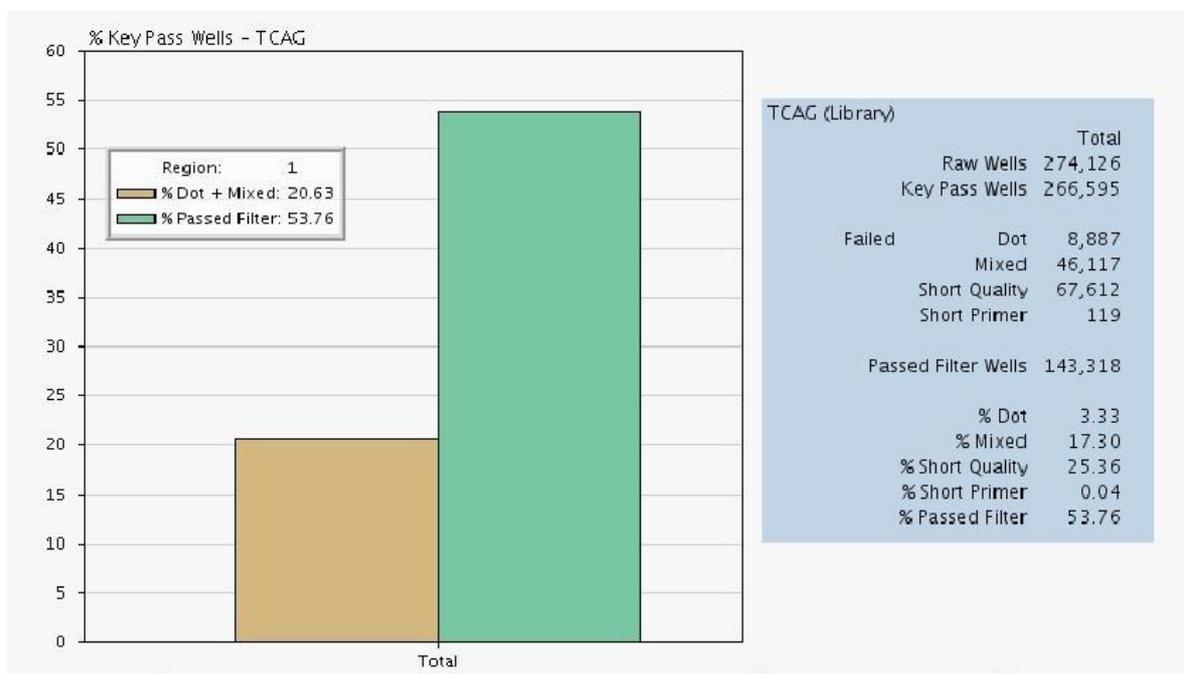


Figura 25: Ejemplo de resultados de los filtros aplicados tras la secuenciación de una carrera en el sistema 454 GS Junior.

Para analizar la calidad de la secuenciación de nuestro sistema, se analizó el porcentaje de lecturas en los filtros aplicados en 30 carreras consecutivas.

Posteriormente, se siguió con el procesado de datos utilizando el programa *Amplicon Variant Analyzer* (AVA, Roche) para llevar a cabo el análisis de las secuencias y detectar las variantes presentes en cada amplicón, las lecturas que le corresponden a cada uno por muestra y en qué sentido se había secuenciado.

Utilizando como referencia la secuencia de cada uno de los amplicones, y tomando como posición 1 la primera base del cebador utilizado en la amplificación, obtuvimos la posición relativa de las distintas variaciones en el programa AVA. Por medio de éste, el nombre de la referencia de la anotación del cDNA y del cambio aminoacídico a nivel de la proteína (**Tabla 2**). En caso de duda, se verificaron en los pirogramas los datos obtenidos y se comprobaron si las variaciones encontradas eran artefactos o reales.

GEN	EXÓN	POSICIÓN AVA	POSICIÓN cDNA	POSICIÓN PROTEÍNA	GEN	EXÓN	POSICIÓN AVA	POSICIÓN cDNA	POSICIÓN PROTEÍNA
BRAF	E15	76G:A	c.1756G>A	p.E586K			57 G:A	c.34G>A	p.G12S
		88G:A	c.1768G>A	p.V590I			57 G:C	c.34G>C	p.G12R
		101A:G	c.1781A>G	p.D594G			57 G:T	c.34G>T	p.G12C
		101A:T	c.1781A>T	p.D594V			58 G:A	c.35G>A	p.G12D
		105T:G	c.1785T>G	p.F595L			58 G:C	c.35G>C	p.G12A
		106G:C	c.1786G>C	p.G596R			58 G:T	c.35G>T	p.G12V
		115A:G	c.1795A>G	p.T599A			60 G:A	c.37G>A	p.G13S
		116C:T	c.1796C>T	p.T599I			60 G:C	c.37G>C	p.G13R
		118G:A; 119T:A	c.1798_1799GT>AA	p.V600K			60 G:T	c.37G>T	p.G13C
		118G:A; 119T:G	c.1798_1799GT>AG	p.V600R			61 G:A	c.38G>A	p.G13D
		119T:A	c.1799T>A	p.V600E			61 G:C	c.38G>C	p.G13A
		121A:G	c.1801A>G	p.K601E			61 G:T	c.38G>T	p.G13V
		71-72 GG:AA	c.34_35GG>AA	p.G12N			63 G:A	c.40G>A	p.V14I
		71G:A	c.34G>A	p.G12S			87A:G	c.64C>A	p.Q22K
		71G:T	c.34G>T	p.G12C			88C:A	c.181C>A	p.Q61K
	72G:A	c.35G>A	p.G12D	89A:G	c.182A>G	p.Q61R			
	72G:C	c.35G>C	p.G12A	89A:T	c.182A>T	p.Q61L			
	72G:T	c.35G>T	p.G12V	90A:C	c.183A>C	p.Q61H			
	74G:C	c.37G>C	p.G13R	90A:T	c.183A>T	p.Q61H			
	75G:A	c.38G>A	p.G13D	153 G:A	c.436G>A	p.A146T			
75G:T	c.38G>T	p.G13V	153 G:C	c.436G>C	p.A146P				
NRAS	E3	30C:T	c.149C>T	p.T50I		E4	154 C:T	c.437C>T	p.A146V
		60 G:A	c.179G>A	p.G60E			64 A:G	c.347A>G	p.N116S
		62 C:A	c.181C>A	p.Q61K			68 A:C	c.351A>C	p.K117N
		64 A:C	c.183A>C	p.Q61H			68 A:T	c.351A>T	p.K117N
		63 A:G	c.182A>G	p.Q61R					
	63 A:T	c.182A>T	p.Q61L						
	59A:G	c.349A>G	p.K117E						
	146G:A	c.436G>A	p.A146T						
	146G:C	c.436G>C	p.A146P						
	147C:T	c.437C>T	p.A146V						

Tabla 2: Posiciones relativas de las variaciones en el programa AVA, nombre de la referencia de la anotación del cDNA y del cambio aminoacídico a nivel de la proteína.

4. Rendimientos de la técnica

Con los rendimientos de la técnica nos referimos a la cobertura, ya sea por carrera, por muestra o por amplicón, que nos facilita alcanzar los parámetros de exactitud, precisión y sensibilidad y especificidad analítica.

- Lecturas por carrera: Para analizar la cobertura obtenida por carrera, se analizaron el número total de lecturas en 30 carreras consecutivas. Se determinó el promedio y la desviación estándar (SD) de las lecturas por carrera.
- Lecturas por muestra: Se seleccionaron 5 muestras de cada una de las 30 carreras anteriores para calcular las lecturas por carrera ($n=150$). Se determinaron los promedios de lecturas por muestra en cada carrera con sus respectivas SD y su coeficiente de variación (CV) en tanto por ciento.
- Lecturas por amplicón: El rendimiento de las coberturas por amplicón se analizó utilizando las 150 muestras del estudio anterior. Con ello se analizó el promedio del número de lecturas y las SD de cada uno de los amplicones.

5. Pruebas de Validación

Las pruebas validación en un proceso determinan la idoneidad de una prueba, que se ha desarrollado, optimizado y estandarizado adecuadamente para un fin concreto.

Para la realización de las pruebas de validación hemos seguido la ISO 15189 (ISO/IEC 15189, 2012) donde se definen algunas de las normas que hemos utilizado para validar nuestra técnica de NGS. Para esto hemos utilizado estimaciones como son la exactitud, precisión, sensibilidad y especificidad analítica y concordancia con otras técnicas.

5.1 Exactitud del método de análisis

La exactitud se define como el grado de concordancia entre el resultado y el valor de referencia aceptado (ISI/IEC 3534-1, 2006; CLSI, 2008). En aplicaciones de NGS, la exactitud se define como “el grado de acuerdo entre las secuencias de ácidos nucleicos derivados del ensayo de nuestra técnica y una secuencia de referencia” (ten Bosch JR *et al*, 2008).

La exactitud, en nuestro caso, la expresamos como el error relativo del valor obtenido con respecto al valor asignado de muestras controles. Para evaluar la exactitud de nuestro método en el sistema 454 GS Junior, se analizaron muestras comerciales de *Horizon Diagnostics (Cambridge, United Kingdom)*, de la siguiente manera:

- Muestras con una sola variación presente al 50%:
 - Muestra HD593, mutación BRAF_E15: p.V600E
 - Muestra HD592, mutación KRAS_E2: p.G13D
 - Muestra HD580, mutación KRAS_E4: p.A146T
 - Muestra HD255, mutación KRAS_E2: p.G12A
 - Muestra HD204, mutación KRAS_E2: p.G12D
 - Muestra HD286, mutación NRAS_E3: p.Q61R

- Muestra HD200, panel con diferentes porcentajes de las mutaciones:
 - BRAF_E15: p.V600E al 10.50%
 - KRAS_E2: p.G12D al 6.00%
 - KRAS_E2: p.G13D al 15%
 - NRAS_E3: p.Q61K al 12.50%
- Y muestra WT HD172, que no presenta ninguna de las mutaciones descritas anteriormente

Todas estas muestras se analizaron por triplicado, en 3 amplificaciones con diferentes MID y en 3 carreras distintas, obteniendo el promedio de los porcentajes con sus correspondientes SD y errores relativos.

5.2 Precisión del método de análisis

La precisión se define como el grado de concordancia entre mediciones repetidas del mismo material. Para analizar la precisión de nuestro método se tuvo en cuenta sus dos componentes: la repetibilidad, similitud entre nuestros datos dentro de una misma carrera; y la reproducibilidad, concordancia entre nuestros datos obtenidos en distintas carreras. (ten Bosch JR *et al.*, 2008).

En este caso, tanto para estudiar la repetibilidad como la reproducibilidad se utilizaron muestras reales de diagnóstico oncológico con concentraciones entre 150 y 200 ng/μl, de calidades próximas a 1.80 de coeficiente de absorbancia (260/280 nm de longitud de onda del espectro).

Para medir la precisión de nuestro método se ha utilizado el CV, que mide la dispersión o variación de nuestras medidas frente a la media de éstas.

5.2.1 Repetibilidad

La repetibilidad se define como la precisión obtenida bajo las mismas condiciones de operación en un intervalo corto de tiempo, en diferentes amplificaciones de la misma muestra, en el mismo ensayo y en el mismo equipo (ten Bosch JR et al., 2008).

Para analizar la repetibilidad de nuestro método, se utilizaron cuatro muestras con una sola variación conocida y analizadas previamente:

- Muestra M1, mutación KRAS_E2: p.G12D (10% aprox.)
- Muestra M2, mutación NRAS_E3: p.Q61K (8% aprox.)
- Muestra M3, mutación BRAF_E15: p.V600E (4% aprox.)
- Muestra M4, mutación KRAS_E2: p.G13D (2% aprox.)

Estas muestras se analizaron por triplicado, en 3 amplificaciones diferentes con diferentes MID, en la misma carrera.

5.2.2 Reproducibilidad

La reproducibilidad se define como la precisión obtenida bajo diferentes condiciones de operación en un intervalo de tiempo más largo, en diferentes amplificaciones de las muestras, en diferentes ensayos y en diferentes equipos (ten Bosch JR et al., 2008).

Para analizar la reproducibilidad de nuestro método, se utilizaron tres muestras reales de diagnóstico oncológico con una sola mutación al 30% aproximado:

- Muestra M5, mutación BRAF_E15: p.V600E
- Muestra M6, mutación KRAS_E2: p.G12D
- Muestra M7, mutación NRAS_E3: p.Q61K

Se analizaron en 6 amplificaciones diferentes con diferentes MID, en 6 carreras y en dos equipos 454 GS Junior distintos.

5.3 Sensibilidad y especificidad analítica

La sensibilidad analítica se define como la proporción de resultados positivos de la prueba y que se clasifican correctamente como positivos (Chen B et al., 2009). La sensibilidad analítica puede expresarse como la probabilidad que, en el ensayo de una técnica, se detecte una variación de la secuencia cuando está presente en la región secuenciada (Gargis AS et al., 2012).

Por otro lado, la especificidad analítica se define como la proporción de resultados negativos de la prueba y que se clasifican correctamente como negativos (Chen B et al., 2009). Por tanto, la especificidad analítica es la probabilidad que el ensayo no detecte la variación cuando ésta no está presente (Gargis AS et al., 2012).

En el caso de la NGS, estos valores de sensibilidad y especificidad son dependientes de la cobertura y el porcentaje de la variación, puesto que a mayor número de lecturas y mayor porcentaje de variación disminuye el error de análisis. Para elegir el rango de lecturas y porcentaje de mutación que fueran representativos y homogéneos, se tuvo en cuenta el CV en las distintas condiciones, tomándose un valor del 20% como referencia.

Para analizar la sensibilidad y especificidad analítica de nuestro método, se escogieron 4 muestras de diagnóstico reales, analizadas previamente:

- Muestra con mutación BRAF_E15: p.V600E al 30%.
- Muestra con mutación KRAS_E2: p.G12D al 30%.
- Muestra con mutación NRAS_E3: p.Q61K al 30%.
- Mezcla de muestras WT, al 0% de todas las mutaciones anteriores.

A partir de estas muestras se prepararon unas mezclas de menor porcentaje. Para el estudio, se igualaron las concentraciones a 100ng/μL en tampón TE (Tris-EDTA, pH 8.0, Sigma-Aldrich), haciendo una mezcla con las tres (**Tabla 3**), y obteniendo una mezcla al 10% de cada una de las mutaciones (M1). Posteriormente se hicieron mezclas con un control WT al 1/2 (5% de mutación, M2), 1/5 (2% de mutación, M3) y 1/10 (1% de mutación, M4).

Preparación MUESTRA WT a 100 ng/μL					
WT	Mutación Portadora MIX-WT	[ADN]ng/μL	Vol muestra (μl)	+ Vol TE1x (μL)	Vol final (μl)
		188.00	53.19	46.81	100.00
Preparación MIX1 a 100 ng/μL y al 10% de cada una de las mutaciones					
M1	Mutación Portadora	[ADN]ng/μL	Vol muestra (μl)	+ Vol TE1x (μL)	Vol final (μl)
	NRAS_E3: p.Q61K	195.22	10.24	9.76	20.00
	KRAS_E2: p.G12D	172.05	11.62	8.38	20.00
	BRAF_E15: p.V600E	149.63	13.37	6.63	20.00
			TOTAL:	24.76	60.00
Preparación MIX2 al 5% de cada una de las mutaciones					
M2	Vol. 0% MUT (WT) (μl)	Vol. Al 10 % (M1) (μl)	Vol final al 5 % de MUT (μl)		
	20	20	40		
Preparación MIX3 al 2% de cada una de las mutaciones					
M3	Vol. 0% MUT (WT) (μl)	Vol. Al 5 % (M2) (μl)	Vol final al 2 % de MUT (μl)		
	24	16	40		
Preparación MIX4 al 1% de cada una de las mutaciones					
M4	Vol. 0% MUT (WT) (μl)	Vol. Al 2 % (M3) (μl)	Vol final al 1 % de MUT (μl)		
	20	20	40		

Tabla 3: Tabla de cálculo de todas las muestras a 100 ng/μL y preparación de las mezclas a distintos porcentajes. MIX-WT: Mezcla de muestras WT.

Se procesaron dichas mezclas en tres amplificaciones diferentes con diferentes MID, y se cargaron en cinco carreras buscando obtener 1000, 750, 500, 250 y 100 lecturas por amplicón, respectivamente. A partir de los resultados, se obtuvo el promedio de los porcentajes de variación con sus correspondientes SD, el CV (%) y los valores de sensibilidad y especificidad analítica (en tanto por ciento) a las lecturas seleccionadas.

5.4 Concordancia con otros métodos

La concordancia es el grado en que dos o más métodos, técnicas u observaciones están de acuerdo sobre el mismo fenómeno observado. Sirve para conocer si con un método o instrumento nuevo se obtienen resultados equivalentes. Hemos comparado el resultado de diferentes mutaciones obtenidos por nuestro método de análisis con otros métodos:

- Comparación de la variación BRAF_E15: P.V600E con el kit *BRAF Mutation Analysis Reagent*.
- Comparación de las variaciones de *KRAS* con el sistema cobas® *KRAS Mutation Test*.
- Comparación de las variaciones de *NRAS* con método de Sanger.

En todos los casos, se seleccionaron una serie de muestras portadoras de mutación junto a algunas no portadoras en los genes a estudiar: *BRAF*, *KRAS* y *NRAS*.

5.4.1 Comparación de la variación BRAF_E15: P.V600E con el Kit *BRAF Mutation Analysis*

El kit *BRAF Mutation Analysis Reagent* (Applied Biosystems, Foster City, CA, USA) detecta y diferencia tres mutaciones en el codón V600 del gen BRAF (V600E, V600A, y V600G) utilizando un ensayo de extensión del cebador de terminación desplazada (STA) y análisis de los fragmentos por electroforesis capilar (ABI 3730 DNA Analyzer, Applied Biosystems) (**Fig 26**).

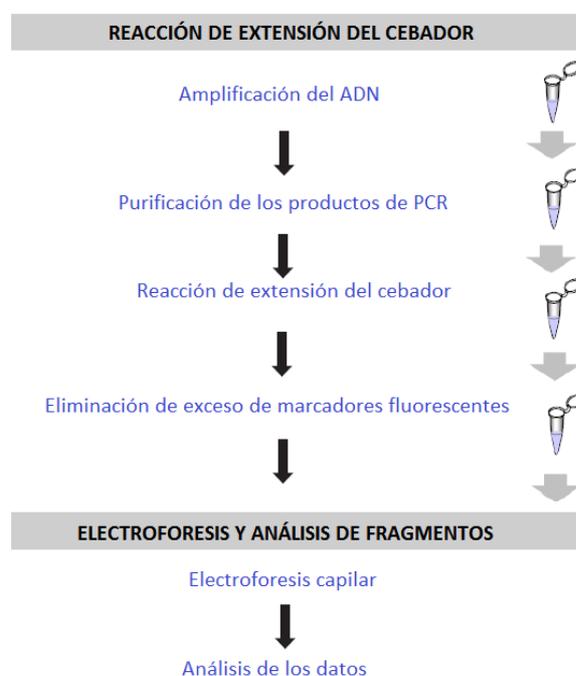


Figura 26: Esquema de pasos a realizar durante el procesamiento de las muestras con el Kit *BRAF Mutation Analysis*.

Los productos de extensión amplificados presentan un patrón de electroferograma distinto al diferenciar entre los WT y los alelos mutantes (**Fig. 27**).



Figura 27: Esquema de un ensayo de la terminación-desplazado (STA) reacción de ampliación del cebador y posterior análisis de los fragmentos por electroforesis capilar.

Se seleccionaron 34 muestras previamente genotipadas mediante este sistema en nuestro laboratorio. Y en paralelo, se introdujeron en los análisis en nuestro sistema (454 GS Junior, Roche). Una vez realizados estos análisis se compararon los resultados obtenidos entre ambos métodos.

Para comparar los resultados de ambas técnicas utilizamos el Test exacto de Fisher para muestras pequeñas, que nos indicaron si las variables eran independientes o no.

5.4.2 Comparación de las variaciones de *KRAS* con el kit *cobas*® *KRAS Mutation Test*

El kit de análisis de mutaciones en *KRAS cobas*® *KRAS Mutation Test* (Roche Molecular Systems, Mannheim, Germany) se basa en amplificación del ADN objetivo mediante PCR, utilizando pares de cebadores complementarios y dos sondas de oligonucleótidos con marcador fluorescente. Una sonda está diseñada para detectar la secuencia del codón 12/13 del exón 2 de *KRAS* y la otra para detectar la secuencia del codón 61 de *KRAS* en el exón 3 del gen *KRAS*. La detección se realiza en el analizador cuantitativo *cobas z 480* (Equipo c4800, Roche Diagnostics). Cada kit incluye un control de mutación, un control negativo y un calibrador para confirmar la validez de la prueba.

Tras la amplificación, cada amplicón generado mediante la prueba de mutación en *KRAS* de cobas® está sujeto a un programa de fusión en el que la temperatura aumenta de 40 °C a 95 °C. La sonda específica para amplicones no mutados se une tanto al amplicón mutado como al no mutado a temperaturas bajas. En el estado de unión, el marcador emisor (*reporter*) de fluoresceína en el extremo 5' de la sonda se encuentra lo suficientemente alejado del marcador silenciador (*quencher*) en el extremo 3', lo que permite al marcador fluorescente emitir una longitud de onda luminosa específica. A medida que aumenta la temperatura, la sonda se disocia del amplicón y permite al marcador silenciador posicionarse cerca del marcador fluorescente, lo que disminuye la cantidad de fluorescencia cuantificable. Los amplicones perfectamente unidos a la sonda (no mutados) se separan a una temperatura más alta que los amplicones con uno o más desapareamientos (mutados). Es posible detectar la presencia de una secuencia *KRAS* mutada en el exón 2 (codones 12 y 13) y en el exón 3 (codón 61) cuando las temperaturas de fusión se encuentran dentro de los intervalos establecidos (**Fig. 28**).

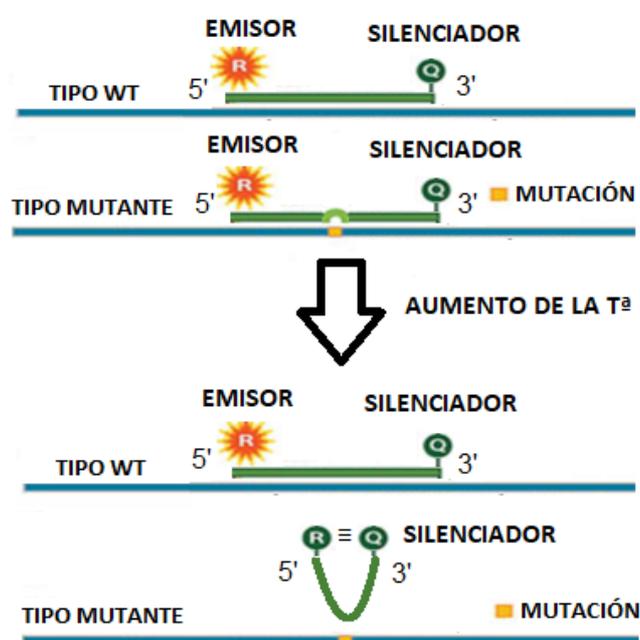


Figura 28. Esquema del proceso tras la amplificación de los fragmentos de ADN con el kit *cobas® KRAS Mutation Test*

Se seleccionaron 24 muestras previamente genotipadas mediante nuestro sistema de NGS en el laboratorio y se analizaron por el test *cobas®*. Una vez

realizados estos análisis se compararon los resultados obtenidos entre ambos métodos. Utilizando el Test exacto de Fisher para muestras pequeñas.

5.4.3 Comparación de las variaciones de NRAS con el método Sanger

Se seleccionaron 35 muestras, analizadas por nuestro sistema NGS, portadoras y no portadoras de mutación en el gen *NRAS* para determinarlas mediante secuenciación Sanger.

Para ello, se amplificó por PCR en tres reacciones uniplex cada uno de los tres exones de *NRAS* (2, 3 y 4). El mix de reacción por muestra consistió en 7.5 µl de master mix, los cebadores, 1.5 µl de DNA molde (a concentraciones entre 8 y 300 ng/ µl) y agua hasta un volumen final de 15 µl. El programa de amplificación fue idéntico al utilizado en las reacciones multiplex, descrito en el apartado 1.1 de la Metodología. Estos productos de PCR fueron purificados con el kit CleanPCR (Biotech B.V., Neherlands) y, posteriormente, se procedió a la reacción de secuenciación.

Para la reacción de secuenciación se utilizó el kit BigDye® terminator v.3.1 cycle Sequencing Kit (Applied Biosystems, Foster City, CA, USA). Se preparó una mezcla de reacción con los siguientes volúmenes por pocillo:

- Tampón de Secuenciación 5X: 1 µl.
- DMSO: 0.30 µl.
- H₂O: 0.30 µl.
- BigDye: 0.40 µl;
- Oligonucleótido: 0.80 µl a 2 µM.
- Producto de PCR purificados: 2.50 µl.

El programa de amplificación consistió en un primer paso de 10 minutos a 96°C, seguido de 45 ciclos que consistían en 96°C, 20 segundos, y 62°C, 1 minuto, con una extensión final de 10 minutos a 62°C

Una vez terminada la reacción se purificaron los productos con el kit de purificación *genMAG-Sequence* (Chemical GmbH, Germany) obtenidos para eliminar los restos de oligonucleótidos, sales y sobre todo ddNTPs marcados.

Finalmente, se cargó todo el volumen obtenido en un equipo de análisis de ácidos nucleicos (ABI 3730 DNA Analyzer).

Y comparamos las diferentes mutaciones de *NRAS* en el sistema 454 GS Junior con la secuenciación Sanger (**Fig. 29**), aplicando el Test exacto de Fisher para muestras pequeñas.

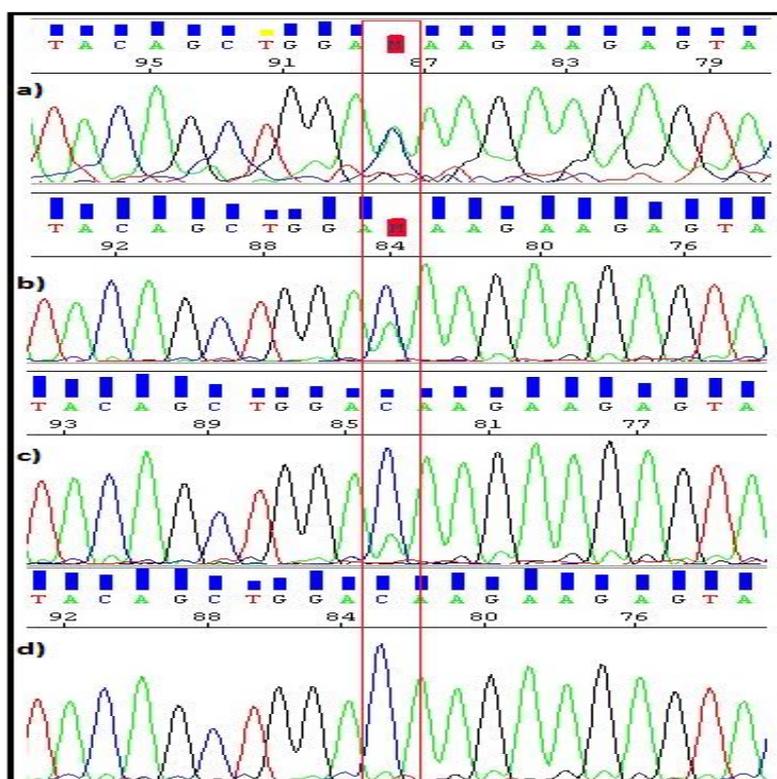


Figura 29. Electroferogramas visualizados con el programa Sequencing Analysis. Aparición de la variación *NRAS_E3*: p.Q61K (c.181 C>A): a) 40%, b) 20%, c) 10% y d) 2%. M: Cambio de C por A.

6. Controles de calidad

La participación en un control de calidad externo es un requisito imprescindible para garantizar la calidad de una determinación. Su principal ventaja es que su diseño, implementación y evaluación son completamente independientes, lo que garantiza la transparencia del proceso (Hernández-Losa J *et al.*, 2012).

El laboratorio donde se ha realizado el presente trabajo está inscrito en el Programa de Garantía de Calidad de la SEAP. Este programa consiste en la participación en un programa de análisis molecular interlaboratorio para la determinación de mutaciones de interés en los genes *BRAF*, *KRAS* y *NRAS*.

Tras la recepción de las muestras y antes de la finalización del plazo convenido, se realizó la determinación siguiendo el proceso de rutina del laboratorio. Finalmente, a partir de los datos recibidos por todos los laboratorios participantes, el laboratorio coordinador emitió los resultados para la evaluación interna de los mismos.

El sistema 454 GS Junior ha sido el utilizado para la realización de los Módulos de Patología Molecular del periodo 2014-2016.

RESULTADOS

1. Diseño del panel con adaptadores 454 universales

Por medio de la aplicación PRIMER3, se obtuvo un panel de siete parejas de cebadores para la amplificación de los fragmentos de PCR que cumplieran las condiciones descritas en la Metodología (**Tabla 4**). Las regiones amplificadas se obtuvieron mediante la aplicación UCSC In-Silico PCR (<https://genome.ucsc.edu>).

AMPLICIÓN	SECUENCIA
BRAF_E15	AGGTCAGGATCAACGC TCAAG ATAATGCTTGCTCTGATAGGAAAATG CATCTTGCATGATCCAACC TTC AAAAATAGCCTCAATTCTTACCATCC chr7:140453055-140453254 200bp ATAATGCTTGCTCTGATAGGAAAATGagatctactgttttcctttactactacacctcagatatattctcatgaag acctcacagtaaaaatagggtattttggtctagctacagtgaaatctcgatggagtggtcccatcagttgaacagttgctggat ccattttgGGATGGTAAGAATTGAGGCTATTTT
KRAS_E2	AGGTCAGGATCAACGC TCAAG TTTTATTATAAGGCCTGCTGAAAATG CATCTTGCATGATCCAACC TTC GTA CT CATGAAAATGGTCAGAGAAAACC chr12:25398122-25398341 220bp TTTTATTATAAGGCCTGCTGAAAATGactgaatataaaactgtggtagttggagctggtggcgtagggcaagagt gccttcagcagatacagctaattcagaatcattttgggacgaatgatccaacaatagaggtaaatctgttttaatatgcatattact ggtgcaggaccattctttgatacagataaaGGTTTCTCTGACCATTTTCATGAGTAC
KRAS_E3	AGGTCAGGATCAACGC TCAAG GTGTTTCTCCCTTCTCAGGATTC CATCTTGCATGATCCAACC TTC TTTATGGCAAATACACAAAGAAAGC chr12:25380204-25380364 161bp GTGTTTCTCCCTTCTCAGGATTCctacaggaagcaagtagtaattgatggagaaacctgtctcttgatattctcg acacagcaggtaagaggagtagcagtgcaatgagggaccagtaacatgaggactggggaggGCTTTCTTTGTGTA TTTGCCATAAA
KRAS_E4	AGGTCAGGATCAACGC TCAAG TCCCAGAGAACAATTAAGAG CATCTTGCATGATCCAACC TTC GTATTTATTTCA GTGTTACTTACCTGTCT chr12:25378525-25378714 190bp TCCCAGAGAACAATTAAGAGGtaaggactctgaagatgtacctatggtcctagtaggaaataaatgtgatt tgcttctagaacagtagacacaaaacaggctcaggactagcaagaagttatggaattcctttattgaaacatcagcaaaag acaAGACAGGTAAGTAACACTGAAATAAATAC
NRAS_E2	AGGTCAGGATCAACGC TCAAG TGATTACTGGTTTCCAACAGGTTT CATCTTGCATGATCCAACC TTC TGGGATCATATTCATCTACAAAAGTGG chr1:115258679-115258818 140bp TGATTACTGGTTTCCAACAGGTTTctgctggtgtgaaatgactgagtacaaactggtggtggggagcagggtg gtttgggaaaagcgcactgacaatccagctaaccagaaCCACTTTGTAGATGAATATGATCCCA
NRAS_E3	AGGTCAGGATCAACGC TCAAG CAGAAAACAAGTGGTTATAGATGGTGA CATCTTGCATGATCCAACC TTC CCTGTAGAGGTTAATATCCGCAAA chr1:115256420-115256591 172bp A CAGAAAACAAGTGGTTATAGATGGTGAaacctgtttgtggacatactggatacagctggacaagaagagta cagtgccatgagagaccaatacatgaggacaggcgaaggctcctctgtgatttgccatcaataatgcaagtcaTTGC GGATATTAACCTCTACAGG
NRAS_E4	AGGTCAGGATCAACGC TCAAG GGAGCAGATTAAGCGAGTAAAAG CATCTTGCATGATCCAACC TTC TGAATATGGATCACATCTCTACCAG chr1:115252113-115252349 237bp GGAGCAGATTAAGCGAGTAAAAGactcggatgatgtacctatggtgctagtgggaaacaagtgtgatttgccaa caaggacagttgatacaaaaacagcccacgaactggccaagagttacgggattccattcattgaaacctcagccaagacc agacaggtatggtacagcttcagcatttgcgaagattgcatcagttgattaactCTGGTAGAGATGTGATCCAT ATTCA

Tabla 4: Secuencia de los adaptadores 454, en negrita, más la secuencia homóloga de los amplicones de interés, el cromosoma y la posición genómica de dicho amplicón, así como su tamaño en pb y la secuencia mediante la aplicación UCSC In-Silico PCR.

1.1 Obtención de los fragmentos por PCR

Se consiguieron unas condiciones de amplificación, en cuanto a la preparación del mix y programa de amplificación, idóneos para la obtención de todos los fragmentos esperados y con una proporción similar de todos ellos. A partir de concentraciones iniciales de 0.4 μM de par de cebadores, éstas llevaron un proceso de puesta a punto para llegar a las concentraciones apropiadas para la consecución de este objetivo detalladas en la **Tabla 5**.

NOMBRE DEL CEBADOR	CONCENTRACIÓN (μM)*	TAMAÑO (pb)**	
NRAS_E4	0.60	307	A
KRAS_E2	0.30	290	B
BRAF_E15	0.24	271	C
KRAS_E4	0.46	260	D
NRAS_E3	0.16	242	E
KRAS_E3	0.20	231	F
NRAS_E2	0.12	210	G

Tabla 5: Concentración (μM) para realizar la PCR y tamaños obtenidos tras el programa de amplificación. * Se refiere a la concentración final de cada uno de los cebadores en la mezcla de reacción de PCR. ** Hace referencia al tamaño de cada fragmento + adaptadores tras la PCR. Las letras corresponden a los picos de cada fragmento representados en la Figura 30.

Se puso a punto el programa de amplificación idóneo que consistió en un primer paso de 15 minutos a 95°C, seguido de 40 ciclos que consistían en 98°C, 20 segundos, 60°C, 30 segundos y 72°C, 1 minuto, con una extensión final de 10 minutos a 72°C.

Para confirmar la amplificación de todos los fragmentos, se visualizaron los resultados con el sistema QiAxcel de Qiagen, que permite separar ácidos nucleicos dependiendo de su tamaño mediante electroforesis capilar en gel. Como puede observarse en la **Figura 30**, se obtuvieron los 7 fragmentos esperados.

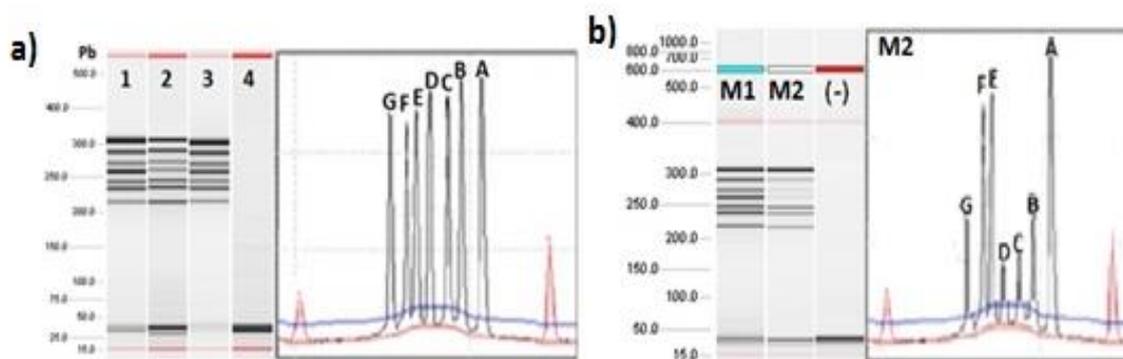


Figura 30: Apariencia de los fragmentos obtenidos (A-G) tras la PCR y posterior electroforesis en capilar en sistema QIAxcel. a) Carreras 1, 2 y 3 pertenecientes a muestras comerciales y carrera 4 control negativo; b) M1: muestra de buena calidad, M2: muestra de calidad baja, control negativo (-).

Este sistema permite también confirmar la calidad de las muestras, ya que las que tienen buena calidad permiten la obtención proporcional de todos los amplicones, dando un electroferograma como el que se muestra en la **Figura 30a**. Las muestras de baja calidad, sin embargo, normalmente no amplifican todos los fragmentos por igual, dando un electroferograma con picos a diferentes alturas, proporcionales a la cantidad de producto amplificado (**Fig. 30b**).

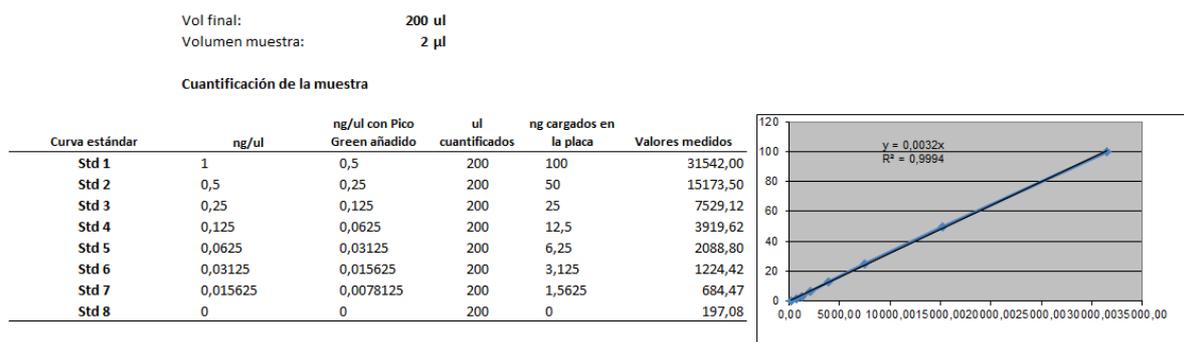
1.2 Estandarización de la librería

Una vez obtenida la librería de ADN purificada con los fragmentos deseados, se procedió al agrupamiento de todas las muestras a secuenciar en la carga del equipo. De este modo, conseguir que el número de lecturas totales (de moléculas iniciales que serán amplificadas) entre individuos sea lo más homogéneo posible.

Cuando tenemos cuantificadas las muestras purificadas (**Tabla 6**), se hacen los cálculos necesarios para pasar las concentraciones de ng/μl a número de moléculas/μl, mediante el promedio del número de pares de bases de los 7 fragmentos amplificados, el peso molecular promedio de un par de bases y el número de Avogadro.

Cuando ya hemos obtenido el número de moléculas por μl , se igualan a la de menor concentración y se obtiene el *pool* de todas las muestras a la misma concentración en moléculas por μl , que como en el ejemplo de la **Tabla 6** es $4,13 \times 10^{10}$ moléculas/ μl . Después se hacen las diluciones necesarias hasta llegar a las $5,0 \times 10^5$ moléculas/ μl .

CUANTIFICACIÓN PARA AGRUPAMIENTO DE MUESTRAS A $5,00\text{E}+05$ moléculas de ADN/ μl



Muestra	Valores medidos	ng en 2 μl (ng/ μl)	Longitud (pb)	Concentración (molec/ μl)	Nº de fragmentos	Vol en base a fragmentos	Vol agua	moleculas totales	Concentracion individual
MUESTRA 1	29004,30	46,43	260	1,64E+11	7	2,52	7,48	4,13E+11	3,44E+09
MUESTRA 2	12430,60	19,84	260	7,00E+10	7	5,90	4,10	4,13E+11	3,44E+09
MUESTRA 3	7358,99	11,70	260	4,13E+10	7	10,00	0,00	4,13E+11	3,44E+09
MUESTRA 4	13559,90	21,65	260	7,64E+10	7	5,41	4,59	4,13E+11	3,44E+09
MUESTRA 5	25946,10	41,52	260	1,46E+11	7	2,82	7,18	4,13E+11	3,44E+09
MUESTRA 6	26613,30	42,59	260	1,50E+11	7	2,75	7,25	4,13E+11	3,44E+09
MUESTRA 7	24291,10	38,87	260	1,37E+11	7	3,01	6,99	4,13E+11	3,44E+09
MUESTRA 8	35714,40	57,19	260	2,02E+11	7	2,05	7,95	4,13E+11	3,44E+09
MUESTRA 9	32642,50	52,26	260	1,84E+11	7	2,24	7,76	4,13E+11	3,44E+09
MUESTRA 10	19550,20	31,26	260	1,10E+11	7	3,74	6,26	4,13E+11	3,44E+09
MUESTRA 11	31193,00	49,94	260	1,76E+11	7	2,34	7,66	4,13E+11	3,44E+09
MUESTRA 12	21193,00	33,90	260	1,20E+11	7	3,45	6,55	4,13E+11	3,44E+09
						46,23	73,77		
						Volumen total	120,00		
						Concentración total en el pool		4,13E+10	
						Volumen a coger para hacer diluciones		5,00	

	Vol final	Añadir al Pool
Vol a añadir para llegar a	1,00E+09	206,45
		201,45
Concentración partida:	1,00E+09	μl Pool
Dilución 1	3,333	6
Dilución 2	50	4
Dilución 3	12	5

Tabla 6: Ejemplo de una tabla de cuantificación de un *pool* de una librería. En gris la concentración (moléculas / μl) de la muestra más diluida, las demás se igualaron a esta concentración de manera que todas quedan a la misma concentración y ya se pueden diluir para llegar a $5 \cdot 10^5$ moléculas / μl .

Las muestras que después de ser amplificadas obtenían una concentración menor o igual a 2 ng/μl no se incluían en el *pool*. Pues, normalmente, ésta eran muestras de mala calidad que amplificaban mal, o por un mal funcionamiento de su amplificación, en estos casos se producen una gran cantidad de fragmentos pequeños que, aun purificando con el kit Magsi-NGS PREP, seguían interfiriendo en la calidad de la secuenciación. Estas muestras se volvían a procesar y, si la amplificación era buena, se seguía con el protocolo habitual; pero, si no amplificaban bien o su concentración era inferior a 2 ng/μl, se daban como no analizables.

2. Calidad de la secuenciación

Para saber la calidad de las lecturas y del proceso en la secuenciación, se estudió el porcentaje de lecturas que han pasado por los distintos filtros aplicados y los rendimientos de la técnica que hemos obtenido durante este proceso.

CARRERA	%Dot+Mixed	% Short Quality	% Passed Filter
1	11.41	31.41	56.51
2	14.91	29.47	55.14
3	14.08	25.68	57.94
4	12.81	40.23	46.35
5	8.32	42.63	48.56
6	8.38	27.81	63.80
7	16.44	42.53	40.96
8	31.58	45.08	22.23
9	10.11	28.73	60.76
10	14.93	36.96	47.69
11	9.67	26.08	63.87
12	33.68	41.78	24.52
13	13.30	37.68	48.27
14	11.25	36.61	51.48
15	11.46	29.66	57.38
16	43.41	38.60	17.50
17	9.76	34.82	54.66
18	45.72	34.17	19.72
19	12.49	40.27	46.33
20	16.22	35.40	48.05
21	25.68	43.35	30.41
22	18.26	45.24	35.30
23	36.73	37.20	25.67
24	19.71	39.13	40.74
25	15.41	32.31	51.99
26	14.31	27.39	57.85
27	19.14	43.84	36.61
28	10.03	32.36	57.12
29	10.39	29.66	59.66
30	12.28	36.24	51.23
PROMEDIO	17.73	35.74	45.94

Tabla 7: Lecturas en porcentajes de los distintos filtros durante la secuenciación de 30 carreras en el Sistema GS 454 Junior.

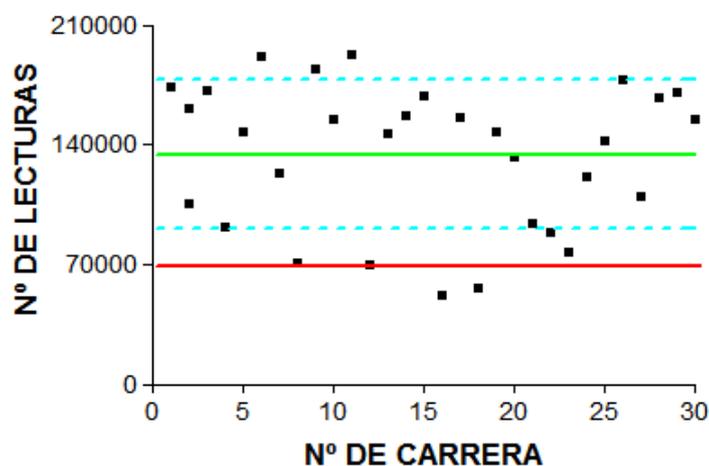
En la **Tabla 7** se puede apreciar que el promedio de lecturas que han pasado los filtros es del 45.94%, con un *Dot+Mixed* del 17.73% que nos indica una buena calidad en la secuenciación en general según los estándares de Roche.

2.1 Rendimientos de la técnica

Nos referimos a los rendimientos como la estabilidad de la técnica, es decir, como al número de lecturas por carrera, por muestra o por amplicón y que este esté dentro de un límite y sea más o menos constante entre ellos.

2.1.1 Lecturas por carrera

Se cargaron 30 carreras en el sistema GS Junior y se obtuvo un promedio de 135.000 lecturas por carrera, con una SD de 42.000 lecturas (Gráfica 1). El fabricante aconseja considerar un número inicial de 70.000 lecturas, superado, como se puede ver, en la mayoría de los casos.



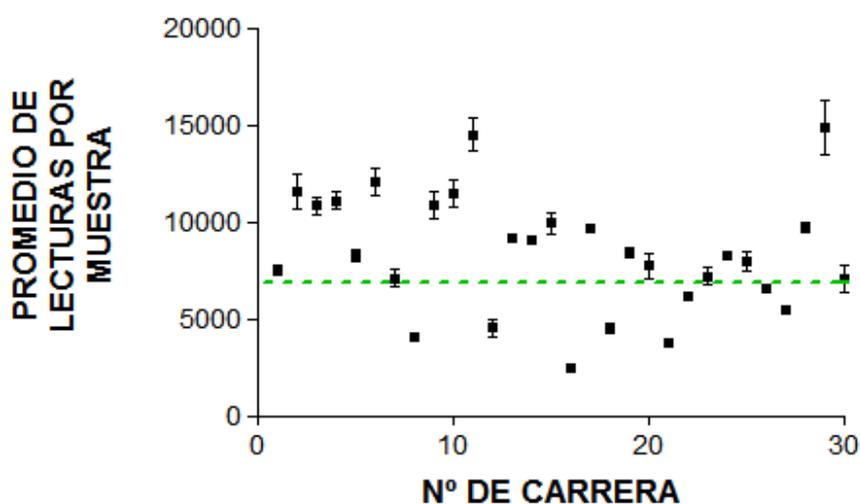
Gráfica 1: Representación gráfica del número de lecturas totales obtenido frente 30 carreras analizadas en el sistema GS Junior.

En la **Gráfica 1**, la línea recta roja indica el mínimo número de lecturas por carrera para poderse considerar que ha tenido un buen rendimiento. La verde indica el promedio de lecturas obtenido tras analizar 30 carreras consecutivas, y las líneas azules discontinuas nos indican las SD. Podemos ver que la mayoría de los resultados obtenidos, en número de lecturas por carrera, están entre el promedio y sus SD. Se puede apreciar que en las carreras 16 y 18, que tienen 52000 y 56000 lecturas, respectivamente, están por debajo de las 70000 lecturas (umbral que el fabricante recomienda). Esto puede ser ocasionado por un error en la cuantificación y posterior agrupamiento de las muestras. De manera, que en algunas microesferas

de emPCR habría más de un amplicón, por lo que en su secuenciación, al pasar los filtros de calidad, muchas de sus lecturas fueran rechazadas. La consecuencia fue que se obtuvieran menos lecturas de las esperadas. Sin embargo, esto no invalida el resultado de la carrera, por lo que estos datos fueron utilizados para dar resultados en los diagnósticos pertinentes.

2.1.2 Lecturas por muestra

Se analizaron 150 muestras en 30 carreras en el sistema 454 GS Junior. Se obtuvo el promedio de lecturas por muestra y su SD en cada carrera. Con un promedio de 8300 lecturas por muestras con una SD de 3200 lecturas.



Gráfica 2: Representación gráfica del promedio de lecturas por muestra obtenido frente a 30 carreras analizadas en el sistema GS Junior.

En la **Gráfica 2** se puede apreciar que el número de lecturas por carrera para muestras individuales presenta una variabilidad notable entre diferentes carreras. Sin embargo, en cada carrera, en la mayoría de los casos los valores de número de lecturas presentan poca dispersión.

También podemos ver, que esta dispersión no es dependiente del número de lecturas entre los individuos analizados, es decir que, aunque haya más o menos lecturas éstas se distribuyen homogéneamente entre las distintas muestras de una misma carrera.

Para ver homogeneidad entre el número de lecturas en las distintas muestras de una carrera analizamos el CV.

CARRERA	PROMEDIO DE LECTURAS POR MUESTRA	CV (%)	CARRERA	PROMEDIO DE LECTURAS POR MUESTRA	CV (%)
1	7525	6.83	16	2496	7.68
2	11535	17.12	17	9696	3.82
3	10821	9.73	18	4499	11.38
4	11107	9.14	19	8443	6.67
5	8278	8.46	20	7720	17.62
6	12060	12.55	21	3791	7.51
7	7104	13.46	22	6177	5.36
8	4021	10.38	23	7189	13.91
9	10884	14.87	24	8295	4.35
10	11423	13.73	25	7946	14.67
11	14478	12.96	26	6533	4.26
12	4536	20.99	27	5427	4.28
13	9189	4.69	28	9743	5.61
14	9106	1.46	29	14864	20.59
15	9930	11.61	30	7068	22.41

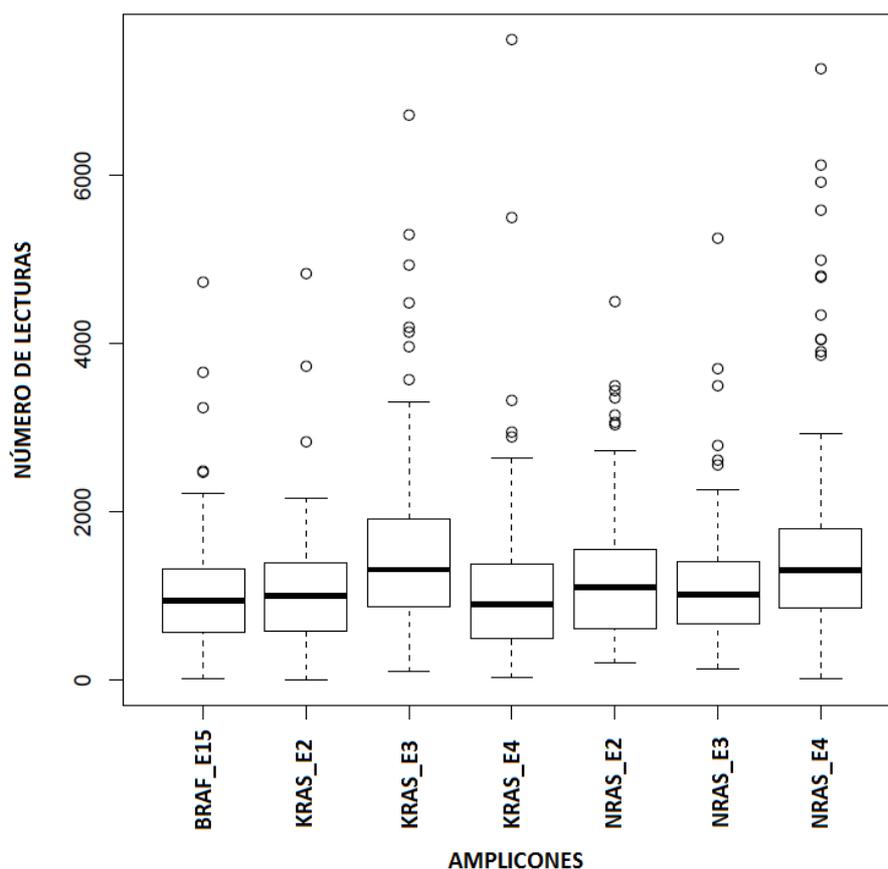
Tabla 8: Promedio de lecturas y CV obtenidos a partir de 150 muestras analizadas en 30 carreras en el sistema GS Junior.

Los CV van del 22.41%, en carreras donde el número de lecturas de las muestras tienen mayor dispersión, a un 1.46%, en carreras donde el número de lecturas son más homogéneas (**Tabla 8**).

2.1.3 Lecturas por amplicón

De las 150 muestras analizadas, se representa el número de lecturas obtenidas agrupadas por amplicón, en un diagrama de cajas (**Gráfica 3**).

Se puede ver que las medianas, en la mayoría de los fragmentos, están en torno a las 1000 lecturas; siendo en los amplicones KRAS_E3 Y NRAS_E4 superiores a éstas. También se aprecia que los valores alejados (*outlier*) también son mayoritarios en estos amplicones.



Gráfica 3: Representación gráfica en diagrama de cajas del número de lecturas por amplicón obtenido de 150 muestras analizadas en el sistema GS Junior. Las líneas discontinuas representan los cuartiles primero y cuarto. Las cajas representan los cuartiles segundo y tercero separados por la mediana (línea más gruesa). (o) valores alejados (outlier).

Podemos observar que salen en todos los fragmentos, en el primer cuartil, valores cercanos a 0 lecturas. Esto sucede porque en algunas muestras estos fragmentos no se amplificaron bien, ya sea por la calidad de las muestras, o porque hay competencia entre los oligonucleótidos de amplificación o por cualquier otra circunstancia estos no amplificaron correctamente. En la práctica, cuando sucede esto, volvemos a repetir todo el análisis de la muestra, si el análisis es correcto en cobertura se informa, si no lo es se da como no analizable.

De la misma manera, como hemos dicho arriba, aparecen valores muy alejados, en torno a 6000 lecturas.

3. Pruebas de validación

Para validar nuestro método, como se describe en el apartado 4 de la Metodología, tuvimos en cuenta la exactitud, precisión, sensibilidad y especificidad analítica, además de los rendimientos de la técnica que acabamos de ver. Por otro lado, también se analizó la concordancia de los resultados mutacionales obtenidos con los de otras técnicas y con controles de calidad interlaboratorio.

3.1 Exactitud del método de análisis

En la **Tabla 9** se representan los resultados obtenidos para las muestras comerciales cuyo porcentaje de mutación es conocido.

Se puede ver que los promedios de los porcentajes de las variaciones encontradas, tras el análisis por triplicado en el total de lecturas, se acercan mucho a los esperados. Las lecturas obtenidas fueron proporcionales en ambas direcciones y el porcentaje de mutación observado fue similar.

En las muestras comerciales con presencia de mutación (desde la HD593 hasta la HD200), con una densidad de lecturas superior a 1000, se observa que la dispersión obtenida es baja con unas SD inferiores al 4%.

En el análisis de la muestra comercial HD172, WT para los genes interrogados por nuestra técnica, en ningún caso se detectó la presencia de mutación.

Los resultados obtenidos para mutaciones descritas fueron los mismos a los esperados. De la misma forma, no se detectaron mutaciones no esperadas en ninguna de las muestras comerciales analizadas.

MUESTRA	GEN_EXON	cDNA	PROT	VARIACIONES ESPERADAS	VARIACIONES ENCONTRADAS			PROMEDIO	SD	ERROR RELATIVO
HD593	BRAF_E15	c.1799T>A	p.V600E	50.00	47.38	46.77	48.02	47.39	0.63	5.22
HD592	KRAS_E2	c.38G>A	p.G13D	50.00	50.14	48.90	51.41	50.15	1.26	0.30
HD580	KRAS_E4	c.436G>A	p.A146T	50.00	55.84	57.63	50.00	54.49	3.99	8.98
HD255	KRAS_E2	c.35G>C	p.G12A	50.00	48.68	52.40	44.89	48.66	3.76	2.69
HD204	KRAS_E2	c.35G>A	p.G12D	50.00	50.74	52.34	48.55	50.54	1.90	1.09
HD286	NRAS_E3	c.182A>G	p.Q61R	50.00	49.02	49.61	48.41	49.01	0.60	1.97
HD200	BRAF_E15	c.1799T>A	p.V600E	10.50	11.07	9.12	10.26	10.15	0.98	3.33
	KRAS_E2	c.35G>A	p.G12D	6.00	5.71	3.39	4.73	4.61	1.16	23.17
	KRAS_E2	c.38G>A	p.G13D	15.00	15.04	12.91	15.61	14.52	1.42	3.20
	NRAS_E3	c.181C>A	p.Q61K	12.50	6.28	5.44	7.55	6.42	1.06	48.61
HD172	BRAF_E15	c.1799T>A	p.V600E	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	KRAS_E2	c.38G>A	p.G13D	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	KRAS_E4	c.436G>A	p.A146T	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	KRAS_E2	c.35G>C	p.G12A	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	KRAS_E2	c.35G>A	p.G12D	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	NRAS_E3	c.182A>G	p.Q61R	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	NRAS_E3	c.181C>A	p.Q61K	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Tabla 9: Representación de los porcentajes de mutación detectados frente a los valores esperados correspondientes en muestras comerciales. El número de lecturas obtenido en este estudio fue superior a 1000 para todos los amplicones analizados (HD: Horizon Diagnostics). Valores expresados en porcentajes.

Vemos que los errores relativos en las muestras comerciales, que solo son portadoras de una mutación al 50%, son inferiores al 9%. En la muestra comercial HD172, que es de WT para todas las mutaciones del estudio, su error relativo es del 0%. Y en la muestra comercial HD200, panel con diferentes porcentajes de las mutaciones, obtenemos SD en torno al 3% para las variaciones BRAF: p.V600E y KRAS: p.G13D. Mientras que el error relativo en la variación KRAS: p.G12D es del 23.17% y para NRAS: p.Q61K es del 48.61%, respectivamente.

3.2 Precisión del método de análisis

Para estudiar la precisión analizamos la repetibilidad (similitud entre nuestros datos dentro de una misma carrera) y reproducibilidad (aproximación entre nuestros datos en distintas carreras). En este caso se utilizaron muestras reales de diagnóstico oncológico.

3.2.1 Repetibilidad

En la **Tabla 10** están representados los resultados de la repetibilidad obtenidos del análisis de cuatro muestras reales, con concentraciones y calidades parecidas, como se indica en el apartado 5.2.1 de la Metodología. Se puede observar que los porcentajes de mutación obtenidos en el total de lecturas para cada uno de los cuatro casos fueron muy similares dentro de la misma carrera, con unas SD inferiores al 2%. Los resultados obtenidos tanto en sentido directo como reverso fueron similares.

De la misma manera los resultados del CV fueron inferiores al 20%, excepto en uno de los casos de la muestra M2 donde se obtuvo un aumento de la dispersión de los resultados

El número de lecturas obtenido en el estudio de la repetibilidad fue aproximadamente de 1000 para todos los amplicones analizados.

MUESTRA	GEN_EXON	cDNA	PROT	% POSITIVAS	% PROMEDIO	SD	CV(%)
M1	KRAS_E2	c.35G>A	p.G12D	10.8	11.08	1.73	15.63
				12.94			
				9.51			
M2	NRAS_E3	c.181C>A	p.Q61K	8.57	7.45	1.73	23.28
				8.32			
				5.45			
M3	BRAF_E15	c.1799T>A	p.V600E	3.74	3.87	0.51	13.25
				3.44			
				4.44			
M4	NRAS_E2	c.38G>A	p.G13D	2.42	2.25	0.18	8.04
				2.06			
				2.27			

Tabla 10: Precisión dentro de la carrera: Repetibilidad. Porcentajes individuales, promedios, SD y CV obtenidos en el análisis en la misma carrera.

3.2.2 Reproducibilidad

En la **Tabla 11** se representan los resultados de la reproducibilidad obtenidos del análisis de tres muestras reales, como se indica en el apartado 5.2.2 de la Metodología.

Se puede observar que los porcentajes de mutación obtenidos en el total de lecturas para cada uno de los seis casos fueron muy similares dentro de las distintas carreras y equipos, con unas SD inferiores al 4%. Los resultados obtenidos tanto en sentido directo como reverso fueron similares.

De la misma manera los resultados de CV fueron inferiores al 15% en los tres casos.

El número de lecturas para el estudio de la reproducibilidad de nuestra técnica fue de aproximadamente de 1000 para todos los amplicones analizados.

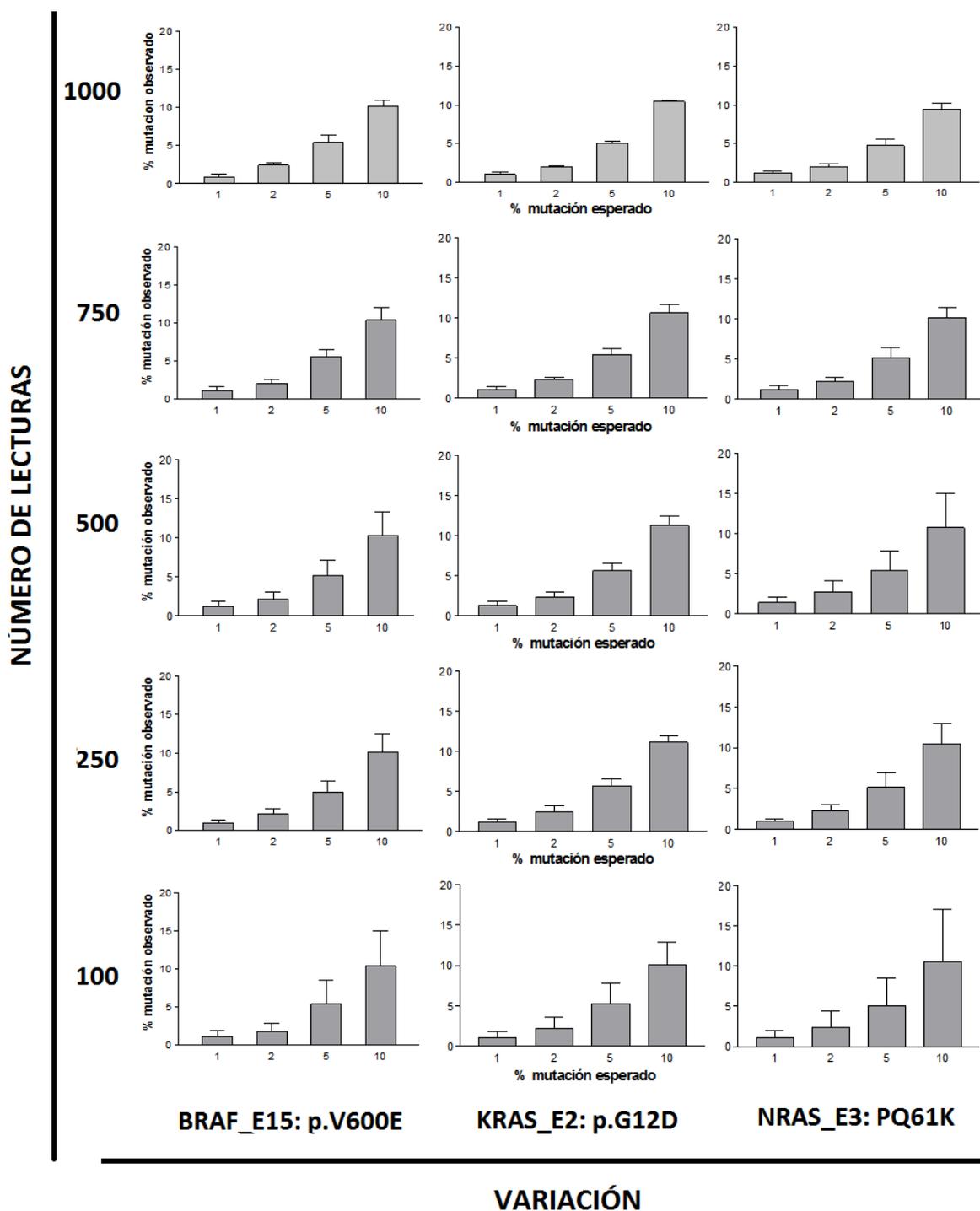
MUESTRA	GEN_EXON	cDNA	PROT	CARRERA	EQUIPO	% POSITIVAS	% PROMEDIO	SD	CV (%)
M5	BRAF E15	c.1799T>A	p.V600E	1	B	31.46	26.96	4.03	14.96
				2	B	24.69			
				3	A	30.67			
				4	B	23.36			
				5	A	29.42			
				6	A	22.15			
M6	KRAS E2	c.35G>A	p.G12D	1	B	31.34	27.98	3.05	10.90
				2	B	25.34			
				6	A	24.12			
				4	B	28.53			
				5	A	31.47			
				6	A	27.05			
M7	NRAS-E3	c.181C>A	p.Q61K	1	B	35.81	33.56	2.63	7.85
				2	B	32.19			
				3	A	33.33			
				4	B	37.59			
				5	A	31.74			
				6	A	30.71			

Tabla 11: Precisión entre carreras: Reproducibilidad. Porcentajes individuales, promedios, SD y CV obtenidos en el análisis de carreras y equipos diferentes.

3.3 Sensibilidad y especificidad analítica

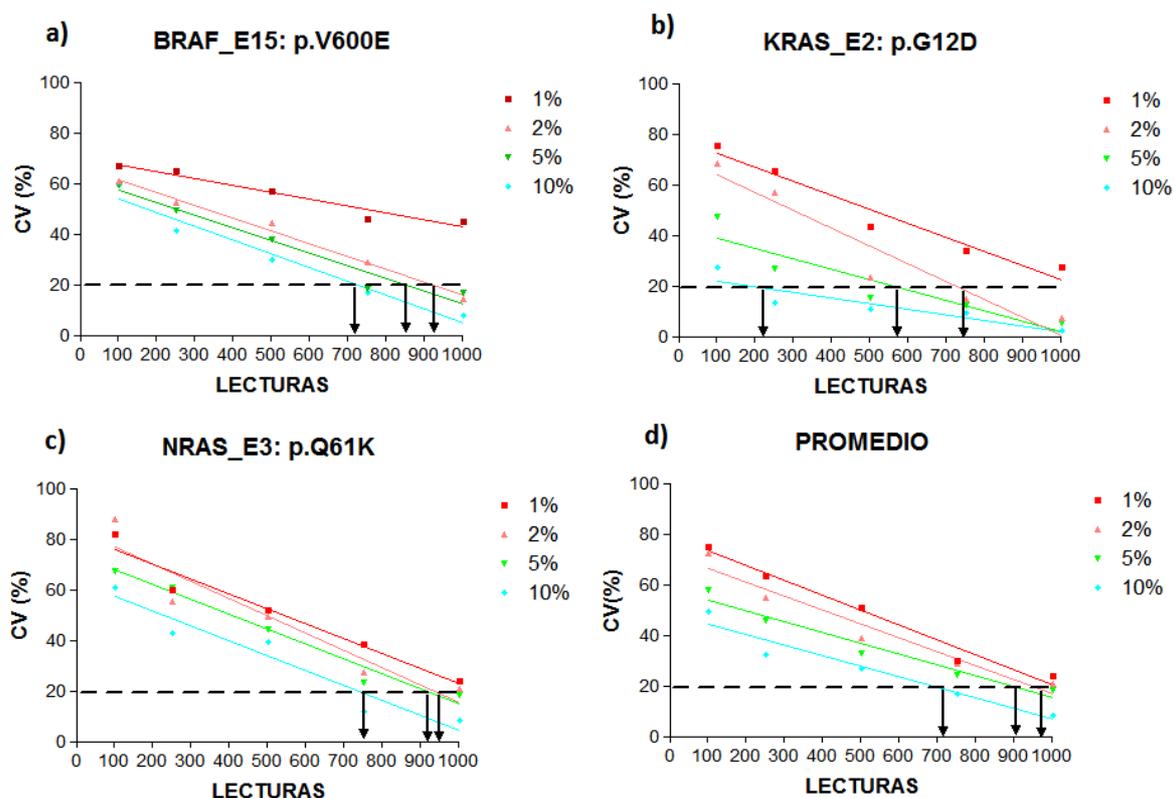
Para estudiar hasta qué porcentaje de mutación nuestro método es capaz de detectar las mutaciones BRAF_E15: p.V600E, KRAS_E2: p.G12D y NRAS_E3: p.Q61K se analizaron porcentajes al 10, 5, 2 y 1%. Se procesaron dichas mezclas, como se ha explicado en la Metodología (**Tabla 3**) y sus secuenciaciones se ajustaron a 1000, 750, 500, 250 y 100 lecturas por amplicón, respectivamente.

Se observa que en todos los casos el promedio del porcentaje obtenido se aproxima mucho al esperado, independientemente del porcentaje de mutación y del número de lecturas. La SD, es mayor en el caso de la mutación NRAS_E3: p.Q61K, algo menor en BRAF_E15: p.V600E y más pequeña en KRAS_E2: p.G12D. En general, a mayor promedio del porcentaje de mutación, la SD aumenta, y cuando llegamos a 1000 lecturas ésta tiende a estabilizarse, ya no es dependiente del porcentaje de mutación. Además, se observa en todos los casos, que a mayor número de lecturas, la SD va disminuyendo (**Gráfica 4**).



Gráfica 4: Representación gráfica del porcentaje de mutación observado frente al porcentaje de mutación esperado, para las variaciones BRAF_E15: p.V600E, KRAS_E2: p.G12D y NRAS_E3: p.Q61K, al 10, 5, 2 y 1% de cada una de ellas y a 1000, 750, 500, 250 y 100 lecturas por amplificación.

Estos resultados repercuten los valores del CV (**Gráfica 5**), se puede ver que el CV disminuye a medida que aumenta el número de lecturas y el porcentaje de mutación.



Gráfica 5: Representación gráfica del CV frente al número de lecturas, para las variaciones a) BRAF_E15: p.V600E, b) KRAS_E2: p.G12D, c) NRAS_E3: p.Q61K y d) el promedio de todas ellas. Al 1, 2, 5 y 10% de cada una de ellas y a 100, 250, 500, 750 y 1000 lecturas por amplicón. Para todas las rectas $r^2 > 0.9$ y $p\text{-valor} < 0.05$.

Como se observa en la **Gráfica 5**, en las distintas mutaciones y porcentajes de aparición de éstas, los CV se comportan de manera diferente. Podemos observar en todos los casos que a un 1% del promedio de estas mutaciones, el CV se aproxima al 20% a 1000 lecturas, aunque no llega a alcanzarse. A medida que aumenta el porcentaje de las mutaciones, podemos ver que, entre 750 y 1000 lecturas, el CV es menor del 20%. Se observa que al 2% de mutación y con más de 1000 lecturas el

CV es inferior al 20%, o lo que es lo mismo, valores de referencia donde el rango de lecturas y porcentaje de mutación son representativos y homogéneos (Dane, 2006).

Teniendo en cuenta esto, se puede calcular la sensibilidad y especificidad analítica, para el análisis de todas las mutaciones, tomando como referencia el 2% de mutación a 750 y 1000 lecturas respectivamente.

SENSIBILIDAD ANALÍTICA			
POSITIVAS \geq 2%			
750 Lecturas			
30%	+	9	100,00%
	-	0	
10%	+	9	100,00%
	-	0	
5%	+	9	100,00%
	-	0	
2%	+	8	88,89%
	-	1	
TOTAL	+	35	97,22%
	-	1	

SENSIBILIDAD ANALÍTICA			
POSITIVAS \geq 2%			
1000 Lecturas			
30%	+	9	100,00%
	-	0	
10%	+	9	100,00%
	-	0	
5%	+	9	100,00%
	-	0	
2%	+	9	100,00%
	-	0	
TOTAL	+	36	100,00%
	-	0	

ESPECIFICIDAD ANALÍTICA			
NEGATIVAS $<$ 2%			
750 Lecturas			
1%	+	2	77,78%
	-	7	
0%	+	0	100,00%
	-	24	
TOTAL	+	2	93,94%
	-	31	

ESPECIFICIDAD ANALÍTICA			
NEGATIVAS $<$ 2%			
1000 Lecturas			
1%	+	1	88,89%
	-	8	
0%	+	0	100,00%
	-	24	
TOTAL	+	1	96,97%
	-	32	

Tabla 12: Tabla de resultados obtenidos al analizar la sensibilidad y especificidad analítica de nuestra técnica, con coberturas de 750 y 1000 lecturas. Porcentajes de mutación esperada: 30, 10, 5, 2, 1 y 0%. El resultado es positivo (+) cuando el porcentaje de mutación obtenido es \geq 2% y negativo (-) cuando es $<$ 2%.

Se observa en la **Tabla 12** que a 750 lecturas la sensibilidad analítica al 2% de aparición de mutación es del 88.9% y al 5% es del 100%. Mientras que a 1000 lecturas la sensibilidad analítica al 2% de aparición de mutación ya es del 100%. Por lo que respecta a la especificidad analítica, el valor de este parámetro aumenta un 3% cuando el número de lecturas pasa de 750 a 1000.

3.4 Concordancia con otros métodos

3.4.1 Comparación de la variación BRAF_E15: P.V600E con el Kit BRAF Mutation Analysis (Applied Biosystems)

Se comparó la detección de la mutación BRAF_E15: P.V600E en 454 GS Junior con el Kit Applied Biosystems BRAF Mutation Analysis Reagents kit, obteniendo los resultados que se muestran en la **Tabla 13**. El test exacto de Fisher para muestras pequeñas obtuvo un p-valor = 5.388e-10 confirmando que las variables no son independientes (**Tabla 14**).

SISTEMA GS JUNIOR 454				KIT APPLIED BIOSYSTEMS
TOTALES				
MUESTRA	% POSITIVAS	LECTURAS	RESULTADOS	RESULTADOS
1	28.79	1195	POSITIVO	POSITIVO
2	0.21	2363	NEGATIVO	NEGATIVO
3	52.45	1630	POSITIVO	POSITIVO
4	0.00	447	NEGATIVO	NEGATIVO
5	35.16	1445	POSITIVO	POSITIVO
6	0.00	445	NEGATIVO	NEGATIVO
7	13.61	1609	POSITIVO	POSITIVO
8	0.25	406	NEGATIVO	NEGATIVO
9	14.28	1737	POSITIVO	POSITIVO
10	0.00	426	NEGATIVO	NEGATIVO
11	56.17	1734	POSITIVO	POSITIVO
12	45.71	1831	POSITIVO	POSITIVO
13	10.12	1877	POSITIVO	POSITIVO
14	0.00	456	NEGATIVO	NEGATIVO
15	15.27	2010	POSITIVO	POSITIVO
16	0.21	484	NEGATIVO	NEGATIVO
17	10.54	1395	POSITIVO	POSITIVO
18	0.00	406	NEGATIVO	NEGATIVO
19	49.27	1995	POSITIVO	POSITIVO
20	0.00	537	NEGATIVO	NEGATIVO
21	8.05	2049	POSITIVO	POSITIVO
22	0.00	409	NEGATIVO	NEGATIVO
23	24.90	2414	POSITIVO	POSITIVO
24	0.59	339	NEGATIVO	NEGATIVO
25	33.67	2076	POSITIVO	POSITIVO
26	0.00	432	NEGATIVO	NEGATIVO
27	19.30	1601	POSITIVO	POSITIVO
28	0.00	517	NEGATIVO	NEGATIVO
29	37.54	1950	POSITIVO	POSITIVO
30	21.78	1763	POSITIVO	POSITIVO
31	35.00	2226	POSITIVO	POSITIVO
32	0.00	518	NEGATIVO	NEGATIVO
33	14.37	2171	POSITIVO	POSITIVO
34	0.00	509	NEGATIVO	NEGATIVO

Tabla 13: Comparación de la mutación BRAF_E15: P.V600E en 454 GS Junior con el Kit Applied Biosystems BRAF Mutation Analysis Reagents kit.

El sistema de Applied Biosystems nos da valores categóricos; sin embargo, con 454 GS JUNIOR, además de saber si es positivo o negativo, también se puede saber cuál es el porcentaje de mutación.

En todos los casos, los resultados obtenidos por el sistema de NGS 454 GS JUNIOR coinciden para la detección de BRAF con el Kit de APPLIED BIOSYSTEMS, obteniendo una concordancia del 100%.

		JUNIOR		
		POSITIVO	NEGATIVO	
KIT APPLIED	POSITIVO	19	0	19
	NEGATIVO	0	15	15
		19	15	34

Tabla 14: Tabla de contingencia de los resultados obtenidos en la comparación de la mutación BRAF_E15: P.V600E en 454 GS Junior con el Kit de Applied

Se dan como positivos todos los resultados de las mutaciones mayores al 2% de aparición, como se ha visto en el punto 2.3 (Límite de detección). Los resultados negativo se dan como negativos con menos de 1000 lecturas porque el porcentaje de mutación obtenido era del 0.00%.

3.4.2 Comparación de las variaciones de KRAS con el cobas® KRAS Mutation Test

Se analizaron diferentes muestras para la detección de mutaciones de KRAS en los codones 12, 13 y 61 en 454 GS Junior y con el sistema cobas® KRAS Mutation Test, obteniendo los resultados que se muestran en la **Tabla 15**. Al igual que en el caso anterior, el test de COBAS® KRAS MUTATION nos proporciona valores solamente categóricos.

Se aplicó el test exacto de Fisher para muestras pequeñas, y se obtuvo un p-valor = 0.005477. Este resultado nos dice que las variables no son independientes (**Tabla 16**).

Podemos ver que hay resultados, como en el caso de las muestras 1 y 6, que en nuestro sistema nos da como positivo a un 2.34% con 1196 lecturas y 2.91% y 2713

lecturas, respectivamente; mientras que con el test de COBAS® KRAS MUTATION, nos da resultados negativos. También podemos observar las muestras 4 y 23, aunque consideradas negativas con nuestro método, ya que no cumple el criterio para ser positivas ($\geq 2\%$ mutación y más de 1000 lecturas de cobertura), nos aparecen al 1.29% y 1.00% y son detectadas por el test de Cobas para KRAS.

Se puede ver que en la mayoría de los casos los resultados obtenidos coinciden en las dos técnicas para la detección de KRAS. Sólo la muestra 11, que era positiva en nuestro sistema, nos dio un resultado no analizable en el test de COBAS® para KRAS. Se obtuvo una concordancia del 81.25%.

RESULTADO JUNIOR 545					RESULTADO COBAS® KRAS MUTATION TEST	
MUESTRA		LECTURAS*	%			
1	KRAS-E2: p.G12A	1196	2.34	+	Mutación no detectada	-
2	KRAS-E2: p.G13D	1467	4.23	+	Mutación detectada en Codón 12/13	+
3	KRAS-E3: p.Q61L	574	38.15	+	Mutación detectada en Codón 61	+
4	KRAS-E2: p.G12V	928	1.29	-	Mutación detectada en Codón 12/13 y 61	+/+
	KRAS-E3: p.Q61H	1193	23.72	+		
5	KRAS-E2: p.G12V	3351	52.43	+	Mutación detectada en Codón 12/13	+
6	KRAS-E2: p.G13D	2713	2.91	+	Mutación no detectada	-
7	KRAS-E2: p.G12A	2648	30.51	+	Mutación detectada en Codón 12/13	+
8	KRAS-E2: p.G12V	1142	3.59	+	Mutación detectada en Codón 12/13	+
9	KRAS-E2: p.G12D	2067	20.85	+	Mutación detectada en Codón 12/13	+
10	KRAS-E2: p.G12V	1328	5.42	+	Mutación detectada en Codón 12/13	+
11	KRAS-E3: p.Q61H	1259	3.42	+	Invalidado	NA
12	KRAS-E3: p.Q61R	1236	40.29	+	Mutación detectada en Codón 61	+
13	KRAS-E2: p.G12C	1108	2.71	+	Mutación detectada en Codón 12/13	+
14	KRAS-E3: p.Q61K	1957	14.31	+	Mutación detectada en Codón 61	+
15	NEGATIVO	1796 ⁽¹⁾	—	-	Mutación no detectada	-
16	KRAS-E2: p.G12D	961	10.09	+	Mutación detectada en Codón 12/13	+
17	NEGATIVO	1764 ⁽¹⁾	—	-	Mutación no detectada	-
18	NEGATIVO	1217 ⁽¹⁾	—	-	Mutación no detectada	-
19	KRAS-E2: p.G12D	3737	31.34	+	Mutación detectada en Codón 12/13	+
20	KRAS-E2: p.G12D	1445	10.00	+	Mutación detectada en Codón 12/13	+
21	KRAS-E2: p.G12D	1321	5.00	+	Mutación detectada en Codón 12/13	+
22	KRAS-E2: p.G12D	1163	2.00	+	Mutación detectada en Codón 12/13	+
23	KRAS-E2: p.G12D	1075	1.00	-	Mutación detectada en Codón 12/13	+
24	NEGATIVO	1345 ⁽¹⁾	—	-	Mutación no detectada	-

Tabla 15: Comparación de las mutaciones de KRAS en los codones 12/13 y 61 en 454 GS Junior con cobas® KRAS Mutation Test. *Lecturas totales del amplicón que soporta la mutación indicada. (1) El número de lecturas se calculó con la media de las lecturas correspondientes a los fragmentos de interés (exones 2 y 3 de KRAS). NA: no analizable.

		JUNIOR		
		POSITIVO	NEGATIVO	
COBAS	POSITIVO	16	2	18
	NEGATIVO	2	4	6
	NA	1	0	1
		19	6	25

Tabla 16: Tabla de contingencia de los resultados obtenidos en la comparación de las mutaciones de KRAS en los codones 12/13 y 61 en 454 GS Junior con cobas® KRAS Mutation Test. NA: no analizable.

La discordancia obtenida ($\approx 20\%$) se debe probablemente a que nuestro sistema es más sensible que el test cobas®. Con nuestra técnica, en la mayoría de los casos, podemos ver porcentajes bajos de variaciones de *KRAS* con coberturas muy inferiores a 1000 lecturas. En este estudio de concordancia estamos por encima de este valor, lo que lo hace muy sensibles, ya que vemos mutaciones en torno al 1%, que las damos como negativas por el criterio descrito en el límite de detección.

3.4.3 Comparación de las variaciones de NRAS con el método Sanger

Hemos comparado diferentes mutaciones de NRAS en los exones 2, 3 y 4 en 454 GS Junior con el método de Sanger (**Tabla 17**). Se aplicó el test exacto de Fisher para muestras pequeñas, y se obtuvo un p-valor = 0.005784. Resultado que nos indica que las variables no son independientes (**Tabla 18**).

También, como en los casos anteriores, se dan como positivos todos los resultados de las mutaciones iguales o mayores del 2% de mutación.

Podemos ver que hay resultados, como en el caso de las muestras 8, 9, 11, 23, 24, 25, 28, 31 y 32, que en nuestro sistema nos da como positivo a valores inferiores al 20% de aparición de la mutación, mientras que el método de Sanger nos lo da como negativo, esto se debe a la baja sensibilidad que tiene el método de Sanger.

Los resultados obtenidos para la detección de NRAS tienen una concordancia del 62.86%.

MUESTRA	MUTACIÓN	RESULTADO (JUNIOR)			ABI 3730
		LECTURAS*	% POSITIVAS	JUNIOR	
1	NRAS_E2: p.G12C	912	22.26	+	+
2	NRAS_E2: p.G12C	396	15.15	+	+
3	NEGATIVO	831 ⁽¹⁾	—	-	-
4	NRAS_E2: p.G12D	1683	61.26	+	+
5	NRAS_E2: p.G12D	1978	42.57	+	+
6	NRAS_E2: p.G12D	1190	31.43	+	+
7	NRAS_E2: p.G12D	1296	22.07	+	NA
8	NRAS_E2: p.G12D	1053	7.98	+	-
9	NRAS_E2: p.G12D	1136	3.43	+	-
10	NEGATIVO	542 ⁽¹⁾	—	-	-
11	NRAS_E2: p.G12S	593	3.20	+	-
12	NEGATIVO	1100 ⁽¹⁾	—	-	-
13	NRAS_E2: p.G12V	2464	42.25	+	+
14	NEGATIVO	596 ⁽¹⁾	—	-	-
15	NRAS_E2: p.G13D	3648	29.11	+	+
16	NEGATIVO	1664 ⁽¹⁾	—	-	NA
17	NRAS_E2: p.G13R	515	27.96	+	+
18	NRAS_E2: p.G13R	1116	20.88	+	+
19	NEGATIVO	1131 ⁽¹⁾	—	-	-
20	NRAS_E3: p.Q61K	1544	43.39	+	+
21	NRAS_E3: p.Q61K	3946	35.81	+	+
22	NRAS_E3: p.Q61K	1211	27.17	+	+
23	NRAS_E3: p.Q61K	814	11.43	+	-
24	NRAS_E3: p.Q61K	1299	6.93	+	-
25	NRAS_E3: p.Q61K	1108	2.71	+	-
26	NEGATIVO	924 ⁽¹⁾	—	-	-
27	NRAS_E3: p.Q61L	881	18.62	+	NA
28	NRAS_E3: p.Q61L	946	10.99	+	-
29	NEGATIVO	437 ⁽¹⁾	—	-	-
30	NRAS_E3: p.Q61R	1351	32.86	+	+
31	NRAS_E3: p.Q61R	1317	21.72	+	-
32	NRAS_E3: p.Q61R	807	2.73	+	-
33	NEGATIVO	480 ⁽¹⁾	—	-	-
34	NRAS_E4: p.A146V	620	2.74	+	NA
35	NEGATIVO	994 ⁽¹⁾	—	-	-

Tabla 17: Comparación de las mutaciones de NRAS en los exones 2, 3 y 4 en 454 GS Junior con el método de Sanger, para todos los porcentajes de las mutaciones obtenidas. *Lecturas totales del amplicón que soporta la mutación indicada. (1) El número de lecturas se calculó con la media de las lecturas correspondientes a los fragmentos de interés (exones 2, 3 y 4 de NRAS). NA: no analizable.

		JUNIOR		
		POSITIVO	NEGATIVO	
ABI	POSITIVO	13	0	13
	NEGATIVO	9	9	18
	NA	3	1	4
		25	10	35

Tabla 18: Tabla de contingencia de los resultados obtenidos en la comparación de las mutaciones de NRAS en los exones 2, 3 y 4 en 454 GS Junior con el método de Sanger, para todos los porcentajes de las mutaciones obtenidas. NA: no analizable.

Teniendo en cuenta sólo los resultados de las mutaciones mayores o igual al 20%, se observa que el porcentaje de concordancia aumenta al 85.71%. Si se aplica el test exacto de Fisher, y se obtiene un p-valor = 0.000129 (**Tabla 19**). En este caso, hay un aumento significativo de la concordancia entre las dos técnicas.

		JUNIOR		
		POSITIVO	NEGATIVO	
ABI	POSITIVO	12	0	12
	NEGATIVO	1	6	7
	NA	1	1	2
		14	7	21

Tabla 19: Tabla de contingencia de los resultados obtenidos en la comparación de las mutaciones de NRAS en los exones 2, 3 y 4 en 454 GS Junior con el método de Sanger, para porcentajes a partir del 20% de las mutaciones obtenidas. NA: no analizable.

4. Resultados en la participación de los controles de calidad

Como hemos dicho anteriormente, nuestro laboratorio está inscrito en el Programa de Garantía de Calidad de la SEAP. Y se reciben una serie de casos mutados y no mutados para *BRAF*, *KRAS* y *NRAS*. Se han obtenido los siguientes resultados durante los años 2014-2016 (**Tabla 20**).

AÑO	CASOS	RESULTADO ESPERADO	RESULTADO OBTENIDO	OTROS
2014	KRAS-9-1	KRAS: p.G13D	KRAS: p.G13D 22.0% (1894)	
	KRAS-9-2	KRAS: p.G12A	KRAS: p.G12A 28.9% (2057)	
	KRAS-9-3	KRAS wt	KRAS: p.G13D 4.1% (4140)*	
	KRAS-9-4	KRAS wt	No Mutación en KRAS (2097)	BRAF: p.V600E 21.65%(2273)
	BRAF-10-1	BRAF wt	No Mutación en BRAF (1425)	
	BRAF-10-2	BRAF: p.V600E	BRAF: p.V600E 21.5% (2634)	
	BRAF-10-3	BRAF: p.V600E	BRAF: p.V600E 9.0% (2645)	
	BRAF-10-4	BRAF: p.V600E	BRAF: p.V600E 44.1% (1335)	
	K-RAS-10-1	KRAS: p.A146T	KRAS: p.A146T 31,2% (2417)	
	K-RAS-10-2	KRAS wt	No Mutación en KRAS (2073)	
N-RAS-10-3	NRAS wt	No Mutación en NRAS (1905)	KRAS: p.Q61H 25.7% (1979)	
N-RAS-10-4	NRAS: p.Q61K	NRAS: p.Q61K 46,8% (1529)	KRAS: p.Q61H 5.1% (2283)	
2015	K-RAS-11-1	KRAS: p.G12C	KRAS: p.G12C 20,9% (1568)	
	K-RAS-11-2	KRAS wt	No Mutación en KRAS (2165)	NRAS: p.G13V 12.6% (2172)
	N-RAS-11-3	NRAS: p.G13V	NRAS: p.G13V 63.5% (2264)	
	N-RAS-11-4	NRAS wt	No Mutación en NRAS (1999)	KRAS: p.G12S 24,9% (2290)
	BRAF-12-1	BRAF: p.V600E	BRAF: p.V600E 13.2% (840)	
	BRAF-12-2	BRAF wt	No Mutación en BRAF (715)	NRAS: p.P140L 2.6% (1945)
	BRAF-12-3	BRAF: p.V600E	BRAF: p.V600E 18.1% (780)	
	BRAF-12-4	BRAF: p.V600E	BRAF: p.V600E 11.9% (488)	
	N-RAS-12-1	NRAS: p.Q61L	NRAS: p.Q61L 47.6% (1513)	
	K-RAS-12-2	KRAS: p.G13D	KRAS: p.G13D 42.4% (1304)	
K-RAS-12-3	KRAS: p.G12D	KRAS: p.G12D 24.5% (1304)		
N-RAS-12-4	NRAS: p. G12C	NRAS: p. G12C 21.2% (830)		
2016	BRAF-13-1	BRAF: p.V600E	BRAF: p.V600E 27.3% (827)	
	BRAF-13-2	BRAF: p.V600E	BRAF: p.V600E 11.6% (762)	
	BRAF-13-3	BRAF wt	No Mutación en BRAF (594)	KRAS: p.G13D 13.6% (770)
	BRAF-13-4	BRAF wt	No Mutación en BRAF (880)	
	N-K-RAS-13-1	KRAS: p.G13D	KRAS: p.G13D 43.4 % (1082)	
	N-K-RAS-13-2	N-K-RAS wt	No Mutación en KRAS (1272) / NRAS (1001)	
	N-K-RAS-13-3	N-K-RAS wt	No Mutación en NRAS (797)	KRAS: p.G12D 6.2 % (825)*
	N-K-RAS-13-4	NRAS: p.G12D	NRAS: p.G12D 16.7% (815)	
	BRAF-14-1	BRAF: p.V600E	BRAF: p.V600E 27.3% (1027)	
	BRAF-14-2	BRAF: p.V600E	BRAF: p.V600E 38.4% (868)	
BRAF-14-3	BRAF: p.V600E	BRAF: p.V600E 18.8% (976)		
BRAF-14-4	BRAF wt	No Mutación en BRAF (1078)		
N-K-RAS-14-1	NRAS: p.Q61K	NRAS: p.Q61K 37.43% (1181)		
N-K-RAS-14-2	KRAS: p.Q61L	KRAS: p.Q61L 27.03% (1258)	KRAS: p.Q61H 15.42% (1258)**	
N-K-RAS-14-3	N-K-RAS wt	No Mutación en KRAS (1427) /NRAS (1718)	BRAF: p.V600E 20.55% (949)	
N-K-RAS-14-4	KRAS: p.G12D	KRAS: p.G12D 14.5% (1317)		

Tabla 20: Tabla de los resultados obtenidos durante los años 2014-2016 en el Programa de Garantía de Calidad de la SEAP. Entre paréntesis, la cobertura en número de lecturas.
*Discrepancias.

Cabe destacar en la **Tabla 20** los resultados dispares obtenidos: Casos KRAS-9-3 y N-K-RAS-13-3, se esperan resultados negativos mientras encontramos mutaciones al 4.1% (p.G13D) y 6.2% (p.G12D), respectivamente. Caso N-K-RAS-14-2, se espera la mutación p.Q61L de KRAS, pero además de detectar ésta (27.03%), también se detecta la p.Q61H (15.42%).

5. Resultados mutacionales de la población analizada

Como hemos visto en la Metodología, las muestras analizadas proceden de un estudio basado en la población afectada por tumores de CCR, de las Comunidad Valenciana, cuyos individuos han sido seleccionados en diferentes Hospitales de dichas comunidades.

Resultados obtenidos al analizar un estudio de muestras de pacientes diagnosticados de CCR metastásico.

Como podemos ver en la **Tabla 21**, en 875 muestras de pacientes con CCRM analizadas. El 41.7% son negativas; el 56.0% son positivas para los diferentes exones de los genes *BRAF*, *KRAS* y *NRAS* analizados en nuestro sistema. Y el 2.3% restante son no analizables por mala calidad de las muestras, éstas no amplificaban bien en la PCR o no se obtenía la cobertura idónea para su análisis.

Variaciones en el gen KRAS

De todas las muestras analizadas, 398 fueron positivas para mutaciones en gen *KRAS*. En los exones 2, 3 y 4 de este gen se aprecia que el más mutado es el exón 2 con un 39.0%, seguido del exón 3 con un 3.1% y el exón 4 con un 3.4%.

En *KRAS*, las mutaciones que aparecen con más frecuencia son las mutaciones *KRAS_E2: p.G12D* (12.6%) y la *KRAS_E2: p.G12V* (10.5%), seguidas de la *KRAS_E2: p.G13D* (8.0%).

Porcentaje de variaciones encontradas en el exón 15 de BRAF

Se encontraron 58 muestras (6.6%) que fueron positivas para mutaciones el exón 15 de *BRAF*. La variación que aparece con más frecuencia es la *BRAF_E15: p.V600E* (6.1%) y el 0.6% restante pertenece a las variaciones *BRAF_E15: p.D594G* (0.5%) y *BRAF_E15: p.E586K* (0.1%).

Variaciones en el gen NRAS

Del resto de las muestras analizadas, 34 muestras fueron positivas para mutaciones en gen *NRAS*. En los exones 2, 3 y 4 de *NRAS*, se puede ver que tanto el exón 2 (1.9%) como el 3 (1.8%) están igualmente afectados, mientras que el exón 4 (0.1%) nada más se encontró una muestra mutada en este exón. En *NRAS* las variaciones que aparecen con más frecuencia son la *NRAS_E2*: p.G12D y la *NRAS_E3*: p.Q61K (ambas al 1.0%).

GEN	%	EXÓN	%	MUTACIÓN	%
BRAF	6.6	15	6.6	p.V600E	6.1
				p.D594G	0.5
				p.E586K	0.1
KRAS	45.5	2	39.0	p.G12D	12.6
				p.G12V	10.5
				p.G13D	8.0
				p.G12C	4.0
				p.G12A	1.8
				p.G12S	1.0
				p.G12R	0.6
				p.G13C	0.2
				p.G13R	0.1
		p.Q22K	0.1		
		3	3.1	p.Q61H	1.4
				p.Q61K	0.7
				p.Q61R	0.5
				p.Q61L	0.3
				p.G60D	0.1
4	3.4	p.R68S	0.1		
		p.A146T	2.4		
NRAS	3.9	2	1.9	p.A146V	0.8
				p.K117N	0.2
				p.G12D	1.0
				p.G12C	0.2
				p.G12V	0.2
				p.G13R	0.2
		3	1.8	p.G12S	0.1
				p.G13D	0.1
				p.Q61K	1.0
		4	0.1	p.Q61R	0.5
				p.Q61L	0.2
				p.A146V	0.1
NEGATIVO	41.7				
NO ANALIZABLE	2.3				

TABLA 21: Resultados obtenidos en porcentaje tras analizar un estudio de 875 muestras de pacientes diagnosticados de CCRM.

DISCUSIÓN

El gran aumento de conocimientos relacionados con los mecanismos patogénicos del cáncer y la descripción de los procesos moleculares de algunos tumores, ha permitido el desarrollo de nuevos fármacos específicos para determinadas “dianas terapéuticas”. Con la gran cantidad de fármacos antitumorales aprobados para su uso, todos ellos desarrollados para estas dianas pero cuya aplicabilidad está en función de la presencia de diferentes mutaciones. Además, han aparecido mutaciones que confieren resistencia a estos tratamientos tras su aplicación. Todo ello ha hecho necesario, en el entorno clínico, el análisis rutinario para detectar la presencia de estas variaciones genéticas en las muestras oncológicas. El estudio de dichos biomarcadores validados es crucial para la aplicación de los tratamientos personalizados en la práctica clínica. Esta necesidad, junto al avance de la tecnología de secuenciación de última generación, ha permitido y estimulado la detección de estas mutaciones y el desarrollo de sistemas rápidos, fiables y cada vez más económicos (LiVolsi V *et al.*, 2012; Golubnitschaja O *et al.*, 2012).

Problemática de las muestras a analizar y de las condiciones de análisis en la rutina clínica

El análisis molecular de variaciones genéticas de rutina está sujeto a una serie de condicionantes que lo hacen especialmente sensible al desarrollo y optimización:

En primer lugar, habitualmente se dispone de tamaños de muestra muy limitados, especialmente en determinados tipos de tumores, o cuando éstos están localizados en zonas de difícil acceso para la realización de biopsias. Esto resulta frecuentemente en cantidades muy pequeñas de material genético de partida para los análisis.

En segundo lugar, tales muestras contienen un número variable de células tumorales frente a células sanas, y por lo tanto, el contenido en ADN tumoral variará entre muestras. Esto se traduce en que los porcentajes con los que una mutación puede estar presente serán a menudo muy bajos, del orden del 1 al 5% (Kerick M *et al.*, 2011)

Tercero, las biopsias tumorales de las que se dispone rutinariamente para su extracción y análisis del ADN, son muestras fijadas con formalina e incluidas en

parafina (muestras FFPE), lo cual implica inevitablemente un proceso de fragmentación y pérdida de calidad del material genético que se obtendrá a partir de ellas (Hadd AG *et al.*, 2013). Una de las consecuencias del proceso de fijación en formalina de las muestras es la adición de grupos hidroximetilo a las bases del ADN lo que acaba provocando reacciones cruzadas con proteínas. Estas reacciones cruzadas pueden acabar produciendo artefactos en las reacciones de PCR donde se producen transiciones artificiales C>T o G>A (Williams C *et al.*, 1999, Lin MT *et al.*, 2014).

Por último, la rutina clínica, cuyo objetivo es la realización de un análisis rápido para proceder a los tratamientos disponibles en cuanto sea posible, no necesita conocer una cantidad elevada de mutaciones, muchas de las cuales son en realidad mutaciones acompañantes (*passenger mutations*) irrelevantes, en principio, para el tratamiento. Por tanto, no es necesario el análisis de grandes paneles de mutaciones que consume muchos recursos, especialmente en el análisis y almacenamiento informático, y requiere la selección de las mutaciones de utilidad presentes. Por el contrario, tampoco resulta interesante la realización de estos análisis de un modo seriado, focalizando la búsqueda mutación tras mutación, o gen tras gen, mediante técnicas analíticas basadas en genotipados directos, hasta dar con un resultado positivo o haber agotado todas las mutaciones de interés. Este último tipo de abordaje también resulta caro en términos económicos, de cantidad de muestra consumida y de tiempo de respuesta (Croce CM, 2008).

El presente trabajo presenta el diseño, desarrollo y optimización de una técnica para el análisis de este tipo de muestras que tiene como premisas iniciales solventar la problemática descrita. Estas son:

- Plantear un sistema de detección de mutaciones que necesite poco material genético de partida, al comenzar con una única reacción de amplificación por PCR.
- Mantener unos parámetros de sensibilidad y exactitud elevados aún en muestras con calidades bajas.
- Que resulte interesante desde un punto de vista económico y de tiempo de respuesta.

- Que sea escalable, modulable, tanto en el número de muestras a introducir como en cuanto a la introducción de nuevas mutaciones a analizar. Esto último, sólo en el caso que la nueva mutación a estudiar no esté dentro de la secuencia de alguno de los fragmentos amplificados, supondría la introducción de nuevos fragmentos de ADN amplificado, sin que esto redunde en un aumento significativo del precio final por muestra, ni del tiempo de realización de los análisis.

Estas características se consiguen mediante el acoplamiento de una sola reacción de PCR multiplex donde se seleccionan las zonas de ADN a analizar, a una reacción de secuenciación en una plataforma NGS de tamaño medio (454 GS Junior, Roche) para una mayor agilidad y menor coste económico (Xuan J *et al.*, 2013).

Estrategia planteada en el diseño del panel: Selección por PCR

Esta estrategia, basada en un enriquecimiento selectivo mediante amplificación previa por PCR, tiene dos principales ventajas: produce una elevada especificidad, así como una cobertura regular en todos los fragmentos analizados y todos los individuos (Mamanova L *et al.*, 2010). Esto se consigue al ser agrupados tras su cuantificación individual (individual en el sentido de individuo a individuo, no en el sentido de fragmento a fragmento) del primer producto de PCR para su posterior amplificación en la reacción específica de la técnica de secuenciación masiva empleada (emPCR).

Otro aspecto interesante a reseñar es que se optimizó la técnica para obtener todos los fragmentos de cada individuo en una sola reacción, y no en dos reacciones consecutivas como se ha de hacer obligatoriamente cuando se trabaja con paneles de gran número de fragmentos. Este único paso de amplificación supone, además del evidente ahorro económico y de tiempo, la minimización de la posibilidad de contaminación al reducirse a la mitad los procesos de manipulación en la amplificación por PCR.

Diseño de los cebadores y selección de las zonas de ADN a analizar

La metodología seleccionada en la que se basa el panel implica una serie de desafíos que redundan en la necesidad de un periodo de diseño y optimización previos a su introducción en la clínica rutinaria:

En primer lugar, la especificidad final, esto es, el enriquecimiento de los fragmentos de ADN seleccionados desde el ADN genómico obtenido para cada muestra, que depende en gran medida de un correcto diseño de los cebadores para la reacción de PCR. Es preciso evitar la aparición de reacciones de interferencia entre cebadores y la producción de dímeros (fragmentos de ADN resultantes de la amplificación de cebadores que se concatenan al hibridar entre sí), estos últimos especialmente problemáticos cuando el proceso continúa con técnicas de secuenciación masiva. También es necesario evitar la hibridación inespecífica de estos cebadores en sitios distintos de los seleccionados, que produce una menor eficiencia de estos cebadores, pero sobre todo posibles sesgos alélicos (Ikegawa S *et al.*, 2002).

Cuanto mayor sea el número de amplicones seleccionados en el panel, mayor es la dificultad en el diseño de estos cebadores y más compleja es la optimización de las condiciones de amplificación. De hecho, si el número de amplicones es muy elevado (en el caso de paneles muy amplios), la reacción de PCR se ha de preparar con técnicas específicas como PCR microfluídica, *microdroplet* PCR, o PCR en emulsiones. Esto implica mayor tiempo de trabajo, mayor coste económico y mayor tiempo de respuesta (Zhang Y *et al.*, 2016). El panel descrito en el presente trabajo, al contar con siete amplicones, no necesita de una amplificación en dispositivos de este tipo.

Con el objetivo de conseguir un número homogéneo de lecturas en todos los amplicones en la secuenciación, es necesario un proceso de optimización de las condiciones de PCR. Esto supone la prueba de diferentes concentraciones relativas de cada uno de las siete parejas de cebadores, así como de los tiempos y temperaturas ideales en la reacción de amplificación.

Otro aspecto importante a tener en cuenta al realizar el diseño de los cebadores de las reacciones de PCR es el tamaño final de cada amplicón. Al tratarse de muestras provenientes de biopsias, que han sufrido un proceso de parafinización-desparafinización que provoca, entre otras cosas, la fragmentación del ADN genómico, los tamaños finales seleccionados para los amplicones resultan del compromiso entre:

- i) El tamaño máximo que va a permitir una cantidad mínima necesaria de producto de PCR debido a esta fragmentación del ADN genómico molde.
- ii) El tamaño máximo que puede ser leído en el proceso final de secuenciación en el tipo de plataforma NGS utilizada.
- iii) La cantidad de información que se busca dentro del fragmento en cuestión (el número de mutaciones de obligado análisis que debe abarcar el fragmento).
- iv) Una correlación de tamaños en los siete fragmentos para poder hacer un seguimiento de su presencia y homogeneidad, tanto en el proceso de optimización, como una vez introducido el panel en la rutina clínica, permitiendo un punto de seguimiento de la calidad del proceso global mediante su observación en equipos de visualización de ADN.

En nuestro caso, teniendo en cuenta todas estas premisas, se obtuvo un panel de cebadores diseñados para la amplificación de siete fragmentos de PCR individuales y no correlativos en secuencia, de modo que se cubrió todas las mutaciones seleccionadas a priori para el análisis descrito.

Establecimiento del número óptimo de lecturas por cada fragmento:

Cuando se analizan muestras de origen oncológico para evitar falsos positivos, se pueden adoptar dos estrategias alternativas:

Una es adoptar un método de alineamiento altamente restrictivo durante el análisis, de modo que descarte un elevado número de secuencias obtenidas donde se observan variaciones presentes con muy poca frecuencia. Estas variaciones pueden ser artefactos de la PCR, errores de la secuenciación e incluso del mismo sistema de detección de la plataforma de secuenciación. Este abordaje tiene como

consecuencia final la obtención de bajas profundidades de lecturas, lo que redundará en la pérdida de información, dando lugar a posibles falsos negativos (Kerick M *et al.*, 2011).

La estrategia alternativa es la obtención de un elevado número de lecturas, de modo que la presencia de estas variaciones resulta en frecuencias despreciables en el alineamiento y posterior análisis de los resultados. En mutaciones somáticas, se sabe que con un número de lecturas mayor de 80 se excluyen los errores por estos procesos de fijación y desparafinización (Wagle N *et al.*, 2012).

Ante las dos estrategias se eligió la segunda, ya que a mayor número de lecturas se minimiza la pérdida de información. En base a nuestros resultados, observamos que a mayor cobertura la sensibilidad y especificidad analítica aumentan.

Calidad de la secuenciación y rendimientos de la técnica

Establecido un rango de valores de profundidad adecuados para una detección fiable de las variaciones presentes, y puesto que la intención del presente trabajo es proporcionar un sistema integrable en la rutina clínica, el siguiente factor de calidad fue establecer la estabilidad de este sistema en términos de cobertura total por carrera, por muestra y por fragmentos amplificados.

Tras la secuenciación de 30 carreras se obtuvieron buenas calidades en la mayoría de los casos, sólo en 6 carreras el porcentaje de *Dots + Mixed Filter* fue superior al 20%. En la mayoría de los casos el número total de lecturas superó ampliamente el valor esperado. Este valor elevado se debe, probablemente, a la sencillez del panel en cuanto al número total de amplicones. Gracias a esto, el sistema finalmente permite el análisis simultáneo de, al menos 20 muestras consiguiendo un compromiso entre una detección fiable y un interés económico y de aprovechamiento del material de partida (Kuo FC *et al.*, 2017).

Respecto a la cobertura por muestra y por fragmento, las variaciones observadas entre amplicones se explicarían por la diferente calidad entre muestras y por la distinta procesividad de los distintos amplicones, que compiten entre ellos de manera que unos se llevan más lecturas frente a otros que apenas tienen (Simen BB *et al.*, 2015; Deeb KK *et al.*, 2015). Una de las necesidades del proceso de optimización fué, asumiendo la inevitabilidad de esta situación, conseguir que los amplicones menos eficientes superen siempre los parámetros de calidad mínimos.

Consideraciones sobre la validación del método

La validación de un método analítico es un paso fundamental para asegurar la fiabilidad de los resultados obtenidos. Aunque el número de biomarcadores relacionados con el diagnóstico, pronóstico y tratamiento en el campo de la oncología continúa en crecimiento, y en paralelo y a la notable velocidad que se desarrollan diferentes paneles para el análisis de estas mutaciones (Zutter MM *et al.*, 2014), no hay todavía un acuerdo para establecer una guía para la validación de todo este tipo de material analítico por parte de las organizaciones reguladoras internacionales. Por tanto, continúa quedando a la discreción de cada laboratorio el diseño de las pruebas orientadas a la validación de los paneles que desarrolle (Hagemann IS *et al.*, 2014; Ananda G *et al.*, 2015).

En el presente trabajo, el diseño de las pruebas necesarias para la validación se hizo teniendo en cuenta las normas de la ISO 15189 (ISO/IEC 15189, 2012) y en el caso particular la particular de las técnicas de NGS en las que se dan un elevado número de pasos que son únicos en esta tecnología (Gargis AS, *et al.*, 2015). En general, las regulaciones relacionadas con la validación de los paneles para análisis en NGS aconsejan cinco conceptos principales en cuanto al desarrollo de estos paneles, que serían los rendimientos que obtenemos de la técnica, la exactitud, la precisión, la sensibilidad y especificidad analítica, y por último, la comparación de nuestro método con otros métodos validados.

Por otro lado, una vez analizadas estas características, es necesaria también la consideración de una validación del *pipeline* bioinformático, así como un seguimiento

de la calidad de los resultados clínicos a lo largo del tiempo cuando el método se introduce en la rutina clínica. Todas las pruebas descritas en este trabajo se realizaron en nuestro laboratorio.

Aspectos técnicos específicos relacionados con la validación

1. Exactitud

Hay varios factores a lo largo de un flujo de trabajo NGS que influyen en la exactitud de los resultados de la prueba. Para las pruebas en las que las muestras de pacientes son multiplexadas, se toman precauciones para minimizar la posibilidad de que un MID se asocie con la muestra de paciente equivocada. La fiabilidad de la multiplexación está fuertemente influenciada por: el diseño del índice, el método mediante el cual se añaden los índices a cada fragmento y el uso de herramientas de software para demultiplexar lecturas de secuencias.

La evaluación de la exactitud de cada lectura depende de varios factores que incluyen las puntuaciones de la calidad por base leída en la plataforma (Q30: tasa de error inherente de la tecnología NGS), la profundidad de cobertura y el contenido de la secuencia (Gargis AS *et al.*, 2012).

Dentro del contexto particular del tipo de detección de este trabajo, consideramos este parámetro como la cercanía en porcentaje de mutación de nuestra medición a los porcentajes reales presentes en muestras de genotipo conocido (ten Bosch JR *et al.*, 2008), comerciales en este caso. Las muestras comerciales seleccionadas fueron lo más parecido posible a las muestras que se analizan en el entorno de análisis oncológico, muestras provenientes de líneas celulares (esto es, no provenientes de biopsias) pero sí sometidas a un proceso de fijación-parafinización y desparafinización.

La exactitud de nuestra técnica, con una cobertura de 1000 lecturas, en muestras que solo son portadoras de una mutación al 50% fue del 91-99%. En muestras WT para todas las mutaciones del estudio, la exactitud fue del 100%. Y en la muestra comercial, con diferentes porcentajes de las mutaciones, obtenemos exactitudes en torno al 97% para las variaciones BRAF: p.V600E y KRAS: p.G13D. Mientras que

en la variación KRAS: p.G12D era del 77% y para NRAS: p.Q61K es del 51%, respectivamente. Se obtuvo una exactitud media de la técnica del 94% para estas variaciones de frecuencia alélica descritas en la Metodología.

En los valores de exactitud obtenidos se observó que hay diferencias dependiendo de la mutación detectada. El error relativo de muestras comerciales aumenta en función del porcentaje de aparición de la mutación, mayor a bajos porcentajes, y probablemente a la cobertura, y también mayor a bajas coberturas, aunque estas últimas son por la limitación de la propia plataforma que debe mantener un compromiso entre número muestras analizables y coberturas aceptables.

Nuestros resultados de exactitud son equiparables a los obtenidos en un trabajo de 2015, donde la detección de las variantes se determinó por medio un ensayo de diagnóstico molecular basado en NGS: *JAX Cancer Treatment Profile*™ (JAX-CTP™), que detecta mutaciones viables en tumores sólidos. Analizaron doce muestras tumorales clínicas frente a muestras control, en 93 genes en común, en un laboratorio externo. El porcentaje de exactitud obtenido de esta técnica en este trabajo fue del 98% (Ananda G. et al., 2015).

2. Precisión

Para estudiar la precisión de nuestra técnica, desde el punto de vista del grado de concordancia entre mediciones repetidas del mismo material, se estableció la precisión del método en términos de:

- i) Repetibilidad entre muestras dentro de un mismo ensayo (dentro de la misma carrera). Se observó que el CV, en la mayoría de los casos, fue inferior al 20%, indicando que la dispersión entre los resultados obtenidos es aceptable entre réplicas de la misma muestra, preparadas desde el principio por una misma persona y con un mismo lote de reactivos (tanto en preparación de la librería, como en todos los siguientes pasos, hasta llegar a la secuenciación, incluida ésta).

ii) Reproducibilidad de una misma muestra entre diferentes ensayos para evaluar la consistencia de los resultados al probar la misma muestra en diferentes condiciones. En este caso, también se obtuvo un CV inferior al 20% al replicar una serie de muestras a lo largo de varios ensayos con diferentes lotes de reactivos. Sugiriendo, como en el caso de la repetibilidad, que los resultados obtenidos son también aceptables.

Nuestros resultados fueron precisos como los obtenidos en un trabajo donde se estudió la precisión de un panel para análisis de 740 variaciones en 46 genes relacionados con el cáncer usando Ion-Torrent AmpliSeq (Singh RR *et al.*, 2013). La repetibilidad de la técnica se obtuvo a partir del análisis de 1 muestra con 7 mutaciones a porcentajes diferentes y 10 índices distintos en la misma carrera. Y la reproducibilidad, analizando durante 25 carreras una muestra con 3 mutaciones distintas y a diferentes frecuencias alélicas. Obtuvieron en el estudio de los resultados una repetibilidad y reproducibilidad con un CV del 7,16% y 1.36%, respectivamente, frente al 15.05% y 11.24% obtenidos con nuestra técnica. Pero con 5 casos negativos en el llamado de variantes, cosa que en nuestra técnica nunca ocurrió.

3. Sensibilidad y especificidad analítica.

Para establecer el número de lecturas y el mínimo del porcentaje de aparición de mutación idóneos, se estudiaron la sensibilidad y especificidad analítica. Estos se expresaron como la probabilidad de que un ensayo detecte una variación de secuencia cuando está presente dentro de la región genómica que está secuenciado; y como la probabilidad de que el ensayo no detecte la variación de secuencia cuando no esté presente en la región genómica que se analiza, respectivamente (Gargis AS *et al.*, 2015).

Se utilizaron ambas para un conjunto de variaciones de secuencia previstas tal que representasen el espectro de posibles hallazgos clínicamente relevantes mediante la comparación de los resultados de la prueba.

Dado que la naturaleza de la técnica nos permite una aproximación a límites de variación bajos, modificando la profundidad de la zona analizada, se buscó el establecimiento de un intervalo aconsejable en cuanto al número de lecturas como paso previo a todo el resto de pruebas. Este paso se realizó mediante experimentos de dilución de ADN portadores de mutación y porcentaje conocidos con ADN sin mutación hasta alcanzar un valor de porcentaje de mutación presente límite.

Se intentó primero utilizando muestras comerciales (*Horizon Diagnostics, Cambridge, United Kingdom*) en las que la mutación de partida, así como su porcentaje, eran asegurados por el comercializador. Desgraciadamente, las diluciones en ADN no mutado no pudieron ser precisas, dado que las cantidades de muestra que la casa comercial que lo surte son tan bajas que los errores en las cuantificaciones de ADN eran excesivamente elevados, haciendo no fiables los resultados finales (Singh RR *et al.*, 2013). Aun así, estas muestras nos permitieron un primer ensayo de exactitud puesto que los valores de porcentaje de mutación presente en las muestras sin diluir, así como en las primeras diluciones, sí fueron muy aproximados a los valores estipulados por la casa comercial.

A falta de muestras comerciales aceptables, los experimentos de dilución se realizaron con ADN de muestras reales, analizadas anteriormente por técnicas alternativas validadas. Así, se eligieron muestras portadoras de las mutaciones *BRAF/KRAS/NRAS* y muestras no portadoras de ninguna de estas mutaciones, y se realizaron las diluciones como se describe en el apartado metodológico de este trabajo. Los resultados nos permitieron establecer los límites de detección, así como establecer también el límite inferior en cuanto al número de lecturas aconsejable. De este modo se estableció como límite inferior de porcentaje de mutación aceptable un 2% y 1000 lecturas de cobertura al ser el porcentaje en el que se observa un CV por debajo del 20% de concordancia en los genotipados obtenidos frente a muestras portadoras para las tres mutaciones analizadas.

En nuestro caso, asumimos que el número de lecturas ha de ser en general de 1000 en total para descartar que la presencia de una mutación obtenida no se deba a artefactos (de la PCR, errores de la secuenciación, etc...). Esto significa que, en la práctica, ningún resultado observado en fragmentos con menos de 1000 lecturas se

aceptará cuando el porcentaje de la variación detectada sea menor del 2%. El criterio del laboratorio rozando por el límite de detección aceptado, nos obliga a la repetición del fragmento en cuestión. No ocurre lo mismo con los resultados negativos con menos de 1000 lecturas, pues teniendo en cuenta el contexto en el que nos movemos, tenemos un compromiso en términos económicos y de rapidez, haciendo que cuando el porcentaje de mutación es inferior al 0.5%, lo damos como negativo, cuando la cobertura, en número de lecturas, es mayor de 200.

Los experimentos de análisis de dilución de muestras con mutación y porcentaje conocidos en muestras *WT* descritos anteriormente nos permitieron, por otro lado, una acotación del intervalo aproximado del número de lecturas necesarias para que los resultados fueran aceptables a la vez que se puedan analizar un mayor número de muestras. En estos experimentos observamos que valores promedio de coberturas iguales o mayores de 1000 no presentan menor desviación del valor real, ni menores SD. Sin embargo, al analizar en conjunto los resultados para las tres mutaciones seleccionadas en estos experimentos si observamos que con coberturas menores de 1000 lecturas, y especialmente si el porcentaje de mutación es muy bajo, los valores de SD de los porcentajes de mutación en los triplicados eran dudosos, de modo que se estableció un valor de referencia de 1000 lecturas como ideal teniendo en cuenta el compromiso entre calidad de los resultados y número de muestras analizadas en cada carga del equipo.

Los resultados obtenidos, al analizar la sensibilidad analítica de nuestra técnica, son comparables a otros estudios. En uno de ellos, se diluyeron una serie de líneas celulares del 20-1% y obtuvieron un límite de detección del 2%, con una cobertura media de 1.000 lecturas para una la técnica de NGS con el sistema Ion Torrent AmpliSeq (Lin MT *et al.*, 2014). En otros, se obtuvieron sensibilidades analíticas del 100% a una frecuencia alélica del 4% con una cobertura de 1000 lecturas (Simen BB *et al.*, 2015; Haene ND *et al.*, 2015). Aunque con nuestro sistema, a coberturas inferiores y a este porcentaje de mutación, la sensibilidad analítica también fue del 100%. Otros grupos de trabajo hicieron pruebas de validación con muestras reales de diagnóstico, que como en este estudio, las mezclaron y diluyeron para obtener sensibilidades analítica similares a las nuestras (Singh RR *et al.*, 2013; Ananda G *et al.*, 2015).

4. Concordancia con otros métodos

Se comparó el resultado de diferentes mutaciones detectadas con nuestro método de análisis frente a otros métodos validados, midiendo la concordancia de los resultados entre ambos. Se dieron como positivos todos los resultados de las mutaciones mayores o iguales al 2% de aparición de la mutación y con más de 1000 lecturas.

Para la detección de mutaciones en el gen *BRAF*, comparamos frente a una técnica basada en PCR alelo-específica y posterior reacción de minisequenciación que detecta tres variaciones en el codón 600 (V600E, V600A y V600G). Se obtuvo una concordancia del 100%, haciendo visible la dependencia de los resultados por ambos métodos. Estos resultados fueron equiparables a los resultados obtenidos en un trabajo donde se comparó este kit con la plataforma 454 GS FLX Titanium (Lopez-Rios F *et al.*, 2013). Cabe destacar que nuestro sistema proporciona valores porcentuales, mientras esta técnica proporciona valores categóricos. Esto supone un valor añadido al método que se presenta en este trabajo, además de saber si es positivo o negativo, también se puede estimar el porcentaje de mutación presente en la muestra.

Para las mutaciones más comunes en *KRAS* se eligió comparar nuestro sistema frente a una técnica basada en PCR alelo-específica semicuantitativa, *cobas® KRAS Mutation Test*. En este caso, se obtuvo una baja discordancia (19.75%), mostrando que había dependencia entre ambos métodos. Esto se debe probablemente a que nuestro sistema es más sensible que el test *cobas®*. Además, al ser un sistema cerrado, sólo informa de la presencia o ausencia de alguna variación en los codones 12/13 y 61 de *KRAS*, pero sin informar de cuál es exactamente la variación en cuestión.

Cabe destacar en los resultados obtenidos por nuestro sistema: porcentajes de mutación detectados inferiores al 2%, aunque son detectados con una sensibilidad discutida, son informados como negativos, pues no cumple criterios límite de detección. Y porcentajes de mutación detectados iguales o superiores al 2% presentan dos tipos de divergencia: primer tipo, en torno al 2%, son detectados en nuestro sistema pero no detectados en el test de *KRAS* de *cobas*. Esto se explica porque el límite de detección en nuestro sistema es superior al de este test. Y

segundo tipo, porcentajes de mutación detectados superiores al 2%, son detectados por los dos sistemas.

Frente a las dos técnicas anteriores, con nuestro sistema somos capaces de informar sobre los porcentajes de las alteraciones genéticas centrales para el desarrollo del tumor (*drivers*). También, es posible detectar aquellas alteraciones *passengers* presentes en las cercanías de éstas mutaciones *drivers*. Esta información puede servir para futuros estudios prospectivos ya que en el *pipeline* de análisis informático, la presencia de estas variaciones acompañantes, así como sus porcentajes, quedan registrados (Hernández-Losa J *et al.*, 2012, Singh RR *et al.*, 2013).

Se planteó también la comparación de nuestro sistema frente al método de secuenciación de Sanger para variaciones presentes en el gen *NRAS*, por ser éste el verdadero *gold estándar* frente a cualquier técnica de detección de variaciones en el genoma. En este caso, observados los resultados para todo el rango de porcentaje de mutaciones (en nuestra técnica), se obtuvo una concordancia baja. Pero si comparamos sólo los resultados de las mutaciones con porcentaje mayor o igual al 20%, se observa que la concordancia aumenta casi en 25 puntos porcentuales. Esto ocurre porque la sensibilidad de nuestro sistema es mayor que la del método de Sanger, que es muy poco fiable para muestras donde el porcentaje de mutación es menor de un 20% (Simen BB *et al.*, 2015). De este modo, el método presentado en este trabajo de investigación comparte con la secuenciación por el método de Sanger su capacidad para detectar cualquier variación presente en la zona estudiada, pero con una mayor sensibilidad, así como un menor coste en términos de tiempo, muestra y economía.

Tras lo descrito anteriormente, un importante aspecto a tener en cuenta de nuestro sistema respecto a las tres técnicas de referencia ampliamente utilizadas en el entorno clínico es que los resultados son comparables; pero, además, el sistema presentado en esta tesis doctoral supone un evidente ahorro en todos los sentidos gracias a la *multiplexación* y el valor de sensibilidad obtenido.

Como aspectos no positivos a tener en cuenta, este método precisa de una buena formación técnica del personal encargado de los análisis un profundo conocimiento del proceso puramente químico de la técnica para decidir dar por

bueno un cambio detectado en caso de posible artefacto, mediante la observación de los flujogramas obtenidos (Singh RR *et al.*, 2013). Así como una buena preparación bioinformática, dado que para este personal técnico resulta muy aconsejable un conocimiento básico del entorno *Linux*. Estos factores han hecho complicada su implementación en el entorno clínico, aunque por conceptos como los descritos en el párrafo anterior, las técnicas basadas en secuenciación masiva son de evidente necesidad en la rutina clínica actual.

5. Controles de Calidad externos

La participación en controles de calidad externos es un requisito imprescindible para garantizar la calidad de un sistema. Su principal ventaja es que su diseño, implementación y evaluación son completamente independientes, lo que garantiza la transparencia del proceso. En España existe desde el año 2010 un Programa de Control de Calidad en Patología Molecular, coordinado por el Laboratorio de Dianas Terapéuticas, integrado en el Programa de Garantía de Calidad de la SEAP (Hernández-Losa J *et al.*, 2012). Nuestro laboratorio ha participado de forma continuada en ellos, antes incluso de la utilización de este sistema de secuenciación.

Los resultados de los controles de calidad externos obtenidos con nuestro sistema de secuenciación fueron satisfactorios. Aunque encontramos resultados positivos a bajo porcentaje y con más de 1000 lecturas donde se esperaban resultados negativos. Esto se debe probablemente a que nuestra técnica es más sensible y abierta en comparación con la técnica utilizada para la caracterización de los controles. Además de ver la mutación esperada, también fuimos capaces de ver otras que con la técnica de caracterización de los controles no se detectaron.

Resultados mutacionales de la población estudiada

Se realizó un análisis de la población en pacientes con CCRM de nuestra Comunidad, buscando la detección de mutaciones en los genes *BRAF/KRAS/NRAS*. Obteniendo como resultado una estratificación de la población en función de la presencia de estas variaciones en las biopsias correspondientes. El análisis global

de estos datos muestra una coincidencia con datos obtenidos en poblaciones equiparables (Sorbye H *et al.*, 2015); así como con los resultados, tanto generales como parciales (en cuanto a cada nodo participante), obtenidos en un estudio donde participamos con los datos obtenidos en el presente trabajo, y cuyo objetivo fue determinar el estado mutacional de *RAS* en pacientes con CCRM a nivel nacional (Hernández-Losa J *et al.*, 2015). Esta concordancia en los resultados es otro factor que apunta a la fiabilidad de nuestro sistema.

Como hemos visto, aproximadamente el 40% de los pacientes con cáncer colorrectal metastásico tienen el gen *KRAS* normal o sin mutación, lo que los hace candidatos a recibir tratamientos anti-EFGR. Sobre este grupo de pacientes, cuando se analiza *NRAS* y *BRAF*, las mutaciones son menos frecuentes.

Estudios realizados con anti-EFGR, panitumumab, en combinación con FOLFOX, revelaron que entre el 17 y el 22% de los pacientes con mutaciones en *KRAS* también tenían otras mutaciones *RAS*. El beneficio para el paciente se demostró superior en pacientes sin mutaciones *RAS*, incluso llegando a señalar la falta de eficacia y/o efectos nocivos cuando se administraba en pacientes con *RAS* mutado, de ahí la importancia de conocer este dato de manera precisa (Douillard JY *et al.*, 2015).

Ante las diferentes opciones disponibles, advierten que la eficacia de los tratamientos disminuye cuando su aplicación es posterior a otros (lo que se conoce como segunda o tercera línea). Según este estudio, el impacto del tratamiento cuando se aplica en primera línea puede suponer hasta 8 meses adicionales para la supervivencia de un paciente. Por todo ello, un mayor estudio de las mutaciones *RAS* y de otros potenciales biomarcadores permitirá una mayor precisión en la determinación genética y, como consecuencia, la aplicación del tratamiento más efectivo para la supervivencia del paciente. (Rivera F *et al.*, 2017).

Nuestros hallazgos muestran la presencia de mutación en el gen *BRAF* en 6.6% de las muestras estudiadas. En otras poblaciones estudiadas, la presencia de mutación en el gen *BRAF* está en torno al 8.5% (Yaeger R *et al.*, 2018). La importancia de la determinación de mutaciones en el gen *BRAF* no sólo radica en su valor como factor pronóstico. En el estudio LBA-07 en pacientes afectados por CCRM se ha demostrado que una combinación de anticuerpos anti-EGFR e

inhibidores de *BRAF* son capaces de bloquear la transducción de la señal en sentido desde el EGFR hacia proteína cinasa activada por mitógeno (MEK) en los casos con *BRAF* mutado (Kassouf E *et al.*, 2016).

En la actualidad, algunas recomendaciones para el cáncer colorrectal avanzado sugieren establecer el status de RAS/*BRAF*. Otros genes que pueden participar en la resistencia a la terapia anti-EGFR, como los de la vía *PIK3CA-AKT-mTOR*, están en estudio. También se han demostrado diferencias en la frecuencia de mutaciones de *BRAF* en distintas etnias, por lo que sería importante determinar su status en nuestra población originaria (Kassouf E *et al.*, 2016).

La selección de pacientes que obtendrán un mayor beneficio del tratamiento por encima de otros, permite evitar toxicidades innecesarias, sin olvidar los criterios de optimización de recursos que aporta al sistema. Por eso, programas para determinar el estado mutacional de RAS/*BRAF*, ayudan a seleccionar y, por tanto, a tratar adecuadamente a todos los pacientes con CCRM, cubren una necesidad social de responsabilidad con el paciente. La realización del estado mutacional de RAS/*BRAF* resulta coste-efectiva y permite optimizar los recursos del Sistema Nacional de Salud (Van Cutsem E *et al.*, 2016).

Actualmente, la clasificación del CCR incluye la determinación de una serie de marcadores moleculares como factores pronósticos y/o predictivos de respuesta a terapia dirigida. Entre los genes candidatos a esta clasificación se encuentran, además de los estudiados en este trabajo, el estudio de la inestabilidad microsatélites, *Tp53*, *PIK3CA*, *AKT*, *mTOR* y *PTEN* entre los más frecuentemente señalados por los estudios publicados. La implementación de estas determinaciones son una necesidad actual y un elemento fundamental en la elección de la terapia dirigida a los portadores de esta frecuente enfermedad tumoral (Yaeger R *et al.*, 2018).

A modo resumen final, el presente panel al ser escalable y modulable permite la introducción de un número mayor de muestras, como en cuanto a la introducción de nuevas mutaciones a analizar, sin que esto redunde en un aumento significativo del precio final por muestra, ni del tiempo de realización de los análisis. También, el proceso de desarrollo y validación de este panel ha propiciado el desarrollo de otros paneles con más genes y variaciones a detectar en la misma plataforma. De hecho,

además del panel descrito en este trabajo, se desarrollaron otros paneles similares donde era posible detectar variaciones en los genes *EGFR*, *AKT1* y *PIK3CA*. De la misma manera, este proceso de desarrollo y validación ha servido para el desarrollo rápido de otros paneles de mayor complejidad en otra plataforma de secuenciación. En definitiva, el presente trabajo permitirá también agilizar la implementación del estudio mutacional de nuevos genes de interés.

CONCLUSIONES

Las conclusiones del presente trabajo son:

- Se ha desarrollado un sistema de detección rápido, sensible, reproducible y fácilmente integrable en la rutina clínica, para la detección de mutaciones de interés oncológico, a partir de biopsias incluidas en parafina de pacientes con carcinoma colorrectal metastásico.
- Este sistema de detección de mutaciones de interés oncológico ha sido validado utilizando las normas aplicables, a este tipo de metodología, de la ISO 15189: 2012 para determinar la idoneidad de esta técnica en el ámbito clínico.
- Mediante el uso de este panel se pueden detectar mutaciones como mínimo al 2% en exón 15 del gen *BRAF* y en los exones 2, 3 y 4 de los genes *KRAS* y *NRAS*.
- Con los datos generales obtenidos por medio de nuestra técnica, en función de las mutaciones halladas, se han utilizado para participar en un estudio cuyo objetivo ha sido determinar el estado mutacional de *RAS* en pacientes con CCRM a nivel nacional (Hernández-Losa J *et al.*, 2015).
- Los procesos del diseño y validación de nuestro sistema de detección de mutaciones de interés oncológico, permite el desarrollo de paneles similares en otras plataformas de secuenciación de nueva generación.

BIBLIOGRAFÍA

1. 454 Sequencing System Software Manual, Version 2.7, Part B: GS Run Processor, GS Reporter, GS Run Browser, GS Support Tool. March 2012.
2. Ahmadian A, Ehn M, Hober S Pyrosequencing: history, biochemistry and future. *Clin Chim Acta*. 2006;363:83–94
3. Amado RG, Wolf F, Peeters M, Van Cutsen E, Siena S, Freeman DJ, et al. Wild-type KRAS is required for panitumumab efficacy in patients with metastatic colorectal cancer. *J. Clin. Oncol*. 2008. 2008; 26:1626-34.
4. Ambardar S, Gupta R, Trakroo D, Lal R, Vakhlu J. High Throughput Sequencing: An Overview of Sequencing Chemistry. *Indian J Microbiol*. 2016 Dec;56(4):394-404.
5. Ananda G, Mockus S, Lundquist M, Spotlow V, Simons A, Mitchell T, Stafford G, Philip V, Stearns T, Srivastava A, Barter M, Rowe L, Malcolm J, Bult C, Karuturi RK, Rasmussen K, Hinerfeld D. Development and validation of the JAX Cancer Treatment Profile™ for detection of clinically actionable mutations in solid tumors. *Exp Mol Pathol*. 2015 Feb;98(1):106-12.
6. Applied Biosystems (2010) BRAF mutation analysis reagents. Foster City, CA, USA. Applied Biosystems.
7. Applied Biosystems SOLiD® 4 System, INSTRUMENT OPERATION GUIDE, 2012.
8. Arnold M, Sierra MS, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global patterns and trends in colorectal cancer incidence and mortality. *Gut*. 2017 Apr;66(4):683-691.
9. Bardeesy N, Kim M, Xu J, Kim RS, Shen Q, Bosenberg MW, et al. Role of epidermal growth factor receptor signaling in RAS-driven melanoma. *Mol. Cell. Biol*. 2005; 25: 4176-88.
10. Beeram M, Patnaik A, Rowinsky EK. Raf: a strategic target for therapeutic development against cancer. *J Clin Oncol*. 2005 Sep. 20;23(27):6771-90. Review.
11. Bokemeyer C, Bondarenko I, Hartmann JT, De Braud FG, Schuch G, Zobel A, et al. Biomarkers predictive for outcome in patients with metastatic colorectal cancer (mCRC) treated with first-line FOLFOX4 plus or minus cetuximab:

- updated data from the OPUS study. American Society of Clinical Oncology (ASCO) Gastrointestinal Cancer Symposium 2010. Proceedings of ASCO GI 2010. 2010. 22-1-2010. Abstract 428.
12. Bokemeyer C, Bondarenko I, Makhson A, Hartmann JT, Aparicio J, de BF, et al. Fluorouracil, leucovorin, and oxaliplatin with and without cetuximab in the first-line treatment of metastatic colorectal cancer. *J. Clin. Oncol.* 2009; 27: 663-71.
 13. Braslavsky, I., Hebert, H., Kartalov, E. and Quake, S.R. «Sequence information can be obtained from single DNA molecules». *Proceedings of the National Academy of Sciences of the United States of America* (2003) 100: 3960–3964.
 14. Chen B, Gagnon M, Shahangian S, Anderson NL, Howerton DA, Boone JD. Good laboratory practices for molecular genetic testing for heritable diseases and conditions, *MMWR Recomm Rep* 2009; 58(RR-6): 1-37.
 15. Chen YJ, Roller EE, Huang X. DNA sequencing by denaturation: experimental proof of concept with an integrated fluidic device. *Lab Chip.* 2010 May 7; 10(9):1153-9.
 16. Church G.M. Genomes for ALL. *Scientific American* 2006; 294 (1): 47–54.
 17. CLSI. [CLSI document MM17-A]. Verification and validation of multiplex nucleic acid assays; approved guideline. Wayne, PA: Clinical and Laboratory Standards Institute; 2008.
 18. Cobas® KRAS Mutation Test (Roche), 2014
 19. Collins FS, Morgan M and Patrinos A. The Human Genome Project: lessons from large-scale biology. *Science*, vol. 300, no. 5617, pp. 286–290, 2003.
 20. Cortés-Reyes E, Rubio-Romero JA, Gaitán-Duarte H. Métodos estadísticos de evaluación de la concordancia y la reproducibilidad de pruebas diagnósticas. *Revista Colombiana de Obstetricia y Ginecología.* 2010; Vol. 61 N° 3: 247-255.
 21. Cripps C, Gill S, Ahmed S, Colwell B, Dowden S, Kennecke H, et al. Consensus recommendations for the use of anti-EGFR therapies in metastatic colorectal cancer. *Curr. Oncol.* 2010; 21:39-45.
 22. Croce CM. Oncogenes and cancer. *N Engl Med* 2008; 358: 502-11.

23. Cunningham D, Atkin W, Ilenz HJ, Lynch HT, Minsky B, Nordlinger B, et al. Colorrectal cancer. *Lancet*. 2010; 375: 1030-47.
24. Cunningham D, Humblet Y, Siena S, Khayat D, Bleiberg H, Santoro A, et al. Cetuximab monotherapy and cetuximab plus iriotecan in iriotecan-refractory metastatic colorectal cancer. *N. Engl. J. Med.* 2004; 351:337-45.
25. Dane. Metodología estadística de diseño y estimación de la muestra cocensal, Censo General 2005, 2006.
26. de Macêdo MP, de Melo FM, Lisboa BC, Andrade LD, de Souza Begnami M Junior SA, Ribeiro HS, Soares FA, Carraro DM, da Cunha IW. KRAS gene mutation in a series of unselected colorectal carcinoma patients with prognostic morphological correlations: a pyrosequencing method improved by nested PCR. *Exp Mol Pathol*. 2015 Jun;98(3):563-7
27. De Roock W, Jonker DJ, Di Nicolantonio F, Sartore-Bianchi A, Tu D, Siena S, et al. Association of KRAS p.G13D mutation with outcome in patients with chemotherapy-refractory metastatic colorectal cancer treated with cetuximab. *JAMA*. 2010; 304: 1812-20.
28. Deeb KK, Hohman CM, Risch NF, Metzger DJ, Starostik P. Routine Clinical Mutation Profiling of Non-Small Cell Lung Cancer Using Next-Generation Sequencing. *Arch Pathol Lab Med*. 2015 Jul;139(7):913-21.
29. D'Haene N, Le Mercier M, De Nève N, Blanchard O, Delaunoy M, El Housni H, Dessars B, Heimann P, Remmelink M, Demetter P, Tejpar S, Salmon I. Clinical Validation of Targeted Next Generation Sequencing for Colon and Lung Cancers. *PLoS One*. 2015 Sep 14;10(9): e0138245.
30. Douillard JY, Siena S, Peeters M, Koukakis R, Terwey JH, Tabernero J. Impact of early tumour shrinkage and resection on outcomes in patients with wild-type RAS metastatic colorectal cancer. *Eur J Cancer*. 2015 Jul;51(10):1231-42.
31. Edlundh-Rose E, Egyhazi S, Omholt K, Mansson-Brahme E, Platz A, Hansson J, et al. NRAS and BRAF mutations in melanoma tumours in relation to clinical characteristics: A study based on mutation screening by pyrosequencing. *Melanoma Res*. 2006; 16: 471-8.

32. Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, Bibillo A, Bjornson K, Chaudhuri B, Christians F, Cicero R, Clark S, Dalal R, Dewinter A, Dixon J, Foquet M, Gaertner A, Hardenbol P, Heiner C, Hester K, Holden D, Kearns G, Kong X, Kuse R, Lacroix Y, Lin S, Lundquist P, Ma C, Marks P, Maxham M, Murphy D, Park I, Pham T, Phillips M, Roy J, Sebra R, Shen G, Sorenson J, Tomaney A, Travers K, Trulson M, Vieceli J, Wegener J, Wu D, Yang A, Zaccarin D, Zhao P, Zhong F, Korlach J, Turner S. Real-time DNA sequencing from single polymerase molecules. *Science*. 2016 Jan 2;323(5910):133-8.
33. emPCR Amplification Method Manual - Lib-A, GS Junior Titanium Series, 2012.
34. Fearnheard NS, Britton MD, Bodmer WF. The ABC of APC. *Hum. Mod. Gene*. 2001 Apr; 10(7): 721-33. Review.
35. Folprecht G, Gruenberger T, Bechstein WO, Raab HR, Lordick F, Hartmann JT, et al. Tumour response and secondary resectability of colorectal liver metastases following neoadjuvant chemotherapy with cetuximab: the CELIM randomised phase 2 trial. *Lancet Oncol*. 2010; 11: 38-47.
36. Gargis AS, Kalman L, Berry MW, Bick DP, Dimmock DP, Hambuch T, et al. Assuring the quality of next-generation sequencing in clinical laboratory practice. *Nat Biotechnol* 2012; 30:1033_6.
37. Gargis AS, Kalman L, Lubin IM. Chapter 21 - Assay Validation*. In: Kulkarni S, Pfeifer J, editors. *Clinical Genomics* Boston: Academic Press; 2015. p. 363-376.
38. Garnett MJ, Rana S, Paterson H, Barford D, Marais R. Wild-type and mutant B-RAF activate C-RAF through distinct mechanisms involving heterodimerization. *Mol. Cell*. 2005; 20: 963-9.
39. General Assembly of the World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *J Am Coll Dent*. 2014 Summer;81(3):14-8.
40. Golubnitschaja O, Costigliola V. General report and amop; recomendation in predictive and peronalised medicine 2012: White paper of the European

- Association for Predictive, Preventive and Personalised Medicine. *EMPA J.* 2012 Nov. 1; 3 (1): 14.
41. Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet.* 2016 May 17;17(6):333-51
42. Grady WM, Carethers JM. Genomic and epigenetic instability in colorectal cancer pathogenesis. *Gastroenterology.* 2008; 135: 1079-99.
43. Guillermo Borga H, Bárbara Martínez A, José Maldonado R. Frecuencia mutacional KRAS-NRAS en cancer de colon metastásico/ Recaída implicaciones en la supervivencia. *Rev Venez Oncol* 2017;29(1):15-2.
44. Guinney J, Dienstmann R, Wang X, de Reyniès A, Schlicker A, Soneson C, Marisa L, Roepman P, Nyamundanda G, Angelino P, Bot BM, Morris JS, Simon IM, Gerster S, Fessler E, De Sousa E Melo F, Missiaglia E, Ramay H, Barras D, Homicsko K, Maru D, Manyam GC, Broom B, Boige V, Perez-Villamil B, Laderas T, Salazar R, Gray JW, Hanahan D, Taberero J, Bernards R, Friend SH, Laurent-Puig P, Medema JP, Sadanandam A, Wessels L, Delorenzi M, Kopetz S, Vermeulen L, Tejpar S. The consensus molecular subtypes of colorectal cancer. *Nature Medicine.* 2015 Nov;21(11):1350-6.
45. Gut IG. DNA analysis by MALDI-TOF mass spectrometry. *Hum Mutat.* 2004 May;23(5):437-41.
46. Hadd AG, Houghton J, Choudhary A, Sah S, Chen L, Marko AC, Sanford T, Buddavarapu K, Krosting J, Garmire L, Wylie D, Shinde R, Beaudenon S, Alexander EK, Mambo E, Adai AT, Latham GJ. Targeted, high-depth, next-generation sequencing of cancer genes in formalin-fixed, paraffin-embedded and fine-needle aspiration tumor specimens. *J. Mol. Diagn.* 2013 Mar; 15(2): 234-47.
47. Hagemann IS, Cottrell CE, Lockwood CM. Design of targeted, capture-based, next generation sequencing tests for precision cancer therapy. *Cancer Genet.* 2013 Dec;206(12):420-31.
48. Hall N. Advanced sequencing technologies and their wider impact in microbiology. *The Journal of Experimental Biology.* 2007; 209: 1518–1525.

49. Hanahan D, Weinberg RA. The hallmarks of cancer. *Cell*. 2000; 100: 57-70.
50. Hernández-Losa J, Landolfi S, Ramón y Cajal S, Caldés T, Bando I, Llovet P, Freire J, Pereda S, García-Berbel P, Patiño-García A, Salgado J, Abdulkader I, Sánchez M, Chaves F J, Blesa S, Seda E, Calzado V. Experiencia de la plataforma NK-RAS en la determinación del estado mutacional de *RAS* (*KRAS/NRAS*) en pacientes con cáncer colorrectal metastásico. Congreso Sociedad Española de Anatomía Patológica 2015.
51. Hernández-Losa J, Sanz J, Landolfi S, López-Ríos F, Palacios J, Bautista MD, Díaz-Rubio E, Tabernero J, García Foncillas J y Ramón y Cajal S. Recomendaciones para la determinación de mutaciones K-RAS en cáncer de colon. *Rev. Esp. Patol*. 2012; 45(2): 76-8
52. <http://bioinfo.ut.ee/primer3-0.4.0/>.
53. <https://genome.ucsc.edu>
54. Ikegawa S, Mabuchi A, Ogawa M, Ikeda T. Allele-specific PCR amplification due to sequence identity between a PCR primer and an amplicon: is direct sequencing so reliable? *Hum Genet*. 2002 Jun; 110(6): 606-8.
55. ISO/IEC 15189. Medical Laboratories—particular requirements for quality and competence. Geneva: International Organization for Standardization; 2012.
56. ISO/IEC 3534-1. Statistics—vocabulary and symbols—Part 1: general statistical terms and terms used in probability. Geneva: International Organization for Standardization, 2006.
57. Johnson B, Cooke L, Mahadevan D. Next generation sequencing identifies 'interactome' signatures in relapsed and refractory metastatic colorectal cancer. *J Gastrointest Oncol*. 2017 Feb;8(1):20-31
58. Karapetis CS, Khambata-Ford S, Jonker DJ, O'Callaghan CJ, Tu D, Tebbutt NC, et al. KRAS mutations and benefit from cetuximab in advanced colorectal cancer. *N. Engl. J. Med*. 2008; 359: 1757-65.
59. Kassouf E, Tabchi S, Tehfe M. Anti-EGFR Therapy for Metastatic Colorectal Cancer in the Era of Extended RAS Gene Mutational Analysis. *BioDrugs*. 2016 Apr;30(2):95-104.
60. Kerick M, Isau M, Timmermann B, Sültmann H, Herwig R, Krobitch S, Schaefer G, Verdorfer I, Bartsch G, Klocker H, Lehrach H, Schweiger MR.

- Targeted high throughput sequencing in clinical cancer settings: formaldehyde fixed-paraffin embedded (FFPE) tumor tissues, input amount and tumor heterogeneity. *BMC Med Genomics*. 2011 Sep 29;4:68.
61. King C, Scott-Horton T. Pyrosequencing: a simple method for accurate genotyping. *J Vis Exp*. 2008 Jan 8;(11). pii: 630.
62. Kohne C, Rougier P, Stroh C, Schlichting M, Bokemeyer C, Van Cutsem E. Cetuximab with chemotherapy (CT) as first-line treatment for metastatic colorectal cancer (mCRC): a metaanalysis of the CRYSTAL and OPUS studies according to KRAS and BRAF mutation status. *American Society of Clinical Oncology (ASCO) Gastrointestinal Cancer Symposium 2010. Proceedings of ASCO GI 2010*. 2010. 22-1-201. Abstract 406.
63. Koressaar T, Remes M. Enhancements and modifications of primer design program Primer3 *Bioinformatics*. 2007; 23(10):1289-91.
64. Kruglyak KM, Lin E, Ong FS. Next-Generation Sequencing and Applications to the Diagnosis and Treatment of Lung Cancer. *Adv Exp Med Biol*. 2016;890:123-36.
65. Kuo FC, Mar BG, Lindsley RC, Lindeman NI. The relative utilities of genome-wide, gene panel, and individual gene sequencing in clinical practice. *Blood*. 2017 Jul 27;130(4):433-439.
66. Łaczmańska I, Pesz K, Łaczmański Ł. Application of selected methods based on the polymerase chain reaction in medical molecular diagnostics. *Adv Clin Exp Med*, 2009
67. Łaczmański Ł, Ślęzak R, Karpiński P, Kolačkov K, Lebioda A, Milewicz A. Validation of the minisequencing method for detection of G1691A (Leiden) factor V mutation. *Gynecol Endocrinol*. 2013 Apr;29(4):319-22.
68. Li J, Makrigiorgos GM COLD-PCR: a new platform for highly improved mutation detection in cancer and genetic testing. *Biochem Soc Trans*. 2009 Apr; 37(Pt 2):427-32.
69. Lin J, Gu Y, Du R, Deng M, Lu Y, Ding Y. *Int J Clin Exp Pathol*. Detection of EGFR mutation in supernatant, cell pellets of pleural effusion and tumor tissues from non-small cell lung cancer patients by high resolution melting analysis and sequencing. *eCollection 2014*. 2014 Dec 1;7 (12):8813-22.

70. Lin MT, Mosier SL, Thiess M, Beierl KF, Debeljak M, Tseng LH, Chen G, Yegnasubramanian S, Ho H, Cope L, Wheelan SJ, Gocke CD, Eshleman JR. Clinical validation of KRAS, BRAF, and EGFR mutation detection using next-generation sequencing. *Am J Clin Pathol.* 2014 Jun;141(6):856-66.
71. Liu L, Li Y, Li S, Hu N, He Y, Pong R, Lin D, Lu L, Law M. Comparison of next-generation sequencing systems. *J Biomed Biotechnol.* 2012; 2012:251364
72. LiVolsi V, Upton MP. The anatomic pathologist meets molecular pathology. *Am. J. Clin. Pathol.* 2012; 137: 341-2
73. Logan CW, Nusse R. The WNT signaling pathway in development and disease. *Cell. Rev. Bio.* 2004, 20: 781-810.
74. Lopez-Rios F, Angulo B, Gomez B, Mair D, Martinez R, Conde E, Shieh F, Vaks J, Langland R, Lawrence HJ, de Castro DG. Comparison of testing methods for the detection of BRAF V600E mutations in malignant melanoma: pre-approval validation study of the companion diagnostic test for vemurafenib. *PLoS One.* 2013;8(1): e53733
75. Loupakis F, Moretto R, Aprile G, Muntoni M, Cremolini C, Iacono D, Casagrande M, Ferrari L, Salvatore L, Schirripa M, Rossini D, De Maglio G, Fasola G, Calvetti L, Pilotto S, Carbognin L, Fontanini G, Tortora G, Falcone A, Sperduti I, Bria E. Clinico-pathological nomogram for predicting BRAF mutational status of metastatic colorectal cancer. *Br J Cancer.* 2016 Jan 12;114(1):30-6.
76. Lu H, Giordano F, Ning Z. Oxford Nanopore MinION Sequencing and Genome Assembly. *Genomics Proteomics Bioinformatics.* 2016 Oct;14(5):265-279.
77. Malumbres M, Barbacid m. Ras oncogenes: The first 30 years. *Nat. Rev. Cancer.* 2003; 3:459-65.
78. Mamanova L, Coffey AJ, Scott CE, Kozarewa I, Turner EH, Kumar A, Howard E, Shendure J, Turner DJ. Target-enrichment strategies for next-generation sequencing. *Nat Methods.* 2010 Feb;7(2):111-8.
79. Mardis ER. Next-generation DNA sequencing methods. *Annu Rev Genomics Hum Genet.* 2008b

-
80. Mardis ER. The impact of next-generation sequencing technology on genetics. *Trends Genet.* 2008 Mar;24(3):133-41
81. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM. «Genome sequencing in microfabricated high-density picolitre reactors». *Nature* (2005) 437: 376–380.
82. Maxam AM, Gilbert W. A new method for sequencing DNA. *PNAS* 74 (1977) 560-564
83. McCormick F. KRAS as a Therapeutic Target. *Clin. Cancer Res.* 2015 Apr 15;21(8):1797-801.
84. Milbury CA, Li J, Makrigiorgos GM. PCR-based methods for the enrichment of minority alleles and mutations. *Clin Chem.* 2009 Apr; 55(4):632-40.
85. Morey M, Fernández-Marmiesse A, Castiñeiras D, Fraga JM, Couce ML, Cocho JA. A glimpse into past, present, and future DNA sequencing. *Mol Genet Metab.* 2013 Sep-Oct;110(1-2):3-24
86. Morlan J, Baker J, Sinicropi D. *PLoS One.* 2009; 4(2) Epub 2009 Feb 25. Mutation detection by real-time PCR: a simple, robust and highly selective method.
87. MRC-Holland MLPA. Salsa MLPA probemix P520-A2 MPN mix2. Description Versión 6, September 2017.
88. Nair P. Epidermal growth factor receptor family and its role in cancer progression. *Current Science.* 2005, 88(6): 890-898; March 25.
89. Ochenduazko SL, Krzemieniecki K. Targeted therapy in advanced colorectal cancer: More data, more questions. *Anticancer Drugs.* 2010; 21: 737-48.

-
90. Olson MT, Harrington C, Beierl K, Chen G, Thies M, O'Neill A, Taube JM, Zeiger MA, Lin MT, Eshleman JR. BRAF pyrosequencing analysis aided by a lookup table. *Am J Clin Pathol*. 2014 May;141(5):639-47
 91. Papke B, Der CJ. Drugging RAS: Know the enemy. *Science*. 2017 Mar 17;355(6330):1158-1163.
 92. pct.mdanderson.org.
 93. Peeters M, Price TJ, Cervantes A, et al. Final results from a randomized phase 3 study of FOLFIRI {+/-} panitumumab for second-line treatment of metastatic colorectal cancer. *Ann Oncol*. 2014 Jan; 25(1):107-16
 94. Peeters M, Price TJ, Cervantes A, Sobrero AF, Ducreux M, Hotko Y, et al. Randomized phase III study of panitumumab with fluorouracil, leucovorin and irinotecan (FOLFIRI) compared with FOLFIRI alone as second-line treatment in patients with metastatic colorectal cancer. *J. Clin. Oncol*. 2010; 28:4706-13.
 95. Perkel J. Making contact with sequencing's fourth generation. *Biotechniques*. 2011 Feb;50(2):93-5.
 96. Popal S, Houlston RS. A systematic review and meta-analysis of the relationship between chromosome 18q genotype, DCC status and colorectal cancer prognosis. *Eur. J. Cancer*. 2005; 41(14): 2060-70.
 97. Pritchard CC, Grady WM. Colorectal cancer molecular biology moves into clinical practice. *Gut*. 2011; 60:116-29.
 98. Rivera F, Karthaus M, Hecht JR, Sevilla I, Forget F, Fasola G, Canon JL, Guan X, Demonty G, Schwartzberg LS. Final analysis of the randomised PEAK trial: overall survival and tumour responses during first-line treatment with mFOLFOX6 plus either panitumumab or bevacizumab in patients with metastatic colorectal carcinoma. *Int J Colorectal Dis*. 2017 Apr 19.
 99. Rizzo S, Bronte G, Fanale D, Corsini L, Silvestris N, Santini D, et al. Prognostic vs. predictive molecular biomarkers in colorectal cancer: is KRAS and BRAF wild type status required for anti-EGFR therapy? *Cancer Treat Rev* 2010; 36 Suppl 3: S56-S61.
 100. Ross JS, Cronin M. Am J Clin Pathol. 9 Whole cancer genome sequencing by next-generation methods. 2011 Oct 136(4):527-3.

101. Saltz LB, Lenz HJ, Kindler HL, Hochrter HS, Wadler S, Hoff PM, et al. Randomized phase II trial of cetuximab, bevacizumab alone in irinotecan-refractory colorectal cancer: The BOND-2 study. *J. Clin. Oncol.* 2007; 25: 4557-61.
102. Saltz LB, Meropol NJ, Loehrer Sr PJ, Needle MN, Kopit J, Mayer RJ. Phase II trial of cetuximab in patients with refractory colorectal cancer that expresses the epidermal growth factor receptor. *J. Clin. Oncol.* 2004; 22: 1201-8.
103. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A.* 1977 Dec; 74(12):5463-7.
104. Schaffer, Amanda; Reyes, Francisco. (mayo, junio de 2012). «TR10: Secuenciación de nanoporos»
105. Shendure J., Porreca G.J., Reppas N.B., Lin X., McCutcheon J.P., Rosenbaum A.M., Wang M.D., Zhang K., Mitra R.D. and Church G.M.. «Accurate Multiplex Polony Sequencing of an Evolved Bacterial Genome». *Science* (2005) 309 (5741): 1728–1732.
106. Sherwood JL, Müller S, Orr MC, Ratcliffe MJ, Walker J. Panel Based MALDI-TOF Tumour Profiling Is a Sensitive Method for Detecting Mutations in Clinical Non-Small Cell Lung Cancer Tumour. *PLoS One.* 2014
107. Simen BB, Yin L, Goswami CP, Davis KO, Bajaj R, Gong JZ, Peiper SC, Johnson ES, Wang ZX. Validation of a next-generation-sequencing cancer panel for use in the clinical laboratory. *Arch Pathol Lab Med.* 2015 Apr;139(4):508-17.
108. Singh RR, Patel KP, Routbort MJ, Reddy NG, Barkoh BA, Handal B, Kanagal-Shamanna R, Greaves WO, Medeiros LJ, Aldape KD, Luthra R. Clinical validation of a next-generation sequencing screen for mutational hotspots in 46 cancer-related genes. *J Mol Diagn.* 2013 Sep;15(5):607-22
109. Skronski M, Chorostowska-Wynimko J, Szczepulska E, Szpechcinski A, Rudzinski P, Orłowski T, Langfort R. Reliable detection of rare mutations in EGFR gene codon L858 by PNA-LNA PCR clamp in non-small cell lung cancer. *Adv Exp Med Biol.* 2013; 756:321-31.

110. Smith LM, Sanders JZ, Kaiser RJ, Hughes P, Dodd C, Connell CR, Heiner C, Kent SB, Hood LE. Fluorescence detection in automated DNA sequence analysis. *Nature*. 1986 Jun 12-18; 321(6071):674-9.
111. Sorbye H, Dragomir A, Sundström M, Pfeiffer P, Thunberg U, Bergfors M, Aasebø K, Eide GE, Ponten F, Qvortrup C, Glimelius B. High BRAF Mutation Frequency and Marked Survival Differences in Subgroups According to KRAS/BRAF Mutation Status and Tumor Tissue Availability in a Prospective Population-Based Metastatic Colorectal Cancer Cohort. *PLoS One*. 2015 Jun 29;10(6):e0131046.
112. ten Bosch JR, Grody ww. Keeping up with the next generation: massively parallel sequencing in clinical diagnostic. *J. Mol Diagn* 2008; 10: 484-92.
113. Tran NH, Cavalcante LL, Lubner SJ, Mulkerin DL, LoConte NK, Clipson L, Matkowskyj KA, Deming DA. Precision medicine in colorectal cancer: the molecular profile alters treatment strategies. *Ther Adv Med Oncol*. 2015 Sep;7(5):252-62.
114. Untergrasser A., Cutcutache I., Koressaar T., Ye J., Faircloth B. C., Remm M., Rozen S. G. Primer3 new capabilities and interfaces. *Nucleic Acids Research*. 2012; 40 (15): e115.
115. Van Cutsem E, Cervantes A, Adam R, Sobrero A, Van Krieken JH, Aderka D, Aranda Aguilar E, Bardelli A, Benson A, Bodoky G, Ciardiello F, D'Hoore A, Diaz-Rubio E, Douillard JY, Ducreux M, Falcone A, Grothey A, Gruenberger T, Haustermans K, Heinemann V, Hoff P, Köhne CH, Labianca R, Laurent-Puig P, Ma B, Maughan T, Muro K, Normanno N, Österlund P, Oyen WJ, Papamichael D, Pentheroudakis G, Pfeiffer P, Price TJ, Punt C, Ricke J, Roth A, Salazar R, Scheithauer W, Schmoll HJ, Tabernero J, Taïeb J, Tejpar S, Wasan H, Yoshino T, Zaanan A, Arnold D. ESMO consensus guidelines for the management of patients with metastatic colorectal cancer. *Ann Oncol*. 2016 Aug;27(8):1386-422.
116. Van Cutsem E, Kohne CH, Hitre E, Zaluski J, Chang Chien CR, Makhson A, et al. Cetuximab and chemotherapy as initial treatment for metastatic colorectal cancer. *N. Engl. J. Med*. 2009; 360: 1408-17.

117. Van Cutsem E, Lang I, Folprecht G, Nowacki M, Cascinu S, Shchepotin I, et al. Cetuximab plus FOLFIRI in the treatment of metastatic colorectal cancer (mCRC): The influence of KRAS and BRAF biomarkers on outcome: Update data from the CRYSTAL trial. American Society of Clinical Oncology (ASCO). Gastrointestinal Cancer Symposium 2010. Proceedings of ASCO GI 2010. 2010; 22-1-2010. Abstract 281.
118. Voelkerding KV, Dames SA, Durtschi JD. Next-generation sequencing: from basic research to diagnostics. *Clin Chem*. 2009 Apr;55(4):641-58.
119. Wagle N, Berger MF, Davis MJ, Blumenstiel B, Defelice M, Pochanard P, Ducar M, Van Hummelen P, Macconail LE, Hahn WC, Meyerson M, Gabriel SB, Garraway LA. High-throughput detection of actionable genomic alterations in clinical tumor samples by targeted, massively parallel sequencing. *Cancer Discov*. 2012 Jan;2(1):82-93.
120. Williams C, Pontén F, Moberg C, Söderkvist P, Uhlén M, Pontén J, Sitbon G, Lundeberg J. A high frequency of sequence alterations is due to formalin fixation of archival specimens. *Am J Pathol*. 1999 Nov;155(5):1467-71.
121. www.cancer.sanger.ac.uk/cosmic.
122. www.ensembl.org/index.html.
123. www.mycancergenome.org.
124. www.ncbi.nlm.nih.gov/.
125. Xuan J, Yu Y, Qing T, Guo L, Shi L. Next-generation sequencing in the clinic: promises and challenges. *Cancer Lett*. 2013 Nov 1;340(2):284-95.
126. Yaeger R, Chatila WK, Lipsyc MD, Hechtman JF, Cercek A, Sanchez-Vega F, Jayakumaran G, Middha S, Zehir A, Donoghue MTA, You D, Viale A, Kemeny N, Segal NH, Stadler ZK, Varghese AM, Kundra R, Gao J, Syed A, Hyman DM, Vakiani E, Rosen N, Taylor BS, Ladanyi M, Berger MF, Solit DB, Shia J, Saltz L, Schultz N. Clinical Sequencing Defines the Genomic Landscape of Metastatic Colorectal Cancer. *Cancer Cell*. 2018 Jan 8;33(1):125-136.
127. Yeo CD, Kim JW, Kim KH, Ha JH, Rhee CK, Kim SJ, Kim YK, Park CK, Lee SH, Park MS, Yim HW. Detection and comparison of EGFR mutations in matched tumor tissues, cell blocks, pleural effusions, and sera from patients

- with NSCLC with malignant pleural effusion, by PNA clamping and direct sequencing. *Lung Cancer*. 2013 Aug; 81(2):207-1.
128. Zhang Y, Jiang HR. A review on continuous-flow microfluidic PCR in droplets: Advances, challenges and future. *Anal Chim Acta*. 2016 Mar 31;914:7-16.
129. Zutter MM, Bloom KJ, Cheng L, Hagemann IS, Kaufman JH, Krasinskas AM, Lazar AJ, Leonard DG, Lindeman NI, Moyer AM, Nikiforova MN, Nowak JA, Pfeifer JD, Sepulveda AR, Willis JE, Yohe SL. The Cancer Genomics Resource List 2014. *Arch Pathol Lab Med*. 2015 Aug;139(8):989-1008.