

Trabajo de Fin de Grado

Análisis de mutaciones del gen de c-MET en pacientes con cáncer de pulmón de células no pequeñas (NSCLC) utilizando Secuenciación de Nueva Generación (NGS).

Marina Murillo Recio

Grado en Biotecnología
Tutor: Josep M. Serrat Jurado
Director: Cristian Tornador Antolin
Vic, Junio de 2018

AGRADECIMIENTOS

Hace cuatro años empecé una etapa, una de las mas importantes. Ahora esta etapa ya ha finalizado y puedo decir que ha sido un camino largo pero lleno de aprendizaje, emociones y cambios. Todo empezó con un yo quiero ser y ahora por fin puedo decir el tan esperado yo soy. Se que volvería a escoger la carrera de biotecnología, no me imagino dedicando mi futuro a algo que no sea la ciencia. Quiero dar las gracias a la Universidad de Vic por formarme y aportarme todos los conocimientos necesarios que me han permitido llegar a ser biotecnóloga. Desde el primer día hasta el último he sentido como si la universidad fuera mi segunda casa. También quiero agradecer a mi tutor, Josep M. Serrat por haber resuelto mis dudas en todo momento y por transmitir su pasión por la genética en cada una de sus clases.

He de decir que una de las mejores cosas que me llevo de esta etapa son las persona que han formado parte de ella, tanto amigos como profesores y profesores que son como amigos. Estos cuatro años me han enseñado realmente el valor de la palabra amistad, he encontrado a personas que comparten mis aficiones y pasiones. Hemos compartido motivaciones, alegrías, risas, locuras, lagrimas, apuntes (muchos apuntes) y hemos creado recuerdos, recuerdos que no cambiaria por nada.

Quiero agradecer a Cristian y al equipo de Whole Genix por darme la oportunidad de realizar el proyecto con ellos. El numero de conceptos que he aprendido en la empresa durante estos meses es incontable. También quiero agradecer a Luca por haber compartido sus conocimientos conmigo y haberme ayudado en todo momento (así que Luca, *print 'Gracias!'*) y al resto del equipo quiero darle las gracias por acogerme desde el primer día.

Todo esto no podría haber sido posible sin la ayuda de mis padres quienes siempre me han apoyado en todas las decisiones que he tomado. Ellos me han aportado las herramientas necesarias que han permitido que yo pueda realizar esta etapa, gracias por enseñarme a no rendirme nunca. Gracias papá por todas las charlas filosóficas que hemos tenido y gracias a mi hermana y a mi madre por hacer el payaso conmigo cuando necesitaba desconectar de todo lo demás. A mi familia y amistades gracias por dejarme compartir esta etapa con vosotros y manteneros a mi lado durante los buenos y los malos momentos.

Puede que este sea el final de una etapa, pero quedan muchas mas por vivir y disfrutar, pero siempre tendré en mente lo importante y impresionante que ha sido la etapa universitaria para mi.

ÍNDICE

ÍNDICE.....	4
RESUMEN DEL TRABAJO DE FIN DE GRADO	5
ABSTRACT	6
1. INTRODUCCIÓN	7
1.1 Características generales de c-Met.	8
1.2 Relación entre c-MET y el cáncer.	12
1.2.1 Angiogénesis y transición epitelio-mesenquimal (EMT).....	12
1.2.2 Cáncer de pulmón de células no pequeñas (NSCLC).	14
1.2.3 Índice de supervivencia en pacientes con NSCLC.	15
1.2.4 C-MET como diana terapéutica.	15
1.3 Lesiones genéticas en c-MET.....	18
1.4 Secuenciación de nueva generación (NGS)	21
2. OBJETIVOS.....	23
3. MATERIALES Y MÉTODOS	24
3.1 Pipeline experimental.....	26
3.1.1 Extracción de DNA.	26
3.1.2 Control de calidad.....	26
3.1.3 Secuenciación por Illumina.....	27
3.2 Pipeline bioinformático.	28
3.2.1 Control y mejora de la calidad.....	29
3.2.2 Alineamiento.	30
3.2.3 Recalibración.	31
3.2.4 Obtención de variantes somáticas (Variant Calling).	32
3.2.5 Anotación.	33
3.2.6 Filtrar las variantes.	34
3.2.7 Programa automático y funciones.....	34
3.2.8 Simulación de la estructura proteica.....	37
4. RESULTADOS Y DISCUSIÓN	38
4.1 Variantes de c-MET observadas en NSCLC.....	38
4.1.1 Variantes estructurales (SV) en el exón 14 de c-MET.....	38
4.1.2 Variantes de un único nucleótido (SNV) y indels observados en el exón 14 de c-MET..	40
4.1.3 Alteraciones en la estructura proteica de c-MET debido a la falta del exón 14.	41
4.2 Verificación de la delección del exón 7 de c-MET en muestras de cáncer de glioma.	43
4.2.1 Alteraciones en la estructura proteica de c-MET debido a la falta del exón 7.	43
4.3 Fenómeno de exclusión.....	45
5. CONCLUSIONES.....	46
BIBLIOGRAFIA	47
APÉNDICE	51

RESUMEN DEL TRABAJO DE FIN DE GRADO

Grado en biotecnología

Título: *Análisis de mutaciones del gen de c-MET en pacientes con cáncer de pulmón de células no pequeñas (NSCLC) utilizando Secuenciación de Nueva Generación (NGS).*

Palabras clave: *Factor de transición epitelio–mesénquimal (c-MET), cáncer de pulmón de células no pequeñas (NSCLC), exón 14, análisis bioinformático.*

Autora: Marina Murillo Recio

Tutores: Josep M. Serrat Jurado (UVIC), Cristian Tornador Antolin (Whole Genix)

Fecha: Junio de 2018

Las terapias actualmente utilizadas para tratar el cáncer de pulmón de células no pequeñas (NSCLC) no funcionan adecuadamente en la mayoría de los pacientes que se encuentran en estados avanzados de la enfermedad, de modo que en estos casos se observa un índice de supervivencia del 26% al 13%. Por este motivo existe un proyecto europeo donde participan el Hospital Universitario Quirón Dexeus, el Hospital Clínico y Pangaea Oncology. El objetivo del proyecto es caracterizar los genes que comparten un patrón de expresión y identificar las mutaciones que se producen en este conjunto de genes para cada subgrupo de pacientes con NSCLC, en otras palabras, se busca caracterizar la firma genética de la enfermedad. El estudio que se realizó en Whole Genix se centró en analizar las mutaciones que se producen en el gen de c-MET (factor de transición epitelio–mesénquimal) en estados avanzados de la enfermedad. C-MET es un receptor de membrana considerado un proto-oncogen debido a que alteraciones en c-MET se relacionan con procesos metastático como la angiogénesis y la transición epitelio-mesénquimal.

En el proyecto se destinaron 900 muestras para estudios de expresión. Diez de las muestras que presentaban alteraciones en la expresión de c-MET se destinaron a Whole Genix donde se realizó el análisis genético. Cinco muestras provenían de pacientes con NSCLC en estados avanzados y las cinco muestras restantes eran de pacientes con cáncer de glioma (para verificar que se produce la delección del exón 7, se utilizó como control). El objetivo principal del trabajo fue utilizar técnicas de secuenciación de nueva generación (NGS) y herramientas bioinformáticas para encontrar alteraciones genéticas en el exón 14, ya que estudios anteriores resaltan la importancia de alteraciones en este exón en pacientes con NSCLC.

El análisis bioinformático realizado se basa en la observación de la cobertura (*depth of coverage*), análisis combinatorio de estrategias de *paired-reads* y herramientas de visualización del genoma para detectar variantes estructurales (SV). Para analizar variantes de un único nucleótido (SNV) y indels se utilizó la metodología de *mpileup*. Los resultados preliminares indicaron que en algunos casos de NSCLC se detectaron delecciones que afectaban al exón 14. En otros casos se observaron SNV y indels, que alteraban el mecanismo del splicing alternativo produciendo la eliminación del exón 14. También se encontraron mutaciones que no afectaban al exón 14 pero también se reportaron en el trabajo. Las mutaciones descritas permitirán ayudar en la caracterización de la firma genética en pacientes con NSCLC en estados avanzados con la finalidad de mejorar el diagnóstico y el tratamiento de la enfermedad para poder aumentar el índice de supervivencia de este grupo de pacientes.

ABSTRACT

Biotechnology Degree

Title: *Analysis of c-MET mutations in patients with non-small cell lung cancer (NSCLC) using Next Generation Sequencing (NGS).*

Key words: *Epithelial-mesenchymal transition factor (c-MET), non-small cell lung cancer (NSCLC), exon 14, bioinformatic analysis.*

Author: Marina Murillo Recio

Tutors: Josep M. Serrat Jurado (UVIC), Cristian Tornador Antolin (Whole Genix)

Date: June 2018

The therapies currently used to treat non-small cell lung cancer (NSCLC) do not work as expected in patients who are in advanced stages of the disease, in these cases the survival rate goes from 26% to 13%. For this reason, exist a European project involving the *Hospital Universitario Quirón Dexeus, Hospital Clínico* and Pangaea Oncology. The aim of the project is characterizing the genes that share an expression pattern and identify the mutations that occur in this set of genes for each subgroup of patients with NSCLC, in other words, it consists in characterize the genetic signature of the disease. The study made in Whole Genix focused on analyzing the mutations that occur in c-MET (epithelial-mesenchymal transition factor) in advanced stages of the disease. C-MET is a membrane receptor considered a proto-oncogene. It has been shown that alterations in c-MET protein are related with metastatic processes such as angiogenesis and epithelial-mesenchymal transition.

In the project 900 samples were assigned for expression studies. Ten of the samples that showed alterations in c-MET expression were destined to Whole Genix where the genetic analysis was carried out. Five samples came from patients with NSCLC in advanced stages and the remaining five samples were from glioma cancer (to verify the deletion of exon 7, it was used as a control). The main objective of the work was to use next generation sequencing (NGS) and bioinformatics tools in order to find genetic alterations in exon 14, since previous studies highlight the importance of alterations in this exon in patients with NSCLC.

The bioinformatic analysis performed is based on studies of *depth of coverage*, combinatorial analysis of *paired-reads* strategies and genome visualization tools to detect structural variants (SV). To analyze variants of a single nucleotide (SNV) and indels, the mpileup methodology was used. The preliminary results indicated that deletion of exon 14 occurred in some cases of NSCLC. In other cases, SNV and indels were observed, which affected alternative splicing, producing the loss of exon 14. Also, mutations that did not affect exon 14 were found and reported in the study. The described mutations will help in the characterization of the genetic signature in patients with NSCLC in order to improve the diagnosis and treatment of the disease and be able to increase the survival rate of this group of patients.

1. INTRODUCCIÓN

El estudio de las vías de señalización donde participa c-MET es de gran relevancia en la investigación de terapias para el cáncer. c-MET (factor de transición epitelio–mesénquimal), es un receptor de membrana considerado un proto-oncogen, debido a que cuando se encuentra sobreexpresado, potencialmente mutado o amplificado en las células cancerígenas, diferentes vías de señalización que tienen un papel clave la división celular se ven alteradas (Zhang *et al.*, 2018). Los procesos celulares que se ven mayoritariamente alterados son la proliferación celular, la supervivencia, la movilidad, la dispersión, la diferenciación, la morfogénesis y la angiogénesis. Estas alteraciones del gen de c-MET se han observado en diferentes tipos de cáncer. Investigaciones actuales indican que se produce la sobreexpresión de la proteína en pacientes que se encuentran en estados más avanzados de la enfermedad, un claro ejemplo es el cáncer de pulmón de células no pequeñas (NSCLC). En estos casos se ha contemplado una relación entre los niveles de c-MET y la supervivencia del paciente, mayores niveles de c-MET disminuyen la supervivencia del paciente, debido a que las vías de señalización alteradas a causa de la sobreexpresión de c-MET derivan a las células cancerígenas a iniciar la metástasis. Por este motivo c-MET es un biomarcador que se está convirtiendo en una diana terapéutica, poder conocer que modificaciones se producen en el gen de c-MET nos permitirá utilizar terapias dirigidas y personalizadas en oncología, con la finalidad de poder aumentar la supervivencia del paciente evitando el crecimiento y dispersión del tumor. Además el hecho de que el tumor sea más agresivo en estados avanzados hace que los pacientes no respondan a terapias que habitualmente se utilizan en pacientes con cáncer de NSCLC. Por esta razón hay investigaciones que contribuyen en la caracterización de la firma genética de NSCLC, con la finalidad de mejorar el diagnóstico y poder encontrar un tratamiento personalizado que se ajuste al historial clínico del paciente (estadio del cáncer, antecedentes, estilo de vida...). Actualmente, se utilizan fármacos capaces de inhibir c-MET y su ligando HGF (factor de crecimiento de hepatocitos), conduciendo a la célula a la apoptosis. Pacientes en estado avanzado de cáncer de glioma respondieron adecuadamente a un tratamiento personalizado. Este tratamiento incluye un conjunto de inhibidores de c-MET que responden a una variante específica de c-MET. Poder comprobar que mutaciones se producen en cáncer de NSCLC en estados avanzados y entender que repercusión biológica tienen estas alteraciones resultaría de gran ayuda a la hora de buscar un tratamiento para este tipo de cáncer, ya que se podrían utilizar los mismos inhibidores que se utilizan para tratar el cáncer de glioma o otros tipos de cáncer. (Salgia, 2009; Garnett *et al.*, 2013).

1.1 Características generales de c-Met.

C-MET es un receptor tirosina quinasa (RTKs) situado en la membrana de células de origen epitelial. El gen de c-MET se encuentra en el cromosoma 7 (7q21-q31), incluye 21 exones, 20 intrones y codifica una proteína de aproximadamente 120 kDa (Zhang *et al.*, 2018). La molécula de unión o ligando de c-MET es el factor de crecimiento de hepatocitos (HGF) el cual se une al dominio tirosina quinasa de c-MET produciendo respuestas biológicas, que incluyen la proliferación, supervivencia, motilidad y morfogénesis (Salgia, 2009). En condiciones normales la interacción entre HGF y c-MET regula la embriogénesis, la regeneración tisular, la curación de heridas, y la formación de nervios y músculos donde se estimula la migración celular y la mitosis (Navis *et al.*, 2015; Van Der Steen *et al.*, 2016).

El gen de c-MET da lugar a un precursor el cual se procesa para formar un heterodímero (α y β) donde se puede diferenciar una región extracelular y una región intracelular. La región extracelular esta constituida por un dominio SEMA (semaforina), un dominio PSI (plexina, semaforina y integrina) y cuatro dominios IPT (inmunoglobulina plexinas y factores de transcripción). Por otro lado, la región intracelular se encuentra formada por un dominio yuxtamembrana, un dominio TK (tirosina quinasa) y un lugar de acoplamiento molecular o *docking* (Figura 1). El dominio SEMA es el que permite la unión de HGF a c-MET mientras que PSI permite estabilizar la interacción. El dominio yuxtamembrana permite la regulación negativa de c-MET mediante la unión de la ubiquitina c-CBL (Linfoma de línea B casitas), concretamente Ser-975 y Tyr-1003 son las posiciones que le otorgan al dominio esta función. Cuando HGF interacciona con c-Met mediante Tyr-1234 y Tyr-1235 del dominio tirosina quinasa se produce la fosforilación de estas posiciones, causando la fosforilación de Tyr-1349 y Tyr-1356 en el lugar de docking dando lugar a una cascada de señalización producida por la fosforilación de varias tirosinas quinasas, provocando el reclutamiento de algunos efectores como pueden ser GRB2, SHC, CRK, CRKL, P13K, PLC γ , SHIP-2, STAT-3, GAB1, Src, MAPK (Organ and Tsao, 2011; Zhang *et al.*, 2018). Cada uno de estos efectores están involucrado en diferentes actividades celulares (Figura 2).

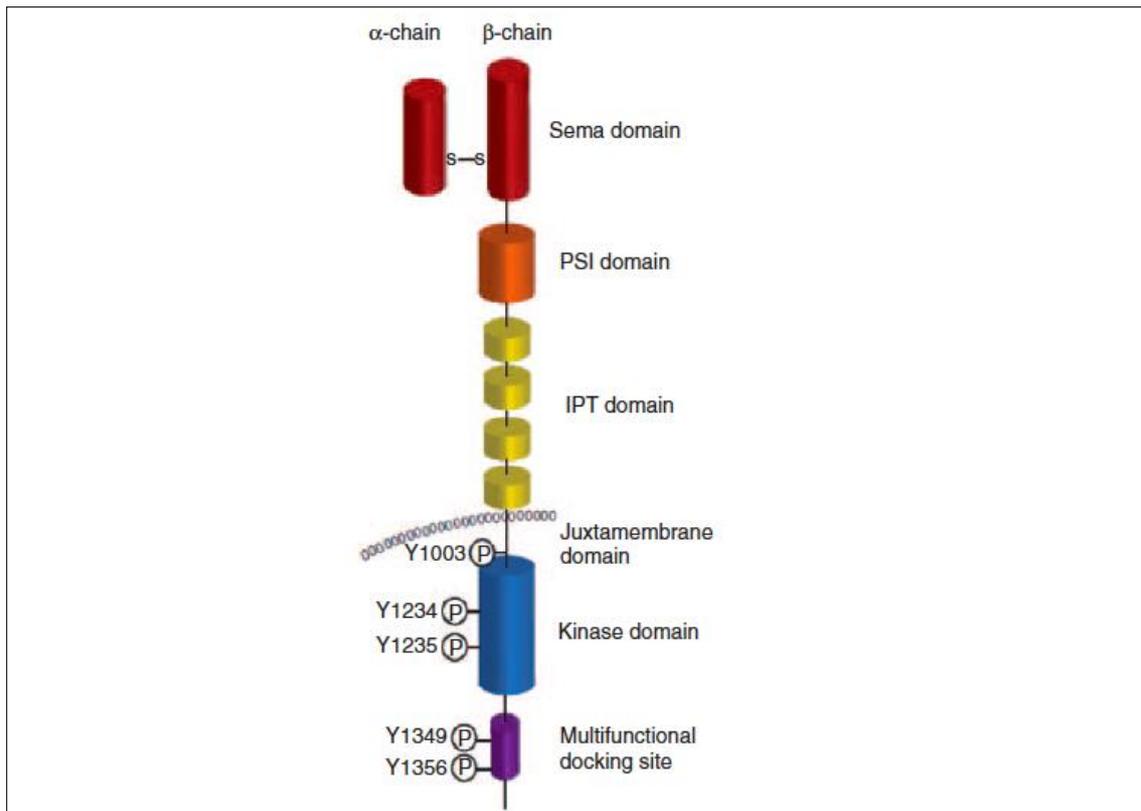


Figura 1. Estructura del receptor c-MET. La figura representa la estructura de c-MET formada por un heterodímero (α y β) donde se pueden ver representados los diferentes dominios. La parte extracelular formada por el dominio SEMA, PSI y los cuatro dominios IPT y la parte intracelular formada por el dominio yuxtamembrana (JM), el dominio de tirosina quinasa y el lugar de acoplamiento molecular o *docking*. El dominio extracelular y el intracelular se encuentran unidos por el dominio de transmembrana (TM). (Organ and Tsao, 2011)

La alteración de algunos efectores de c-MET y de sus correspondientes vías de señalización también se relacionan con mecanismos cancerígenos, algunas de estas vías son (Figura 2):

- **PI3K/AKT:** Cuando HGF se une a c-MET se produce la fosforilación del dominio C-terminal o dominio *docking* permitiendo la unión de PI3K. Seguidamente se inicia la fosforilación de muchos intermediarios que forman parte especialmente de cascadas de señalización asociadas a la aparición de tumores. Uno de estos intermediarios es AKT, cuando se activa se dirige al núcleo donde inhibe la expresión de supresores tumorales, aumentando la supervivencia de las células cancerígenas. Asimismo, se ha observado que AKT fosforila la enzima óxido nítrico sintasa (NOS) para producir NO, favoreciendo la angiogénesis. (Rosário and Birchmeier, 2003; Zhang *et al.*, 2018)

- **MAPK/Ras:** Se observan mutaciones en el gen Ras en el 25% de casos de cáncer. Una vez producida la unión de HGF a c-MET se produce la unión de SHC y de GRB2 en el dominio *docking*. Este último permite el reclutamiento de SOS el cual seguidamente permite la activación de Ras. Ras activa sucesivamente a MAPK produciendo su entrada en el núcleo de la célula para activar varios factores de transcripción (por ejemplo, Elk1, Ets1, c-Myc) a través de la fosforilación. Este proceso, puede interferir con el ciclo celular e inducir la transformación celular, promoviendo la carcinogénesis. MAPK también inducen la degradación de proteínas de la matriz, promoviendo la migración celular y la proliferación tumoral. (Liebmann, 2001; Organ and Tsao, 2011; Zhang *et al.*, 2018)

- **JAK/STAT:** STAT3 se ha visto implicado en procesos donde interviene la transformación celular. La unión directa de STAT3 a c-MET da como resultado la fosforilación de STAT3 produciendo la unión de CRK y activación de JAK. Esta cascada de señalización produce un aumento de la división celular y de la formación de tumores. (Organ and Tsao, 2011; Thomas *et al.*, 2015)

- **WNT/ β -CATENIN:** El receptor c-MET está relacionado con la señalización de Wnt/ β -catenina. La señalización mediante Wnt es uno de los factores fundamentales en los mecanismos que dirigen la proliferación celular, la polaridad celular, la determinación del destino celular y la homeostasis tisular. β -catenina actúa como coactivadora de Wnt y además participa en la adhesión célula-célula. En circunstancias normales β -catenina se encuentra en el citoplasma en una forma fosforilada y acetilada impidiendo su entrada en el núcleo. Pero en las células tumorales con una sobreexpresión de c-MET la señalización de Wnt/ β -catenina se ve alterada ya que se bloquea la fosforilación y acetilación de β -catenina provocando su entrada en el núcleo y como consecuencia reclutando a un seguido de factores de transcripción que facilitan la proliferación, la invasión y la metástasis. Por lo tanto, inhibiendo c-MET también controlamos la señalización de Wnt/ β -catenina. (Bryan T. MacDonald; Keiko Tamai and Xi He, 2010)

- **C-CBL (Regulación negativa):** c-MET es negativamente regulado mediante la ligasa de ubiquitina E3, (c-CBL). Una vez producida la activación de c-MET la proteína c-CBL se une al dominio yuxtamembrana a partir de Tyr1003. C-CBL induce a la poliubiquitinación de c-MET produciendo su degradación dependiente de proteosoma (complejo proteico que permite la degradación de las proteínas). Modificaciones en los procesos que permiten la regulación negativa puede causar alteraciones en la concentración de c-MET y por lo tanto se relaciona con su sobreexpresión. (Thien and Langdon, 2005; Qamsari *et al.*, 2017)

Por lo tanto, alteraciones en c-MET causar la desregulación de varias vías de señalización cuya alteración favorece el crecimiento y dispersión del tumor aumentando el carácter metastático de las células cancerígenas. Este es otro factor que nos indica que mutaciones en c-MET son clave para la aparición y supervivencia del tumor.

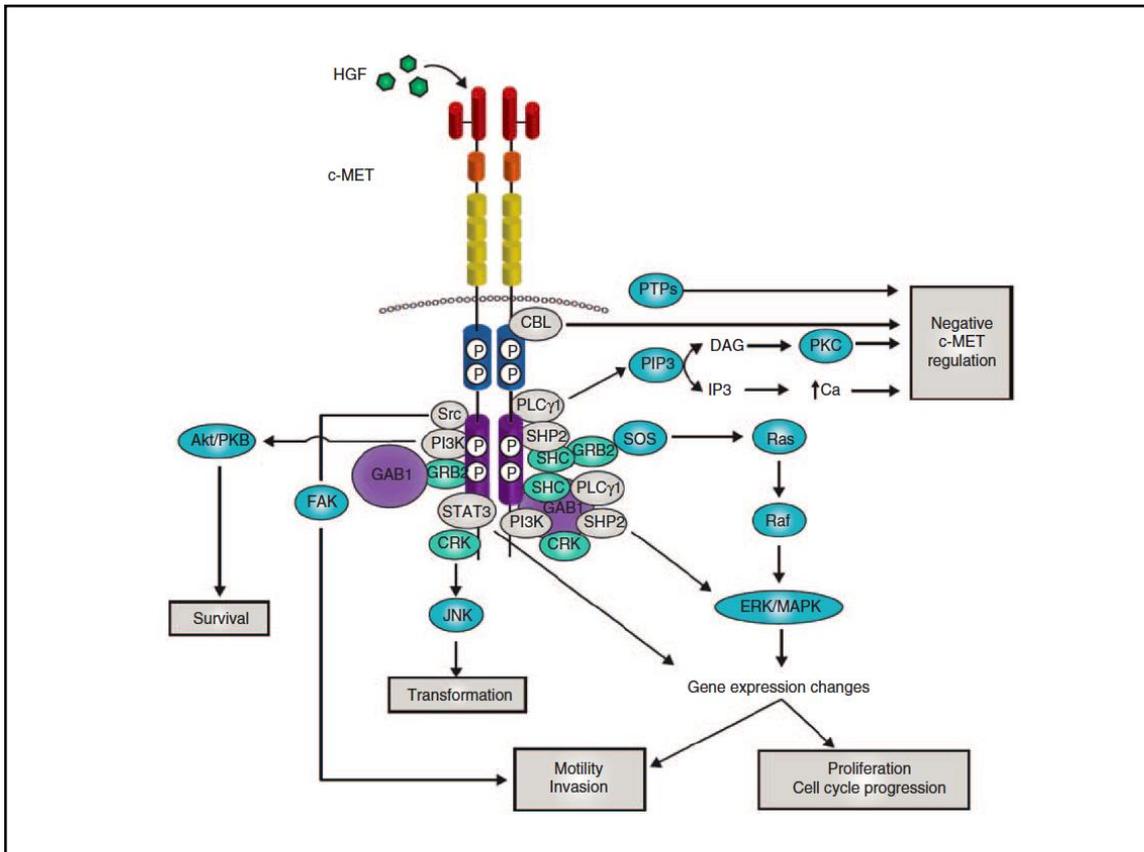


Figura 2. Cascada de señalización de c-MET. La unión de HGF y c-Met induce a cambios conformacionales en c-Met, lo que resulta en la activación de un conjunto de mediadores y efectores. Cuando el lugar de *docking* queda fosforilado se produce el reclutamiento de moléculas como la proteína de unión al receptor para el factor de crecimiento 2 (GRB2), el dominio Src de homología 2 (SHC), el oncogén CRK y CRKL. También se unen efectores como Fosfoinositol 3-quinasa (PI3K), la fosfolipasa Cg (PLCg), el proto-oncogen tirosina quinasa (SRC), 5'-inositol fosfatasa (SHIP-2), el transductor de señal y activador de la transcripción 3 (STAT3). Además, se une la proteína de unión asociada a GRB2 1 (GAB1), la cual crea lugares de unión para más moléculas y puede unirse a c-MET de manera directa o indirecta. Una vez producida la unión de los diferentes efectores se activan varias vías de señalización como PI3K/AKT, MAPK/Ras y JNK/STAT, Src/FAK. Además, el esquema muestra la regulación negativa de c-MET, un proceso crucial para controlar su actividad. El dominio yuxtamembrana de c-MET permite la regulación negativa mediante la unión de c-CBL a Y1003. Las quinasas fosfatasa (PTPs), los niveles de calcio y la proteína quinasa C (PKC) mediante el inositol trifosfato (PIP3) también participan en la regulación negativa. (Organ and Tsao, 2011)

1.2 Relación entre c-MET y el cáncer.

Tal y como hemos visto en el apartado anterior c-MET interacciona con varios oncogenes. La actividad prolongada de c-MET se encuentra relacionada con el crecimiento tumoral y por lo tanto con la progresión y supervivencia de las células cancerígenas. El receptor c-MET se encuentra activado de manera anómala en la mayoría de los tumores sólidos debido a la presencia de mutaciones en el gen de c-MET, convirtiendo a c-MET en una proteína causante del fenotipo de dispersión celular que se observa en algunos tumores. Además también se ha observado que tiene una gran relación con la angiogénesis y la transición epitelio–mesenquimal. Hoy en día se utilizan tratamientos que utilizan como diana c-MET debido a que su inhibición atenúa el carácter cancerígeno del tumor. (Salgia, 2009; Organ and Tsao, 2011)

1.2.1 Metástasis.

La metástasis se compone por un conjunto de fases (*Figura 4*), angiogénesis, transición epitelio-mesenquimal (EMT), intravasación, supervivencia en la circulación, extravasación y crecimiento del tumor secundario. Estos procesos son posibles gracias a la colaboración de un conjunto de proto-oncogenes. En este caso c-MET se relaciona con procesos clave para la metástasis como son la angiogénesis y la transición epitelio-mesenquimal (EMT)

1.2.1 Angiogénesis y transición epitelio-mesenquimal (EMT).

Varios estudios indican que hay una relación directa entre c-MET y la formación de nuevos vasos sanguíneos (angiogénesis). Se ha observado como c-MET interviene en la regulación directa de varios genes relacionados con la angiogénesis como el gen VEGF y trombospondina-1, un factor supresor de la angiogénesis (Piguet *et al.*, 2015).

El proceso a partir del cual se produce la metástasis se describe como una cascada de sucesos interrelacionados, donde el proceso de angiogénesis es el desencadenante. La angiogénesis (*Figura 3b*) es uno de los factores clave que permite el escape de las células tumorales al torrente sanguíneo y la formación de colonias metastásicas en órganos secundarios. Durante el proceso de angiogénesis se forma una red vascular dando lugar a nuevos vasos sanguíneos (neovasos). Alrededor del tumor los neovasos no solo proporcionan nutrientes para que el tumor crezca, sino que también proporcionan una ruta de escape para que las células tumorales lleguen a la circulación. Seguidamente después de la angiogénesis se produce la intravasación (*Figura 3d*) donde las células tumorales deben desprenderse de las células vecinas y la matriz, pasar a través de la membrana basal capilar y migrar a través de la barrera endotelial para llegar al torrente sanguíneo. Este proceso migratorio es posible

gracias a la transición epitelio-mesénquimal (EMT), donde c-MET también tiene una gran relevancia (Figura 3c) (Bielenberg, Zetter and Program, 2016).

Los epitelios se establecen como capas unicelulares o tejidos multicapa con diversas funciones. Las células epiteliales muestran una polaridad apical-basal, se adhieren y se comunican entre sí a través de uniones intercelulares especializadas. Los epitelios funcionan como barreras entre órganos y tejidos. Las células epiteliales que se someten a EMT pierden las características propias de una célula epitelial y adquiere un carácter mesénquimal caracterizado por la capacidad de migración e invasión. Este proceso esta caracterizado por la perdida de la polaridad epitelial, pérdida de la adherencia de célula a célula, degradación de la lámina basal y aumento de migración e invasión. Además, las células que se han sometido a EMT adquieren resistencia a la senescencia y a la apoptosis. Las células cancerígenas deben sobrevivir en el torrente sanguíneo (Figura 3e) el cual es un ambiente nuevo para este conjunto de células y por lo tanto la gran mayoría morirán antes de las 24h. Las que sobreviven proceden a realizar la extravasación (Figura 3f), por lo tanto saldrán de los vasos sanguíneos y se instalar en un nuevo órgano (Figura 3g) donde gracias a procesos de adaptación de las células cancerígenas y la angiogénesis (estimulada c-MET) se desarrollará un tumor secundario (Lamouille, Xu and Derynck, 2014; Sylvester, 2014).

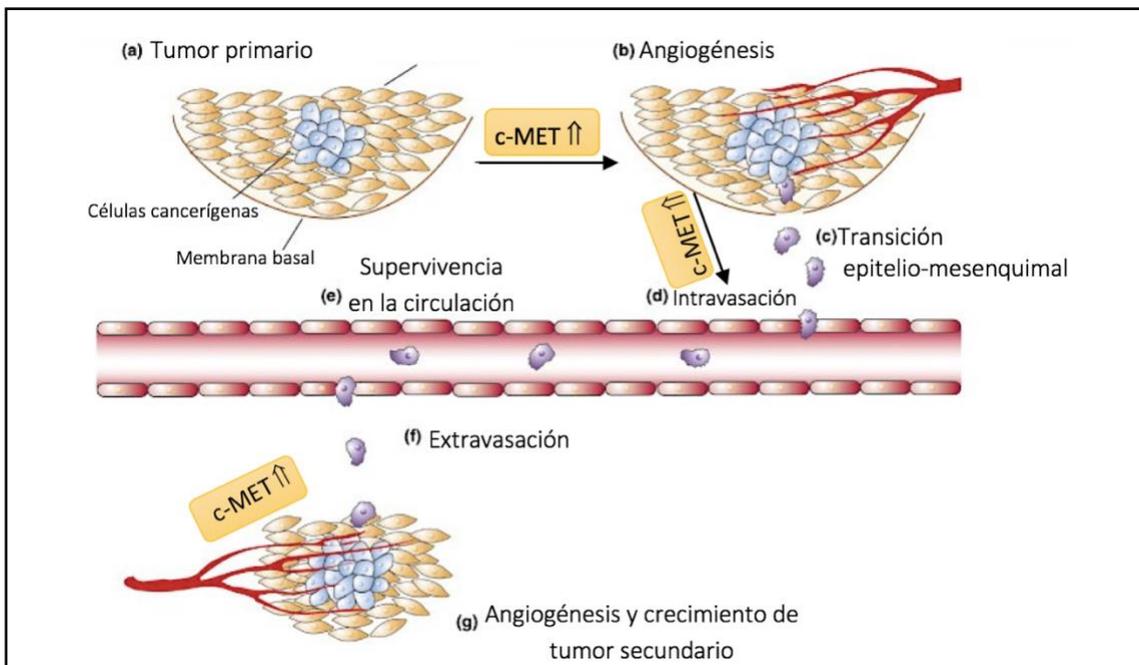


Figura 3. Esquema de la relación entre c-MET y la metástasis. En la imagen se describen los diferentes procesos que permiten la metástasis y se destacan en los que c-MET tiene una gran relevancia como la angiogénesis (a y g) y la transición epitelio-mesénquimal (c) (Ma and Weinberg, 2008).

1.2.2 Cáncer de pulmón de células no pequeñas (NSCLC).

El cáncer de pulmón es uno de los cánceres mas mortales. Más de la mitad de los pacientes diagnosticados con cáncer de pulmón mueren el año posterior al diagnóstico. Existen dos subtipos principales de cáncer de pulmón, carcinoma de pulmón de células pequeñas y carcinoma de pulmón no microcítico o cáncer de pulmón de células no pequeñas (NSCLC), este ultimo representa el 85% de todos los cánceres de pulmón. Además, el NSCLC se divide en tres subtipos (Navis *et al.*, 2015; Zappa and Mousa, 2016):

- **Carcinoma de células escamosas:** Aproximadamente el 25-30% de todos los casos de cáncer de pulmón. Surge de células escamosas que se sitúan en las células epiteliales de los tubos bronquiales en el centro de los pulmones. Este subtipo de NSCLC está fuertemente correlacionado con el tabaquismo.
- **Adenocarcinoma:** El tipo más común de cáncer de pulmón es el adenocarcinoma, comprende alrededor del 40% de todos los cánceres de pulmón. Se genera en el epitelio de las vías respiratorias. El adenocarcinoma es el tipo más común de cáncer de pulmón en fumadores y no fumadores, en hombres y mujeres independientemente de su edad. En comparación con otros tipos de cáncer de pulmón, el adenocarcinoma tiene un crecimiento mas lento y una mayor probabilidad de ser encontrado antes de que se expanda fuera de los pulmones.
- **Carcinoma de células grandes:** El carcinoma de células grandes (indiferenciado) representa el 5-10% de los cánceres de pulmón. El carcinoma de células grandes suele desarrollarse en la parte central de los pulmones, a veces en los ganglios linfáticos cercanos o incluso en la pared torácica. Los tumores de carcinoma de células grandes están asociados con el tabaquismo.

Algunos de los tratamientos actuales que se utilizan son: cirugía, quimioterapia, radioterapia, inmunoterapia y medicina personalizada. La medicina personalizada se dirige a biomarcadores específicos, es decir, a moléculas o cambios celulares que únicamente se producen durante el cáncer. Esta técnica ha ayudado a mejorar la supervivencia en pacientes con NSCLC. Actualmente existen terapias dirigidas que han tenido éxito contra las mutaciones del receptor del factor de crecimiento epidérmico (EGFR) y los reordenamientos del linfoma quinasa anaplásico (ALK). (Hammerschmidt and Wirtz, 2009; Zappa and Mousa, 2016).

1.2.3 Índice de supervivencia en pacientes con NSCLC.

El índice de supervivencia se define como el porcentaje de personas en un estudio o grupo de tratamiento que todavía están vivas durante un determinado período después del diagnóstico o el tratamiento de una enfermedad como el cáncer. La tasa de supervivencia a menudo se establece como una tasa de supervivencia de cinco años, que es el porcentaje de personas en un estudio o grupo de tratamiento que están vivos cinco años después de su diagnóstico o del inicio del tratamiento. También se llama tasa de supervivencia general (Instituto Nacional del Cáncer, 2018b).

El índice de supervivencia en NSCLC varía según el estadio de la enfermedad (*Tabla 1*). En estadios primarios como en el estadio I la supervivencia es del 92% al 77% mientras que en estadios avanzados como en el III la supervivencia es del 26% al 13%. (Sociedad Americana del Cáncer, 2018)

Tabla 1. Estadios del cáncer. Tabla con la clasificación de los diferentes estadios del cáncer y las características más destacadas de cada uno de los estadios (Instituto Nacional del Cáncer, 2018 a).

Estadio	Características
Estadio 0	Hay células anormales presentes, pero no se han diseminado al tejido cercano. Se llama también carcinoma in situ, o CIS. El CIS no es cáncer, pero puede convertirse en cáncer.
Estadio I, Estadio II y Estadio III	Hay cáncer presente. En cuanto más grande es el número, mayor es el tumor y más se ha extendido en los tejidos cercanos.
Estadio IV	El cáncer se ha diseminado a partes distantes del cuerpo (metástasis).

Uno de los objetivos principales en el avance de terapias para cáncer de pulmón NSCLC es buscar tratamientos que permitan aumentar la supervivencia de los pacientes sobre todo en estadios avanzados de cáncer ya que el índice de supervivencia en estos casos es muy bajo. Uno de los factores que produce una disminución del índice de supervivencia es la sobreexpresión c-MET debido a su relación con procesos metastásicos.

1.2.4 C-MET como diana terapéutica.

Actualmente existen tratamientos para diferentes tipos de cáncer que utilizan a c-MET como diana terapéutica. Las terapias dirigidas a c-MET que actualmente más se utilizan en ensayos clínicos pueden clasificarse en dos tipos de terapias, las que utilizan anticuerpos monoclonales y las que utilizan inhibidores de moléculas pequeñas. En este caso, los inhibidores de moléculas pequeñas pueden unirse al dominio tirosina quinasa de c-MET, inhibiendo su activación. Además, se pueden clasificar en inhibidores competitivos de ATP y inhibidores no competitivos de ATP. En varios estudios se observó que el dominio tirosina quinasa necesitaba la unión de ATP y que inhibiendo el lugar de unión de ATP se evita la activación y el reclutamiento de efectores. (Moran-Jones, Brown and Samimi, 2015; Zhang *et al.*, 2018). Se ha observado que las variaciones de c-MET afectan a diferentes dominios de la proteína, como el dominio quinasa o el dominio IPT. Estos pacientes podrían responder a

tratamientos que combinan la quimioterapia con fármacos que utilizan a c-MET o a su ligando HGF como diana terapéutica.(Reungwetwattana *et al.*, 2017) Algunos de los fármacos mas utilizados actualmente son los siguientes:

Inhibidores de moléculas pequeñas:

- **Crizotinib (Xalkori, PF-02341066):** Es uno de los medicamentos aprobados por la FDA. *Crizotinib* es un inhibidor competitivo de ATP con actividad inhibitoria contra c-MET y otros oncogenes. La diana de *Crizotinib* es el dominio tirosina quinasa de c-MET interfiriendo con el lugar de unión de ATP impidiendo la autofosforilación de cMET, que a su vez conduce a la inhibición de la señal de transducción y proliferación celular produciendo la inducción de la apoptosis.
- **Capmatinib (INC280; Novartis):** Es otro inhibidor competitivo de ATP con actividad inhibitoria contra c-MET que se utiliza en ensayos clínicos. Tiene actividades inhibitoras extremadamente altas. Capmatinib consigue inhibir vías de señalización y la proliferación celular inhibiendo la fosforilación de c-MET, induce la apoptosis en líneas celulares que dependen de c-MET. A demás hay estudios que demuestra que podría utilizarse como terapia única.
- **Cabozantinib (XL184):** Es otro Inhibidor efectivo contra MET. Cabozantinib tiene una actividad contra quinasas, por lo tanto, también afecta al dominio tirosina quinasa de c-MET. Actualmente se han realizado ensayos de fase 2 donde se demostró que el fármaco podría ser utilizado como tratamiento único en pacientes con NSCLC.

Hay un conjunto de fármacos que también actúan en contra de c-MET y se ha visto en estudios clínicos de fase II que podrían utilizarse para tratamientos de pacientes con NSCLC, como, por ejemplo, **Tepotinib** (MSC2156119J, EMD 1214063), **Savolitinib** (AZD6094, volitinib, HMPL-504), **AMG-337**, **Glesatinib** (MGCD265), **Merestinib** (LY2801653). (Navis *et al.*, 2015; Zhang *et al.*, 2018)

Este conjunto de fármacos no da resultado en pacientes que se encuentran en estados mas avanzados de NSCLC, por este motivo se buscan terapias dirigidas específicamente para pacientes que se encuentren en este estado, de aquí la importancia de la firma genética.

1.2.5 Terapia personalizada y firma genética.

La mayoría de los estudios que exploran datos genómicos dan como resultado una o más firmas de genes, es decir, una lista de genes que comparten un patrón de expresión común que puede emplearse para clasificar grupos de pacientes (Chibon, 2013). Por este motivo realizar la estratificación de los pacientes según su historial clínico es de gran importancia. Poder obtener la firma genética de cada subgrupo permitirá saber que alteraciones genéticas se producen en cada caso con la finalidad de mejorar el diagnóstico, el pronóstico o predicción de respuesta terapéutica y obtener una terapia personalizada para cada subgrupo. (Cantini *et al.*, 2018). Actualmente estudios indican que no es suficiente encontrar un patrón de expresión para cada subgrupo, sino que la importancia esta en encontrar las mutaciones causantes de la alteración en la expresión. Según las mutaciones que afectan a un gen las terapias utilizadas serán unas u otras.

1.3 Lesiones genéticas en c-MET.

Se han observado un conjunto de alteraciones genéticas que afectan a c-MET como la translocación, la amplificación y mutaciones activadoras. Estas alteraciones se encuentran en diferentes tipos de cáncer.

c-MET puede sufrir una reordenación cromosómica produciendo la translocación del dominio tirosina quinasa de c-MET. En humanos, la translocación se ha relacionado con carcinomas gástricos. Por otro lado, la amplificación y sobreexpresión de c-MET se ha observado en algunos tumores primarios humanos incluyendo carcinomas gástricos, de esófago, meduloblastomas y metástasis hepáticas de carcinoma de colon. Por lo tanto, este es otro indicio que sugiere que la amplificación del gen c-MET se puede adquirir durante el curso de la progresión tumoral (Organ and Tsao, 2011).

Además, se han descubierto mutaciones activadoras en algunos de los dominios de c-MET como en el dominio tirosina quinasa donde se observaron mutaciones en carcinoma papilar renal, en el lugar dominio yxtamembrana de c-MET observado en pacientes con cáncer de glioma y de pulmón, en el lugar de unión a HGF del dominio sema en otros tipos de cáncer. También se han observado mutaciones que afectan al dominio de IPT en el cáncer de glioma (Kong-Beltran *et al.*, 2006; Organ and Tsao, 2011).

1.3.1 Splicing alterativo.

La mayoría de los genes eucariotas tienen exones y intrones. Cuando se produce la transcripción se genera una molécula de pre-mRNA que dará lugar al mRNA después del splicing. El splicing alternativo es un proceso específico para cada tejido y es el proceso a partir del cual se eliminan los intrones y se realiza la unión de los exones con la finalidad de obtener diferentes secuencias de mRNA maduro y por consecuencia más variabilidad de proteínas. El mRNA una vez realizado el splicing contiene únicamente exones, los intrones son eliminados durante el proceso. El mecanismo de splicing mayoritariamente funciona a partir de las secuencias GU-AG y AU-AC situadas en los intrones (Figura 4). La mayoría de los intrones comienzan con la secuencia GU y terminan con la secuencia AG (en la dirección 5' a 3'). Estas secuencias se las conocen como el sitio donador (posición +1 y +2) y el sitio aceptor respectivamente. Otra secuencia necesaria es el sitio de ramificación (*branch site*) situado entre 20 - 50 bases *upstream* del sitio del aceptor. Este conjunto de secuencias indica a la maquinaria de splicing, el spliceosoma (complejo formado por proteínas y RNA, concretamente por RNA nuclear pequeño llamado snRNA) donde debe cortar y unir. El inicio y el final del exón también son importantes para el splicing, en el 60 % de los casos la secuencia del exón es (A/C) AG cerca del sitio donador y G cerca del sitio aceptor. (Watson, 2013).

La regulación del splicing alternativo es necesaria para cumplir con los requisitos funcionales de cada tejido, por este motivo cuando se realiza un proceso de splicing aberrante se produce una disfunción conduciendo a la aparición de enfermedades. Las alteraciones que afectan al splicing alternativo (como mutaciones en intrones) pueden suponer la aparición de un carácter cancerígeno a la célula, incluyendo descontrol de la proliferación, la migración, y la resistencia a la apoptosis y la quimioterapia. El splicing alternativo se ha visto implicado en el desarrollo del cáncer, por lo tanto, podríamos decir que es un desencadenante principal en la enfermedad. Algunas de estas alteraciones estas alteraciones producen deleciones las cuales pueden llevar a la perdida completa de exón.

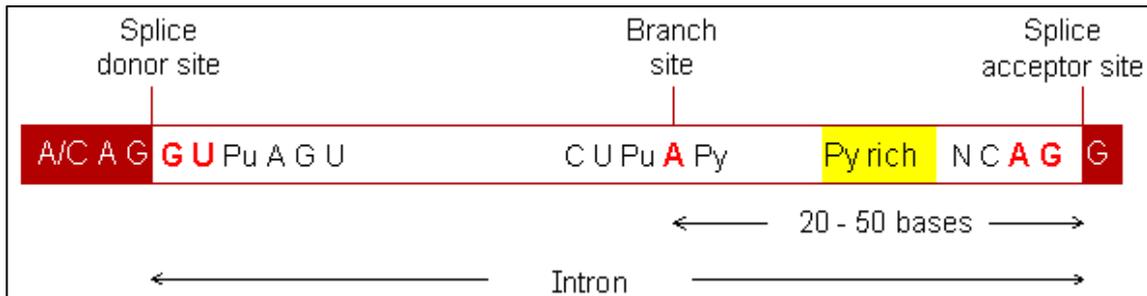


Figura 4. Esquema de la secuencia consenso para el splicing. En el esquema se representa la secuencia consenso para el splicing la cual esta formada por el sitio donador (GU), el sitio aceptor (AG) y el sitio de ramificación (branch site). También se muestra la importancia de las bases que se encuentran al inicio y al final del exón A/C AG y G. Py = C o U. (Molecular Biology Web Book, 2018)

1.3.2 Alteración del exón 14 en c-MET (c-MET^{A14}).

Las alteraciones del exón 14 en c-MET ocurren en aproximadamente el 3% de los adenocarcinomas de pulmón y en un 0,5% de gliomas (Frampton *et al.*, 2015). Se han observado mutaciones somáticas en intrones de c-MET las cuales causan la eliminación del exón 14 durante el *splicing*. Alteraciones en el exón 14 no permiten la unión de c-CBL a c-MET debido a que este exón codifica para el dominio de yuxtamembrana el cual contiene Y1003 que permite la fosforilación necesaria que posibilitara la unión de c-CBL (Van Der Steen *et al.*, 2016). Este fenómeno causa un aumento de la tumorigenicidad ya que la proteína c-CBL está implicada en la regulación negativa de c-MET mediante la ubiquitinación, por lo tanto c-MET^{A14} se encuentra sobreexpresado en las células tumorales que presentan estas mutaciones. (Kong-Beltran *et al.*, 2006; Reungwetwattana *et al.*, 2017)

En el contexto del cáncer de glioma específicamente de NSCLC las alteraciones en el exón 14 se consideran mutaciones conductoras (*driver mutations*), las cuales son las responsables de la progresión del cáncer en estados avanzados.

1.3.3 Alteración de los exones 7 i 8 en c-MET (MET^{Δ7-8}).

El 6% de gliomas en estado avanzado presentan esta variante de c-MET. La caracterización de esta mutación en una línea celular de glioma reveló que se producía la unión del exón 6 directamente con el exón 9, dando lugar a la delección del exón 7 y 8. En este caso MET es nombrada MET^{Δ7-8}, esta variante de c-MET tiene la carencia de partes del dominio extracelular IPT. Concretamente se observo la pérdida de C-terminal de IPT1 y del N-terminal de IPT2, esta alteración hace que c-MET no dependa de HGF para su activación y que por lo tanto sea autoactiva. Por lo tanto, las terapias que usan anticuerpos en contra de HGF no son útiles para pacientes con MET^{Δ7-8}. (Navis *et al.*, 2015)

El dominio IPT se encuentra relacionado con el reconocimiento célula-célula, con la unión a la membrana celular. Además IPT3 y IPT4 tienen una gran importancia en la transmisión de la señal del dominio quinasa después de su activación. (Kong *et al.*, 2016; Suzuki *et al.*, 2016)

Por consiguiente, alteraciones en el dominio IPT causan un aumento de c-MET intracelular ya que c-MET pierde la capacidad de ser una proteína de transmembrana por este motivo el uso de anticuerpos como terapia no es útil ya que estos no detectan a c-MET como diana. (Navis *et al.*, 2015).

1.4 Secuenciación de nueva generación (NGS)

Identificar las variaciones que se producen en c-MET implica la obtención de la secuencia de ADN para poder buscar las posibles variantes. Por lo tanto, la secuenciación es un punto clave en el estudio. La secuenciación de nueva generación (NGS) está transformando rápidamente la investigación biomédica, especialmente en las aplicaciones de diagnóstico clínico. Las tecnologías NGS, ahora permiten la secuenciación completa del exoma (*Exome-Seq*), la secuenciación del genoma completo (WGS), la secuenciación del transcriptoma completo (*RNA-Seq*) incluso la secuenciación de una región concreta del DNA (*Targeted-Seq*) o una combinación.

La tecnología de NGS incluye diferentes metodologías, algunas de las más utilizadas hoy en día son:

- Illumina (Solexa) (Illumina, 2013)
- Roche 454 (Roche Applied Science, 2006)
- Ion torrent (Ion Torrent by Life Technologies, 2011)
- SOLiD (Solid and Delivers, 2008)

A demás NGS posibilita la detección de variantes genéticas causantes de enfermedades humanas, especialmente en el cáncer. Las variantes genéticas identificadas por las tecnologías NGS incluyen (Behjati and Tarpey, 2013):

- **Variantes de un único nucleótido (SNV):** Se consideran variantes de un único nucleótido aquellas variantes que se observan menos del 1% en la población, si superan el 1% se consideran polimorfismos de un solo nucleótido (SNP). Las SNV son mutaciones puntuales y por lo tanto suelen estar relacionadas con enfermedades.
- **Pequeñas inserciones y deleciones (indels):** En el genoma se puede ocurrir la inserción o deleción de uno o varios nucleótidos, se observan variaciones en la longitud del ADN debido a este fenómeno.
- **Variaciones en de número de copias (CNV):** Es el resultado de tener más o menos de las dos copias esperadas de una región de ADN. Se pueden dividir en amplificaciones y eliminaciones.
- **Variantes estructurales grandes (SV):** Se producen mediante un reordenamiento de las regiones del ADN. Se puede dividir en inserciones, deleciones, inversiones y translocaciones.

Este grupo de variantes se puede clasificar en:

- Equilibradas (*Balanced*): Este tipo de variantes no producen alteraciones en los valores de *depth of coverage*¹, por lo tanto, son neutrales. Como inversiones y translocaciones de una región de DNA dentro o entre cromosomas.
- Desequilibradas (*Unbalanced*): Una de las características de estas variantes es que van acompañados de un aumento o una disminución de *depth of coverage* donde una parte del genoma se pierde dando lugar a una delección. (Escaramís *et al.*, 2013)

Para poder observar cada una de estas variantes genéticas se utilizan diferentes metodologías.

¹ *Depth of coverage*: Numero de reads (secuencias que se obtienen después de la secuenciación) que cubren una región del genoma. (Explicado mas detalladamente en el apartado de métodos).

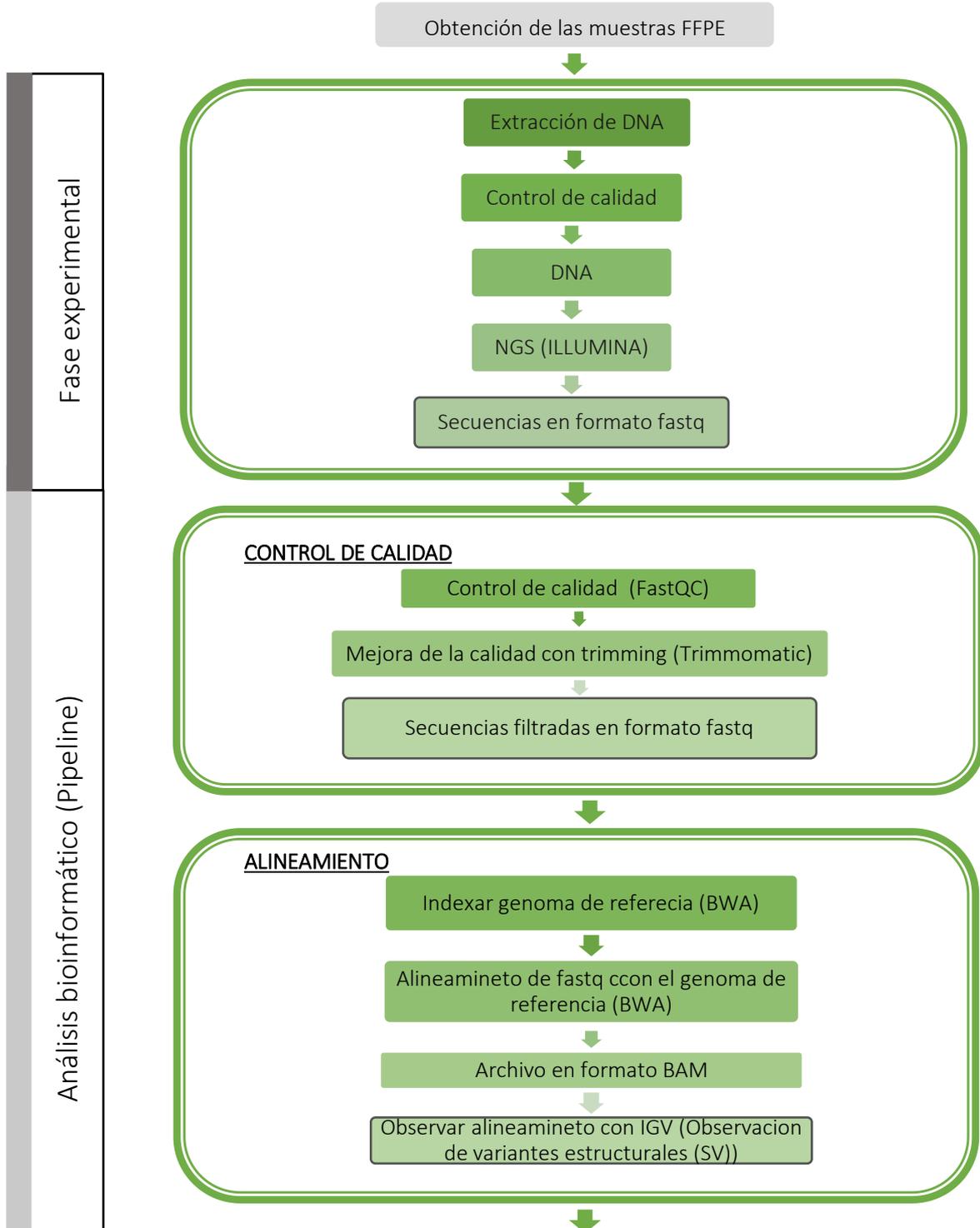
2. OBJETIVOS

Actualmente varios tipos de cáncer incluyen en su tratamiento terapias dirigidas a c-MET debido a su gran relación con procesos relacionados con la metástasis como la angiogénesis y la transición epitelio-mesénquimal. Se han descrito varias mutaciones que afectan a c-MET produciendo alteraciones en su expresión, por este motivo poder caracterizar nuevas mutaciones en el gen de c-MET será de gran ayuda para poder entender los mecanismos a partir de los cuales c-MET contribuye en procesos metastáticos. Estudios actuales se centran en la investigación de alteraciones de c-MET en NSCLC debido a su importancia en estados avanzados de la enfermedad. Como se a descrito anteriormente la firma genética es de gran relevancia en estos casos, por esta razón el proyecto se basa en poder caracterizar las variantes genéticas que se producen en el gen de c-MET en pacientes con NSCLC en estados avanzados con la finalidad de poder ayudar en la caracterización de la firma genética. Se han descrito variantes de c-MET producidas a partir de la alteración de exones como c-MET^{Δ14} y c-MET^{Δ7-8}. Especialmente el análisis de las variantes que producen c-MET^{Δ14} es de gran importancia debido a su relación con el cáncer de NSCLC.

- 1- Identificar y verificar si se producen alteraciones en el exón 14 y en el exón 7 del gen de c-MET tanto en pacientes con cáncer de NSCLC como en pacientes con cáncer de glioma.
- 2- Caracterizar las mutaciones que se producen en el exón 14 de c-MET en pacientes con NSCLC y entender el mecanismo a partir del cual estas alteraciones participan en procedimientos cancerígenos.
- 3- Desarrollar un pipeline bioinformático con la finalidad de poder analizar los dos puntos anteriores.

3. MATERIALES Y MÉTODOS

La metodología se divide en métodos experimentales y en métodos de análisis bioinformático. La fase experimental recae en la extracción y obtención de DNA de pacientes de cáncer de glioma y pacientes con cáncer de NSCLC para realizar la secuenciación mediante Illumina (NGS). El análisis bioinformático se centra en el estudio de las variantes que se producen en el gen de c-MET. A continuación, en el diagrama se describen los diferentes pasos que se realizaron para poder obtener los resultados finales y cumplir los objetivos. Cada apartado del diagrama será explicado detalladamente.



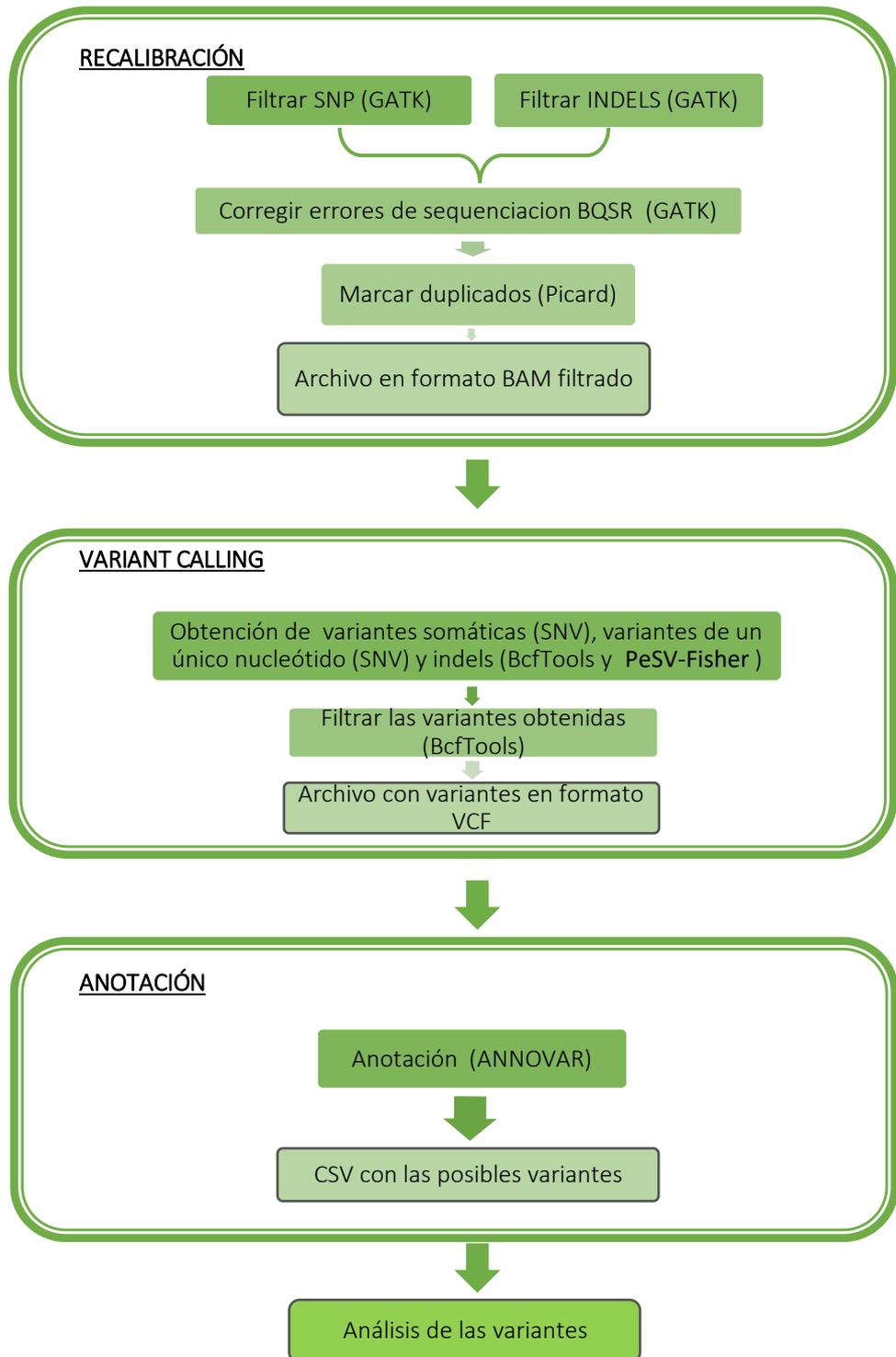


Figura 5. Esquema de la metodología utilizada. Los métodos utilizados se dividen en métodos experimentales (fase experimental) y métodos bioinformáticos (análisis y pipeline bioinformático).

3.1 Pipeline experimental.

Las muestras analizadas provienen de un proyecto europeo donde participan el Hospital Universitario Quirón Dexeus, el Hospital Clínico y Pangaea Oncology. En concreto se analizaron 5 muestras de pacientes con NSCLC y 5 muestras de cáncer de glioma (4 provenían de pacientes y 1 de la línea celular E98). Previamente se realizó un análisis de expresión de las muestras y se utilizaron aquellas donde se observaron alteración en la expresión del gen de c-MET. El estudio se centra en la investigación de NSCLC, pero se utilizaron muestras de cáncer de glioma con delección del exón 7 para realizar la validación de la delección mediante NGS y como control del análisis. Una vez realizada la extracción de las muestras (biopsia), estas fueron fijadas con formaldehído y parafina (FFPE), un proceso que permite mantener la morfología celular y archivar muestras de cada paciente para futuras investigaciones. El procedimiento de extracción de la muestra y fijación fue realizado por los respectivos hospitales.

La parte experimental explicada a continuación fue desarrollada y puesta en práctica por el equipo de laboratorio de la empresa, se explicará de manera breve en que consisten cada uno de los pasos experimentales.

3.1.1 Extracción de DNA.

La extracción de DNA de muestras FFPE se realiza con la finalidad de poder obtener el genoma de las diferentes muestras para poder amplificarlo y secuenciarlo posteriormente. El hecho de que las muestras sean FFPE dificulta la extracción y disminuye el rendimiento de las extracciones, es decir la cantidad de DNA obtenido es menor a la que se obtiene cuando se extrae DNA de tejido no fijado, esto es debido a que el DNA puede fragmentarse en muestras FFPE. Para extraer DNA de muestras fijadas en parafina se utilizó el Kit de extracción de Qiagen QIAamp DNA FFPE Tissue para muestras FFPE (Quiagen, 2012).

3.1.2 Control de calidad.

Una vez realizada la extracción se procedió a realizar el control de calidad el cual consistió en:

- **Integridad del DNA:** Comprobar la calidad del DNA utilizando electroforesis mediante un gel de agarosa². La finalidad es poder observar si se ha producido la degradación o no del ADN. Si la muestra no está degradada se observa una única banda muy bien definida en el gel, en cambio si se observa una banda poco definida podemos decir que se ha degradado parte del DNA.

² Protocolo de gel de agarosa y electroforesis en apéndice.

- **Cuantificación de DNA:** La cuantificación se realizó con la tecnología de Quantus (Promega Corporation, 2017) con el objetivo de saber la cantidad de DNA que se ha obtenido durante la extracción. La concentración esperada tiene que ser alrededor de 50 ng/μl.

3.1.3 Secuenciación por Illumina.

La secuenciación de las muestras se realizó mediante el método de Illumina y se utilizó la maquinaria de secuenciación MiniSeq (MiniSeq™ System, 2006). Illumina utiliza la amplificación clonal y la secuenciación por síntesis (SBS) de esta manera identifica simultáneamente las bases del DNA mientras se sintetizan las cadenas. Cada base emite una señal fluorescente única a medida que se une a la nueva cadena. Este proceso nos permite identificar que base hay en cada posición del genoma de interés. El tipo de secuenciación realizada fue *target sequencing*, mediante este método se secuenciaron únicamente la región de interés, es decir, el gen de c-MET completo. Se secuenciaron las 10 muestras y también se realizó la secuenciación de una muestra control proporcionada por Illumina.

La secuenciación por Illumina requiere de varios pasos:

- **Preparación de librerías:** En este paso se realiza la fragmentación del DNA seguida por la unión de los adaptadores (secuencias que permitirán la generación de clúster) que se unirán a los extremos 5' y 3' de los varios fragmentos de DNA. El protocolo utilizado para la preparación de librerías fue diseñado por MiniSeq Systems (MiniSeq™ System, 2018).
- **Dilución:** La concentración de las librerías tiene que ser la adecuada. Todas las muestras deben tener la misma concentración. Por este motivo se realizó un protocolo que permite poner a punto la concentración de las librerías³.
- **MiniSeq:** Las librerías son introducidas dentro del secuenciador donde se producirá la generación de los clústers y la secuenciación (Illumina, 2013). Durante la secuenciación se pueden obtener *paired-end reads (PE)* o *single-end reads (SE)*. Los *paired-end* se producen a partir de la secuenciación desde los dos extremos del DNA 5' y 3' de los fragmentos de DNA en cambio el método de *single-end* se realiza la secuenciación de uno de los extremos. En este caso se obtuvieron *paired-end reads*.

Al final del proceso se obtienen secuencias *paired-end* de 150 pb, también llamadas *reads* en un archivo en formato fastq.

³ Protocolo de dilución de las librerías en apéndice.

3.2 Pipeline bioinformático.

A continuación, se muestra el análisis bioinformático realizado para poder encontrar mutaciones somáticas que afectan al gen de c-MET. Las variantes analizadas fueron: variantes de un único nucleótido (SNV), indels y variantes estructurales (SV). La secuenciación realizada permitió obtener la secuencia completa de la proteína c-MET. Pero las regiones en las que nos centramos durante el análisis fueron el exón 14 (posición en el genoma 116,411,903-116,412,043) y el exón 7 (posición en el genoma 116,397,491-116,397,593) teniendo como base la isoforma de c-MET NM_000245.2 (*Genome Browser*) y siempre teniendo como referencia la versión del genoma humano GRCh37 (hg19). En el procedimiento descrito a continuación se utilizó como referencia las prácticas de Variant Calling desarrolladas por GATK⁴, de manera que durante el uso de las herramientas bioinformáticas se utilizaron los parámetros estándar excepto en algunos casos. El programa fue desarrollado utilizando los comandos de lenguaje de *BASH* (.sh) y *Python* (.py). Se realizaron scripts⁵ diferentes para cada una de las partes del programa que están descritas a continuación.

Las herramientas bioinformáticas durante el proceso pretenden aumentar la sensibilidad del análisis, puesto que se analizan datos clínicos y por lo tanto uno de los factores más importantes es que el número de falsos negativos y falsos positivos sea nulo. De esta manera se aumenta la fiabilidad de los resultados (mutaciones) reportados al final del análisis. Por ese motivo se han escogido las herramientas teniendo en cuenta que se realiza el estudio de muestras clínicas que provienen de pacientes con cáncer y por lo tanto se buscan mutaciones somáticas. También se tiene en cuenta que las muestras eran FFPE (parafina) lo cual afecta a la calidad de las muestras y que la secuenciación se realizó utilizando la plataforma de Illumina.

⁴ **GATK:** *Genome Analysis Toolkit*, contiene un conjunto de herramientas que permiten realizar el análisis de datos que provienen de NGS. (McKenna *et al.*, 2010)

⁵ **Script:** documento que contiene instrucciones escritas en códigos de programación.

3.2.1 Control y mejora de la calidad.

Una vez realizada la secuenciación mediante Illumina se debe realizar un control de calidad de las secuencias obtenidas también llamadas *reads*. Como el análisis realizado fue de *paired-reads* se obtuvieron dos archivos de *reads*, uno llamado *forward* y otro *reverse*, la calidad se observó en los dos archivos. El control de calidad nos indicara si las muestras se encontraban en buen estado, si la calidad del ADN utilizado era la óptima y si la secuenciación y la parte experimental se han realizado de manera correcta.

Al finalizar el procedimiento de secuenciación obtenemos un archivo en formato *fastq* el cual contiene los *reads* y nos indica la calidad para cada base mediante una puntuación que muestra la fiabilidad de que la base que se ha obtenido sea la base correcta. Para poder analizar los datos obtenidos con el formato de *fastq* se utilizó el programa de *FastQC* versión 0.11.5 (Andrews, 2010) este programa nos permite analizar diferentes parámetros que nos indican el índice de calidad de los resultados. Los parámetros analizados fueron:

- La calidad por base de cada secuencia.
- La puntuación de calidad para cada secuencia.
- El porcentaje de cada base que tienen las secuencias.
- El contenido de GC.
- La longitud de cada secuencia.
- Niveles de duplicación y de sobreexpresión de secuencias.

A continuación, para mejorar la calidad de las secuencias se realiza el proceso de *trimming* donde se tienen en cuenta los resultados obtenidos con el programa de *FastQC*. En el primer análisis con *FastQC* se observó que las primeras 75 bases de cada *read* no tenían la calidad esperada (este fenómeno se observó en las 10 muestras analizadas) además, había un gran número de secuencias sobrerrepresentadas, este análisis nos sugirió un posible error experimental.

Para poder mejorar la calidad se realizó el *trimming* utilizando como herramienta *Trimmomatic* (Bolger, Lohse and Usadel, 2014) (versión 0.36). Primero se eliminaron las primeras 75 bases de cada *read*, utilizando la función *CROP* de *Trimmomatic*. Después se volvió a realizar el *trimming* eliminando los adaptadores y utilizando las funciones por defecto de *Trimmomatic*, las cuales permiten eliminar bases del principio y del final del *read* que tienen una calidad inferior a 3, mediante las funciones *LEADING:3 TRAILING:3*. También miraba las bases de cuatro en cuatro eliminando el conjunto de cuatro bases cuando el valor de calidad promedio era de 15 y se eliminaron los *reads* con una longitud inferior a 36 estos dos pasos se pudieron realizar gracias a las funciones *SLIDINGWINDOW:4:15 MINLEN:36*.

A demás se creo un programa con Python el cual filtraba las secuencias sobrerrepresentadas obtenidas en el archivo de calidad y las eliminaba del archivo que contenía los *reads*, *fastq*. Concretamente se eliminaron las que tenían un porcentaje de repetición del 0,12 %.

Una vez mejorada la calidad de los archivos de *fastq* se volvió a realizar el control de calidad mediante FastQC para comprobar que la calidad había mejorado. Se pudo observar que la calidad de los *reads* mejoro así que este método nos permitió eliminar los errores experimentales.

3.2.2 Alineamiento.

El alineamiento se realizó utilizando la versión hg19 del genoma humano de referencia obtenido de la base de datos de *GATK*. El alineamiento nos permitirá ordenar los *reads* obtenidos durante la secuenciación y por lo tanto mapearlo/alinearlo con el genoma de referencia para después poder identificar las posibles mutaciones. Antes de realizar el alineamiento primero se realizo la indexación del genoma de referencia la cual permite marcar el genoma de referencia para permitir que el alineamiento se realice de manera mas eficaz. Los programas empleado para realizar la indexación fueron BWA⁶ (Li and Durbin, 2009), versión 0.7.12-r1039 y Samtools, versión 0.1.18. También se obtuvo un diccionario el cual contiene la posición de cada cromosoma creado mediante Picard (versión 2.18.5).

Para realizar el alineamiento se utilizó BWA. Previamente hay que pasar los archivos *fastq* (tanto el archivo *forward* como el *reverse*) a formato *.sai* utilizando la función de BWA *aln*. El formato *.sai* contiene las coordenadas para cada *read* formando una matriz de sufijos la cual permitirá realizar el alineamiento de una manera mas rápida. Durante el alineamiento se compararon los resultados obtenidos con dos comandos diferentes BWA *sampe* y BWA *mem*. Una vez finalizado el alineamiento se obtiene un archivo en formato *.sam* con los *reads* que se han alineado al genoma de referencia. Seguidamente se convierte a formato *.bam* y se ordena el archivo (proceso de *sorting reads*) utilizando Samtools y Picard.

Finalmente, se observaron los resultados con la herramienta de visualización de IGV (Robinson *et al.*, 2011), la cual nos permite observar la cobertura (cantidad de *reads* que representan una parte del genoma) y también nos permite identificar variantes estructurales.

Uno de los parámetros que se tuvo en cuenta a la hora de analizar con que comando se obtenían mejores resultados fue la cobertura. Los resultados indicaron que el programa que nos permitió observar una mejor cobertura fue BWA *mem*. También se tuvo en cuenta la velocidad con la que el

⁶ BWA: Burrows-Wheeler Aligner

programa realizaba el alineamiento y se observó que también era el mejor en este sentido por lo tanto facilitaba el proceso computacional. De este modo el comando que se utilizó para realizar los alineamientos durante el proceso de análisis fue BWA mem. Al finalizar se obtiene un archivo BAM que contiene los *reads* alineados al genoma de referencia.

3.2.3 Recalibración.

La recalibración nos permite reducir el nivel de error durante la obtención de las variantes (Variant Calling). Estos errores son debidos a desalineaciones producidas por *indels* (inserciones o deleciones) (Tian *et al.*, 2016), duplicaciones ópticas (en este caso fueron marcadas pero no eliminadas) y errores sistemáticos producidos por el secuenciador cuando estima la calidad y otorga una puntuación a cada base. Para realizar la recalibración se utilizaron los softwares de GATK y Picard versión 2.18.5, además se empleó la base de datos de *indels* (*Mills_and_1000G_gold_standard.indels.hg19.sites.mod.vcf*) proporcionada por GATK y la base de datos de SNPs (*common_all_20170710.vcf*) sin relevancia clínica. Las funciones utilizadas de los diferentes softwares fueron: (Mckenna *et al.*, 2010; Tian *et al.*, 2016):

➤ **Recalibración de errores producidos por *indels*:**

- GATK *RealignerTargetCreator*: Define los intervalos donde se realizará el realineamiento local.
- GATK *IndelRealigner*: Realiza un realineamiento local de los *reads* alrededor de los *indels*.

➤ **Recalibración de errores sistemáticos mediante *BQSR* (*Base Quality Score Recalibration*) :**

- GATK *BaseRecalibrator*: Detecta errores sistemáticos en las puntuaciones de calidad. En este paso también se filtran los *indels* y los SNPs utilizando las bases de datos descritas anteriormente.

➤ **Recalibración de duplicaciones ópticas:**

- Picard *MarkDuplicates*: Identifica *reads* duplicados.

La recalibración se aplica sobre el archivo .bam creado durante el alineamiento así que al final de la recalibración obtenemos un archivo BAM filtrado.

3.2.4 Obtención de variantes somáticas (Variant Calling).

Este procedimiento nos permite obtener mutaciones somáticas mediante la detección de diferencias entre el genoma de referencia y el genoma de cada una de las muestras de NSCLC y cáncer de glioma. Tal y como se ha descrito existen diferentes tipos de variantes. En este caso el estudio se centra en observar variantes de un único nucleótido (SNV), indels y variantes estructurales (SV). Las variantes se analizaron con diferentes tipos de métodos (*Figura 6*):

➤ Análisis de variantes de SNV y indels:

Los programas utilizados durante el análisis fueron:

- La herramienta de *SamTools BcfTools mpileup*, versión 1.6
- *Delly*, versión 1.1, ((Rausch *et al.*, 2012)
- *BreakDancer*, versión 1.1.2, (Chen *et al.*, 2013)

Finalmente se utilizó BcfTools como programa para obtener las variantes. Este método compara el genoma de manera vertical y busca diferencias entre las bases del genoma de referencia y el genoma que se quiere analizar. Esta misma herramienta permitió filtrar los resultados y eliminar los que fueron marcados con baja calidad (se eliminaron los que tenían una calidad inferior a 10).

➤ Análisis de SV:

Inicialmente las variantes estructurales como las deleciones se pueden observar con IGV, pero se utilizan otros métodos que permiten verificar la presencia de estas variantes. Las variantes estructurales se identificaron con el programa de PeSV-Fisher versión 0.93 (Escaramís *et al.*, 2013). El programa utiliza varios algoritmos. PeSV-Fisher se basa en un análisis combinatorio de estrategias de *paired-reads* (observa alteraciones en la distancia en la que se encuentran los *reads*). Este método permite analizar varios tipos de variantes estructurales como deleciones, duplicaciones, translocaciones e inversiones. El método tiene en cuenta la distancia entre los *reads*, por ejemplo cuando se produce una deleción la distancia entre los *reads* aumenta en una región debido a la falta o disminución de la cobertura. También se realizaron estrategias de *depth of coverage* para detectar variantes estructurales, es decir se observó si el número de *reads* disminuye o aumenta en una región del DNA. PeSV-Fisher también fue utilizado para calcular y obtener datos de *depth of coverage*.

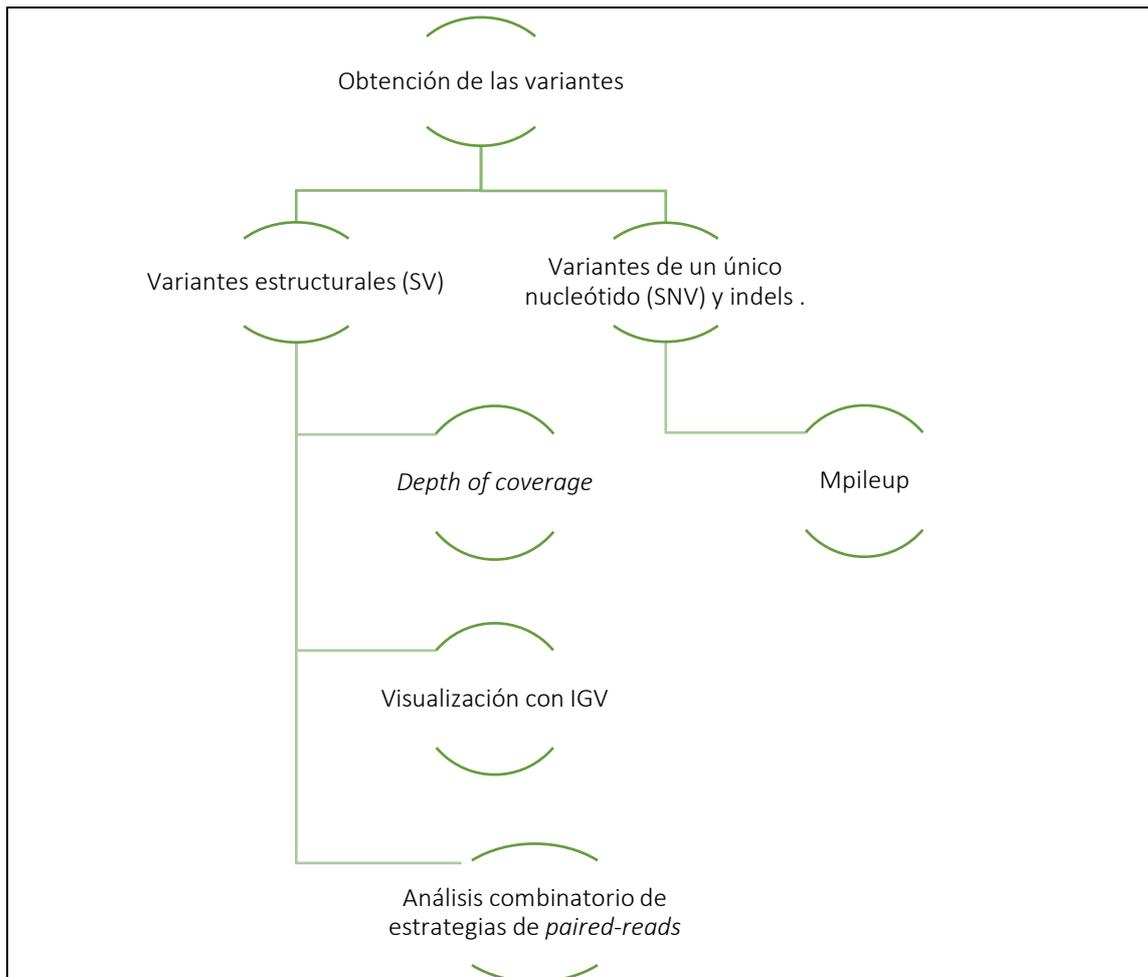


Figura 6. Métodos de análisis de las variantes. En el esquema se indican los diferentes métodos que nos permiten detectar variantes estructurales (SV) y variantes de un único nucleótido (SNV).

3.2.5 Anotación.

La anotación fue realizada con el programa de ANNOVAR, versión 2018Apr16. La anotación nos permite obtener información detallada sobre cada una de las variantes encontradas mediante el uso de varias bases de datos. Las bases de datos utilizadas fueron refGene, cytoBand, exac03, avsnp147, dbnsfp30a, dbnsfp33a, intervar_20170202, cosmic70, esp6500siv2_ea, 1000g2015aug, snp138, avsnp15, clinvar_20170905, popfreq_max_20150413, popfreq_all_20150413, regsnpintron y cadd obtenidas de la pagina de ANNOVAR.

Una vez realizada la anotación se obtuvo un archivo en formato CSV con la información de cada variante. Algunos de los parámetros básicos obtenidos mediante la anotación fueron: la posición de la variante, si se encuentra en un exón o en un intrón, si es una mutación sinónima o no (no afecta a la secuencia proteica se traduce el mismo aminoácido), si es un SNP y en el caso de que lo sea indica su código rs, nos indica si son variantes benignas y en algunos casos da información sobre el gen donde se produce la mutación.

3.2.6 Filtrar las variantes.

El archivo CSV obtenido contiene las posibles variantes del genoma de cada muestra, pero hay que analizarlas para saber si realmente son mutaciones, es posible que el programa utilizado para la obtención de las variantes nos de falsos positivos o falsos negativos. Por ejemplo, podría presentarnos un SNP (polimorfismo de un único nucleótido) como una mutación dando un falso positivo. También hay que observar si la mutación encontrada altera o no vías de señalización relacionadas con el cáncer de pulmón o cáncer de glioma y como la mutación puede afectar a la señalización.

Algunos de los resultados obtenidos se validaron utilizando Sanger, pero la validación utilizando Sanger no se realizó en la empresa.

Los parámetros que se tienen en cuenta para analizar las variantes fueron, *frame-stop-gain* y *frame-stop-loss*, ya que son parámetros que indican cambios en el pauta de lectura y que por lo tanto *afectarán* a la estructura de la proteína. También se analizan *missense*. Durante el análisis se descartaron las variantes que correspondían a SNP, las que estaban categorizadas como benignas por la base de datos de NCBI ClinVar o las que tenían un MAF (frecuencia del alelo menos común) superior a 2,3. También se utilizó la base de datos de COSMIC la cual contiene información sobre las mutaciones que se han encontrado en anteriores estudios.

3.2.7 Programa automático y funciones.

Se construyó un programa el cual permitía ejecutar partes por separado del pipeline bioinformático utilizado, con la finalidad de facilitar el uso ya que hay procedimientos donde únicamente se busca realizar una parte del proceso. Se creó un programa en Python para unir las diferentes partes del pipeline (calidad, indexación, trimming, alineamiento, recalibración, obtención de variables y anotación). De esta manera el usuario puede realizar el análisis vía terminal y puede escoger que partes del análisis quiere que se realicen. Se utilizó la función de *parser* de Python para crear las diferentes opciones de análisis. Las funciones se llaman mediante *flags*, estos son símbolos, en este caso letras que representan una función específica. Además están acompañados por un seguido de argumentos los cuales permiten dar información al programa sobre las variables que se emplearán durante el análisis.

Finalmente, los *flags* utilizados se dividieron en dos grupos, los *flags* individuales que permiten realizar cada etapa del análisis por separado y los *flags* que permitían la combinación de funciones. El programa está diseñado para trabajar sobre archivos de secuenciación en formato de *fastq*. Según el *flag* utilizado el programa llama a los scripts determinados.

Los argumentos son los siguientes:

- **Chr:** Esta opción permite introducir el cromosoma de interés en el caso de querer realizar el análisis únicamente en una región específica. De esta forma el programa tarda menos tiempo en obtener los resultados ya que debe analizar menos datos.
- **Href:** Esta opción permite introducir el genoma de referencia que se utilizara durante el análisis, según el tipo de muestras que se analizan el genoma de referencia es uno u otro.
- **Sample:** Permite indicar al programa sobre que muestra queremos que realice en análisis.
- **Stype:** El programa esta diseñado para poder trabajar con *paired-ends reads* (PE) o con *single-ends reads* (SE), por lo tanto, también hay que indicar al programa que tipo de análisis queremos realizar ya que el script de alineamiento es diferente para cada una de las opciones.
- **Out:** El nombre del archivo de salida. Una vez acabado el análisis el archivo se guarda utilizando el nombre de salida que se ha otorgado.

➤ **Funciones individuales**

Utilizando las funciones individuales (*Tabla 2*) se puede realizar cada uno de los puntos que contiene el pipeline.

Tabla 2. Funciones individuales. Lista con las diferentes funciones individuales que incluye el programa indicando su correspondiente flag, los argumentos que necesita y el Script que llama el programa cuando se utiliza.

Función	Flag	Argumentos	Script
Indexación	-i	'Chr' 'href'	rindex.sh
Trimming	-t	'Sample'	rtrimmer.sh
Calidad	-q	'sample'	rfastqc.sh
Alineamiento SE	-s	'sample' 'href' 'stype'	rbam_SE.sh
Alineamiento PE	-p	'sample' 'href' 'stype'	rbam_PE.sh
Recalibración	-r	'out' 'href' 'stype'	recalbam.sh
Obtención de Variantes (VC)	-c	'out' 'href' 'rtype'	rcalling.sh
Anotación	-a	'out' 'href' 'rtype'	ann.sh

➤ **Combinación de funciones.**

Las combinaciones de funciones (*Tabla 3*) permiten realizar mas de un paso del pipeline con la finalidad de realizar mas de un procedimiento utilizando un único flag. Este método facilita el uso y además agiliza el trabajo. Las combinaciones realizadas se realizaron teniendo en cuenta su utilidad.

Tabla 3. Combinación de funciones. Lista con las diferentes funciones individuales que incluye el programa indicando su correspondiente flag, los argumentos que necesita y el Script que llama el programa cuando se utiliza.

Función	Flag	Argumentos
Trimming + Calidad	-tq	'sample'
Alineamiento SE/PE	-sp	'sample' 'href' 'stype'
Indexación + Calidad + trimming + Calidad	-iqtq	'Chr' 'href' 'sample'
Indexación + Calidad	-iq	'Chr' 'href' 'sample'
Alineamiento+Recalibración+Calling+ Anotación	-arca	'sample' 'href' 'stype' 'out'
All	-all	'Chr' 'href' 'Sample' 'stype' 'out'

También contiene una función de ayuda (*-help*), al ser llamada se obtienen las instrucciones de como utilizar el programa y también aporta una lista con las posibles opciones del análisis de Variant Calling que se pueden utilizar.

Para utilizar el programa (script de Python) el cual finalmente fue llamado *pmain.py* únicamente hay que indicar el flag que se desea utilizar seguido de los argumentos específicos en cada flag tal y como se indica en la ayuda del programa. Como, por ejemplo: *Python pmain.py -i hg19 chr7*, mediante este comando el programa utilizará el script de índice y realizará la indexación únicamente del cromosoma 7 utilizando el genoma de referencia hg19. En cambio, el comando *Python pmain.py -iqtq S1 hg19 PE S1_PE* utilizará los scripts de alineamiento SE o PE, realizará la recalibración, *Variant Calling* y anotación teniendo en cuenta *hg19* como genoma de referencia, realizará el análisis de la muestra S1, el alineamiento de *paired-ends reads* y el archivo de salida tendrá el prefijo de S1_PE.

3.2.8 Simulación de la estructura proteica.

Una vez encontradas las mutaciones producidas en c-MET se realizo la simulación de la estructura proteica para observar de manera computacional como afectaban las mutaciones en la configuración de la proteína. Primero se busco en UniProt la proteína c-MET (P08581). A continuación, se observaron las secuencias cristalizadas de la proteína, se escogieron las secuencias que eran de nuestro interés y se obtuvo la secuencia mediante PDB (*Protein data bank*).

Una de las secuencias de c-MET utilizada fue 2UZY, esta representa desde la posición 22 a la 722 (contiene dominio SEMA, PSI y los dominios IPT1 IPT2), esta fue utilizada para hacer la simulación de la proteína de c-MET con delección del exón 7.

Para poder eliminar el exón 7 de la proteína se busco la secuencia en formato fasta del exón 7 en NCBI (NM_000245.3), se realizo la traducción a proteína con el Software de ENBOSS Transeq. Teniendo en cuenta las tres diferentes pautas de lectura (Seq A, Seq B y Seq C) se obtuvieron 3 secuencias y se escogió la secuencia correcta (seq B) descartando aquellas que contenían señales de STOP(*).

-Seq A: IEMHSWSCHE*AFQYVHNYFKWPRDNTIQYILLCX
-Seq B: LKCTVGPAMNKHFNMSIIISNGHGTTQYSTFSYV
-Seq C: *NAQLVLP*ISISICP*LFQMATGQHNTVHSPMX

A continuación, se realizo un alineamiento con Clustal Omega entre la secuencia obtenida de 2UZY y la secuencia del exón 7 para comprobar que la secuencia era correcta y que por lo tanto se encontraba en 2UZY. Una vez comprobado la secuencia B fue eliminada de 2UZY, con la finalidad de obtener c-MET^{A7}.

Se realizo en mismo procedimiento para la simulación de c-MET con delección del exón 14. En este caso se utilizo la secuencia 5UAB que engloba la posición 1023-1360 (contiene parte del dominio yuxtamembrana y tirosina quinasa). El resultado final no se muestra en el trabajo ya que el dominio yuxtamembrana esta cerca del dominio transmembrana, una región difícil de cristalizar. Por lo tanto, en PDB no se encuentra ninguna secuencia proteica que represente completamente a la región del dominio yuxtamembrana.

Finalmente se visualizaron la estructura de las proteínas utilizando el programa de Chimera, versión 1.12.0.

4. RESULTADOS Y DISCUSIÓN

En el siguiente apartado se describen los resultados obtenidos al finalizar el análisis bioinformático. Algunos de los resultados obtenidos no son reportados en el trabajo debido a que estos serán publicados en un futuro y por lo tanto son datos confidenciales de la empresa.

4.1 Variantes de c-MET observadas en NSCLC.

El análisis bioinformático nos permitió observar la presencia de alteraciones en el exón 14 en la isoforma NM_000245 de c-MET en pacientes con NSCLC. El análisis bioinformático llevado a cabo permitió observar variantes estructurales (SV) en dos de las muestras de NSCLC. Concretamente se identificaron caídas de la cobertura en regiones que involucraban al exón 14 indicando posibles deleciones. La detección de este tipo de variantes se pudo realizar gracias al análisis de *depth of coverage* y el análisis combinatorio de estrategias de *paired-reads* (PeSV-Fisher). En cambio, en las tres muestras restantes no se observó una caída de cobertura, sino que se observaron variantes de un único nucleótido (SNV) y indels, mediante el análisis de *mpileup*. También se realizó el análisis de alteraciones en el exón 7 para poder observar si sucede el mismo fenómeno que en los casos de cáncer de glioma, donde el funcionamiento de la proteína c-MET se ve afectado a causa de la deleción del exón 7, tal y como se describió en la bibliografía (Navis *et al.*, 2015). Al finalizar del análisis se pudo verificar que no se produjo la deleción del exón 7 en la isoforma NM_000245 en ninguno de los pacientes con NSCLC. Las mutaciones reportadas durante el análisis ayudarán en la investigación de la firma genética de NSCLC.

4.1.1 Variantes estructurales (SV) en el exón 14 de c-MET.

Finalizado el análisis se pudo distinguir variantes estructurales que afectaban al exón 14 en c-MET en dos de las muestras analizadas de NSCLC. Las SV reportadas fueron:

- Caída en la cobertura (*depth of coverage*) desde el intrón 13 al intrón 14, incluyendo el exón 14 en una de las muestras NSCLC.

Se observó una caída de la cobertura en parte del intrón 13, el exón 14 y parte del intrón 14 (*Figura 7A*). La disminución de la cobertura en esta región nos indica la presencia de una deleción. Estos resultados indican que es posible que la deleción se encuentre en heterocigosis ya que en este caso se observó una caída de la cobertura y no una pérdida total. De esta forma el número de *reads* que representan al exón 14 disminuye, debido a que en el genoma de partida hay menos cantidad de copias que representan al exón 14 a causa de la heterocigosis, es decir la deleción solo se ha producido en uno de los cromosomas. También existe la posibilidad de que la muestra de células extraída del paciente contuviera células en las que se produjo la deleción y células que

todavía mantienen el exón 14. Este fenómeno se justifica debido a que el DNA proviene de células cancerígenas las cuales se someten continuamente a mutaciones, por lo tanto, los pacientes analizados podrían tener células en el mismo tumor con distintas alteraciones.

- Caída de cobertura de 70 pb de distancia en el exón 14 en una de las muestras NSCLC. En este caso no se observó la delección completa del exón 14 sino que únicamente se observó la caída de cobertura en una región dentro del exón, la cual era de 70pb. La caída de cobertura nos indica la presencia de una pequeña delección en el exón 14. Tal y como se explicó anteriormente la caída de cobertura indica que posiblemente la delección se encuentra en heterocigosis.

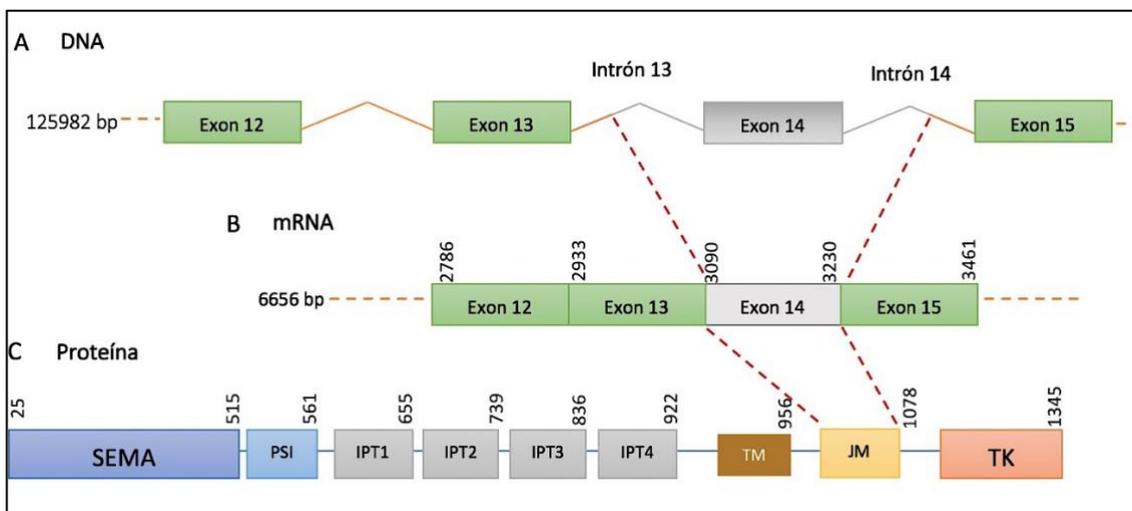


Figura 7. Variantes estructurales (SV) y indels observados en c-MET. A Esquema del DNA desde el exón 12 al exón 15. En gris se representa la región del DNA (parte del intrón 13, el exón 14 y el intrón 14) donde se produce una caída de cobertura indicando la presencia de una delección en heterocigosis. **B** Esquema del mRNA con las posiciones de los diferentes exones, donde se puede apreciar que la delección del exón 14 producirá la eliminación de las bases 3090 hasta la 3230 del mRNA. **C** la proteína de c-MET donde se puede observar como el exón 14 codifica la secuencia del dominio yuxtamembrana, JM (posición 957 a 1078).

4.1.2 Variantes de un único nucleótido (SNV) y indels observados en el exón 14 de c-MET.

Mediante el método de mpileup, después de validar y filtrar las diferentes mutaciones encontradas se observaron SNV y indels (*Figura 8A*) en 3 de las muestras restantes de NSCLC. Las SNV observadas fueron:

- Un indel en la posición +2 del intrón 14.
- Una mutación en la última posición del exón 14.
- Un indel en la posición +1 del intrón 15 junto mutaciones en el exón 21.

En la posición +2 del intrón 14 se ha observado un indel. Concretamente en esta posición se encuentra una de las bases del sitio donador, tal y como se describe en la introducción (*Figura 4*). La maquinaria de splicing no será capaz de reconocer la secuencia donadora y por lo tanto se realizará un splicing aberrante ya que la maquinaria de splicing no es capaz de distinguir en que posición del DNA se realizará el corte durante el splicing, produciendo un splicing aberrante que podrá dar lugar a la pérdida del exón 14 en el mRNA.

También se ha observado una mutación en la última posición del exón 14. Alteraciones en las últimas posiciones de un exón también perjudican al proceso de splicing alternativo ya que las bases que se encuentran al final o al principio de un exón ayudan a determinar que regiones hay que cortar y unir durante el splicing (*Figura 4*). Por lo tanto, esta mutación producirá la pérdida del exón 14 durante el splicing y el mRNA obtenido no contendrá el exón 14.

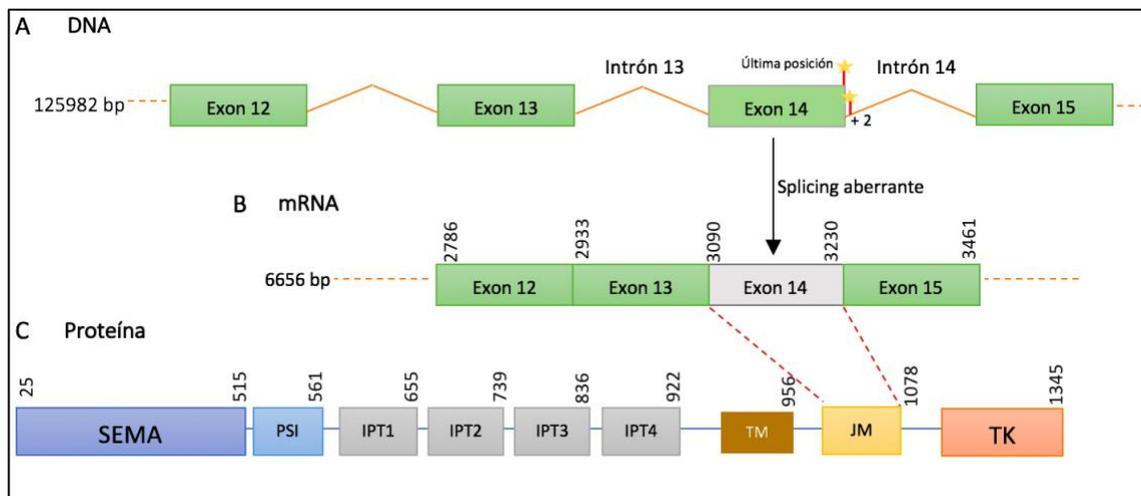


Figura 8. Variantes de un único nucleótido (SNV) y indels que afectan al exón 14 en c-MET en NSCLC. **A** Esquema del DNA desde el exón 12 al exón 15. Se indican las posiciones donde se producen SNV y indels (en la última posición del exón 14 una mutación y en la posición +2 del intrón 14 un indel) que afectan al splicing y por lo tanto conllevan la pérdida del exón 14 **B** Esquema del mRNA con las posiciones de los diferentes exones, donde se puede apreciar que la delección del exón 14 producirá la eliminación de las bases 3090 hasta la 3230 del mRNA. **C** la proteína de c-MET donde se puede observar como el exón 14 codifica la secuencia del dominio de yuxtamembrana, JM (posición 957 a 1078).

Como realizo la secuenciación de todo el gen de c-MET también se reportaron mutaciones que no afectaban al exón 14 como es el caso de la identificación de un indel en la posición +1 del intrón 15 junto con mutaciones en el exón 21 (en la misma muestra). El proyecto se enfocaba en el análisis del exón 14 pero la identificación de estas mutaciones podrá utilizarse para futuros estudios. Además esto nos indica que un proto-oncogen puede tener más de una mutación en un mismo tipo de cáncer, demostrando la gran importancia de utilizar terapias personalizadas durante el tratamiento de enfermedades como el cáncer.

4.1.3 Alteraciones en la estructura proteica de c-MET debido a la falta del exón 14.

Se ha podido observar que tanto las SV, SNV y indels reportados se relacionan con la pérdida del exón 14 en c-MET (c-MET^{Δ14}) en pacientes con NSCLC en estados avanzados. La pérdida del exón 14 conlleva a la pérdida de las bases 3090 a la 3230 del mRNA. El exón 14 codifica el dominio yuxtamembrana. La pérdida del exón 14 afecta directamente a la estructura del dominio yuxtamembrana afectando al funcionamiento y regulación de c-MET (*Figura 8 y Figura 7*). El dominio yuxtamembrana permite la regulación negativa de c-MET mediante la unión de c-CBL a Tyr1003 (*Figura 9*). Cuando c-CBL se une al dominio yuxtamembrana se produce la ubiquitinación de c-MET permitiendo su degradación (Thien and Langdon, 2005). Pero cuando se produce la delección del exón 14 c-CBL no puede unirse a c-MET debido a que la pérdida del dominio yuxtamembrana comporta la pérdida de Tyr1003, produciendo una desregulación de c-MET causada por la alteración del proceso de regulación negativa. La desregulación de c-MET producirá alteraciones de varios procesos celulares relacionados con la angiogénesis, la supervivencia, la movilidad, la dispersión, la diferenciación, la morfogénesis y la angiogénesis tal y como se ha descrito en la bibliografía (Salgia, 2009; Garnett *et al.*, 2013). El análisis realizado nos permite decir que en estudios preliminares de pacientes con NSCLC en estados avanzados se produce la alteración de estos procesos biológicos favoreciendo los mecanismos de metástasis debido a la pérdida del exón 14.

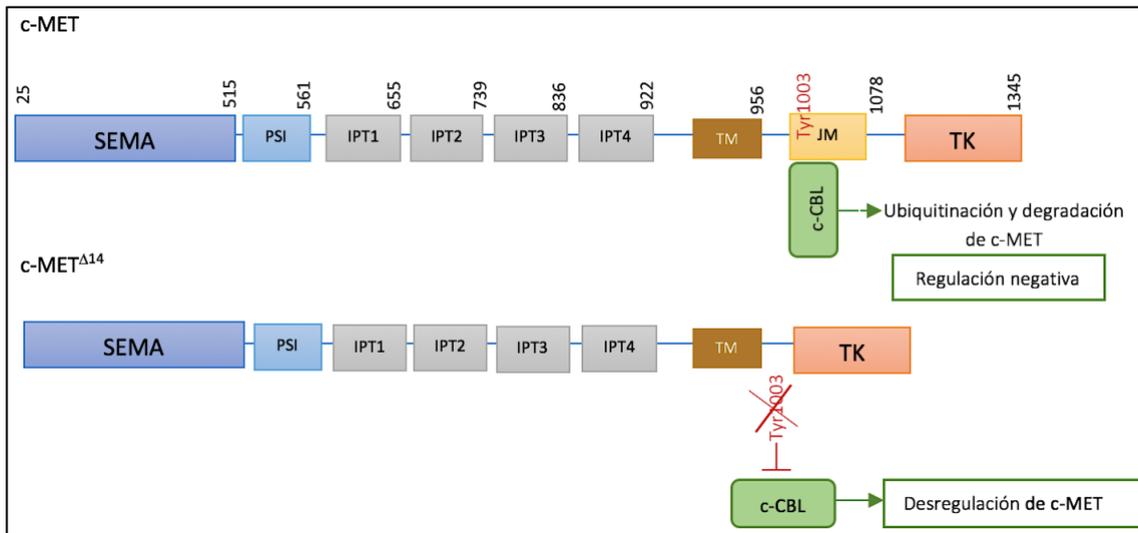


Figura 9. Alteración de la regulación negativa de c-MET debido a la pérdida del exón 14. Se representa la estructura proteica de c-MET esquemáticamente donde se puede observar que la pérdida del exón 14 produce la pérdida del dominio yuxtapmembrana (JM), por lo tanto, el efector c-CBL no podrá unirse a c-MET causando su desregulación.

Gracias al pipeline bioinformático utilizado se han podido caracterizar SV, SNV y indels que producen alteraciones en el exón 14 de la proteína de c-MET en pacientes con NSCLC. Se ha podido observar diferentes alteraciones en las muestras de NSCLC indicando que el exón 14 se ve afectado por más de un tipo de mutación. La pérdida del exón 14 se puede producir a partir de varios procesos, algunos donde se produce directamente la deleción del exón 14 y en otros casos donde se producen SNV o indels que afectan al proceso de splicing causando la eliminación del exón 14 durante el proceso. En futuros estudios sería conveniente realizar el análisis de más muestras de NSCLC en estados avanzados para poder verificar que la pérdida del exón 14 en c-MET sea una alteración común en este grupo de pacientes. Además, realizar más estudios aumentando el número de muestras permitirá caracterizar mejor las mutaciones que se producen en pacientes con esta enfermedad. El hecho de que la deleción de este exón produzca alteraciones en la regulación negativa de c-MET será un punto clave a la hora de buscar un fármaco o tratamiento para la enfermedad, pero determinar las mutaciones que afectan al gen hará que el tratamiento pueda ser mucho más específico y selectivo.

4.2 Verificación de la delección del exón 7 de c-MET en muestras de cáncer de glioma.

Este procedimiento se realizó con la finalidad de verificar la delección del exón 7 de c-MET (c-MET^{A7}) en muestras de cáncer de glioma. Es decir, se utilizaron como control para saber si la metodología experimental y bioinformática utilizada fue la correcta. El análisis bioinformático utilizado nos permitió observar la delección (SV) en el exón 7 en la isoforma NM_000245 de c-MET en pacientes con cáncer de glioma (*Figura 10A*). Se identificó la falta completa de cobertura en el exón 7 o del exón 7 y parte del 8 en las 4 muestras que provenían de pacientes con cáncer de glioma y en la línea celular E98. Estos resultados eran los esperados tal y como se describe en la bibliografía (Navis *et al.*, 2015). Así que, podemos decir que durante el análisis se identificó y verificó la una delección en homocigosis del exón 7 en pacientes con cáncer de glioma. Estos resultados nos indican que la metodología utilizada fue la correcta ya que las herramientas bioinformáticas utilizadas nos permitieron detectar la delección. También se realizó el análisis del exón 14 pero no se encontraron alteraciones en muestras de cáncer glioma.

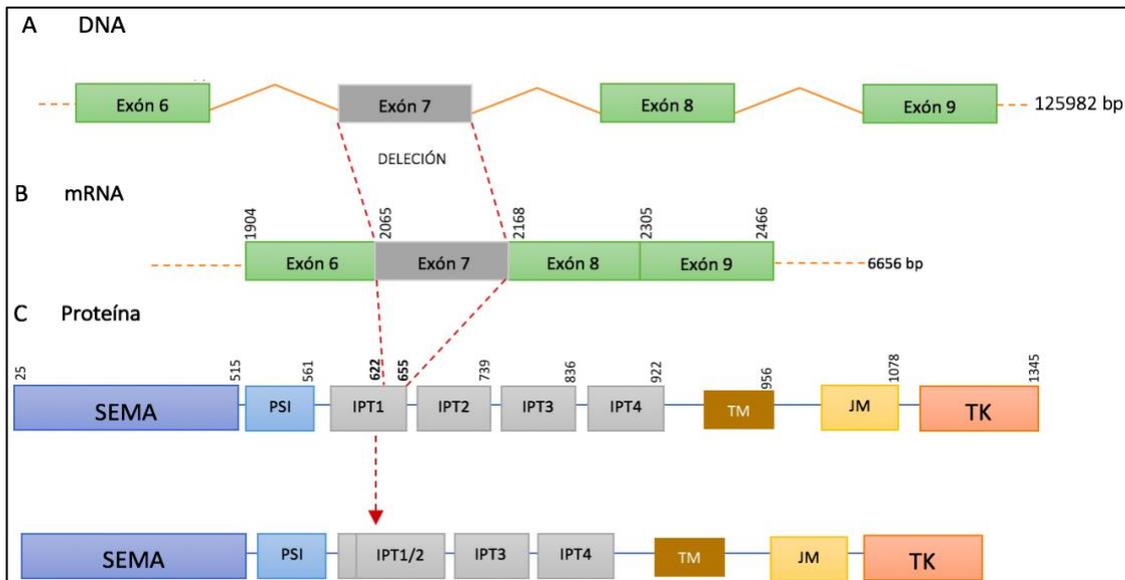


Figura 10. Delección del exón 7 en muestras con cáncer de glioma. **A** Representación de un fragmento de DNA de c-MET desde el exón 6 hasta el exón 9. Se puede observar la delección del exón 7. **B** La pérdida del exón 7 a nivel de mRNA. **C** Estructura de la proteína de c-MET. Debido a la delección del exón 7 se produce la pérdida de parte de la secuencia proteica (de la posición 622 a la 655) afectando parte del dominio IPT1.

4.2.1 Alteraciones en la estructura proteica de c-MET debido a la falta del exón 7.

El exón 7 codifica parte del dominio IPT1 de manera que la delección del exón 7 produce alteraciones en este dominio. Concretamente se pierden de la posición 622 a la 655 de la secuencia proteica de c-MET, dando lugar a la fusión del dominio IPT1 con el dominio IPT2 (*Figura 10C*). Los dominios IPT se relacionan con el reconocimiento celular y con la unión a la membrana celular. La alteración del dominio IPT1 causa un aumento de c-MET a nivel intracelular, pero todavía no está del todo claro el mecanismo a partir del cual se produce este fenómeno. Lo que sí se ha podido observar es que c-

MET queda autoactivada y no depende de HGF para su activación. Reconocer esta delección es de gran importancia para decidir el tipo de tratamiento que se utilizara, ya que los pacientes con esta delección no responden a terapias que utilizan a HGF como diana (Navis *et al.*, 2015).

También se pudo observar mediante el diseño de proteínas como afecta la pérdida del exón 7 en c-MET a nivel estructural. Se puede observar como la estructura de IPT1 se ve afectada causando también alteraciones en IPT2 (Figura 11).

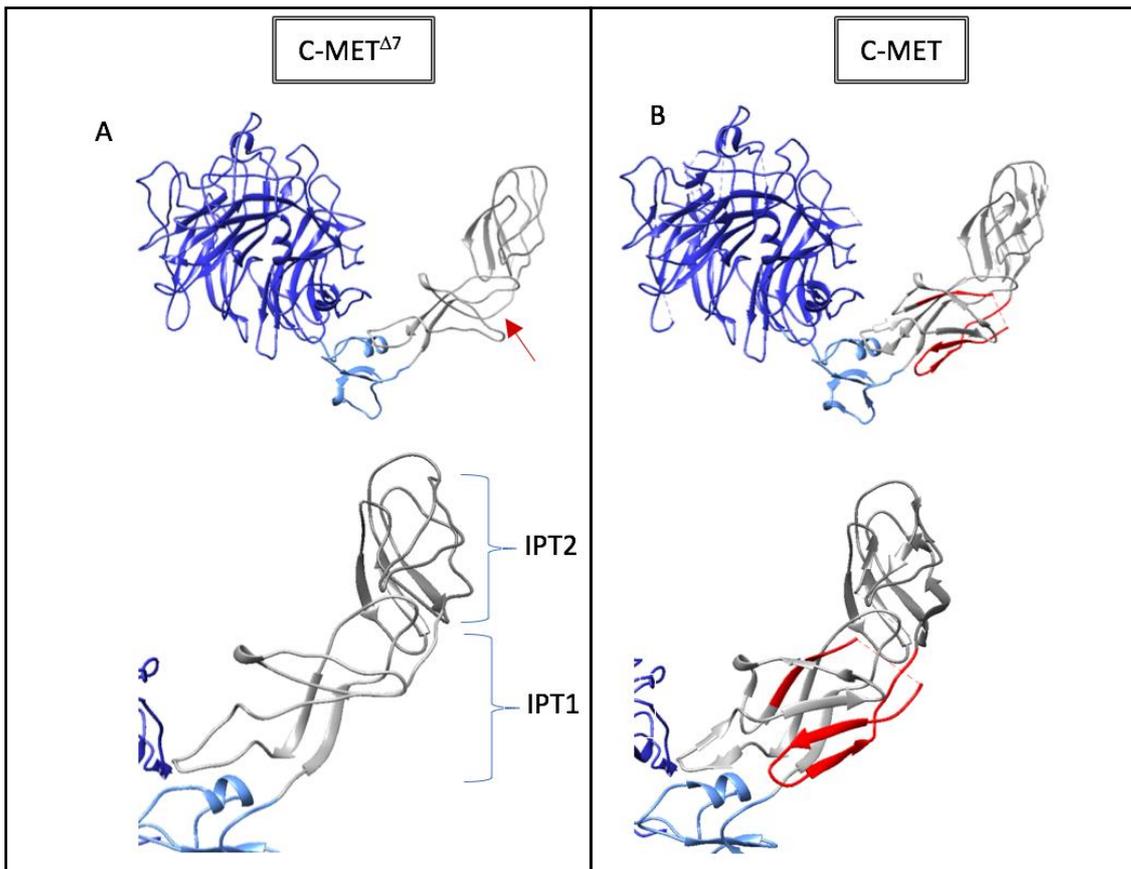


Figura 11. Simulación de parte de la estructura proteica de c-MET y c-MET^{Δ7}. **A** Representación de parte de la estructura proteica de c-MET con la pérdida del exón 7. Se observa que se pierde parte de dominio IPT1 produciendo la alteración de IPT2. **B** Representación de parte de la estructura de c-MET sin alteraciones. En rojo se muestra la parte del dominio IPT1 que se elimina en c-MET^{Δ7}. Dominio SEMA (azul oscuro), dominio PSI (azul claro), dominios IPT1 y IPT2 (gris).

4.3 Fenómeno de exclusión.

El estudio preliminar realizado indica que las alteraciones que se producen en el exón 7 y en el exón 14 de c-MET son excluyentes, es decir, cuando se producen alteraciones en el exón 7 no se observan alteraciones en el exón 14 y viceversa. No podemos atribuir este fenómeno a ningún perfil de pacientes o a una causa concreta ya que no tenemos el historial clínico de los pacientes, lo único que sabemos es que la pérdida del exón 14 se produce en estados avanzados de NSCLC y que la pérdida del exón 7 se observa en pacientes con cáncer de glioma. Sería necesario volver a realizar el estudio aumentando el número de muestras y comprobar si se en algunos de los casos se produce la eliminación tanto del exón 7 como del exón 14. En todo caso sería interesante realizar un estudio funcional in vitro para poder entender porque son excluyentes o porque la delección del exón 14 únicamente se produce en casos de cáncer de NSCLC y la delección del exón 7 en cáncer de glioma. Si se observara la eliminación de los dos exones también sería conveniente realizar un estudio funcional para entender que repercusión tienen a nivel biológico la pérdida o alteración de los dos exones.

Este fenómeno también nos indica que probablemente el tipo de tratamiento que debe de administrarse en casos de NSCLC y casos de glioma sea distinto, ya que en los dos casos se observan alteraciones en el gen de c-MET, pero se producen mutaciones diferentes lo cual puede condicionar el efecto del fármaco que se administra, debido a las diferencias de conformación que adopta la proteína según las mutaciones que se producen. Por este motivo realizar más estudios es un punto clave para poder saber el mecanismo a partir del cual las mutaciones descritas afectan al funcionamiento celular. Además en el caso de que se realizaran más estudios y estos indicaran que se produce la exclusión sería interesante observar cual es la causa ya que pueden ser varias causas por ejemplo que la evolución del tumor sea diferente según el tejido o por la farmacología administrada durante el tratamiento previo.

5. CONCLUSIONES

- Los genes relacionados con enfermedades cancerígenas pueden tener mas de un tipo de mutación y estas mutaciones puede ser diferente según el tipo de cáncer y según el historial clínico del paciente por ese motivo es importante caracterizar la firma genética en cada caso.
- Para poder analizar las variantes somáticas que se producen en cáncer se utilizan diferentes metodologías. El análisis de *depth of coverage*, la visualización con IGV y análisis combinatorio de estrategias de *paired-reads* que utiliza PeSV-Fisher permiten obtener variantes estructurales (SV). En cambio el método de mpileup permite obtener variantes de un único nucleótido (SNV) y indels.

Las conclusiones obtenidas forman parte de un estudio preliminar donde únicamente se analizaron diez muestras. Para poder obtener conclusiones mas robustas y concluyentes es necesario realizar futuros estudios donde se analice un mayor numero de muestras.

- En el gen de c-MET se observan deleciones que afectan al exón 14 y variantes de un único nucleótido o indels que afectan al proceso de splicing alternativo produciendo la perdida del exón 14 en casos de NSCLC.
- La perdida del exón 14 de c-MET produce alteraciones en el dominio yuxtamembrana alterando la regulación negativa de c-MET en casos de NSCLC (Van Der Steen *et al.*, 2016).
- En pacientes con NSCLC no se observan alteraciones en el exón 7 de c-MET, pero si se observan alteraciones en este exón en pacientes con cáncer de glioma.
- Se observar que las alteraciones que se producen en el exón 7 y en el exón 14 son excluyentes pero la confirmación de este fenómeno requiere de su validación.
- En el gen de c-MET se observan alteraciones en el intrón 15 y con mutaciones en el exón 21 en casos de NSCLC.

6. BIBLIOGRAFIA

- Andrews, S. (2010) *A quality control tool for high throughput sequence data*. <<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>>, [28 de mayo de 2018].
- Behjati, S. and Tarpey, P. S. (2013) 'What is next generation sequencing?', *Archives of Disease in Childhood: Education and Practice Edition*, 98(6), pp. 236–238. doi: 10.1136/archdischild-2013-304340.
- Bielenberg, D. R., Zetter, B. R. and Program, V. B. (2016) 'HHS Public Access', 21(4), pp. 267–273. doi: 10.1097/PPO.000000000000138.The.
- Bolger, A. M., Lohse, M. and Usadel, B. (2014) 'Trimmomatic: A flexible trimmer for Illumina sequence data', *Bioinformatics*, 30(15), pp. 2114–2120. doi: 10.1093/bioinformatics/btu170.
- Bryan T. MacDonald; Keiko Tamai and Xi He (2010) 'Wnt/ β -catenin signaling: components, mechanisms, and diseases', *Developmental biology*, 17(1), pp. 9–26. doi: 10.1016/j.devcel.2009.06.016.Wnt/.
- Cantini, L. *et al.* (2018) 'Classification of gene signatures for their information value and functional redundancy', *npj Systems Biology and Applications*. Springer US, 4(1), p. 2. doi: 10.1038/s41540-017-0038-8.
- Chen, K. *et al.* (2013) 'BreakDancer - An algorithm for high resolution mapping of genomic structure variation', *Nature Methods*, 6(9), pp. 677–681. doi: 10.1038/nmeth.1363.BreakDancer.
- Chibon, F. (2013) 'Cancer gene expression signatures', *European Journal Of Cancer*, 49(8), pp. 2000–2009. doi: <https://doi.org/10.1016/j.ejca.2013.02.021>.
- Escaramís, G. *et al.* (2013) 'PeSV-Fisher: Identification of Somatic and Non-Somatic Structural Variants Using Next Generation Sequencing Data', *PLoS ONE*, 8(5). doi: 10.1371/journal.pone.0063377.
- Frampton, G. M. *et al.* (2015) 'Activation of MET via diverse exon 14 splicing alterations occurs in multiple tumor types and confers clinical sensitivity to MET inhibitors', *Cancer Discovery*, 5(8), pp. 850–860. doi: 10.1158/2159-8290.CD-15-0285.
- Garnett, J. *et al.* (2013) 'Regulation of HGF Expression by Δ EGFR-Mediated c-Met Activation in Glioblastoma Cells', *Neoplasia*, 15(1), pp. 73–IN21. doi: 10.1593/neo.121536.
- Hammerschmidt, S. and Wirtz, H. (2009) 'Lung cancer: current diagnosis and treatment.', *Deutsches Ärzteblatt international*, 106(49), pp. 809–18; quiz 819–20. doi: 10.3238/arztebl.2009.0809.

- Illumina (2013) 'Illumina Sequencing Technology'. doi: 10.1016/S0167-7799(03)00189-6.
- Instituto Nacional del Cáncer (2018a) *Estatificación. 2018* <<https://www.cancer.gov/espanol/cancer/diagnostico-estadificacion/estadificacion>>, [Consulta: 29 de mayo de 2018].
- Instituto Nacional del Cáncer (2018b) *Tasa de supervivencia general.* <<https://www.cancer.gov/espanol/publicaciones/diccionario/def/tasa-de-supervivencia-general>>, [Consulta: 29 de mayo de 2018].
- Ion Torrent by Life Technologies (2011) 'Ion Torrent Amplicon Sequencing', *Application Note*, pp. 1–5. doi: 10.7171/jbt.12.
- Kong-Beltran, M. *et al.* (2006) 'Somatic mutations lead to an oncogenic deletion of Met in lung cancer', *Cancer Research*, 66(1), pp. 283–289. doi: 10.1158/0008-5472.CAN-05-2749.
- Kong, Y. *et al.* (2016) 'Structural Basis for Plexin Activation and Regulation', *Neuron*, 91(3), pp. 548–560. doi: 10.1016/j.neuron.2016.06.018.
- Lamouille, S., Xu, J. and Derynck, R. (2014) 'Molecular mechanisms of epithelial-mesenchymal transition.', *National Review Molecular Cell Biology*, 15(3), pp. 178–196. doi: 10.1038/nrm3758.Molecular.
- Li, H. and Durbin, R. (2009) 'Fast and accurate short read alignment with Burrows-Wheeler transform', *Bioinformatics*, 25(14), pp. 1754–1760. doi: 10.1093/bioinformatics/btp324.
- Liebmann, C. (2001) 'Regulation of MAP kinase activity by peptide receptor signalling pathway: Paradigms of multiplicity', *Cellular Signalling*, pp. 777–785. doi: 10.1016/S0898-6568(01)00192-9.
- Ma, L. and Weinberg, R. A. (2008) 'Micromanagers of malignancy: role of microRNAs in regulating metastasis.', *Trends in Genetics*, 24(9), pp. 448–456.
- Mckenna, A. *et al.* (2010) 'The Genome Analysis Toolkit : A MapReduce framework for analyzing next-generation DNA sequencing data', pp. 1297–1303. doi: 10.1101/gr.107524.110.20.
- MiniSeq™ System (2006) 'MiniSeq™ System'.
- MiniSeq™ System (2018) 'Denature and Dilute Libraries Guide Overview', pp. 1–13.
- Molecular Biology Web Book (2018) *RNA Splicing* <<http://www.web-books.com/MoBio/Free/Ch5A4.htm>>, [Consulta 27 de mayo de 2018].
- Moran-Jones, K., Brown, L. M. and Samimi, G. (2015) 'INC280, an orally available small molecule inhibitor of c-MET, reduces migration and adhesion in ovarian cancer cell models', *Scientific Reports*, 5. doi: 10.1038/srep11749.

- Navis, A. C. *et al.* (2015) 'Identification of a novel MET mutation in high-grade glioma resulting in an auto-active intracellular protein', *Acta Neuropathologica*. Springer Berlin Heidelberg, 130(1), pp. 131–144. doi: 10.1007/s00401-015-1420-5.
- Organ, S. L. and Tsao, M. S. (2011) 'An overview of the c-MET signaling pathway', *Therapeutic Advances in Medical Oncology*, 3(1), pp. S7–S19. doi: 10.1177/1758834011422556.
- Piguet, A.-C. *et al.* (2015) 'Impact of MET targeting on tumor-associated angiogenesis and growth of MET mutations-driven models of liver cancer', *Genes & Cancer*, 6(7–8), pp. 317–327. doi: 10.18632/genesandcancer.74.
- Promega Corporation (2017) 'Quantus™ Fluorometer'.
- Qamsari, E. S. *et al.* (2017) 'The c-Met receptor: Implication for targeted therapies in colorectal cancer', *Tumor Biology*, 39(5). doi: 10.1177/1010428317699118.
- Quiagen (2012) 'October 2012 Sample & Assay Technologies QIAGEN Sample and Assay Technologies'.
- Rausch, T. *et al.* (2012) 'DELLY: Structural variant discovery by integrated paired-end and split-read analysis', *Bioinformatics*, 28(18), pp. 333–339. doi: 10.1093/bioinformatics/bts378.
- Reungwetwattana, T. *et al.* (2017) 'The race to target MET exon 14 skipping alterations in non-small cell lung cancer: The Why, the How, the Who, the Unknown, and the Inevitable', *Lung Cancer*. Elsevier Ireland Ltd, 103, pp. 27–37. doi: 10.1016/j.lungcan.2016.11.011.
- Robinson, J. T. *et al.* (2011) 'Integrative Genome Viewer', *Nature Biotechnology*, 29(1), pp. 24–6. doi: 10.1038/nbt.1754.Integrative.
- Roche Applied Science (2006) 'Genome Sequencer 20 System First to the Finish', *Genome*.
- Rosário, M. and Birchmeier, W. (2003) 'How to make tubes: Signaling by the Met receptor tyrosine kinase', *Trends in Cell Biology*, pp. 328–335. doi: 10.1016/S0962-8924(03)00104-1.
- Salgia, R. (2009) 'Role of c-Met in Cancer: Emphasis on Lung Cancer', *Seminars in Oncology*, 36(SUPPL. 1), pp. 1–12. doi: 10.1053/j.seminoncol.2009.02.008.
- Sociedad Americana del Cáncer (2018) *Tasas de supervivencia por etapa para el cáncer de pulmón no microcítico*. <<https://www.cancer.org/es/cancer/cancer-de-pulmon-no-microcitico/deteccion-diagnostico-clasificacion-por-etapas/tasas-de-supervivencia.html>>, [Consulta 29 de mayo de 2018].
- Solid, T. and Delivers, G. (2008) 'Revolutionary Technology SOLID Platform', pp. 1–16. Available at: [papers2://publication/uuid/FBD481BE-0408-45FD-AFF4-B4E7FC29E6F3](https://pubs2://publication/uuid/FBD481BE-0408-45FD-AFF4-B4E7FC29E6F3).

- Van Der Steen, N. *et al.* (2016) 'CMET Exon 14 Skipping: From the structure to the clinic', *Journal of Thoracic Oncology*. Elsevier Inc, 11(9), pp. 1423–1432. doi: 10.1016/j.jtho.2016.05.005.
- Suzuki, K. *et al.* (2016) 'Structure of the plexin ectodomain bound by semaphorin-mimicking antibodies', *PLoS ONE*, 11(6). doi: 10.1371/journal.pone.0156719.
- Sylvester, P. W. (2014) 'Targeting met mediated epithelial-mesenchymal transition in the treatment of breast cancer.', *Clinical and translational medicine*, 3(1), p. 30. doi: 10.1186/s40169-014-0030-5.
- Thien, C. B. F. and Langdon, W. Y. (2005) 'c-Cbl and Cbl-b ubiquitin ligases: substrate diversity and the negative regulation of signalling responses', *Biochemical Journal*, 391(2), pp. 153–166. doi: 10.1042/BJ20050892.
- Thomas, S. J. *et al.* (2015) 'The role of JAK/STAT signalling in the pathogenesis, prognosis and treatment of solid tumours', *British Journal of Cancer*. Nature Publishing Group, 113(3), pp. 365–371. doi: 10.1038/bjc.2015.233.
- Tian, S. *et al.* (2016) 'Impact of post-alignment processing in variant discovery from whole exome data', *BMC Bioinformatics*. BMC Bioinformatics, 17(1), pp. 1–13. doi: 10.1186/s12859-016-1279-z.
- Watson, J. D. (2013) *Molecular Biology of the Gene*. 7th edn.
- Zappa, C. and Mousa, S. A. (2016) 'Non-small cell lung cancer: current treatment and future advances', *Translational Lung Cancer Research*, 5(3), pp. 288–300. doi: 10.21037/tlcr.2016.06.07.
- Zhang, Y. *et al.* (2018) 'Function of the c-Met receptor tyrosine kinase in carcinogenesis and associated therapeutic opportunities', *Molecular Cancer*. Molecular Cancer, 17(1), pp. 1–14. doi: 10.1186/s12943-018-0796-y.

APÉNDICE

Electroforesis y preparación del gel de agarosa.

MATERIALES Y REACTIVOS:

EDTA 1X (tampón de electroforesis)

Agarosa 1%

SYVR Green

Marcador de peso molecular

Cubeta de electroforesis

PROCEDIMIENTO:

Preparación del gel: Para realizar un gel se necesita un volumen de 150 mL de EDTA 1X (tampón de electroforesis) al 1 % de agarosa. Por lo tanto, se pesan 1,5 gramos de agarosa y se mezclan en un erlenmeyer con los 150 mL de EDTA 1X. Se calienta en el microondas aproximadamente durante dos minutos. Se deja enfriar (sin que solidifique) y se añaden 8 µl de SYVER green para cada 100 ml en este caso se añadieron 12 ml. A continuación, se vuelca dentro de la cubeta (con el peine colocado para después poder realizar la carga) y se deja gelificar. Una vez gelificado se llena la cubeta con el tampón de electroforesis y con cuidado se saca el peine. Seguidamente se mezclan tanto las muestras de ADN como el marcador de peso molecular con 0.2 volúmenes de tapón de electroforesis. El volumen total estará determinado por el tamaño de los pocillos, habitualmente 15-30 µl. Y finalmente se corre el gel.

Dilución de librerías

MATERIALES Y REACTIVOS:

- 1 N NaOH,
- Solución tampón para la hibridación (Componente del kit de MiniSeq)
- 200 mM Tris-HCl, pH 7.0
- RSB
- Papeles indicadores de pH

PROCEDIMIENTO:

Método de normalización estándar:

- a) Preparar una dilución fresca de NaOH

1. Descongelar a temperatura ambiente.
2. Combinar los siguientes volúmenes en un tubo de 1.5 ml.
 - Agua de laboratorio (950 μ l)
 - Stock 2.0 N NaOH (50 μ l)
 - El volumen total es 1 ml de NaOH 0,1 N.

3. Invertir el tubo varias veces para mezclar.

NOTA: Utilizar la dilución nueva después de 12 horas.

- b) Preparar la solución tampón de hibridación

1. Descongelar a temperatura ambiente la solución tampón de hibridación.
2. Una vez descongelado almacenarlo entre 2 ° C y 8 ° C hasta que se realice su uso.
3. Agitar brevemente en el vortex antes de usar.

- c) Preparar RSB

NOTA En lugar de RSB, se puede utilizar Tris-HCl 10 mM, pH 8.5 con Tween 20 al 0.1%.

1. Retirar el tubo de RSB de -25 ° C a -15 ° C y descongelarlo a temperatura ambiente.
2. Cuando se haya descongelado, almacenarlo entre 2 ° C y 8 ° C hasta que se realice su uso.
3. Verificar el pH con los papeles indicadores de pH.

Diluir la librería a 1 nM:

1. En base a la concentración de la librería, transfiera la librería un nuevo tubo de microcentrífuga y agregue RSB.

Concentración de la librería	Volumen de la librería	Volumen de RSB
10 nM	10 μ l	90 μ l
4 nM	10 μ l	30 μ l
2 nM	10 μ l	10 μ l

NOTA. Nuestra biblioteca debe estar a 4nM

2. Vortex brevemente y luego centrifugar a 280 revoluciones durante 1 minuto.

a) Desnaturalizar la librería

1. Combinar los siguientes volúmenes 1 nM de librería y 0.1 N de NaOH en un tubo de microcentrifuga.

- Biblioteca de 1 nM (5 µl)
- 0.1N NaOH (5 µl)

2. Vortex brevemente y luego centrifugar a 280 × g durante 1 minuto.

3. Incubar a temperatura ambiente durante 5 minutos.

4. Añadir 5 µl 200 mM Tris-HCl, pH 7.0.

NOTA. Verificar el pH con los papeles indicadores de pH.

7. Vortex brevemente y luego centrifugar a 280 revoluciones durante 1 minuto.

b) Diluir la biblioteca a la concentración de carga.

1. Agregar 985 µl de tampón de hibridación al tubo de la librería desnaturalizada. El volumen total es de 1 ml a 5 pM.

2. Vortex brevemente y luego centrifugar a 280 revoluciones durante 1 minuto.

3. Transferir 140 µl de la librería diluida a un nuevo tubo de microcentrifuga.

4. Agregar 360 µl de tampón de hibridación. El volumen total es de 500 µl a 1.4 pM.

5. Vortex brevemente y luego centrifugar a 280 revoluciones durante 1 minuto.

NOTA. La concentración diluida final de la biblioteca debe optimizarse para evitar el agrupamiento excesivo.

$$V(\text{Lib } 5\text{pM}) = \frac{C_f * 500\mu\text{l}}{5\text{pM}}$$

$$V(\text{Hyb Buff}) = 500\mu\text{l} - V(\text{Lib } 5\text{pM})$$

C_f = Concentración final (1.4 pM en este momento)

V (Lib 5pM) = Volumen final de la librería diluida 5pM

V (Hyb Buff) = Volumen final de tampón de hibridación

Desnaturalizar y diluir el control PhiX:

a) Diluir PhiX a 4 nM

1. Descongelar un tubo de reserva PhiX 10 nM (10 µl/tubo).

2. Combinar los siguientes volúmenes en un tubo de microcentrifuga.

- 10 nM PhiX (10 µl)
- RSB (15 µl)

El volumen total es de 25 µl a 4 nM.

3. Vortex brevemente y luego pulsar en la centrífuga.

NOTA [Opcional] Almacenar el PhiX 4 nM a -25 ° C a -15 ° C hasta 3 meses.

b) Desnaturalizar PhiX

1. Combine los siguientes volúmenes en un tubo de microcentrífuga.

- nM PhiX (5 µl)
- NaOH 0.1 N (5 µl)

2. Vortex brevemente y luego pulsar en la centrífuga.

3. Incubar a temperatura ambiente durante 5 minutos.

4. Añadir 5 µl de Tris-HCl 200 mM, pH 7,0.

5. Vortex brevemente y luego centrifugar a 280 revoluciones durante 1 minuto.

c) Diluir PhiX desnaturalizado a concentración de carga

1. Agregue 985 µl de tampón de hibridación al tubo de PhiX desnaturalizado. El volumen total es de 1 ml a 20 pM.

2. Diluir la PhiX 20 pM desnaturalizada a 1.4 pM de la siguiente manera.

- PhiX desnaturalizado (35 µl)
- Tampón de hibridación preenfriada (465 µl) El volumen total es de 500 µl a 1,4 pM.

NOTA. La concentración final de PhiX debe ser la misma que la de nuestra librería diluida.

3. Invertir para mezclar y luego centrifugar a 280 revoluciones durante 1 minuto. ¡NO VORTEX!

NOTA [Opcional] Almacenar el PhiX de 1,4 pM desnaturalizado a -25 ° C a -15 ° C durante un máximo de 2 semanas.

d) Combinar la librería y el control PhiX

librería y el control PhiX	Librería
Control PhiX Desnaturalizado y diluido)	5 µl
Librería (Desnaturalizada y diluida)	500 µl

NOTA. Solo se agregarán 500 µl de mezcla al cartucho de MiniSeq.

2. Reservar en hielo hasta que esté listo para cargar en el cartucho de reactivo de MiniSeq.

