



FACULTAD DE MEDICINA
UNIVERSIDAD DE CANTABRIA

GRADO EN MEDICINA

TRABAJO FIN DE GRADO

Identificación molecular de mutaciones hot-spot en genes frecuentemente mutados en cáncer de pulmón y colon mediante secuenciación masiva.

Molecular identification of hot-spot mutations in genes frequently mutated in lung and colon cancer through massive sequencing

Autor: Dña. Inés Díez Fernández

Director: D. Ignacio Varela Egocheaga

Co-director: D. Javier Freire Salinas

Santander, junio 2018

AGRADECIMIENTOS

A Nacho, por todo el tiempo y dedicación que ofrece a sus alumnos tanto en las clases, las prácticas, como en actividades externas a sus asignaturas; y sin duda, por haberme dado la oportunidad de hacer este trabajo de fin de grado en su laboratorio y haberme ayudado tanto en él.

A Javier por haber hecho posible la realización de este proyecto, sin las muestras que ha facilitado y su asesoramiento con la secuenciación, no habría sido posible.

A Laura, por todas las horas de laboratorio que me ha acompañado y ayudado con el proyecto.

A Ana, mi madre, por haberme acompañado en estos seis largos años de carrera.

A mis amigos, amigas y pareja por estar siempre a mi lado y apoyarme en todo.

RESUMEN

El cáncer se produce por la acumulación de mutaciones somáticas en el ADN. La identificación de estas mutaciones es esencial para la caracterización de los mecanismos moleculares implicados en el desarrollo tumoral y en muchos casos, condiciona el pronóstico y el posible tratamiento que hay que administrar al paciente. Sin embargo, la caracterización de las mutaciones presentes en una muestra en el ámbito clínico presenta muchas limitaciones debido a una alta cantidad y variedad de los sitios a estudiar, a una necesidad de un diagnóstico rápido y a trabajar con muestras de ADN que puedan estar degradadas por el proceso de fijación del tejido. En este trabajo nos proponemos hacer un repaso del conocimiento hasta la fecha de las bases moleculares de dos tipos tumorales: cáncer de pulmón y cáncer de colon así como poner a punto una metodología rápida basada en el enriquecimiento mediado por MIPs y NGS para el diagnóstico molecular de varias mutaciones hot-spot de gran importancia en estos tumores. Esta metodología se puede ampliar y adaptar a diversos tipos tumorales y a cualquier mutación de interés, suponiendo una mejora muy significativa de los métodos actuales de diagnóstico molecular.

SUMMARY

Cancer is caused by the accumulation of somatic mutations in DNA. The identification of these mutations is essential for the characterization of the molecular mechanisms involved in tumor development and in many cases, it determines the prognosis and the possible treatment that must be administered to the patient. However, the characterization of mutations present in a sample in the clinical setting presents many limitations due to a high number and variety of locations to study, a need for a rapid diagnosis and to work with DNA samples that may be degraded by the process of tissue fixation. In this work we propose to review the knowledge to date of the molecular basis of two tumor types: lung cancer and colon cancer as well as to develop a rapid methodology based on MIPs-mediated enrichment and NGS for molecular diagnosis of several hot-spot mutations of great importance in these tumors. This methodology can be extended and adapted to different tumor types and to any mutation of interest, becoming a very significant improvement of the current methods of molecular diagnosis.

ÍNDICE

1	ABREVIATURAS.....	5
2	INTRODUCCIÓN.....	7
2.1	Evolución hacia una medicina de precisión	7
2.2	Carcinogénesis	7
2.3	Cáncer de pulmón	10
2.3.1	Subtipos histológicos	10
2.3.1.1	Adenocarcinoma	10
2.3.1.1.1	Prevalencia.....	10
2.3.1.1.2	Mutaciones y tratamientos.....	10
2.3.1.2	Carcinoma epidermoide.....	15
2.3.1.2.1	Prevalencia.....	15
2.3.1.2.2	Mutaciones	15
2.3.1.2.3	Tratamientos.....	15
2.3.1.3	Microcítico.....	16
2.3.1.3.1	Prevalencia.....	16
2.3.1.3.2	Mutaciones	16
2.3.1.3.3	Tratamiento	16
2.4	Cáncer colorrectal	16
2.4.1	Carcinogénesis	16
2.4.2	Subtipos histológicos	18
2.4.2.1	Adenocarcinoma	18
2.4.2.1.1	Prevalencia.....	18
2.4.2.1.2	Mutaciones y tratamientos.....	18
2.5	Técnicas de secuenciación de nueva generación	23
2.5.1	Evolución de la secuenciación.....	23
2.5.1.1	Secuenciación primera generación (Sanger).....	23
2.5.1.2	Secuenciación segunda generación	24
2.5.1.2.1	Generación de librerías genómicas.....	25
2.5.1.2.2	Análisis de datos	27
2.5.2	Problemas en la incorporación de NGS a la práctica clínica	27
3	OBJETIVOS.....	28
4	MÉTODO	28
4.1	Protocolo MIPs	28
4.1.1	Diseño de los MIPs	28
4.1.2	Fosforilación.....	28
4.1.3	Ligación	29
4.1.4	Digestión exonucleasas.....	29
4.1.5	MIP PCR.....	29

4.2	Purificación de los fragmentos de PCR	30
4.3	Electroforesis	31
4.4	Protocolo extracción de ADN a partir de muestras en parafina	31
4.5	Secuenciación	32
5	RESULTADOS	33
5.1	Puesta a punto del protocolo	33
5.2	Generación de librerías	35
5.3	Resultados de la secuenciación	38
6	DISCUSIÓN	43
6.1	Control de calidad en el DNA	43
6.2	Problemas en la generación de librerías	44
6.3	Resultado secuenciación	44
7	CONCLUSIÓN.....	46
8	BIBLIOGRAFÍA.....	47

1 ABREVIATURAS

ADN	Ácido desoxirribonucleico
ALK	Kinasa del linfoma anaplásico
ATP	Adenosin trifosfato
BCG	Bacilo Calmette Guerin
CCD	Dispositivo de carga acoplada
CCR	Carcinoma colorrectal
CPNM	Carcinoma de pulmón no microcítico
CTLA-4	<i>Cytotoxic T-Lymphocyte Antigen 4</i>
CXCL	<i>Chemokine ligand</i>
CXCR2	<i>Chemokine receptor 2</i>
Del	Delección
EGFR (ErbB1)	Receptor del factor de crecimiento epidérmico tipo 1
FAP	Poliposis adenomatosa familiar
FDA	<i>US Food and drug administration</i>
FGFR1/2/3	<i>Fibroblast growth factor receptor 1/2/3</i>
FISH	<i>Fluorescent in situ hybridation</i>
HER2 (ErbB2)	Receptor del factor de crecimiento humano 2
HER4 (ErbB4)	Receptor del factor de crecimiento humano 4
HGFR	<i>c-Met/hepatocyte growth factor receptor</i>
HRM	<i>High resolution melt</i>
IFN- γ	Interferón gamma
IGF1R	Factor de crecimiento tipo insulina-1
IgG1/2	Inmunoglobulina G1/2

MAPK	<i>Mitogen-activated protein kinase</i>
MEK1/2	<i>Meiotic chromosome axis associated kinase 1/2</i>
MHC	Complejo mayor de histocompatibilidad
MMR	<i>Mismatch repair genes</i>
MIP	<i>Molecular Inversion Probe</i>
NGS	<i>Next Generation Sequencing</i>
Pb	Pares de bases
PD1	<i>Programmed cell death protein 1</i>
PDGF	<i>Platelet derived growth factor</i>
PDGFR	<i>Platelet derived growth factor receptor</i>
PD-L1/ L2	<i>Programmed death-ligand 1/2</i>
PGM	Personal genoma machine
PI3K	Fosfoinositol-3 quinasa
RAF	<i>Kinase of the rapidly accelerated fibrosarcoma</i>
PCR	Reacción en cadena de la polimerasa
RT-PCR	PCR con transcriptasa reversa
qPCR	PCR a tiempo real
SNV	<i>Single nucleotide variation</i>
T-CD4	Linfocito T CD4
T-CD8	Linfocito T CD8
TCR	Receptor linfocito T
Treg	Linfocito T regulador
UMI	Identificador molecular único
VEGFR-1/2/3	<i>Vascular endothelial growth factor receptor 1/2/3</i>

2 INTRODUCCIÓN

2.1 EVOLUCIÓN HACIA UNA MEDICINA DE PRECISIÓN

Actualmente el cáncer es una de las principales causas de morbilidad y mortalidad en el mundo. Así, en 2015, se posicionó como la segunda causa de muerte en el mundo con 8,8 millones de defunciones, siendo el responsable de 1 de cada 6 defunciones¹. Además, cabe destacar su gran impacto social debido a factores como su historia natural, que a menudo va deteriorando la salud del paciente de una forma progresiva, y el temor a los fármacos que se usan para tratarlo por sus efectos secundarios. Es por esto que los factores poblacionales y psicosociales han hecho de esta enfermedad un objeto de investigación muy importante. Pero dentro del extenso campo de investigación que abarca el mundo de la oncología están dos de los tipos de cáncer más prevalentes: el cáncer de pulmón y el colorrectal. El cáncer de pulmón a día de hoy es el que mayor mortalidad presenta con 1.1 millones de muertes anuales y una tasa de supervivencia de 5 años en el 10% de pacientes en la mayoría de países². Por otra parte, al cáncer colorrectal se le atribuyen más de 50.000 muertes anuales, siendo la tercera causa de muerte por cáncer en los Estados Unidos³.

En las últimas décadas de la investigación en cáncer, los adelantos tecnológicos han facilitado el estudio de la biología celular tumoral, lo cual ha conducido al descubrimiento de posibles marcadores tumorales y ha permitido evidenciar la existencia de una heterogeneidad celular y mutacional tanto intertumoral como intratumoral⁴. Esto ha permitido la creación de terapias dirigidas y la posibilidad de utilizar ciertos genes como marcadores diagnósticos y de seguimiento. Es decir, ha conducido a lo que hoy en día conocemos como “medicina de precisión”. En este planteamiento de la medicina, los profesionales en salud seleccionan los tratamientos más apropiados para los pacientes basándose en sus conocimientos sobre las anomalías moleculares presentes en las células tumorales⁵.

2.2 CARCINOGENÉISIS

El principio básico que rige la tumorigénesis es que la naturaleza clonal de los tumores es una característica crítica de la mutación somática. De acuerdo con este modelo, la ventaja adquirida para el crecimiento por una única célula mutada permite a su descendencia superar en número a sus células vecinas sanas. Según este modelo, con el crecimiento del tumor van surgiendo distintas poblaciones celulares, relacionadas genéticamente, que van acumulando distintas alteraciones que les ofrecen ventajas selectivas. A este modelo, responsable de la heterogeneidad intratumoral, se le denomina “crecimiento tumoral evolutivo ramificado”^{6,7}.

Dentro de los tumores, nos podemos encontrar tanto con mutaciones de la línea germinal como con mutaciones somáticas⁷:

- Mutaciones de línea germinal: son las que se generan previas a la fertilización y que son transmitidas a la descendencia dando lugar a síndromes de cáncer hereditarios.

- Mutaciones somáticas: son las más frecuentes, y se producen durante el desarrollo y/o crecimiento de un determinado tejido u órgano. Por lo tanto, este tipo de mutaciones dan lugar a cáncer de carácter esporádico. Dentro de estas debemos distinguir 2 subtipos:

- Mutaciones conductoras (*driver*): son aquellas que confieren una ventaja en el crecimiento celular.
- Mutaciones pasajeras (*passenger*): son aquellas biológicamente neutras, producidas como resultado de una inestabilidad genética inherente al tumor y que no confieren una ventaja de crecimiento.

Las células cancerígenas mutadas no son las únicas responsables de la tumorigénesis, y en este proceso se encuentran estrechamente asociadas con la matriz extracelular, fibroblastos, células inmunes infiltrantes y células de los vasos sanguíneos. De manera que el micromambiente puede tanto facilitar como prevenir el crecimiento tumoral a través de los siguientes procesos⁸:

- Generación de nuevos vasos sanguíneos y linfáticos que son necesarios para la llegada de nutrientes y células inmunes como:

- Macrófagos asociados al tumor: son imprescindibles para la progresión del tumor, ya que se encargan de la secreción de pro-angiogénicos (VEGF, PDGF).
- Neutrófilos asociados al tumor: la presencia de un número aumentado indica mal pronóstico debido a su gran capacidad para degradar la matriz extracelular con elastasa. A veces se reclutan por secreción de la quimiocina CXCL que se une al receptor CXCR2.
- Células mieloides supresoras: promueven la progresión del tumor aumentando la degradación de la matriz, la proliferación, las metástasis y la angiogénesis.

- Expresión de proteínas de regulación del sistema inmune:

- Expresión de PDL1 (ligando de muerte celular programada 1): Esta molécula se une al PD1 (receptor inmune de muerte programada 1, alias PDCD1) inhibiendo así el ataque del sistema inmune al tumor mediante los siguientes mecanismos (Figura 1):
 - Atenuar la proliferación, supervivencia, efectividad e incluso inducir apoptosis de las células T CD8.
 - Diferenciación de TCD4 a T regulador que expresa *FOXP3* (Forkhead box 3), los cuales permiten una situación de inmuno tolerancia con el tumor.

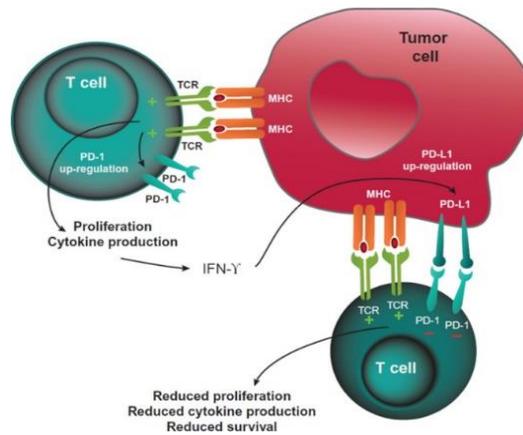


Figura 1 | Inhibición mediada por PD-1 de células T. La estimulación prolongada de TCR durante la respuesta inmune puede estimular la expresión de PD-1. Las células tumorales pueden expresar PD-L1 (y PD-L2) como consecuencia de las citoquinas inflamatorias y/o rutas de señalización oncogénicas. La unión de PD-1: PD-L1 inhibe la señalización positiva mediada por TCR, llevando a una disminución de la proliferación, secreción de citoquinas y supervivencia de los linfocitos T. IFN- γ : interferón- γ ; MHC; PD-1, proteína 1 de muerte programada ; PD-L1, ligando 1 de muerte programada; TCR, receptor de célula T⁹.

- Expresión de CTLA-4 (Antígeno 4 del linfocito T): este se encuentra aumentado en las células T tras la exposición al antígeno y lo que hace es competir con CD28 para unirse a B7.1 y B7.2 con mayor afinidad para así bloquear la señal co-estimuladora que necesita el linfocito T para activarse y atacar (Figura 2).

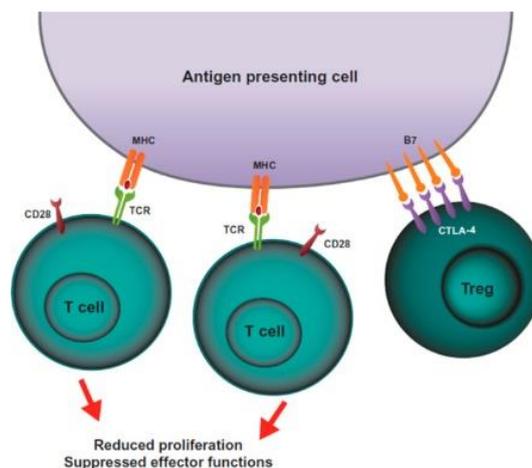


Figura 2 | Inhibición de células Treg mediada por CTLA-4. Esta es una de las hipótesis actuales. La expresión constitutiva de CTLA-4 en linfocitos Treg puede secuestrar o causar la internalización de las moléculas B7 en las células presentadoras de antígeno. La falta de la co-estimulación CD28: B7 lleva a una reducción en la proliferación y funciones efectoras de las células Treg. CTLA-4: antígeno 4 asociado a linfocito T citotóxico; MHC: complejo mayor de histocompatibilidad; TCR: receptor de células T; Tregs: células T reguladoras⁹.

- Cambio en los patrones de secreción de citoquinas y componentes de la matriz extracelular que modifican la composición de la misma permitiendo una matriz más “permeable” a la invasión tumoral. Algunos ejemplos son la expresión de queratinas en los carcinomas epidermoides de pulmón o de fibronectina en los adenocarcinomas de pulmón.

2.3 CÁNCER DE PULMÓN

2.3.1 Subtipos histológicos

En el cáncer de pulmón nos podemos encontrar dos grandes grupos histológicos conocidos como cáncer microcítico o de células pequeñas, el cual es de origen neuroendocrino, y el cáncer no microcítico, que está presente en un 85% de los casos de cáncer de pulmón y su tasa de supervivencia a los 5 años es de tan solo el 15,9%. A su vez, dentro de los tumores no microcíticos, nos vamos a encontrar con dos tipos predominantes, que son el adenocarcinoma y el carcinoma epidermoide⁸.

2.3.1.1 Adenocarcinoma

2.3.1.1.1 Prevalencia

Actualmente se trata del subtipo histológico más frecuente en cáncer de pulmón y constituye aproximadamente el 50% de los tumores no microcíticos⁸. Una de las características más importantes de este tumor radica en que aunque la mayoría de los tumores de pulmón son vistos en fumadores, esta subtipo se da más frecuentemente que el resto en individuos no fumadores, especialmente en mujeres².

2.3.1.1.2 Mutaciones y tratamientos

2.3.1.1.2.1 Receptor del Factor de Crecimiento Epidérmico (EGFR)

Se trata de una molécula encargada de regular procesos como la angiogénesis o la proliferación e invasividad celulares. Esta mutación presenta una incidencia del 10-15% en la población caucásica y del 50% en la población asiática, siendo más frecuente en mujeres no fumadoras¹⁰. El pronóstico es favorable por la posibilidad de emplear inhibidores de la tirosín quinasa^{11,12}.

La mayoría de las mutaciones descritas clínicamente relevantes ocurren dentro de los primeros cuatro exones que codifican el sitio de unión a ATP (exones 18 a 21), pero el 80% de ellas se encuentran en los exones 19 y 21. Se han agrupado en 3 clases¹⁰:

- Clase I: las deleciones en el exón 19 (Del19).
- Clase II: mutaciones puntuales. La más frecuente es la sustitución de una leucina por arginina en el codón 858 del exón 21 (L858R). Menos frecuentes son la mutación L861Q (2%), exón 21; y G719X (3%), exón 18.

- Clase III: duplicaciones o inserciones en el exón 20.

La Secuenciación o la PCR en tiempo real (qPCR) son los principales procedimientos de diagnóstico utilizados hasta la fecha. También se han desarrollado diversos anticuerpos monoclonales para la identificación de la mutación de EGFR por técnicas de inmunohistoquímica cuyos resultados preliminares confirman un buen rendimiento de este procedimiento como método de cribado¹³.

Las terapias actuales dirigidas frente a las mutaciones en este gen son:

- **Erlotinib y Gefitinib**

Ambos son fármacos selectivos para EGFR de primera línea que se diferencian en su estructura molecular. Su mecanismo de acción es formar un complejo con el sitio de unión a ATP, actuando como inhibidores competitivos del mismo¹⁰.

Son eficaces en la mayoría de los pacientes con CPNM (carcinoma pulmonar no microcítico) que presentan las mutaciones comunes activadoras del *EGFR* (Del19 y L858R) y la adquisición de resistencias ocurre aproximadamente en 10-13 meses debido a la adquisición de, principalmente, dos mutaciones¹⁰:

- *EGFR* T790M (treonina por metionina en posición 790) en el exón 20: esta es la mutación más importante. Se produce cerca del sitio de unión a ATP, siendo capaz de desplazar a los inhibidores reversibles no unidos covalentemente al ATP. Está presente en el 50-60% de los pacientes con resistencia adquirida a inhibidores de EGFR reversibles y en el 38% de los pacientes no tratados previamente, asociándose con un peor pronóstico.
- *HER2* sobreexpresado: presente en el 12% de las resistencias a erlotinib/gefitinib que carecen de la mutación *EGFR* T790M.

- **Afatinib**

Se trata de un fármaco selectivo dual EGFR/HER2 de segunda línea. Actúa sobre los residuos de cisteína conservados en los dominios catalíticos de EGFR, HER2 y ErbB4 produciendo una unión covalente que bloquea la actividad tirosín quinasa de estos receptores de forma irreversible¹⁴.

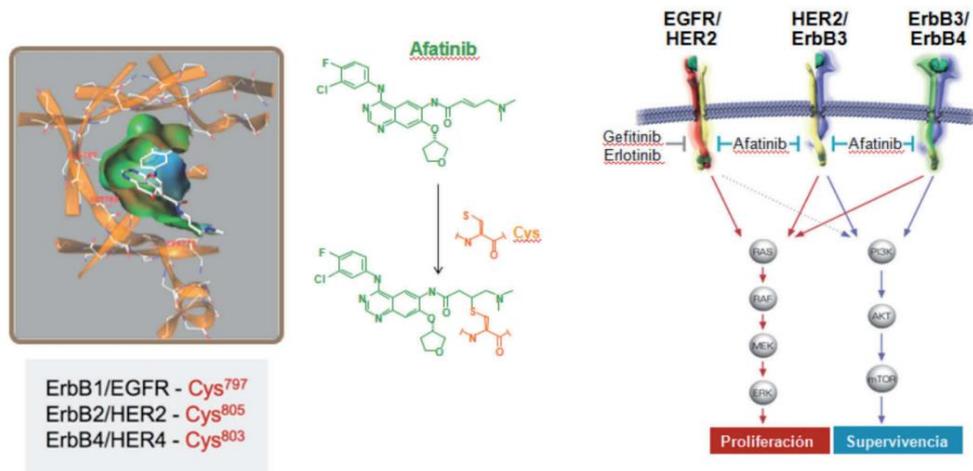


Figura 3 | Mecanismo de acción del afatinib. Este se une covalentemente a los residuos de cisteína (naranja) de EGFR, HER2 y HER4 dentro de su dominio catalítico inhibiendo las vías de señalización intracelular de proliferación y supervivencia¹⁰.

- **Osimertinib**

Se trata de un inhibidor, potente e irreversible de la tirosina quinasa de EGFR selectivo que se puede emplear frente a la mutación de resistencia T790M. Está indicado en pacientes que progresan a un tratamiento previo con inhibidores tirosina quinasa de primera o segunda generación¹⁵.

- **Necitumumab**

Se trata de un anticuerpo monoclonal IgG1 humanizado recombinante diseñado para bloquear la interacción del ligando de EGFR¹⁶.

2.3.1.1.2.2 *Kinasa del linfoma anaplásico (ALK)*

La incidencia de traslocación *ALK* en cáncer no microcítico tiende a ser del 3%-5%. Suelen darse en pacientes de unos 50 años, no fumadores (70%-75%) o fumadores leves. Es típica de adenocarcinoma y se distribuye por igual en hombres y mujeres.

ALK se trata de un receptor que en su forma inactiva (sin ligando) es proapoptótico mientras que si está activo es antiapoptótico, por lo que la presencia de mutaciones activadoras otorga a las células características antiapoptóticas. Esto se consigue a través de dos mutaciones principales^{11,12}:

- Traslocación t(2;15) con el gen *NPM*: presente en el 92-95% de los linfomas anaplásicos.
- Traslocación con el gen *EML4* t(2;2)(p24;p21): presente en un 2-5% de pacientes con cáncer no microcítico.

La prueba recomendada para su diagnóstico es la hibridación fluorescente in situ (FISH), aunque también puede emplearse PCR con transcriptasa inversa (RT-PCR), técnicas de inmunohistoquímica para detección de la proteína aberrante y secuenciación de nueva generación¹².

Las terapias dirigidas de las que disponemos actualmente frente a mutaciones en este gen son:

- **Crizotinib (1ª generación)**

Es un inhibidor competitivo del ATP frente a *c-Met/hepatocyte growth factor receptor* (HGFR), ALK y ROS¹⁷.

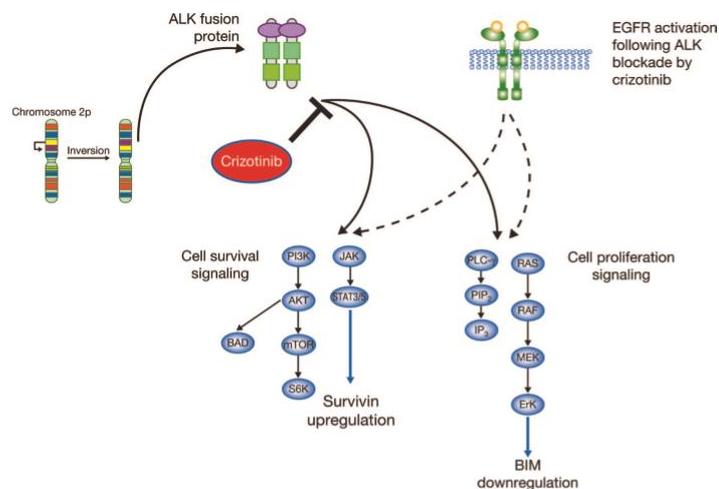


Figura 4 | Rutas de señalización empleadas por EML-ALK. La fusión de *ALK* con *EML4* crea un receptor tirosín quinasa ALK constitutivamente activado, que a través de vías como PI3K/AKT, RAS-RAF-MEK-ERK y JAK/STAT promueve la supervivencia celular. Crizotinib actúa impidiendo la fosforilación de la tirosín quinasa y con ello la transducción de la señal. Estas rutas de señalización son comunes a EGFR, por lo que el bloqueo de las ALK puede activar EGFR y con ello la transducción a través de estas vías¹⁸.

La resistencia en este fármaco suele deberse a las siguientes mutaciones¹⁹:

- La más frecuentes son: L1196M y G1269A.
- Amplificación del gen de fusión *ALK*.
- Activación de vías de señalización alternativas: anomalías en *EGFR*, *KIT* y factor de crecimiento tipo insulina- 1 (*IGF1R*).
- Mutación G1202R: la cual confiere resistencia a los de primera y segunda generación.

- **Ceritinib, alectinib, brigatinib (2ª generación)**

Son también inhibidores de la tirosín quinasa de ALK, pero estos se emplean tras la adquisición de mutaciones de resistencia a crizotinib.

Aunque hay estudios que apoyan la utilización de alectinib como tratamiento de primera línea¹⁹.

- **Lorlatinib (3ª generación)**

Se trata de un inhibidor de la tirosín quinasa de ALK que aún se encuentra en estudio, pero que ha demostrado eficacia frente a la mutación G1202R¹⁹.

2.3.1.1.2.3 *ROS1*

La prevalencia de la mutación es de entre el 0,6% y el 3% en población no seleccionada y se estima que la prevalencia en adenocarcinoma es de aproximadamente el 1-2%. Este tipo de mutación es más frecuente en no fumadores y tiende a darse en pacientes jóvenes.

El oncogén *ROS1* da lugar a un receptor tirosín quinasa que no posee un ligando conocido y está relacionado con ALK.

Este gen sufre reordenamientos que dan lugar a una activación constitutiva en los que el punto de ruptura suele producirse en los exones 32,34 o 35, de manera que el fragmento *ROS1* 3' que va a formar el gen de fusión contiene el dominio tirosín quinasa. Este fragmento va a formar genes de fusión junto a genes como por ejemplo: *CD74*, *SLC34A2*, *TMP3*.

La detección de esta mutación se suele llevar a cabo con FISH, pero también puede detectarse con RT-PCR o secuenciación masiva²⁰.

2.3.1.1.2.4 *RET*

RET es un receptor tirosín quinasa que puede presentar translocación en un 1-2% de los CPNM. Algunos estudios describen que estas translocaciones recurrentes de fusión de *RET* con varios genes pueden llegar a estar presente hasta en un 12% de los CPNM. Habitualmente el perfil clínico de los pacientes portadores de esta translocación se caracteriza por ser jóvenes y no fumadores, tener histología de adenocarcinoma y no presentar ninguna de las otras alteraciones genéticas frecuentes¹³.

La activación de *RET* lleva a la autofosforilación de los residuos de tirosina y con ello a la activación de vías que señalizan la proliferación celular, la migración y la diferenciación⁷. La ganancia de función en línea germinal de mutaciones *RET* puede conducir a una neoplasia endocrina múltiple; mientras que una ganancia somática de función puede llevar también a un carcinoma medular de tiroides esporádico u otros tumores^{10,21}.

2.3.1.1.2.5 *BRAF*

Las mutaciones activadoras de *BRAF* se observan en el 1-3% de los cánceres no microcíticos y suele asociarse a sujetos fumadores.

Se trata de un mediador de la señalización de KRAS que activa la vía de la MAPK (*mitogen-activated protein kinase*). Las mutaciones pueden producirse tanto en la posición V600 del exón 15 (como en el melanoma), como fuera de esta.

La detección se realiza mediante PCR cuantitativa o secuenciación¹¹.

El tratamiento dirigido actual para esta mutación es el dabrafenib. Este es un inhibidor competitivo con el ATP de carácter reversible que inhibe específicamente a la quinasa BRAF²².

2.3.1.1.2.6 KRAS

Es la mutación oncogénica más frecuentemente detectada en pacientes con CPNM, y es posible encontrarla entre el 25%-30% de los adenocarcinomas de pulmón en pacientes fumadores²³.

Actualmente se están estudiando varias mutaciones y su respuesta a las terapias inhibitoras de MEK. Los resultados revelan que las mutaciones G12C y G12V están asociadas a una mayor respuesta con selumetinib; y que los pacientes con mutaciones Q61 no se benefician de la adición de selumetinib al tratamiento²³.

Se prevé que la mutación de este gen puede ser una importante diana por su prevalencia y existen ensayos clínicos con selumetinib, un inhibidor de MEK1 y MEK2, que son elementos esenciales de la ruta de señalización de KRAS²³.

2.3.1.2 Carcinoma epidermoide

2.3.1.2.1 Prevalencia

Constituyen aproximadamente el 40% de los tumores no microcíticos⁸ y están fuertemente asociados al consumo de tabaco, tanto que el 90% de los pacientes que presentan este tumor son fumadores; aunque también existen otras asociaciones como es la exposición a arsénico².

2.3.1.2.2 Mutaciones

Las mutaciones en *DDR2*, *FGFR1*, *2*, *3* y *PI3K* son típicas de este subtipo histológico y se tratan de genes driver que podrían ser potenciales dianas terapéuticas. Es por esto que se están llevando a cabo ensayos clínicos con inhibidores de la FGFR en pacientes con *FGFR1* mutado (en 2014 se obtuvieron resultados preliminares positivos)⁸.

2.3.1.2.3 Tratamientos

Hasta hace un tiempo, docetaxel y el inhibidor de la tirosín quinasa del EGFR erlotinib han sido los dos únicos fármacos aprobados para el tratamiento de segunda línea del carcinoma epidermoide avanzado. En diciembre de 2014, la *US Food and Drug Administration* (FDA) aprobó en segunda línea de tratamiento del CPNM ramucirumab,

un anti-VEGFR-2, en combinación con docetaxel, y en marzo de 2015 aprobó también en segunda línea de pacientes con carcinoma escamoso, el inhibidor de PD1 nivolumab. Asimismo también existen estudios favorables con atezolizumab, otro inhibidor de PD1^{24,25}.

2.3.1.3 Microcítico

2.3.1.3.1 Prevalencia

Representa en torno al 10-15% de todos los cánceres de pulmón²⁶.

2.3.1.3.2 Mutaciones

En este tipo de tumores nos encontramos con la presencia de numerosos genes driver como: *TP53*, *RB1*, *PTEN*, *CREBBP*, *EP300*, *SLIT2*, *MLL*, *COBL* y *EPHA7*. Pero de estos sin duda los más importantes son los genes supresores de tumores *TP53* y *RB1*, ya que la mutación de los mismos es un hallazgo común a todos los tumores microcíticos. Asimismo la mutación (p.Val73fs) en *TP53* está presente tanto en las muestras de adenocarcinoma como en las de microcítico; al contrario que las alteraciones en la argenina 251 de *RB1*, que solo se encuentran en el carcinoma microcítico, lo cual nos indica que podría ser un marcador de diferenciación de adenocarcinoma a cáncer microcítico.

También es importante *SLIT2*, ya que este está involucrado en la polimerización de la actina mediada por Cdc42, lo cual se cree que podría proporcionar a las células capacidades invasivas que las hacen más agresivas²⁷.

2.3.1.3.3 Tratamiento

Este tipo de tumores por lo general son tratados con un régimen quimioterápico de cisplatino y etopósido asociado a radioterapia; aunque esto variará según el estadiaje tumoral. En cuanto a las terapias dirigidas, existen ensayos clínicos con vandetanib (inhibidor de la tirosín quinasa de VEGFR-2), bevacizumab (anticuerpo anti VEGF), tamoxifeno (quimioterápico modulador selectivo de los receptores estrogénicos), inhibidores de las metaloproteinasas de matriz y la vacuna Bec2/BCG (vacuna anti gangliósidos)^{28,29}.

2.4 CÁNCER COLORRECTAL

2.4.1 Carcinogénesis

En 1990, Fearon y Vogelstein propusieron un modelo genético de tumorigénesis conocido como secuencia adenoma-carcinoma, que propone que la acumulación de mutaciones germinales y somáticas condiciona las características del tumor. Los genes que participan en este modelo genético se dividen en dos clases: genes supresores de tumores (*APC*, *DCC* y *TP53*) y oncogenes (*K-RAS* y *CTNNB1*)³⁰.

Existen dos mecanismos que pueden conducir al desarrollo de carcinoma colorrectal (CCR) (Figura 5):

- Vía supresora o de inestabilidad cromosómica: es iniciada por la inactivación mutacional del gen supresor de tumor APC, el cual es responsable de la poliposis adenomatosa familiar (FAP) y de aproximadamente el 85% de los CCR esporádicos. Algunos de estos carcinomas se desarrollan tras la activación mutacional de beta catenina (*CTNNB1*) (actividad normalmente regulada por *APC*)³⁰. Tras la mutación siguen una serie de pasos hasta dar lugar al tumor³¹:

- Ambos genes producen una alteración de la vía de señalización Wnt/ β -catenina, generan una resistencia a la degradación de β -catenina, por lo que se acumula en el citoplasma y se trasloca al núcleo, donde actúa promoviendo la sobreexpresión de oncoproteínas. Esto genera un fenotipo hiperproliferativo que favorece la ganancia de mutaciones en otros genes, permitiendo la progresión a adenoma temprano.

- El adenoma temprano progresa al intermedio por mutación de ganancia de función en el gen *KRAS*, lo que conduce a la activación constitutiva de este oncogén.

- En el adenoma intermedio, el patrón hiperproliferativo es muy marcado, lo que permite el desarrollo de un fenotipo de inestabilidad cromosómica que lleva a la pérdida de copias de genes ubicados en el brazo largo del cromosoma 18, como: *DCC*, *SMAD4*, *SMAD2* e *ITF-2*. Lo anterior contribuye a la inmortalización de las células y favorece la progresión al adenoma tardío.

- Por último, el fenotipo de inmortalización celular del adenoma tardío conduce a la adquisición de más mutaciones, incluidas las del gen *TP53*, evento que dispara el desarrollo del adenocarcinoma.

- Vía mutadora o de inestabilidad de microsatélites: se ha asociado con el 15% de los CCR esporádicos y también, con el síndrome de Lynch. Consiste en la inactivación de una familia de genes supresores de tumores involucrados en la reparación del daño al ADN, conocidos como genes MMR o *Mismatch repair genes*, en donde se incluyen *MSH2*, *MSH6*, *MLH1* y *PMS2*. Si alguno de estos genes se inactivan por cualquier mecanismo (deleción, mutación o metilación) no se repararán los errores de replicación, afectando fundamentalmente a las regiones que contienen microsatélites, lo que dará lugar a una inestabilidad de estas regiones. Algunos genes importantes para el desarrollo tumoral contienen microsatélites en su región codificante, con lo que son dianas preferentes de este fenotipo mutador (ej.: *CTNNB1*, *TGF- β IIR*, *IGF-IIR*, *BAX* y *PTEN*)³¹.

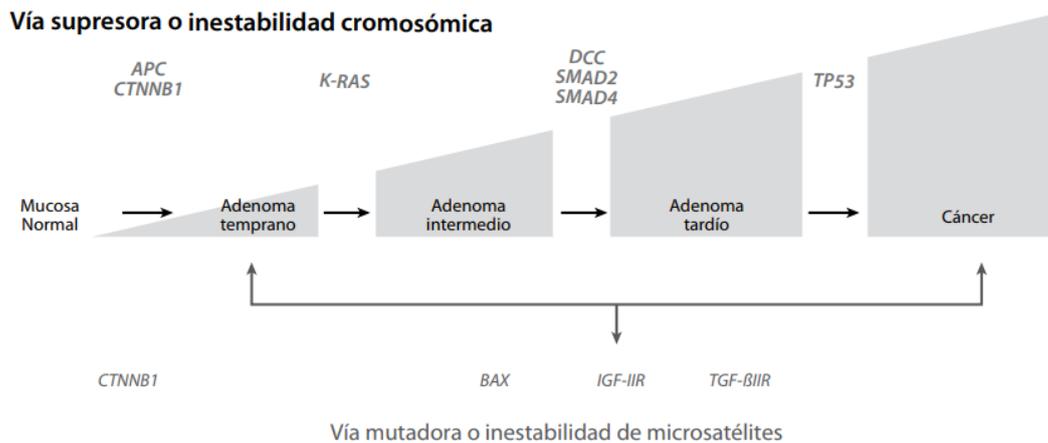


Figura 5 | Vías clásicas de carcinogénesis colorrectal a partir de adenomas: vía supresora y vía mutadora. En la parte superior se visualizan las principales mutaciones que se van sucediendo en la vía supresora para pasar del adenoma temprano al cáncer. En la zona inferior se observan las alteraciones que sufre el tejido en la vía mutadora, la cual se basa en “dos golpes”, una primera mutación de los genes MMR y una alteración secundaria de genes como *CTNNB1*, *TGF-βIIR*, *IGF-IIR* y *BAX*³¹.

2.4.2 Subtipos histológicos

La mayoría de los tumores en esta región son adenocarcinomas originados en las células epiteliales de la mucosa colorrectal, aunque menos frecuentemente pueden observarse carcinomas de tipo neuroendocrino, de células escamosas, adenoescamosos, indiferenciados y de células fusiformes³².

2.4.2.1 Adenocarcinoma

2.4.2.1.1 Prevalencia

Este subtipo es el más relevante, ya que representa aproximadamente el 90% de los tumores colorrectales³³.

2.4.2.1.2 Mutaciones y tratamientos

En estos tumores el análisis de mutaciones es muy relevante, ya que existen estudios que indican que el 52,5% de los cánceres colorrectales presentan al menos una con posibilidad de tratamiento médico dirigido. Las mutaciones genéticas más prevalentes se dan en *TP53* y *APC*, que se encuentran mutados en el 80% y 67,5% de los tumores respectivamente, aunque a día de hoy no constituyen una diana terapéutica³⁴.

2.4.2.1.2.1 *KRAS* y *NRAS*

KRAS se trata de un protooncogén frecuentemente mutado en el CCR (35-45% de los casos), constituyendo el 85% de las mutaciones RAS³⁵; mientras que *NRAS* tan solo constituye el 15% de las mutaciones RAS y se encuentra en un 3% de los tumores de colon³⁶. La determinación de ambos va a ser muy importante,

ya que las mutaciones activadoras de este son el principal factor predictivo de eficacia en pacientes que reciben cetuximab y panitumumab³⁵.

Las mutaciones somáticas en este gen provocan la expresión de proteínas RAS constitutivamente activadas, lo cual permite una activación intrínseca de la cascada de señalización dependiente de EGFR, de forma que esta activación es independiente de la expresión de EGFR y, por tanto, no puede inhibirse por anticuerpos anti-EGFR como son cetuximab y panitumumab, de manera que estos fármacos están contraindicados en presencia de estas mutaciones^{32,35,37,38}.

En el caso de *KRAS*, más del 95% de las mutaciones se producen en los codones 12 y 13, por los que son las que han de ser estudiadas previas al tratamiento con anti-EGFR; aunque también pueden existir mutaciones en los codones 61 y 146³². Así mismo, las mutaciones más frecuentes en *NRAS* se encuentran en los codones 12, 13, 61³⁶.

Actualmente para la detección de estas mutaciones contamos con los métodos de qPCR con cebadores específicos de la mutación y la secuenciación Sanger, con la cual podemos detectar todas las mutaciones posibles³².

Como se ha mencionado, estos datos son imprescindibles para decidir si dar los siguientes fármacos anti-EGFR:

- **Cetuximab**

Es un anticuerpo monoclonal quimérico IgG1 que se une específicamente al EGFR e inhibe competitivamente la unión de los ligandos endógenos y promueve la internalización del EGFR³⁹.

- **Panitumumab**

Se trata de un anticuerpo monoclonal IgG2 completamente humano que es altamente selectivo para el factor de crecimiento epidérmico (EGFR)⁴⁰.

Actúa como antagonista evitando la unión de los ligandos endógenos e induciendo la internalización de EGFR. De esta manera conduce a un aumento de la apoptosis, proliferación reducida de las células tumorales y angiogénesis reducida⁴¹.

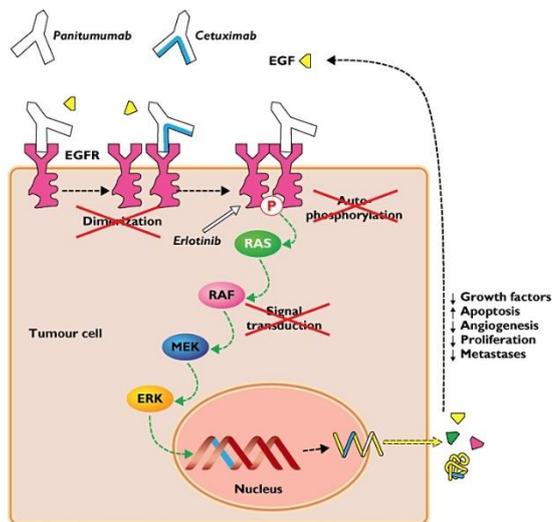


Figura 6 | Mecanismo de acción de panitumumab. Bloquea EGFR previniendo la unión de factores de crecimiento e inhibiendo la transducción de señales y con ello la disminución en la expresión de factores de crecimiento⁴¹.

2.4.2.1.2.2 BRAF

Las mutaciones somáticas en *BRAF* ocurren en una tasa menor que las de *KRAS* en el CCR (5-10%). Se ha demostrado que las mutaciones de *BRAF* también afectan de forma negativa en la respuesta a los anti-EGFR. Así mismo se sabe que las mutaciones *KRAS* y *NRAS*, y *BRAF* son mutuamente excluyentes, y estos dos marcadores juntos pueden identificar hasta un 50-55% de los no respondedores a los anti-EGFR^{35 32,42}.

El 90% de las mutaciones encontradas producen la sustitución V600E, provocando una activación constitutiva de la molécula que lleva a un incremento en la actividad de la vía de señalización MAPK^{32,42}.

Más específicamente, la activación de la mutación del gen *BRAF* se asocia con un alto nivel de metilación del ADN global y silenciamiento epigenético del gen *MLH1*, que se encuentra en el 70-90% de los tumores colorrectales esporádicos con un fenotipo microsatélite inestable. Por lo tanto, la prueba adicional de la mutación *BRAF* en un tumor con inestabilidad de microsatélites ayudará a aclarar la naturaleza esporádica o sindrómica del tumor³².

El fármaco que a menudo se emplea para la mutación de este gen, en especial la V600E es:

- **Regorafenib**

Este es un fármaco antitumoral oral que bloquea varias proteínas quinasa, incluidas las quinasa implicadas en la angiogénesis tumoral (VEGFR 1/2/3, TIE2), la oncogénesis (KIT, RET, RAF-1, BRAF, BRAF con alteración por mutación V600E) y el microambiente tumoral (PDGFR,

FGFR). Este ha sido autorizado en el tratamiento de pacientes adultos con CCR metastásico que han sido previamente tratados con las terapias disponibles o no se les considera candidatos adecuados a dichos tratamientos. Ello incluye quimioterapia basada en fluoropirimidinas, terapia anti-VEGF y terapia anti-EGFR⁴³.

2.4.2.1.2.3 *PI3K y PTEN*

El EGFR activa la vía de señalización *PI3K*, pero esta puede estar anormalmente activada por mutaciones en la unidad P110 de *PIK3CA* o por la inactivación de *PTEN*, un conocido gen supresor de tumores³⁵.

En concreto las mutaciones de los exones 9 y/o 20 están presentes en el 10-20% de los cánceres colorrectales y se encuentran asociadas con otras alteraciones moleculares como son las mutaciones en *KRAS*⁴⁴.

El papel de la desregulación de esta vía de señalización en la respuesta a cetuximab y panitumumab ha mostrado resultados conflictivos. Sin embargo, parece ser que determinadas mutaciones activadoras en *PIK3CA* están asociadas a la resistencia al tratamiento, aunque la correlación dista mucho de la observada con *KRAS*. Por otro lado, la inactivación de *PTEN* también se ha asociado a falta de respuesta al tratamiento. Las mutaciones en *PI3K* y la inactivación de *PTEN* pueden coexistir con las mutaciones en *KRAS* y *BRAF*, hecho que puede explicar la dificultad en demostrar la contribución individual de estas alteraciones en la falta de respuesta. A la espera de estudios que evalúen específicamente la contribución de la alteración de esta vía en la respuesta a anti-EGFR, *PI3K* y *PTEN*, no se utilizan como marcadores en la práctica clínica³⁵.

2.4.2.1.2.4 *EGFR*

Pese a que la positividad para *EGFR* por inmunohistoquímica ha demostrado no correlacionarse con la respuesta al tratamiento con anticuerpos monoclonales anti-EGFR, el incremento en el número de copias del gen *EGFR* se ha asociado a una mayor respuesta a cetuximab. Sin embargo, hasta el momento no se cuenta con los datos suficientes para establecer un punto de corte fiable en el número de copias del gen³⁵.

2.4.2.1.2.5 *FBXW7*

Este gen puede encontrarse mutado en aproximadamente el 15– 20% de los CCRs⁴⁵.

Este gen codifica para la ubiquitin ligasa E3, la cual se encarga de promover la degradación proteasomal de moléculas como el Mcl-1 fosforilado, de manera que se puede alterar la respuesta a terapias dirigidas contra Mcl-1. Esto hace a las células resistentes a regorafenib⁴⁵.

Las mutaciones heterocigotas son habitualmente detectadas en tres residuos de arginina (R465, R479, R505)⁴⁵.

2.4.2.1.2.6 GNAS

Este gen se encuentra mutado en aproximadamente el 2% de los carcinomas colorrectales. A menudo se asocia a mutaciones en la vía de la MAPK, específicamente mutaciones BRAF⁴⁶.

Este gen codifica la subunidad estimulante alfa que forma parte de un complejo proteínico conocido como proteína de unión a nucleótidos de guanina, lo cual estimula una adenilato ciclasa ubicua para todas las células que media la señalización de diversos estímulos extracelulares⁴⁷. El resultado final es una mayor transcripción y proliferación celular. Las mutaciones activadoras más comunes afectan al codón 201, en el cual un residuo de arginina es cambiado por uno de cisteína (R201C) o histidina (R201H)⁴⁶.

El análisis de las 2 mutaciones más comunes se lleva a cabo mediante GNAS análisis *high-resolution melt* (HRM), una técnica basada en la hibridación del ADN frente a sondas específicas de las mutaciones⁴⁶.

2.4.2.1.2.7 Mutaciones MMR

El mecanismo patogénico de estos genes ya ha sido explicado en el apartado de la vía de la inestabilidad cromosómica en cáncer de colon⁴⁸.

La detección de las mutaciones en estos genes es llevada a cabo mediante amplificación de regiones que contienen microsatélites y la comprobación del tamaño de los mismos mediante electroforesis. Si el 30% o más de los marcadores presentan expansión o acortamiento de las secuencias repetitivas en comparación con la mucosa normal del mismo paciente, se dice que el tumor tiene un nivel elevado de inestabilidad de microsatélites. Este procedimiento puede llevarse a cabo también mediante paneles de secuenciación masiva. Además las mutaciones de estos genes que causan síndrome de Lynch pueden ser detectadas mediante métodos de inmunohistoquímica⁴⁸.

A día de hoy para estas mutaciones tenemos los siguientes tratamientos específicos disponibles:

- **Pembrolizumab y nivolumab**

Están dirigidos a la proteína PD-1, la cual es una proteína presente en los linfocitos T que favorece la tolerancia del sistema inmune frente a las células tumorales⁴⁹.

2.5 TÉCNICAS DE SECUENCIACIÓN DE NUEVA GENERACIÓN

2.5.1 Evolución de la secuenciación

2.5.1.1 Secuenciación primera generación (Sanger)

En 1977 Allan M. Maxam y Walter Gilbert publican “A New method for sequencing ADN” en el describían un proceso químico, mediante el cual era posible determinar la secuencia de nucleótidos del ADN. Frederick Sanger, en el mismo año, desarrolla un método de secuenciación de pequeños fragmentos de ADN cuyo enfoque diferencial, frente a la propuesta de Maxam y Gilbert, fue la utilización de síntesis enzimática de la cadena complementaria al fragmento de ADN a secuenciar. Este segundo método, popularizado posteriormente como secuenciación Sanger, se estableció como el método de referencia debido a su simplicidad y a que permitía secuenciar regiones mayores de ADN. En 1982 Marvin Caruthers y Leroy Hood desarrollan el primer método automatizado para secuenciar ADN basado en el método Sanger, el cual era capaz de secuenciar fragmentos de 5 a 75 pares de bases (pb) y, en 1986, Leroy Hood y Lloyd Smith diseñan el primer secuenciador automático que utiliza rayos láser que reconocen marcadores de fluorescencia en el DNA. La descripción de la reacción en cadena de la polimerasa (PCR) por Kary Mullis permite además la realización de reacciones de secuenciación a partir de muy poco material de partida⁵⁰.

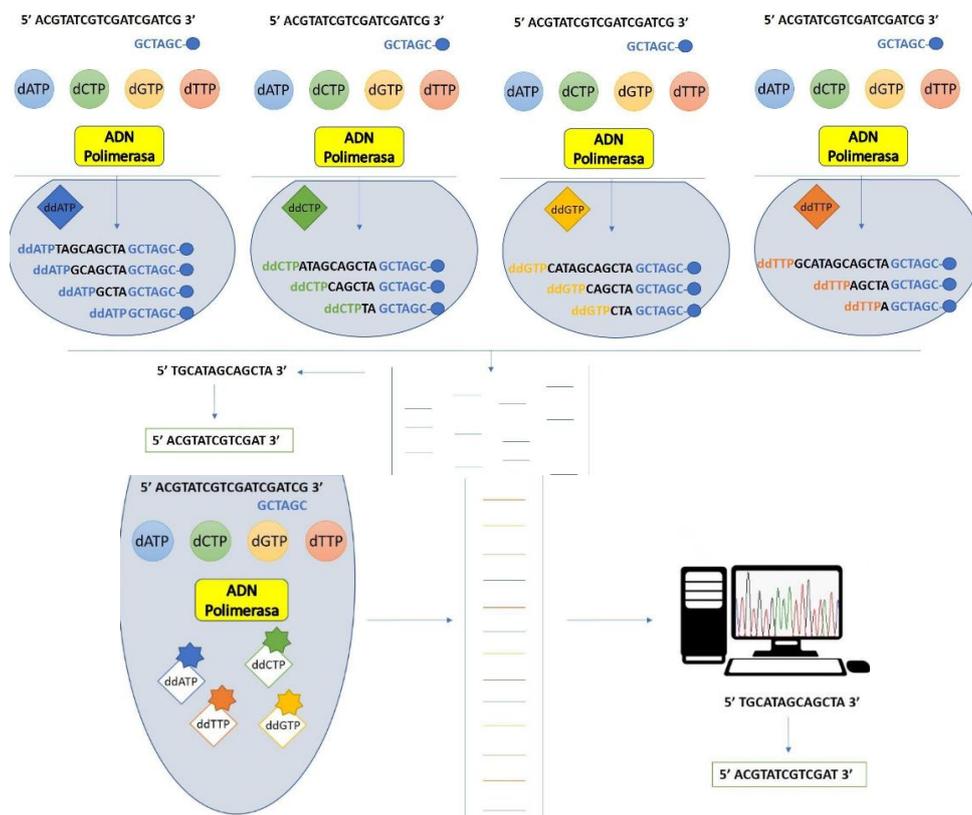


Figura 7 | Fundamento del método de Sanger. Se basa en la síntesis, de forma secuencial, de una hebra de cuatro dideoxinucleótidos marcados con fluoróforos distintos que terminan la cadena al incorporarse. La electroforesis capilar de los fragmentos generados permite detectar la posición de cada nucleótidos en la secuencia⁵¹.

2.5.1.2 *Secuenciación segunda generación*

La secuenciación de segunda generación nació en el año 2005 sobre las bases de la secuenciación Sanger, y se basa en la generación aleatoria de fragmentos cortos de ADN y su secuenciación masiva paralela sobre una superficie sólida. La gran ventaja de estas nuevas técnicas es que son capaces de secuenciar el genoma completo de un humano en tan solo un experimento con un coste mucho menor que la tecnología Sanger.

Esta tecnología no sólo permite la identificación de pequeños cambios en el genoma como deleciones e inserciones, sino que también permite la identificación de grandes reordenamientos genómicos^{52,53}.

Entre las primeras plataformas que ya han quedado obsoletas y descatalogadas cabe destacar las siguientes:

- **Roche 454 (pirosecuenciación)**

Se trata de una estrategia de secuenciación del ADN en tiempo real basada en la detección de la luz emitida cada vez que se incorpora un nucleótido a la cadena y se libera un grupo pirofosfato que se convertirá en ATP gracias a una ATP-sulfurilasa y en luz gracias al enzima luciferasa. La luz emitida es detectada por un sistema CCD de manera proporcional a la cantidad de nucleótidos incorporados⁵².

- **SOLiD (*Sequencing by Oligo Ligation Detection*)**

En cada ciclo de secuenciación se emplean octámeros marcados con cuatro colores diferentes que detectan dos di-nucleótidos al mismo tiempo que son incorporados gracias a la presencia de una ligasa de ADN. Debido a que hay más combinaciones posibles de di-nucleótidos que colores, se debe de llevar a cabo varias reacciones de secuenciación con el desplazamiento de un nucleótido. Este método ofrecía una mayor fiabilidad de secuencia que los sistemas basados en síntesis pero no permitía generar lecturas de gran longitud⁵².

Entre los sistemas que todavía se utilizan en la actualidad cabe destacar los siguientes:

- **Illumina (MiniSeq, MiSeq, NextSeq, HiSeq...)**

Es un sistema basado en síntesis similar al seguido por el sistema 454. Como novedad esta plataforma incluye nucleótidos modificados que solo permiten la adición de un nucleótido en cada ciclo acoplado a una señal fluorescente característica. De esta manera en las plataformas originales la secuenciación era más lenta y permitía lecturas más cortas que la plataforma 454 pero la secuencia tenía mayor calidad. En los últimos años, múltiples mejoras en los tiempos de secuenciación y en la longitud de las mismas han permitido que la tecnología Illumina sea la más utilizada por la comunidad científica.

El proceso comienza con una PCR puente y se sigue de ciclos de secuenciación que consisten en la adición de una mezcla de los 4 tipos de deoxinucleótidos, cada uno de ellos marcado con un fluoróforo diferente y un resto terminador reversible en la posición 3' hydroxyl. Una ADN polimerasa modificada lleva a cabo la adición de nucleótidos, lo cual produce una escisión de marcadores fluorescentes junto con el resto terminador que darán lugar a la formación de imágenes en cuatro canales⁵².

- **Ion Torrent (PGM(*personal genome machine*) y proton)**

Esta metodología se basa en la tecnología del secuenciador 454 de Roche con la modificación de que la incorporación del nucleótido se detecta directamente mediante la liberación del protón durante la unión del nucleótido mediante una placa semiconductor, lo que hace prescindible la utilización de caros y pesados sistemas ópticos de medición. La máquina detecta en cada ciclo el número de protones liberado mediante el cambio de pH producido en el medio, lo que es directamente proporcional al número de bases que se han incorporado⁵⁴.

2.5.1.2.1 Generación de librerías genómicas

Las plataformas de secuenciación masiva requieren colecciones de fragmentos de ADN de un tamaño concreto y con unas secuencias comunes en los extremos (adaptadores) que se utilizan para inmovilizar estos fragmentos al soporte sólido y su secuenciación en paralelo. La generación de estos fragmentos a partir de la muestra de partida es lo que se denomina generación de librerías genómicas.

El protocolo tradicional para la generación de estas librerías consiste en la fragmentación aleatoria del ADN del paciente seguida de la ligación de los adaptadores. Además, el carácter cuantitativo de la secuenciación permite utilizar ADN complementario (cDNA) generado a partir del ARN del paciente para realizar análisis de expresión. Sin embargo en los últimos años ha habido una auténtica explosión de protocolos distintos de generación de librerías que permiten extraer distinta información biológica de las muestras.

Por último, diversas estrategias permiten centrar los esfuerzos de secuenciación en regiones de interés en el ADN. Entre estos métodos cabe destacar la generación de amplicones de las regiones de interés mediante PCR con cebadores conteniendo los adaptadores, o la captura del ADN fragmentado con sondas sintéticas complementarias a las regiones de interés.

En este aspecto, cabe destacar el desarrollo reciente de los métodos de captura mediante la utilización de sondas moleculares de inversión (MIP: *molecular inversion probes*). Esta estrategia se basa en la utilización de moléculas de ADN lineales que, tras hibridar con el ADN, permiten la copia de la región intermedia entre los brazos de homología y la ligación en una molécula circular. Esta molécula es finalmente amplificada mediante PCR para generar un amplicón con las regiones adaptadoras en

ambos extremos. La ventaja de este método frente a la PCR consiste en la baja cantidad de ADN de partida necesario y la posibilidad de introducir cientos de sondas diferentes en la misma muestra ya que no muestran competencia entre ellas (a diferencia de las PCR multiplex). Además, este método incluye la utilización de códigos moleculares únicos de molécula (UMI: *unique molecular identifier*); secuencias de ADN aleatorias introducidas en cada sonda para identificar los eventos iniciales de captura y eliminar posibles errores generados en pasos posteriores durante la generación de librerías (Figura 8).

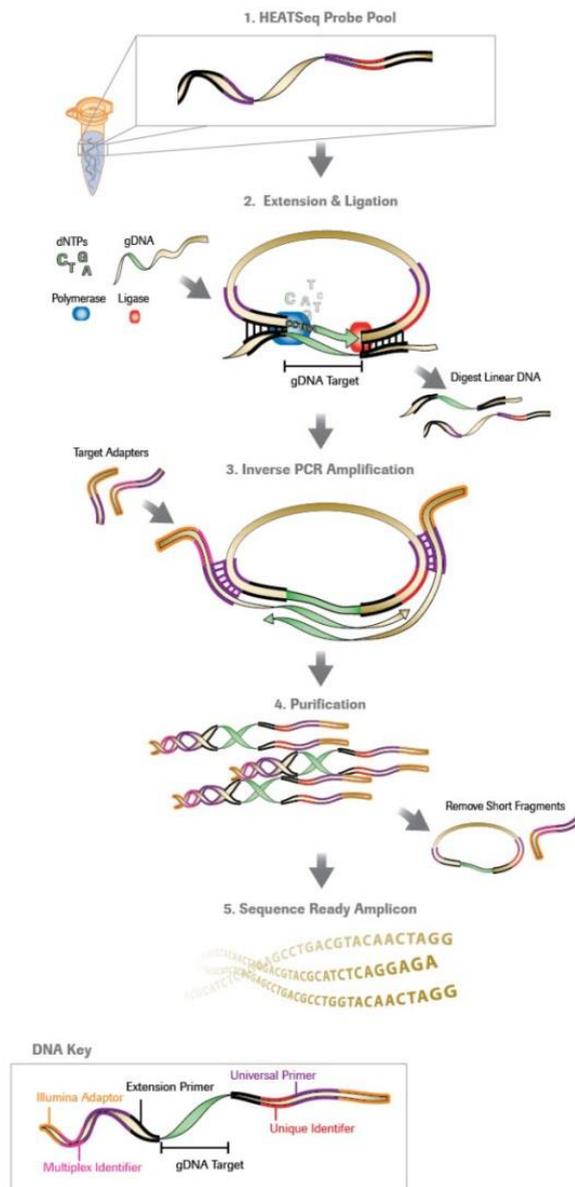


Figura 8 | Sondas de inversión molecular (MIP). Las sondas de captura se componen de dos secuencias diana específicas del gen unidas por una secuencia común y 5 bases aleatorias (identificador molecular, UMI). Las sondas completamente hibridadas se someten a extensión y ligadura para formar estructuras circulares. El tratamiento con exonucleasa elimina el ADN lineal antes de la amplificación por PCR⁵⁵.

2.5.1.2.2 Análisis de datos

Una de las grandes dificultades en la utilización de las nuevas tecnologías de secuenciación masiva es la gran cantidad de datos de secuencia generados que necesitan un procesamiento informático específico. Por lo tanto, acompañando a estos grandes avances en el campo de la secuenciación aparece una nueva generación de herramientas bioinformáticas que se emplean para el análisis de los datos obtenidos en la secuenciación. Existen muchas herramientas bioinformáticas que podemos emplear dependiendo de cuál sea la aplicación que estamos dando a la secuenciación⁵³.

Aunque cada aplicación concreta de análisis requiere un protocolo diferente, la mayoría de los protocolos de análisis comienzan con el alineamiento de las secuencias frente a un genoma de referencia. Este paso consiste en identificar la región concreta del ADN a la que pertenece una determinada secuencia. Este suele ser el paso más lento del análisis y el que requiere mayores recursos informáticos, sin embargo todos los pasos posteriores dependen de un correcto alineamiento de las secuencias.

A continuación distintos programas utilizan la información alineada para identificar cambios en la secuencia del paciente. Los pequeños cambios de secuencia (sustituciones así como pequeñas inserciones y deleciones) se detectan teniendo en cuenta el número de secuencias correctamente alineadas en una determinada región del genoma y que contienen una secuencia distinta al genoma de referencia. Normalmente se utiliza una muestra de ADN no tumoral del mismo paciente para identificar aquellas alteraciones presentes únicamente en el ADN tumoral (somáticas). Para la detección de cambios en el número de copias o grandes reorganizaciones genómicas, diversos autores han desarrollado multitud de estrategias basadas en el carácter cuantitativo de la secuenciación o en las características de posición y orientación de las dos secuencias generadas a partir del mismo fragmento (secuenciación pareada).

2.5.2 Problemas en la incorporación de NGS a la práctica clínica

Aunque estos nuevos avances técnicos llevan a una mejor comprensión de la patogenia de la enfermedad llevando así a una nueva era en la que la investigación en patología molecular y medicina personalizada se complementan, la implementación de estas tecnologías a la práctica clínica se ha encontrado con multitud de dificultades. En primer lugar se necesita una nueva infraestructura de secuenciación e informática así como personal especializado en el uso de estas nuevas técnicas. En segundo lugar, las necesidades de cantidad y calidad del ADN de partida para la generación de las librerías han limitado mucho el uso de ADN extraído de muestras clínicas fijadas y parafinadas. Por último, es imprescindible la adaptación de estas tecnologías a un diagnóstico rápido para la práctica clínica.

Por todas estas razones, en los últimos años se han ido desarrollando protocolos de aplicación a la práctica clínica que permitan la utilización de muestras de menor calidad o menor cantidad de partida, rápidas y de fácil aplicación por los profesionales del diagnóstico clínico^{56,57}.

3 OBJETIVOS

- Poner a punto la técnica de enriquecimiento basada en la utilización de MIPs para su empleo en el diagnóstico molecular del cáncer.
- Demostrar la eficacia de la utilización de UMIs para aumentar la sensibilidad de la detección y eliminar errores cometidos por la polimerasa.
- Comprobar el comportamiento de la técnica en ADN extraído de muestras conservadas en parafina.

4 MÉTODO

Este experimento toma como punto de partida un protocolo diseñado y puesto a punto en el Instituto Sanger (UK) que ha sido ligeramente modificado durante la puesta a punto del mismo.

4.1 PROTOCOLO MIPs

4.1.1 Diseño de los MIPs

Para el diseño de los MIPs utilizamos el programa MIPgen⁵⁸. Diseñamos 5 MIPs para enriquecer las regiones correspondientes a los codones 12-13 de KRAS y NRAS, la región correspondiente al codón 61 de NRAS, así como las regiones correspondientes al exón 19 y al codón 79 de EGFR.

4.1.2 Fosforilación

Mezclamos el MIP/oligo con una quinasa para obtener el extremo 5'-fosfato

1 µL MIP 1 (100µM)	
1 µL MIP 2 (100µM)	
1 µL MIP 3 (100µM)	
1 µL MIP 4 (100µM)	
1 µL MIP 5 (100µM)	
2 µL T4 Ligasa Buffer 10x	
1 µL T4 PNK	
12 µL H ₂ O	
<hr/>	
20 µL	

Termociclador:

37°C 45'
65°C 20'
4°C ∞

4.1.3 Ligación

Copia de nuestro ADN de interés entre ambos adaptadores y circularización del MIP.

200ng DNA
1 μ L mix-MIPs 0,1 nM
0,8 μ L dNTPs (0,01mM)
2,5 μ L Buffer Ampligasa 10x
0,16 μ L NEB Klentaq
0,20 μ L Ampligasa
Hasta 25 μ L H₂O

25 μ L

Termociclador:

98°C 3'
60°C 24h

4.1.4 Digestión exonucleasas

Digestión con exonucleasas para eliminar todo el ADN que no haya circularizado.

0,25 μ L Exo I
0,25 μ L Exo III
0,2 μ L Buffer Ampligasa 10x
1,3 μ L H₂O

2 μ L

Termociclador:

37° 2h
95° 2'
4° ∞

4.1.5 MIP PCR

Amplificación de los productos de la captura.

5 μ L Ligación+exonucleasa
2,5 μ L Buffer DreamTaq polimerasa (x10)
0,25 μ L DreamTaq polimerasa
2,5 μ L Oligo 1 (10 μ M): Sanger_MIP_F
2,5 μ L Oligo 2 (10 μ L): SangeriPCRtag1
0,5 μ L dNTPs
11,75 μ L H₂O

25 μ L

Termociclador:

98°C 3'

98°C 15"

60°C 30"

72°C 30"

35 Ciclos

72°C 5'

4°C ∞

4.2 PURIFICACIÓN DE LOS FRAGMENTOS DE PCR

Se llevó a cabo con Agencourt AMPure XP e hizo posible eliminar las bandas inespecíficas que alteraban los resultados.

1. Calentar las bolas de Ampure a temperatura ambiente y mezclar bien antes de usar.
2. Preparar 3.6 ml de etanol al 70% para cada muestra. La solución de etanol al 70% se debe preparar fresca.
3. Agregar el volumen adecuado de bolas de Ampure a cada muestra, mezclar por inversión durante 10 minutos a temperatura ambiente. La relación de volumen de bolas/volumen de DNA determinará el tamaño de fragmentos a purificar. Como regla general una relación 1.8:1 bolas:DNA (1.8x) purifica DNA de cualquier tamaño. En muchas aplicaciones en este trabajo hemos purificado con una relación 0.8:1 (0.8x) o 0.7:1 (0.7x) para eliminar fragmentos de DNA de pequeño tamaño.
4. Colocar las muestras en un separador magnético. Cuando las bolas se hayan acumulado en la pared del tubo y la solución esté limpia, transferir el líquido a un tubo nuevo sin mover las bolas. El líquido contiene el ADN separado en el extremo.
5. Agregar 300 µl de etanol al 70%. Lavar las cuentas invirtiendo el tubo y permitiendo que estas vuelvan a recogerse en el costado del tubo. Invertir el tubo 6 veces.
6. Retirar y desechar el etanol.
7. Repetir los pasos 5 y 6 dos veces más.
8. Retirar los tubos del separador magnético, invertir el tubo rápidamente una vez más, volver a colocar el imán y eliminar el líquido restante.

9. Mantener los tubos en el imán y las tapas abiertas, secar las perlas a temperatura ambiente durante 20-30 minutos. Se observarán grietas en el pellet de las perlas cuando se complete el secado.
10. Agregar 30 μL de agua a las perlas secas y agitar en vórtex para mezclar bien.
11. Colocar las muestras en un separador magnético, cuando las cuentas se hayan acumulado en la pared del tubo y la solución esté clara, transferir el líquido a un tubo nuevo. El líquido contiene los fragmentos de DNA purificados

4.3 ELECTROFORESIS

Para comprobar los resultados tras cada PCR, fueron realizadas electroforesis en gel de agarosa al 2% en las que se cargaron 10 μL y se programaron con una duración de 30 min y una potencia de 120 mV.

Si la técnica ha funcionado correctamente, hemos de observar bandas alrededor de 300 pb. Esto se debe a la suma de los siguientes elementos:

- La zona que tiene que sintetizar la polimerasa para cerrar el MIP (en el caso del primer MIP del Sanger) es de unos 120pb
- El tamaño en si del MIP (~90 pb).
- El resto del adaptador de Illumina que se añade en el paso de PCR (~80 pb)

4.4 PROTOCOLO EXTRACCIÓN DE ADN A PARTIR DE MUESTRAS EN PARAFINA

Tras las pruebas con líneas celulares se procede a la utilización de ADN extraído de muestras en parafina cuyas mutaciones habían sido analizadas previamente por qPCR en el servicio de anatomía patológica del Hospital Universitario Marqués de Valdecilla (HUMV).

La Extracción del DNA se llevó a cabo con el kit comercial Cobas *DNA Sample Preparation Kit* según el siguiente protocolo:

1. Corte de cada bloque de parafina de 5 secciones de 5 micras de grosor e introducción en un eppendorf de 2 ml previamente identificado.
2. Anadir 1000 μL de xilol y mezclar en vórtex hasta completa disolución. Centrifugar a máxima potencia y temperatura ambiente durante 5 minutos. Eliminar el xilol con cuidado de no arrastrar el precipitado.
3. Anadir 500 μL de xilol y 500 μL de etanol absoluto y mezclar en vórtex hasta completa disolución. Centrifugar a máxima potencia y

temperatura ambiente durante 5 minutos. Eliminar con cuidado de no arrastrar el precipitado.

4. Añadir 1000 μL de etanol absoluto sobre el pellet y mezclar en vórtex. Centrifugar a máxima potencia y temperatura ambiente durante 5 minutos. Descartar el etanol con cuidado de no arrastrar el pellet.

5. Incubar a 56 $^{\circ}\text{C}$ durante 5 minutos para eliminar los restos de etanol.

6. Añadir 180 μL de buffer ATL y 70 μL de Proteinasa K. Mezclar por agitación durante 15 segundos. Incubar a 50 $^{\circ}\text{C}$ en agitación durante 60 minutos hasta lisis completa de los tejidos, e incubar a 90 $^{\circ}\text{C}$ en agitación durante 60 minutos.

7. Dar un pulso de centrifuga. Añadir 200 μL de buffer PBB y mezclar por agitación durante 15 segundos. Incubar a RT $^{\circ}\text{C}$ durante 10 minutos.

8. Dar un pulso de centrifuga. Añadir 100 μL de isopropanol y mezclar por agitación. Filtrar la mezcla obtenida a través de una columna. Centrifugar a 8000 r.p.m. durante 1 minuto.

9. Pasar la columna a un tubo limpio y desechar el tubo con el eluido. Añadir sobre la columna 500 μL de Wash buffer 1. Centrifugar a 8000g durante 1 minuto.

10. Pasar la columna a un tubo limpio y desechar el tubo con el eluido. Añadir sobre la columna 500 μL de Wash buffer 2. Centrifugar a 8000g durante 1 minuto.

11. Pasar la columna a un tubo limpio y desechar el tubo con el eluido. Centrifugar a máxima velocidad durante 2 minutos.

12. Colocar la columna sobre un nuevo tubo eppendorf con tapa debidamente rotulado. Añadir 50 μL de buffer EB. Incubar durante 5 minutos a temperatura ambiente y centrifugar a 8000 r.p.m. durante 1 minuto.

Medir la concentración del DNA obtenido mediante Qubit[®]

4.5 SECUENCIACIÓN

Las librerías genómicas se secuenciaron en un plataforma MiSeq[®] de Illumina utilizando un protocolo de 150 PE con un kit Illumina Nano que produce 1 millón de secuencias por carrera. Durante la secuenciación se mezclaron las librerías con un 25% de una librería control generada a partir del fago PhiX que funciona como calibrador interno.

5 RESULTADOS

5.1 PUESTA A PUNTO DEL PROTOCOLO

En primer lugar, para poner a punto el protocolo de captura mediante MIPs llevamos a cabo una prueba con tres MIPs diferentes por separado (A, B y C), los cuales presentaban una misma región de reconocimiento genómica publicada por el Instituto Sanger y que denominaremos de aquí en adelante MIP Sanger⁵⁹, pero diferentes regiones adaptadoras para realizar la PCR. Además en vez de empezar probando con muestras de ADN conservadas en parafina, se decidió comenzar con DNA extraído de una línea celular de carcinoma de pulmón humana A549. En el análisis de los resultados de la PCR podemos ver que no existe la banda esperada cerca de los 300 pb, pero sí que existe una banda inespecífica de unas 120 pb (Figura 9).

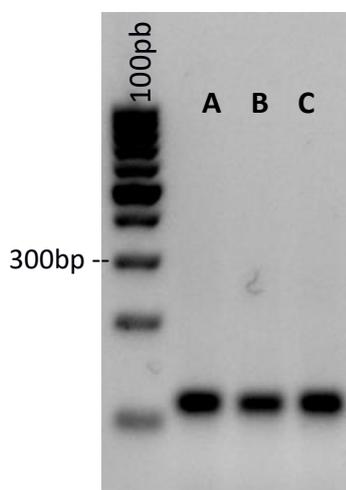


Figura 9 | Electroforesis en gel de agarosa de los productos PCR obtenidos con tres tipos de MIPs (A, B y C). Se muestra la ausencia de la banda esperada ~300pb y presencia de una banda inespecífica de ~120pb.

Dados los resultados en la fase anterior, para descartar un posible problema con la calidad del ADN extraído, se decidió emplear ADN de otras líneas celulares (Caco2, HL60, H1299, A549) que ya habían sido empleadas en el laboratorio para otros fines con buenos resultados. Además debido a los anteriores experimentos fallidos con el MIP Sanger (-S), se decidió también realizar el protocolo con una mezcla de 5 MIPs diferentes (-M) diseñados en nuestro laboratorio para las mutaciones de interés (*KRAS*, *NRAS* y *EGFR*), observándose en el gel que sí que aparece la banda esperada de ~300 pb en aquellos casos que se ha empleado la mezcla de MIPs. De esta manera se confirma que el MIP_Sanger (-S) que se estaba utilizando no era eficaz. Así mismo, aparecía una banda inespecífica de ~120 pb que aparece también en el control negativo con MIP_Sanger. Hipotetizamos que esa banda podría deberse a la amplificación de productos de ligación inespecíficos de pequeños tamaño, por lo que decidimos purificar la mitad del producto de la ligación mediante bolas Agencourt AMPure XP (#093A63881, Beckman Coulter) al

1x antes de realizar la PCR. Sin embargo perdíamos demasiado ADN y la banda se seguía viendo, por lo que se estableció que era mejor no purificar (Figura 10).

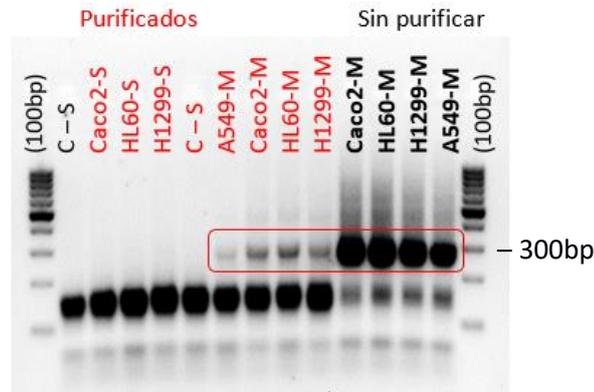


Figura 10 | Electroforesis en gel de agarosa comparativa del MIP Sanger y mezcla de MIPs. Utilización de las líneas celulares Caco2, HL60 y H1299 con MIP Sanger (-S) y una mezcla de 5 MIPs diferentes (-M). En aquellas muestras con la mezcla de MIPs se observan tanto la banda esperada de ~300pb como la banda inespecífica de ~120pb, esta anterior con bastante intensidad, por lo que se purifica el producto de PCR, viendo que esto amplificaba la banda inespecífica y atenuaba la de interés. Control negativo con MIP Sanger.

Se tenía la duda de si realmente era posible la utilización de un solo MIP para la captura o si era imprescindible utilizar diversos MIPs en la misma. Es por ello que se hizo una prueba con el MIP_Sanger y los 5 MIPs que componían el mix de MIPs por separado. Aquí se observa que el problema estaba únicamente en el MIP_Sanger, ya que en las reacciones con los MIPs por separado sí que aparece la banda esperada. En el control negativo con MIP_Sanger, como es de esperar tampoco aparece la banda de ~300pb. Además se observa una intensidad diferente en la banda de los MIPs, lo cual nos indica que tienen un rendimiento diferente. (Figura 11).

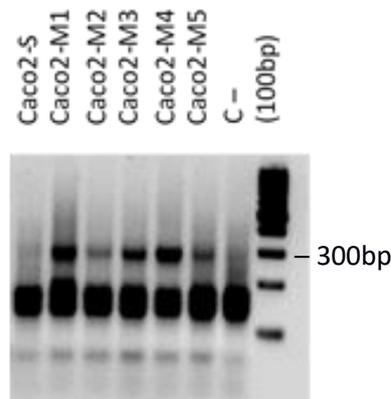


Figura 11 | Electroforesis en gel de agarosa utilizando los MIPs de la mezcla por separado. Se observa que los MIPs que han sido empleados en la mezcla también funcionan si se utilizan por separado. Las bandas esperadas presentan intensidades diferentes para los diferentes MIPs, mientras que están ausentes en el control negativo y en el MIP Sanger.

5.2 GENERACIÓN DE LIBRERIAS

Llegados a este punto, y con una versión del protocolo que da buenos resultados se procede a la realización del mismo empleando las muestras de ADN en parafina. Para extraerlo se sigue el protocolo mencionado previamente y una vez extraído, se lleva a cabo una cuantificación de las muestras con Qubit 4 Fluorometer[®] (Tabla 1).

<u>Nº muestra</u>	<u>ng/μl</u>
<u>Control ADN parafina</u>	
1	3,58
2	<
3	<
4	5,9
5	<
6	2,83
7	<
8	2,75
9	4,99
10	<
11	<
12	20,8
<u>Muestras EGFR</u>	
13	3
14	14,8
15	<
16	4,31
17	14,4
18	14,5
19	120
20	11,8
<u>Muestras KRAS, NRAS</u>	
21	46,9
22	<
23	<
24	3,44
25	7,21
26	2,39
27	6,59
28	87,6
29	17,2
30	25,9
31	32,9
32	4,36

Tabla 1 | Muestras de ADN extraídas de parafina con su cuantificación obtenida a través de Qubit[®]. En verde, las muestras que finalmente se enviaron a secuenciar.

Empezamos empleando aquellas con mayor cuantificación, las cuales presentaban las siguientes mutaciones (Tabla 2):

Muestra	Tag	Mutación detectada previamente	Cuantificación
19	1	EGFR (Del19)	120
14	2	EGFR (Del19)	14,8
18	3	EGFR (Del19, T790)	14,5
17	4	EGFR (Del19)	14,4
28	5	NRAS G13R/V	87,6
32	6	KRAS 61	4,36
29	8	NRAS G13R	17,2
31	9	KRAS G12D 30% GGT>GAT	32,9
27	10	KRAS G12C	6,59

Tabla 2 | Datos de las muestras enviadas a secuenciar. Aquí podemos observar el número de las muestras, el etiquetado que se le añadió por cada paciente (tag) y la mutación que había sido detectada previa a esta secuenciación.

Para la generación de librerías se llevaron a cabo tres PCRs de las cuales se obtuvo el ADN para la secuenciación (Figura 12):

En el gel de arriba a la izquierda (a) se visualiza una primera prueba con 35 ciclos de PCR en la cual existe claramente la banda esperada en la muestra 29 (M-29) y un hallazgo inesperado en la muestra 24 (24-M), en la cual no aparece la banda entre 100-200pb, ni la banda esperada. Este último hallazgo nos indica que quizá ese tag no funcione bien.

En el gel b, se pueden observar los resultados de la PCR con muestras que presentan una buena cuantificación en el análisis con Qubit 4 Fluorometer®, en la que se decidió reducir el número de ciclos a 30, ya que a 35 ciclos las bandas que se veían eran demasiado intensas. Además empleamos Caco2 como control positivo y realizamos una prueba con esta muestra empleando un menor número de ciclos (x25) y un menor volumen (19µl), en la cual se obtiene una banda pero mucho menos intensa. En cuanto a los resultados en las muestras se observa que las muestras 19, 27 y 32 presentan claramente la banda esperada, y una vez más destaca el hallazgo de que en la muestra 24 con el tag 7, en la cual no se observa la banda esperada ni la existente entre 100 y 200 pb.

En aquellos del gel b que no se observa la banda esperada (en dorado) se decide realizar otra PCR con 35 ciclos (c) en vez de 30, para ver si así se produce una mayor amplificación y se observa banda. Especialmente en las muestras 18 y 28 sí que se observa que esta vez aparece la banda.

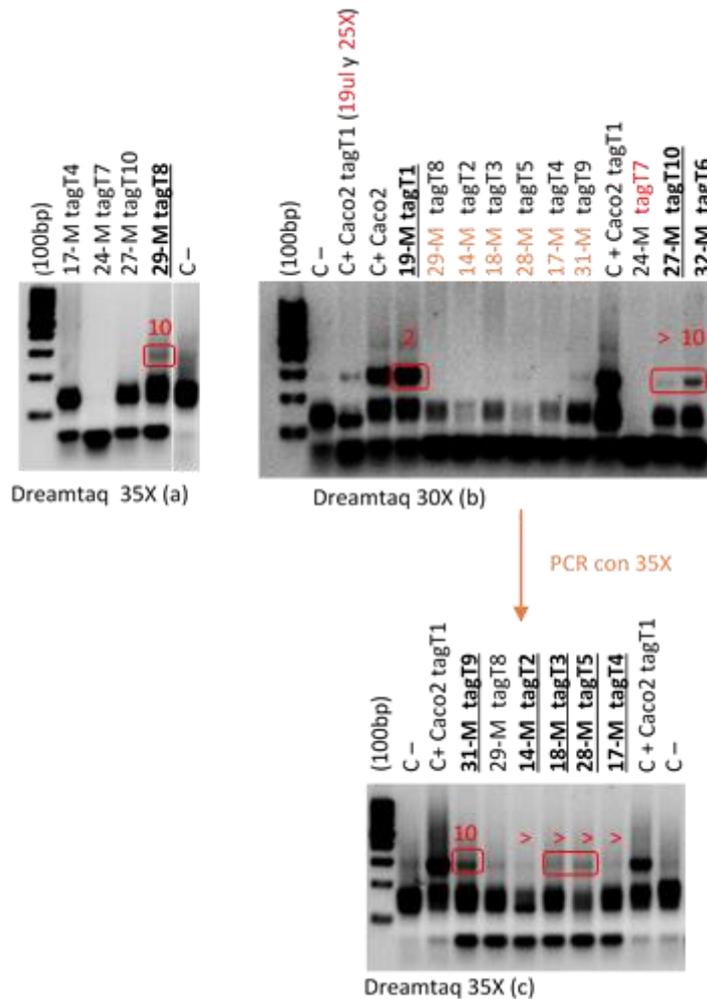


Figura 12 | Electroforesis en gel de agarosa de los productos de tres PCRs realizadas para la generación de la librerías. Muestras extraídas de parafina marcadas con su tag correspondiente, en negrita y subrayadas, aquellas que fueron enviadas a secuenciar. Señalado en rojo podemos ver las bandas esperadas que se detectan y los volúmenes que serán introducidos de cada muestra a la secuenciación (con el símbolo >, aquellos de los cuales introduciremos el máximo volumen que tengamos).

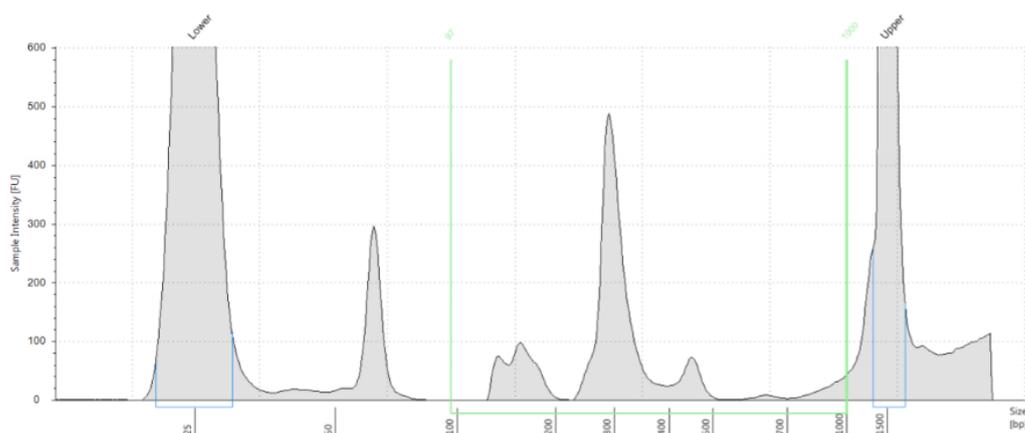
Por último se lleva a cabo la mezcla de las muestras para enviar a secuenciar, de manera que se decide el volumen de cada una en función de la intensidad de la banda que presentan (Figura 12, en rojo). Por ello se decide meter en el tubo de secuenciación:

- 2µl: 19-M (tag T1).
- 10µl: 29-M (tag T8); 32-M (tag T6); 31-M (tag T9).
- Máximo disponible (~25µl): 14-M (tag T2); 18-M (tag T3); 28-M (tag T5); 17-M (tag T4).

A continuación se añadieron los adaptadores de Illumina a la mezcla y se incorporaron mediante una PCR.

5.3 RESULTADOS DE LA SECUENCIACIÓN

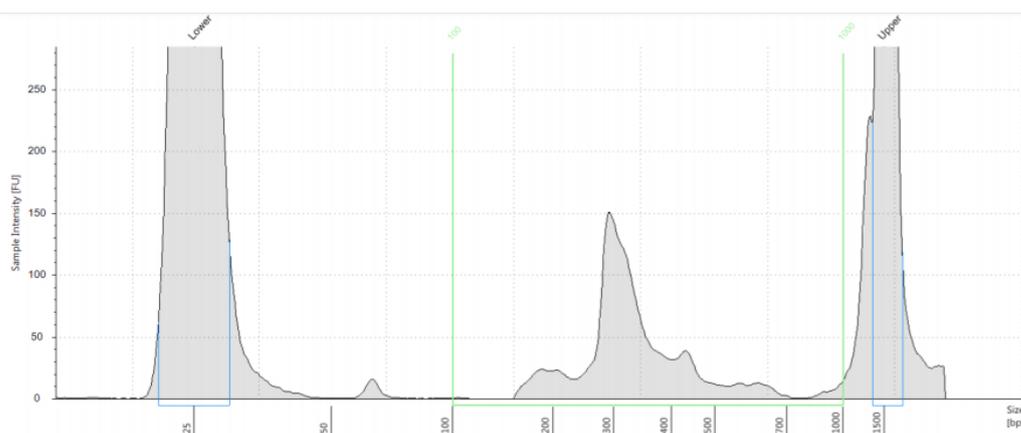
En primer lugar se secuenció la mezcla de librerías purificada al 0,8x con Agencourt, en la cual se detectaron interferencias en la secuenciación que se cree son debidas al pico que aparece en la imagen de ScreenTape® de fragmentos de pequeño tamaño (Figura 13), por lo que se decide purificar más la muestra al 0,7x consiguiéndose eliminar casi por completo esta interferencia, aunque eliminando parte de nuestro ADN de interés (Figura 14).



Region Table

From [bp]	To [bp]	Average Size [bp]	Conc. [ng/μl]	Region Molarity [nmol/l]	% of Total	Region Comment	Color
97	1000	322	2.32	13.1	51.55		■

Figura 13 | ScreenTape® con muestra purificada al 0,8x. Se observan bandas inespecíficas entre 50 y 200 pb, y nuestra banda esperada a ~300pb.



Region Table

From [bp]	To [bp]	Average Size [bp]	Conc. [ng/μl]	Region Molarity [nmol/l]	% of Total	Region Comment	Color
100	1000	350	1.16	5.67	59.22		■

Figura 14 | ScreenTape® con muestra purificada al 0,7x. Desaparición casi por completo de las bandas inespecíficas que se encontraban entre 50 y 200pb. Se observa la banda esperada de 300pb.

Tras este paso, se secuencian las librerías de ADN obteniéndose buenos resultados y llevando a cabo un correcto análisis de control de calidad del mismo, en el cual se registran los siguientes datos:

En primer lugar obtuvimos un total de 1.3 millones de secuencias y el porcentaje de lecturas identificadas fue del 78,22%, correspondiendo a cada muestra uno de los siguientes porcentajes (Figura 15):

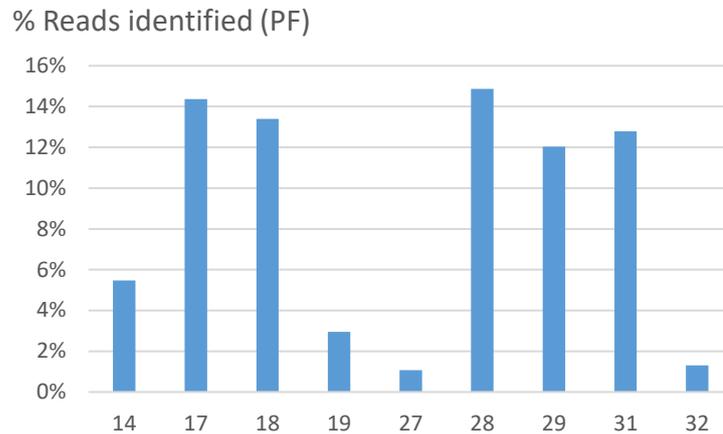


Figura 15 | Porcentajes de lecturas identificadas que provienen de las diferentes muestras.

Así mismo analizamos también la calidad de las secuencias obtenidas y observamos buenos resultados, ya que las secuencias mantenían una alta calidad (Phred score > 20, 99% fiabilidad) en prácticamente la totalidad de la secuencia y teníamos alrededor de un 90% de secuencias con un valor de calidad medio >30 (99.9 % fiabilidad).

Per base sequence quality

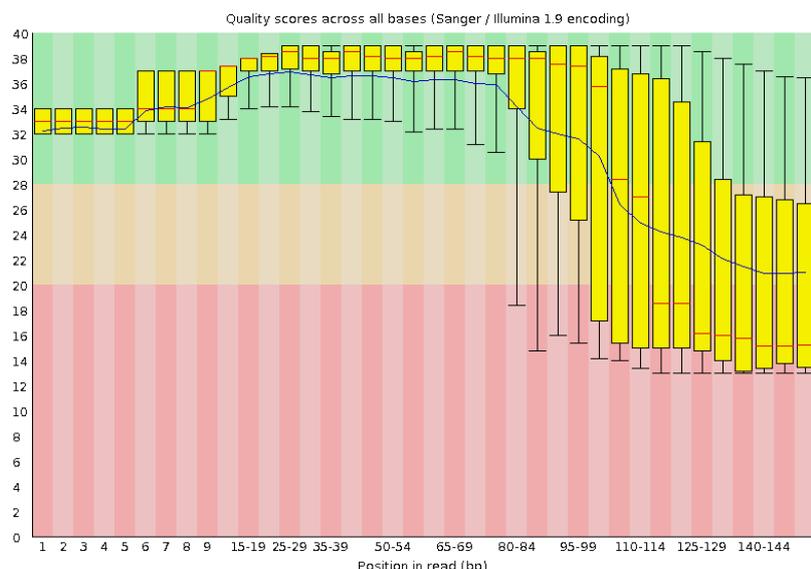


Figura 16 | Calidad de la secuenciación. Con una línea roja la mediana. En amarillo los valores entre percentiles 25 y 75. Al final de los bigotes, valores máximo y mínimo de cada serie.

Con un análisis de control de calidad favorable procedemos al análisis de las variantes genómicas:

- Análisis de sustituciones de base EGFR T790M, NRAS G13R y KRAS (G12C, G12D y Q61L)

En estos resultados lo que hicimos fue observar el número de lecturas de cada base y el número de UMIs diferentes que contenían una concreta base en cada muestra, de manera que si un solo UMI contenía la base de la mutación, este cambio de base se consideraba un error de la polimerasa cometido durante la PCR, mientras que si encontrábamos la base característica de la mutación en muchos UMIs diferentes sabemos que esta se trata de una mutación.

Muestra	Mutación	Sustitución	Lecturas_A	Lecturas_C	Lecturas_G	Lecturas_T	UMIs_A	UMIs_C	UMIs_G	UMIs_T	Frecuencia
18	EGFR-T790M	C→T	3	10761	2	324	0	326	0	23	0,02921551
	KRAS-G12D	C→T	23	5315	5	10	3	267	0	3	0,00186741
	KRAS-G12C	C→A	10	3720	3	1617	1	228	0	106	0,00186916
	NRAS-G13R	C→G	10	6267	38	8	2	233	8	3	0,00600506
	KRAS-Q61L	T→A	12	88	4	6411	5	11	2	276	0,00184162
19	EGFR-T790M	C→T	0	11176	0	34	0	953	0	9	0,00303247
	KRAS-G12D	C→T	33	18544	6	62	6	1003	0	18	0,00332386
	KRAS-G12C	C→A	18	18557	7	71	5	1003	0	22	0,00096499
	NRAS-G13R	C→G	9	9174	5	12	2	934	3	5	0,00054348
	KRAS-Q61L	T→A	16	57	2	7654	6	27	0	853	0,00206932
14	EGFR-T790M	C→T	2	3834	0	9	0	137	0	2	0,0023407
	KRAS-G12D	C→T	12	3445	3	6	1	178	1	2	0,0017311
	KRAS-G12C	C→A	6	2652	4	803	1	153	0	40	0,0017316
	NRAS-G13R	C→G	9	2963	18	1	1	149	4	1	0,00601604
	KRAS-Q61L	T→A	1	26	3	4961	0	6	1	142	0,00020036
32	EGFR-T790M	C→T	0	3261	0	4	0	423	0	1	0,00122511
	KRAS-G12D	C→T	10	4151	0	35	2	512	0	7	0,0083254
	KRAS-G12C	C→A	2	3826	2	374	0	489	1	60	0,00047574
	NRAS-G13R	C→G	2	3230	3	10	0	483	1	3	0,0009245
	KRAS-Q61L	T→A	103	24	2	1975	15	7	0	297	0,04895437
28	EGFR-T790M	C→T	4	20574	3	101	0	636	0	8	0,00488347
	KRAS-G12D	C→T	34	17028	12	102	7	748	0	6	0,00593852
	KRAS-G12C	C→A	13	16202	5	956	0	737	0	77	0,00075687
	NRAS-G13R	C→G	30	11796	8034	91	6	490	411	11	0,40262604
	KRAS-Q61L	T→A	6	202	6	9714	1	21	0	445	0,00060435
29	EGFR-T790M	C→T	1	9292	5	81	0	326	0	9	0,00863632
	KRAS-G12D	C→T	16	5538	2	17	2	347	0	0	0,00304714
	KRAS-G12C	C→A	11	4864	5	695	2	327	1	68	0,00197309
	NRAS-G13R	C→G	13	4768	93	11	2	244	16	1	0,01902619
	KRAS-Q61L	T→A	3	101	1	5565	1	12	0	297	0,00052901
27	EGFR-T790M	C→T	1	2684	0	3	0	146	0	0	0,00111607
	KRAS-G12D	C→T	11	3652	2	30	0	189	0	3	0,00811908
	KRAS-G12C	C→A	20	2487	2	1186	2	146	0	59	0,00541272
	NRAS-G13R	C→G	0	1601	1	3	0	100	1	1	0,00062267
	KRAS-Q61L	T→A	2	34	4	1753	0	9	1	106	0,00111545
31	EGFR-T790M	C→T	3	12664	1	23	0	395	0	6	0,00181231
	KRAS-G12D	C→T	19	9201	6	509	2	412	0	72	0,05228557
	KRAS-G12C	C→A	9	9040	4	678	1	426	0	72	0,00092488
	NRAS-G13R	C→G	8	8021	26	6	3	314	8	2	0,00322421
	KRAS-Q61L	T→A	141	97	9	7975	13	24	1	302	0,01714703
17	EGFR-T790M	C→T	2	4832	1	2	0	134	0	0	0,00041348
	KRAS-G12D	C→T	17	4396	1	4	2	152	0	2	0,00090539
	KRAS-G12C	C→A	10	3300	4	1097	0	129	0	52	0,00226706
	NRAS-G13R	C→G	9	3399	35	7	1	105	6	2	0,01013611
	KRAS-Q61L	T→A	0	14	2	3122	0	6	1	100	0

Tabla 3 | Resultados de la secuenciación de mutaciones con cambio de base. En naranja las mutaciones que presenta cada muestra.

La descripción de los hallazgos es la siguiente (Tabla 3):

- Muestra 18
En EGFR T790M nos encontramos con un número de 324 lecturas con la base T y para la misma 23 UMIs diferentes, lo cual nos indica que esto se trata de una mutación, que además se encuentra en una frecuencia del 2,92%.
- Muestras 19 y 14
No presentan mutaciones de cambio de base.
- Muestra 32
Presenta la mutación KRAS Q61L con presencia de la base A característica de la mutación en un total de 103 lecturas acompañada de 15 UMIs diferentes y en una frecuencia del 4,89%.
- Muestra 28
Existe mutación en NRAS G13R en 8034 lecturas con un total de 411 UMIs diferentes, constituyendo una mutación con frecuencia del 40,26%.
- Muestra 29
Se puede observar la mutación NRAS G13R con un total de 93 lecturas con base G y 16 UMIs diferentes asociados a la misma. Se observa con una frecuencia del 1,9%.
- Muestra 27
Presenta la mutación KRAS G12C con cambio de C por A en 1186 lecturas, pudiendo encontrarse este cambio para un total de 59 UMIs diferentes. La frecuencia de mutación es de un 0,54%.
- Muestra 31
Nos encontramos con dos mutaciones:
 - KRAS G12C: se observa la base T característica de la mutación en 509 lecturas asociada a un total de 72 UMIs diferentes. La frecuencia de la mutación es de 5,23%.
 - KRAS Q61L: la base A de la mutación se identifica en 141 lecturas asociada a 13 UMIs diferentes en una frecuencia mutacional del 1,71%.
- Muestra 17
Se identifica la mutación NRAS G13R con un total de 35 lecturas que presentan G asociadas a un total de 6 UMIs diferentes; encontrando una frecuencia de mutación del 1,01%.

En resumen nos encontramos con la presencia de mutaciones puntuales en los siguientes casos (Figura 17):

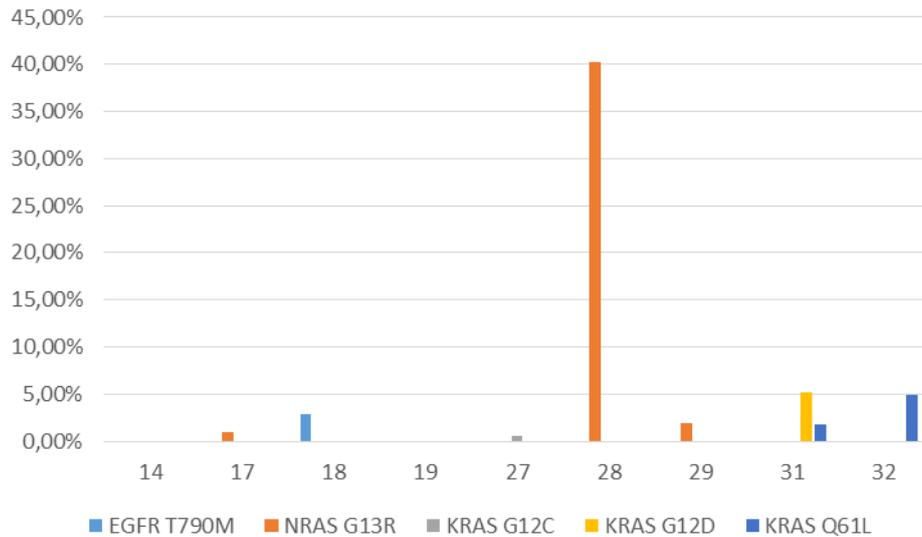


Figura 17 | Frecuencia de las mutaciones detectada en cada una de las muestras.

- **Análisis EGFR (Del 19)**

En este caso el análisis es diferente, ya que se lleva a cabo de forma cuantitativa. Lo que se hace es mirar del total de lecturas de cada muestra, en cuántas de estas se encontraban en el exón 19, estableciendo así la frecuencia del mismo en las lecturas (Figura 18), de manera que aquellas muestras con una frecuencia de lectura del exón 19 baja probablemente presenten células con delección del exón 19. En este caso se llevó a cabo el análisis de manera convencional y también cuantificando el número de UMIs diferentes que hay para el exón19, pero los resultados fueron muy similares.

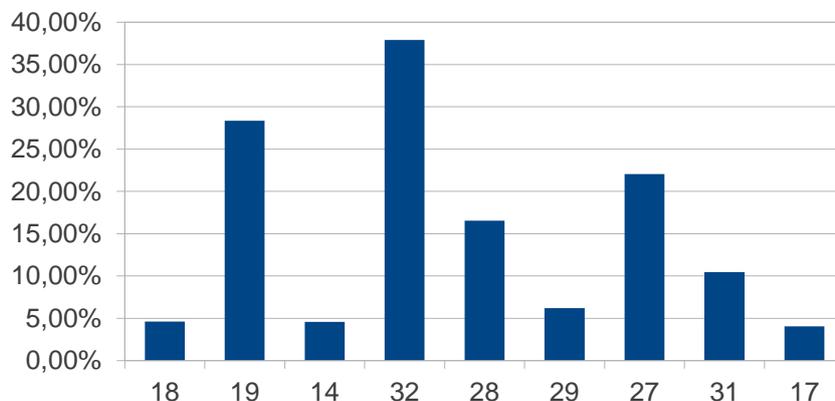


Figura 18 | Porcentaje de lecturas que presentan el exón 19.

Estos resultados nos confirman la presencia de la delección del exón 19 en 14,18 y 17, ya que el número de lecturas es mucho menor que en el resto de las muestras. Sin embargo, en la muestra 19, a pesar de haberse encontrado la mutación por qPCR, se observa un alto número de lecturas en este exon. Como hallazgo podemos añadir que podría existir la mutación en la muestra 29.

6 DISCUSIÓN

6.1 CONTROL DE CALIDAD EN EL DNA

En la cuantificación de las muestras en parafina se evidencia uno de los grandes problemas que limitan la aplicación de esta técnica a la práctica diaria, y es la gran degradación del ADN que provoca este método de conservación. Como se puede observar en la tabla 4, partimos de concentraciones bastante bajas, de hecho, antes de comenzar, se esperaba una concentración aproximada de 200 ng/μl por muestra (Tabla 4).

Cabe destacar que no existe una correlación muy clara entre el número de lecturas identificadas en la secuenciación y la cuantificación previa del ADN. Esto puede deberse a que la cuantificación no identifica bien la calidad que puede tener el ADN, este podría ser de mala calidad o estar modificado químicamente, haciendo imposible una correcta elaboración del molde.

Muestra	Cuantificación	Lecturas identificadas (%)
14	14,8	5,4645%
17	14,4	14,3637%
18	14,5	13,3917%
19	120	2,9438%
27	6,59	1,0667%
28	87,6	14,8626%
29	17,2	12,0391%
31	32,9	12,7814%
32	4,36	1,3064%

Tabla 4 | Cuantificación y número de lecturas identificadas de cada muestra. Se observa que no existe una proporcionalidad. El ejemplo más claro es 19, que a pesar de tener una cuantificación muy alta, presenta uno de los porcentajes de lectura más bajos. En rojo aquellos con porcentajes bajos para su cuantificación.

6.2 PROBLEMAS EN LA GENERACIÓN DE LIBRERÍAS

- **Fallo MIP Sanger**

Este MIP fue enviado desde el Instituto Sanger para la realización de la puesta a punto del protocolo, pero no funcionaba, no aparecía la banda de ~300pb. Por este motivo se decidió probar con 2 rediseños en los que lo que cambiaba era la región adaptadora al suponerse que el problema se producía en el último paso de PCR. Pero estos tampoco funcionaban, lo cual demuestra que el problema era seguramente la secuencia adaptadora que hibrida con el genoma de referencia (Figuras 9, 10, 11).

- **Bandas inespecíficas**

Estas son fragmentos de ADN que se observan tanto en el ScreenTape® como en las electroforesis, que no corresponden al de interés y que nos interferían en la secuenciación, requiriendo por ello purificación para obtener así unos resultados más fiables. La optimización en el diseño del loop adaptador del MIP, podría acabar con la aparición de estas bandas, ya que estos son probablemente el resultado de amplificaciones inespecíficas durante el último paso de PCR.

- **Tag7 no funciona**

Podría ser un problema en el primer de la PCR, el cual no se une a esta secuencia y por lo tanto no permite su replicación (Figura 12).

6.3 RESULTADO SECUENCIACIÓN

- **Cuántas de las mutaciones conocidas se han validado y en cuántos casos han aparecido nuevas mutaciones**

Los resultados obtenidos en la detección de mutaciones puntuales han sido muy positivos, ya que todas las mutaciones han sido confirmadas (sensibilidad 100%) e incluso han sido detectadas algunas no detectadas previamente por los métodos de laboratorio (Tabla 5):

- NRAS G13R en la muestra 17
- KRAS Q61L en la muestra 31
- EGFR Del19 en la muestra 29

Por otro lado nos encontramos con que la delección 19 en EGFR de la muestra 19 no ha sido confirmada, suceso sobre el que se hablará a continuación.

La utilización de los UMIs claramente nos ayuda en el análisis de secuenciación a localizar la presencia de mutaciones en frecuencias incluso menores al 1%, como ocurre en la muestra 27. Esto aumenta la sensibilidad de los métodos actuales de diagnóstico y permite la

caracterización de muestras con una alta contaminación de ADN de tejido normal.

Muestra	Mutación detectada previamente	Mutaciones confirmadas	Mutaciones nuevas
19	EGFR (Del19)		
14	EGFR (Del19)	EGFR Del19	
18	EGFR (Del19, T790)	EGFR (T790M,Del19)	
17	EGFR (Del19)	EGFR Del19	NRAS G13R
28	NRAS G13R/V	NRAS G13R	
32	KRAS Q61L	KRAS Q61L	
29	NRAS G13R	NRAS G13R	EGFR Del19
31	KRAS G12D	KRAS G12D	KRAS Q61L
27	KRAS G12C	KRAS G12C	

Tabla 5 | Resumen de mutaciones previas detectadas por qPCR y mutaciones encontradas y confirmadas con la secuenciación.

- **Problema en la detección de EGFR (Del19)**

En este caso el problema que tenemos es que se lleva a cabo un análisis cuantitativo en el cual tan sólo se puede comparar la cantidad de copias del exón 19 de EGFR que se encuentra en las muestras que ya tenían la mutación en comparación con las que no. Este método probablemente no sea demasiado adecuado teniendo en cuenta que las deleciones en el exón 19 pueden ser lo suficientemente pequeñas como para no interferir con el alineamiento de las lecturas y por lo tanto es necesario diseñar un protocolo de identificación más adecuado para este tipo de mutaciones.

En el caso de las muestras 14, 18 y 17 sí que coincide la presencia previa de la mutación con la obtención de un bajo número de copias del exón 19 en la secuenciación. Por el contrario en muestras como la 19, que debería de tener niveles bajos, nos encontramos con que estos son bastante altos (Figura 18). La identificación de pequeñas inserciones y deleciones es un problema de gran complejidad en los datos de secuenciación masiva en donde hay un intenso debate. Por esa razón, para este tipo de mutaciones no parece que la secuenciación masiva ofrezca en estos momentos una gran ventaja frente a los métodos actuales de genotipado mediante qPCR.

7 CONCLUSIÓN

En este TFG hemos demostrado que la estrategia de utilizar MIPs para el diagnóstico genético combinados con NGS funciona especialmente en el análisis de mutaciones puntuales. Esto ofrece una gran ventaja frente a los métodos tradicionales en los siguientes aspectos:

1. Se pueden caracterizar diversas regiones del ADN de un paciente en un experimento único y con un tiempo de diagnóstico y una cantidad de muestra mucho menor que con los métodos tradicionales utilizados actualmente.
2. Se puede utilizar ADN extraído de muestras fijadas en donde la cantidad y la calidad de material limita la utilización de otras estrategias basadas en secuenciación masiva.
3. Se pueden incluir en un mismo experimento de secuenciación caracterizaciones de decenas de pacientes al mismo tiempo, marcadas cada una con códigos de barras específicos.
4. Gracias a la utilización de UMIs se pueden descartar los errores de la PCR con lo que podemos identificar mutaciones presentes en poblaciones muy minoritarias de la muestra.

Sin embargo hemos comprobado que todavía hace falta bastante trabajo de puesta a punto para que este método sea fiable en la identificación de otras alteraciones genómicas como la presencia de pequeñas inserciones y deleciones.

8 BIBLIOGRAFÍA

1. OMS. Cancer. (2017). Available at: <http://www.who.int/mediacentre/factsheets/fs297/es/>.
2. No Title WHO Classification of Tumours of the Lung, Pleura, Thymus and Heart. in (ed. Travis WD, Brambilla E, Burke AP, Marx A, N. A.) 26–44
3. American Cancer Society. Estadísticas importantes sobre el cáncer colorrectal. Available at: <https://www.cancer.org/es/cancer/cancer-de-colon-o-recto/acerca/estadisticas-clave.html>.
4. McGranahan, N. & Swanton, C. Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future. *Cell* **168**, 613–628 (2017).
5. Staudt L. Cómo está la genómica dando forma a la medicina de precisión en oncología. *NIH* (2015). Available at: <https://www.cancer.gov/espanol/noticias/temas-y-relatos-blog/2015/concepto-medicina-precision>.
6. Izumchenko, E. *et al.* Targeted sequencing reveals clonal genetic changes in the progression of early lung neoplasms and paired circulating DNA. *Nat. Commun.* **6**, (2015).
7. Frucht, H., Lucas, A. L. Molecular genetics of colorectal cancer. *Uptodate* (2017). Available at: <http://www.uptodate.com/contents/molecular-genetics-of-colorectal-cancer>.
8. Chen, Z., M.Fillmore, C., S.Hammerman, P., F.Kim, C., & Wong, K.-K. Non-small-cell lung cancers: a heterogeneous set of diseases. *Nature Reviews* **14**, 535–546 (2014).
9. Desai, E. I. B. A. CTLA-4 and PD-1 Pathways. *Am J Clin Oncol* **39**, 98–106 (2016).
10. Diz Taín, P., González, A. L. & García-Palomo, A. Mecanismo de acción y desarrollo preclínico de afatinib. *Med. Clin. (Barc.)* **146**, 7–11 (2016).
11. Sequist, L. V., Neal, J. W. Personalized genotype-directed therapy for advanced non-small cell lung cancer. *Uptodate* (2017). Available at: <https://www.uptodate.com/contents/personalized-genotype-directed-therapy-for-advanced-non-small-cell-lung-cancer>.
12. Javier Freire, J. J. G.-R. Biomarcadores en el carcinoma pulmonar. **10**, 17–23

13. Javier de Castro Carpeño. Cáncer de Pulmón. *Instituto Roche* Available at:
https://www.instituto Roche.es/oncobylg/marcadoresmoleculares/cancer_de_pulmon/investigacion/104/#info_marcador.
14. Yang, Z. *et al.* Comparison of gefitinib, erlotinib and afatinib in non-small cell lung cancer: A meta-analysis: Gefitinib vs . erlotinib vs . afatinib in NSCLC. *Int. J. Cancer* **140**, 2805–2819 (2017).
15. SEOM. Informe SEOM de evaluación de fármacos: Osimertinib (Tagrisso[®]) en cáncer de pulmón no microcítico. (2017). Available at:
https://seom.org/seomcms/images/stories/Informes_SEOM/IPT_OSIMERTINIB.pdf.
16. Lilly, E. La FDA aprueba Portrazza[®] (necitumumab) para cáncer de pulmón metastásico de células no pequeñas escamosas. (2015).
17. Crizotinib en el tratamiento del cáncer no microcítico de pulmón (CPNM) avanzado y / o metastásico ALK -positivo 2 . TÍTULO : Informe SEOM de Evaluación de Crizotinib en el tratamiento del cáncer no microcítico de pulmón (CPNM) avanzado y / o metastás. 1–22
18. Ou, S.-H. I., Bartlett, C. H., Mino-Kenudson, M., Cui, J. & Iafrate, A. J. Crizotinib for the Treatment of *ALK* -Rearranged Non-Small Cell Lung Cancer: A Success Story to Usher in the Second Decade of Molecular Targeted Therapy in Oncology. *The Oncologist* **17**, 1351–1375 (2012).
19. Benjamin Solomon, A. T. S. Anaplastic lymphoma kinase (ALK) fusion oncogene positive non-small cell lung cancer. (2018). Available at:
<https://www.uptodate.com/contents/anaplastic-lymphoma-kinase-alk-fusion-oncogene-positive-non-small-cell-lung-cancer>.
20. Kerr, K. ROS1 in Lung Cancer: ESMO Biomarker Factsheet. *19 Abril* (2016). Available at:
<http://oncologypro.esmo.org/Education-Library/Factsheets-on-Biomarkers/ROS1-in-Lung-Cancer>.

21. V., C. T., Subbiah V., Heymach. .. J. Targeting RET Rearrangements in Non–Small Cell Lung Cancer. (2017).
22. Cebollero, A., Puértolas, T., Pajares, I., Calera, L. & Antón, A. Comparative safety of BRAF and MEK inhibitors (vemurafenib, dabrafenib and trametinib) in first-line therapy for BRAF-mutated metastatic melanoma. *Mol. Clin. Oncol.* **5**, 458–462 (2016).
23. Kaufman, J. & Stinchcombe, T. E. Treatment of *KRAS*- Mutant Non–Small Cell Lung Cancer: The End of the Beginning for Targeted Therapies. *JAMA* **317**, 1835 (2017).
24. Cobo, M. *et al.* Afatinib en pacientes con carcinoma escamoso de pulmón: contexto actual y opción de un tratamiento oral. *Med. Clínica* **146**, 25–29 (2016).
25. Rittmeyer, A. *et al.* Atezolizumab versus docetaxel in patients with previously treated non-small-cell lung cancer (OAK): a phase 3, open-label, multicentre randomised controlled trial. *The Lancet* **389**, 255–265 (2017).
26. American Cancer Society. ¿Qué es el cancer microcítico de pulmón? *American Cancer Society* (2016). Available at: <https://www.cancer.org/es/cancer/cancer-de-pulmon-microcitico/acerca/que-es-cancer-de-pulmon-microcitico.html>.
27. Peifer, M. *et al.* Integrative genome analyses identify key somatic driver mutations of small-cell lung cancer. *Nat. Genet.* **44**, 1104–1110 (2012).
28. Gregory P Kalemkerian, E. H. B. Limited-stage small cell lung cancer: Initial management. *Uptodate* (2018). Available at: https://www.uptodate.com/contents/limited-stage-small-cell-lung-cancer-initial-management?search=limited-stage%20small%20cell%20lung&source=search_result&selectedTitle=1~150&usage_type=default&display_rank=1.
29. Karen Kelly. Extensive stage small cell lung cancer: Initial management. *Uptodate* (2017). Available at: <https://www.uptodate.com/contents/extensive-stage-small-cell-lung-cancer-initial->

- management?search=small%20cell%20lung&source=search_result&selectedTitle=1~150&usage_type=default&display_rank=1.
30. Clara Ibet Juárez-Vázquez, M. A. R.-R. Cáncer colorrectal (CCR): alteraciones genéticas y moleculares. *Gaceta Médica de México* **150**, 154–164 (2014).
 31. Sanabria, M. C. *et al.* Vías de carcinogénesis colorrectal y sus implicaciones clínicas. *Rev. Colomb. Cancerol.* **16**, 170–181 (2012).
 32. Fleming, M., Ravula, S., Tatishchev, S. F. & Wang, H. L. Colorectal carcinoma: Pathologic aspects. *J. Gastrointest. Oncol.* **3**, 153–173 (2012).
 33. Carolyn C Compton. Pathology and prognostic determinants of colorectal cancer. *Uptodate* Available at: <https://www.uptodate.com/contents/pathology-and-prognostic-determinants-of-colorectal-cancer>.
 34. Lipson, D. *et al.* Identification of new ALK and RET gene fusions from colorectal and lung cancer biopsies. *Nat. Med.* **18**, 382–384 (2012).
 35. Francesc Balaguer, L. M. Inhibidores del EGFR: ¿ qué pacientes se benefician de estos fármacos? *GH Contin.* **9**, 233–237 (2010).
 36. Irahara, N. *et al.* NRAS Mutations Are Rare in Colorectal Cancer: *Diagn. Mol. Pathol.* **19**, 157–163 (2010).
 37. Agencia Europea de medicamentos. Ficha técnica Vectibix, INN-panitumumab.
 38. Agencia Europea de medicamentos. Ficha técnica Erbitux, INN-cetuximab.
 39. Cetuximab. *Offarm* **27**, 126–30 (2005).
 40. Peeters, M., Karthaus, M., Rivera, F., Terwey, J.-H. & Douillard, J.-Y. Panitumumab in Metastatic Colorectal Cancer: The Importance of Tumour RAS Status. *Drugs* **75**, 731–748 (2015).
 41. Dubois, E. A. & Cohen, A. F. Panitumumab. *Br. J. Clin. Pharmacol.* **68**, 482–483 (2009).
 42. Strickler, J. H., Wu, C. & Bekaii-Saab, T. Targeting BRAF in metastatic colorectal cancer: Maximizing molecular approaches. *Cancer Treat. Rev.* **60**, 109–119 (2017).

43. AEMPS. Informe de posicionamiento terapéutico de regorafenib (Stivarga) en cancer colorrectal. (2015).
44. Hamada, T., Nowak, J. A. & Ogino, S. *PIK3CA* mutation and colorectal cancer precision medicine. *Oncotarget* **8**, (2017).
45. Tong, J., Tan, S., Zou, F., Yu, J. & Zhang, L. FBW7 mutations mediate resistance of colorectal cancer to targeted therapies by blocking Mcl-1 degradation. *Oncogene* **36**, 787–796 (2017).
46. Liu, C. *et al.* *GNAS* mutations are present in colorectal traditional serrated adenomas, serrated tubulovillous adenomas and serrated adenocarcinomas with adverse prognostic features. *Histopathology* **70**, 1079–1088 (2017).
47. NIH. *GNAS* gene. Available at: <https://ghr.nlm.nih.gov/gene/GNAS>.
48. Aung Ko Win, N. M. L. Lynch syndrome (hereditary nonpolyposis colorectal cancer): Clinical manifestations and diagnosis. (2017). Available at: https://www-uptodate-com.scsalud.a17.csinet.es/contents/lynch-syndrome-hereditary-nonpolyposis-colorectal-cancer-clinical-manifestations-and-diagnosis?sectionName=IDENTIFICATION%20OF%20INDIVIDUALS%20AT%20RISK%20FOR%20LYNCH%20SYNDROME&topicRef=2484&anchor=H541769197&source=see_link#H541769197.
49. American Cancer Society. Tratamiento cáncer colorrectal: Inmunoterapia. (2018). Available at: <https://www.cancer.org/es/cancer/cancer-de-colon-recto/tratamiento/inmunoterapia.html>.
50. Reza, M. T. R. La secuenciación del DNA: consideraciones históricas y técnicas. *Rev. Colomb. Biotecnol.* **16**, 5–8 (2014).
51. Fran Garrigues. Sanger: Estrategia de secuenciación de Primera Generación. *Genética Médica* (2017). Available at: <https://revistageneticamedica.com/blog/sanger/>.

52. Shendure, J. & Ji, H. Next-generation DNA sequencing. *Nat. Biotechnol.* **26**, 1135–1145 (2008).
53. Thaidy Moreno RodríguezGUE. Molecular characterization of the role of chromatin remodeling complexes in tumor progression. (Universidad de Cantabria, 2017).
54. Salipante, S. J. *et al.* Performance Comparison of Illumina and Ion Torrent Next-Generation Sequencing Platforms for 16S rRNA-Based Bacterial Community Profiling. *Appl. Environ. Microbiol.* **80**, 7583–7591 (2014).
55. Sanger Institute. Molecular Inversion Probes protocol.
56. Singh, R. R., Luthra, R., Routbort, M. J., Patel, K. P. & Medeiros, L. J. Implementation of next generation sequencing in clinical molecular diagnostic laboratories: advantages, challenges and potential. *Expert Rev. Precis. Med. Drug Dev.* **1**, 109–120 (2016).
57. Kerick, M. *et al.* Targeted high throughput sequencing in clinical cancer Settings: formaldehyde fixed-paraffin embedded (FFPE) tumor tissues, input amount and tumor heterogeneity. *BMC Med. Genomics* **4**, (2011).
58. Boyle, E. A., O’Roak, B. J., Martin, B. K., Kumar, A. & Shendure, J. MIPgen: optimized modeling and design of molecular inversion probes for targeted resequencing. *Bioinformatics* **30**, 2670–2672 (2014).
59. Gozdecka, M. *et al.* UTX-mediated enhancer and chromatin remodeling suppresses myeloid leukemogenesis through noncatalytic inverse regulation of ETS and GATA programs. *Nat. Genet.* (2018). doi:10.1038/s41588-018-0114-z