

DETECCIÓN DE MUTACIONES DEL GEN PI3K EN CÁNCER DE MAMA MEDIANTE PIROSECUENCIACIÓN

FACULTAD DE CIENCIAS DE LA SALUD

SECCIÓN DE MEDICINA
[DEPARTAMENTO DE CIENCIAS MÉDICAS BÁSICAS]
[SERVICIO DE ANATOMÍA PATOLÓGICA]

TRABAJO DE FIN DE GRADO presentado por

Ginés Koy Salido
Ismael Gómez Pérez

[\(alu0101127934@ull.edu.es\)](mailto:alu0101127934@ull.edu.es)

[\(alu0101098647@ull.edu.es\)](mailto:alu0101098647@ull.edu.es)

TUTORA

Dra. Rosa Nieves Rodríguez Rodríguez

Índice

Agradecimientos	2
1. Resumen	2
2. Palabras clave	3
3. Introducción	3
4. Hipótesis de trabajo y objetivos	9
4.1 Objetivo primario.....	10
4.2 Objetivos secundarios.....	10
5. Material y métodos	10
5.1 Endopredict High/Low Risk.....	11
5.1.1 Obtención de datos.....	11
5.1.1.1 Purificación ARN.....	11
5.1.1.2 Quality Control.....	13
5.1.1.3 Test EndoPredict (EP).....	14
5.1.2 Hoja de cálculo: Excel.....	15
5.2 PI3K mutated/wild type.....	17
5.2.1 Purificación ADN.....	17
5.2.1.1 Cromatografía por afinidad.....	17
5.2.1.2 Con fenol-cloroformo (PCI).....	19
5.2.2 Cuantificación del ADN mediante el NanoDrop.....	20
5.2.3 Diseño de primers.....	21
5.2.4 PCR e9, e20.....	22
5.2.5 Pirosecuenciación.....	23
6. Resultados	25
7. Discusión	28
8. Conclusiones	31
9. ¿Qué hemos aprendido durante este TFG?	32
10. Bibliografía	32
11. Figuras	35
12. Tablas	45
13. Anexos	46

Agradecimientos

Nos gustaría manifestar nuestro agradecimiento a Ayelen por darnos su apoyo incondicional a lo largo de este proyecto.

A la Dra. Rosa y al Dr. Salido por despertarnos el interés por la Anatomía Patológica.

Y a Maruca por despertarnos la pasión por la ciencia.

1. Resumen

La proliferación celular consiste en cuán rápidamente una célula cancerosa copia su ADN y se divide en dos células. En el cáncer de mama esta capacidad se incrementa por mutaciones en oncogenes tales como el de la fosfoinositol 3-cinasa (*PI3K*), y se refleja en resultados elevados del *Ki67* o de alto riesgo en el test EndoPredict (EP). Este resultado de alto riesgo implica una mayor probabilidad de recidiva del cáncer y determina la utilización de quimioterapia en el tratamiento del cáncer de mama Luminal B. En este estudio hemos explorado si existe una correlación entre las mutaciones más frecuentes de *PI3K* y el resultado del test EP (o alguno de los genes analizados en este score) así como con la proliferación celular (evaluada mediante porcentaje de *Ki67*). Hemos encontrado una asociación significativa entre el *Ki67* y el resultado del test EP; sin embargo, mutaciones en *PI3K* no se asocian significativamente al score global endopredict (EPclin), aunque sí lo hacen con un parámetro de cálculo intermedio del test EP (el *score* de proliferación), a causa, en mayor medida, de la expresión de uno de los genes que utiliza en su algoritmo (el gen *UBE2C*). Esta es una enzima, implicada en la cascada de *PI3K-akt-mTOR*, que podría tener trascendencia en la proliferación, el pronóstico y el tratamiento en futuros estudios del cáncer de mama.

Abstract

The proliferation of cancer cells consists of DNA replication and division into more cells at a rapid pace. In breast cancer this capacity can be enhanced by mutations in oncogenes like phosphatidylinositol 3-kinase (*PI3K*), and it is reflected in the presence of elevated levels of *Ki67* or a high risk score in the EndoPredict (EP) test. A high risk score implies a bigger likelihood of recurrence and helps determine the use of chemotherapy to treat type B luminal breast cancer. In this study we wanted to see if there was a correlation between the most common

mutations in *PI3K* and the EP score (or some of the genes analyzed in this test), and with the cellular proliferation (determined by a high Ki67 percentage). We found an important association between Ki67 and the EP score; however, mutations in *PI3K* don't seem to have a great impact on the total EP score (EPclin), although they do seem to have it on a parameter used to calculate the EP score (the proliferation score), likely caused by the expression of one of the genes used in its algorithm (*UBE2C* gene). This is an enzyme, implicated in the *PI3K-akt-mTOR* cascade, that could play an important role in proliferation, prognosis, and in the treatment for future studies on breast cancer.

2. Palabras clave

Cáncer de mama, *PI3K*, test EndoPredict, Ki67, *UBE2C*.

3. Introducción

El gen *PI3K* (*PIK3CA*) se sitúa en el cromosoma 3, concretamente en el brazo largo y en la banda 26.32 (3q26.32). Está compuesto de 91.980 nucleótidos (nt), empezando en el nucleótido 179.148.114 y acabando en el nucleótido 179.240.093¹. Según “Ensembl”², *PI3K* contiene 23 exones (es posible que en algunas bases de datos como ClinVar o OMIM se vean otros resultados debido al splicing alternativo). En total se identificaron 5 transcritos diferentes: PIK3CA-201, PIK3CA-205, PIK3CA-203, PIK3CA-204, y PIK3CA-202; este último no codifica para ninguna proteína. Esto evidencia el splicing alternativo ya mencionado; salen varias especies de mRNA (cada una de distinto tamaño y secuencia codificante) de un solo gen. El mRNA de referencia (PIK3CA-201) mide 9.104 nt, pero su secuencia codificante mide 3.207 nt. Su traducción da lugar a la proteína PI3K conformada por 1068 aminoácidos (aa) y con un peso molecular de 124.284 Da. Se trata de una enzima con una estructura cuaternaria compleja formada por: 1) una subunidad catalítica y 2) una subunidad reguladora que se une a p85 (también llamada PIK3R1, PIK3R2 o PIK3R3)³.

Las fosfoinositol-3-quinazas (PI3K), son enzimas que forman parte de la cascada de señalización que regula una gran serie de fenómenos celulares normales, tales como son el crecimiento, proliferación, diferenciación, motilidad (migración), y la supervivencia celular. Esta

familia de quinasas, actúa fosforilando los grupos OH 3, 4 y 5 del anillo inositol del lípido PIP2 unido a la membrana celular, dando lugar al lípido ya modificado PIP3. La activación de las mismas depende fundamentalmente del dominio citoplasmático del receptor Tyr quinasa (RTK) estimulado por un factor de crecimiento. Tras este paso inicial, PIP3 y mTOR2 activan a AKT (o PKB), que en conjunto constituyen la cascada de señalización *PI3K-AKT*. PTEN es una proteína capaz de regular directamente la señalización *PI3K-AKT* mediante la desfosforilación de los grupos fosfatos añadidos a PIP3. Ante la ausencia de PTEN, AKT seguirá siendo estimulada la cual bloqueará a FOXO (factor de transcripción clave en la senescencia y muerte celular) y activará a mTOR1 (potenciador del metabolismo y, por lo tanto, del crecimiento celular).

El cáncer de mama se origina en la unidad terminal del ducto-lobulillo de la glándula mamaria. Si el tumor está confinado por la membrana basal, se considera un carcinoma in situ. En cambio, si ha traspasado la membrana basal e infiltra el estroma circundante, se considera un carcinoma infiltrante. Tanto el carcinoma in situ como el infiltrante pueden tener un origen ductal o lobulillar. También se puede clasificar el tumor según sus características moleculares: ER positivo (expresa receptores de estrógeno), HER2 positivo (mutación en los receptores con actividad Tyr quinasa para el crecimiento celular), y triple negativo (no expresa los receptores hormonales ni está mutado HER2). Gracias a esta última clasificación, se han identificado 4 tipos de cánceres de mama en: 1) Luminal A (ER positivos y HER2 negativos: menos agresivos); 2) luminal B (ER positivos: más agresivos); 3) HER2 positivo; 4) basal o triple negativo (ER negativos y HER2 negativos: los de peor pronóstico). Asimismo, se han visto grandes avances en el manejo y en el tratamiento de este tumor mediante el uso de nuevas técnicas de laboratorio en la práctica clínica diaria⁴.

En 2020, se ha convertido en el cáncer más diagnosticado, seguido por el cáncer de pulmón, y la quinta causa de muerte por cáncer a escala mundial. También constituye el tumor maligno más incidente y prevalente en la mujer en la mayoría de los países⁵. La enfermedad tiene una incidencia de 128,3 casos por cada 100.000 habitantes por año. Ocurre normalmente en mujeres mayores de 40 años, siendo la edad media de presentación los 62 años⁶. Entre sus factores de riesgo, cabe destacar: la radiación ionizante (rayos X y gamma), ser de raza caucásica, la edad avanzada (a partir de los 50 años), los antecedentes familiares (entre el 5-10%

de los casos, las mutaciones heredables se localizan en los genes *BRCA1* y *BRCA2*), y situaciones que favorecen una mayor exposición a los estrógenos (nuliparidad, embarazo tardío, anticonceptivos orales, etc.). Excepcionalmente, existen casos muy raros de cánceres de mama en varones.

Se sabe que la mayor parte de estos cánceres expresan los ER (aprox. 70%), un 20% están mutados para HER2, mientras que los de tipo basal o triple negativos son menos frecuentes (aprox. 10%)⁷. Las mutaciones que son con más frecuencia responsables de los cánceres de mama hormonodependientes se encuentran en el gen que centra este proyecto: *PI3K*. Y en cuanto al cáncer de mama familiar, que se transmite a la descendencia, se han detectado alteraciones en los genes supresores tumorales como *p53* y *PTEN*, y en otros genes como *BRCA1/2*.

Las mutaciones somáticas más frecuentes en *PI3K* en el cáncer de mama aparecen en los exones 9 y 20: en el 55% de los casos consisten en una mutación missense de adenina (A) por una guanina (G) en la posición 3.140 de la secuencia codificante en el exón 20, mientras que las otras 2 más comunes son cambios de una G por una A en las posiciones 1.633 (20%) y 1.624 (11%) del exón 9. La primera mutación produce un cambio localizado en el dominio quinasa de PI3K: H1047R, mientras que las otras 2 afectan el dominio helical: E542K y E545K (**figura 1**). La mutación más importante, H1047R, aparece en otra serie de enfermedades aparte del cáncer de mama, como por ejemplo en el cáncer de ovario, de colon, de estómago, queratosis seborreica, y en algunas malformaciones vasculares.

En un estudio realizado por Mirzaa et al., se observó que las mutaciones de *PI3KCA* que estaban más implicadas en el crecimiento celular descontrolado podían provocar otras enfermedades en el cerebro (megalocefalia-capilar MCAP) y en otras partes del cuerpo que tengan mucho tejido conectivo (hiperplasia fibroadiposa FH, crecimiento asimétrico del tronco, algunas anomalías esqueléticas, etc.). Ellos también han podido detectar otras variantes patogénicas somáticas, no heredadas, de *PI3K* en el cáncer de mama⁸.

En los estadios iniciales, el cáncer de mama cursa de manera asintomática. Es por este motivo que la detección precoz de esta patología es fundamental para reducir la morbimortalidad. Con este fin, se han implantado programas de cribado con mamografía a mujeres entre 50-69 años con cánceres de mama que no sean palpables o que no tengan otros signos sospechosos de malignidad. En la mamografía se pueden definir varias características radiológicas sospechosas: 1) lesiones espiculadas, 2) lesiones con aumento de densidad, 3) lesiones de bordes mal definidos, 4) microcalcificaciones agrupadas, finas e irregulares, 5) distorsiones del tejido fibroglandular, 6) retracción y/o engrosamiento cutáneo, 7) ectasia ductal asimétrica. Aunque la mamografía es considerada el *gold standard* para el diagnóstico precoz de la enfermedad, su eficacia se ve limitada principalmente en aquellas mujeres con aumento de la densidad mamaria. En estos casos, la ecografía debe considerarse el método de screening mamario de elección, en la que se deben buscar nódulos sólidos con ecos internos, sugiriendo malignidad⁹. Cuando la clínica se hace notable, el primer signo de alarma suele ser un “bulto” detectado por la paciente, que, al tacto, se nota diferente del tejido mamario que lo rodea. De hecho, el estudio *Experiencias de las mujeres con cáncer de mama en España*, realizado por el Observatorio del Cáncer AECC, revela que el 65% de las mujeres encuestadas fueron diagnosticadas de cáncer de mama tras notarse un bulto en el pecho y acudir a sus médicos de cabecera o ginecólogos. Estas mujeres tenían un rango de edad tanto dentro (51-70 años) como fuera de los programas de cribado¹⁰. Este hecho pone de manifiesto la importancia de la autopalpación regular de las mamas para el diagnóstico de la enfermedad. También puede presentarse como una secreción sanguinolenta por el pezón o como una retracción/ulceración de piel o pezón, incluso sin ninguna masa palpable asociada en la mama. Existen algunas formas especiales de cáncer de mama que presentan algunos signos clínicos característicos, como el carcinoma inflamatorio (edema y eritema cutáneos) o la enfermedad de Paget (lesión eczematosa del pezón y areola). La presencia de adenopatías axilares normalmente indica metástasis, el tumor metastatiza inicialmente en los ganglios linfáticos, pero puede diseminarse a cualquier órgano, con especial preferencia por el hueso, hígado, pulmón y el sistema nervioso central (SNC).

En cualquier caso, hace falta un diagnóstico mediante la anatomía patológica. Si la mujer padece de una lesión palpable o identificada por la imagen, se toma la biopsia de la lesión a

través de una aguja gruesa (BAG o trucut) guiada por ecografía o asistida por vacío (BAV). En cambio, si lo que presenta es una adenopatía sospechosa, se debe realizar una punción-aspiración con aguja fina (PAAF) del ganglio afecto para estudiar la citología. En ocasiones, cuando la lesión sólo se ve en pruebas de imagen (y no es palpable), se puede marcar con un marcador metálico mediante ecografía para que sea más fácil localizarla durante una intervención quirúrgica. Es también digno de reseña el uso del marcaje radioisotópico denominada ROLL (del acrónimo en inglés: *Radioguided Occult Lesion Localization*), una técnica que le permite al cirujano localizar el tumor oculto usando una sonda gammadetectora, aunque no se han visto diferencias significativas respecto a la eficacia entre este y el marcador metálico¹¹. Además, se puede aprovechar la misma inyección del radiotrazador para identificar el ganglio centinela para saber si el cáncer se ha extendido, esta combinación se denomina SNOLL (del acrónimo en inglés: *Sentinel Node and Occult Lesion Localization*). Una vez que el patólogo observa el resultado de la biopsia bajo el microscopio, realiza un informe en el que se pone claramente el diagnóstico establecido. Después, como es el caso en el Hospital Universitario de Canarias, el ginecólogo le transmite a la paciente el diagnóstico de su lesión y se le plantean las diferentes opciones terapéuticas.

Cada caso debe ser discutido en un comité de tumores de mama en el que deben participar varios especialistas (cirugía, radioncología, oncología médica, radiodiagnóstico y anatomía patológica) para decidir el tratamiento más adecuado. Esto va a depender de varios factores: 1) Subtipo del tumor, que incluye el estado de los receptores hormonales (RE, RP), el estado del HER2 y la actividad proliferativa (Ki67); 2) estadio del tumor; 3) detección de biomarcadores moleculares a través de plataformas genéticas como Oncotype DX™ o Endopredict; 4) factores dependientes de la paciente como la edad, el estado de salud general, el estado de la menopausia y sus preferencias; 5) la presencia de mutaciones conocidas en los genes heredados del cáncer de mama, como *BRCA1* o *BRCA2*¹². Las pacientes con cánceres de mama luminal A localmente avanzados, luminal B, HER2+ o triple negativo por lo general reciben quimioterapia neoadyuvante¹³. El objetivo principal de esta es disminuir el tamaño del tumor para su posterior extirpación por parte del cirujano. Si se decide realizar una cirugía conservadora (tumorectomía), se debe añadir radioterapia adyuvante. Por otro lado, la mastectomía se ofrece en casos de tumores no infiltrantes (DCIS o LCIS del inglés: *Ductal*

Carcinoma In Situ y *Lobular Carcinoma In Situ* respectivamente) extensos, multifocales o multicéntricos o en tumores centrales en los que no se considere adecuado el tratamiento más conservador por otros motivos¹⁴. También se precisa la estadificación axilar, siendo la biopsia selectiva del ganglio centinela (BSGC) la técnica de referencia. Las únicas contraindicaciones de la BSGC son el carcinoma inflamatorio y los cánceres localmente avanzados. Para estos casos, estaría indicada la linfadenectomía¹³. Y a las pacientes con factores de mal pronóstico o a las que estaba indicada la linfadenectomía, se les da quimioterapia y hormonoterapia adyuvantes según sus estados menopáusicos.

La genética ha supuesto un gran avance para el diagnóstico y tratamiento de muchas enfermedades, incluido el cáncer de mama. Esto dio lugar a lo que se conoce hoy en día como medicina de precisión o medicina personalizada. Unas de las primeras técnicas que se desarrollaron fueron las tecnologías de secuenciación *Sanger*, que detectan cambios de un nucleótido o pequeñas inserciones o deleciones en uno o varios genes implicados en la biología molecular del cáncer. Posteriormente, se conoció la secuenciación de nueva generación o de segunda generación (*Next Generation Sequencing*, NGS), que constituye una técnica de mayor rendimiento. Una de las grandes ventajas de la NGS con respecto a la secuenciación de *Sanger* es que permite obtener muchas secuencias de manera paralela en millones de fragmentos de ADN¹⁵. Por todas estas características, ambas tecnologías han permitido identificar nuevas variaciones de secuencia específicas en tumores que son potencialmente tratables, ya que muchos de los tratamientos van dirigidos a las cascadas de transducción de señales de las células tumorales donde aparecen estas mutaciones¹⁶. Un ejemplo de este tipo de tratamiento es el que va dirigido al gen que centra este trabajo: *PI3K*, como el alpelisib.

El alpelisib es un inhibidor selectivo de la isoforma α de p110 (subunidad catalítica de PI3K) aprobado, hoy en día, en combinación con fulvestrant para el tratamiento de mujeres posmenopáusicas y hombres con cáncer de mama localmente avanzado o metastásico RH+/HER2- con mutación en *PIK3CA* tras la progresión de la enfermedad después de recibir un régimen de terapia endocrina en monoterapia.

Además, un estudio de noviembre de 2021 afirmó que podría resultar beneficioso, en cuanto a la evaluación de la respuesta o resistencia terapéutica del alpelisib, la clasificación en ensayos clínicos de los cánceres de mama según el número de copias de *PI3K* mutante (o frecuencia alélica), y no según la clasificación categórica clásica de *wild type* o *mutated*. De hecho, el pirograma que obtenemos de la pirosecuenciación aporta la frecuencia alélica, pudiendo utilizarla como variable cuantitativa con la que formular otras hipótesis (en las **figuras 5, 6 y 7** se explica su lectura y entendimiento).

Existen dos conceptos actuales que están cobrando importancia en el estudio de *PI3K*: el *stemness* (estado celular asociado a agresividad y fracaso terapéutico) y la señalización. Éstos parecen influir más sobre el pronóstico de los cánceres de mama en comparación con la expresión o no de receptores hormonales. De hecho, la expresión de receptores hormonales se asocia con un menor *stemness* y señalización, razón que explicaría por qué el alpelisib muestra limitaciones en el cáncer de mama triple negativo¹⁷. Un estudio de 2020 sugiere que terapias inhibitoras de *PI3K* serían más favorables en cánceres de mama que expresan altamente la estatmina (una importante proteína reguladora de la dinámica de los microtúbulos), con lo que la inmunohistoquímica de esta proteína podría ser un biomarcador de activación de la ruta PI3K-AKT¹⁸.

4. Hipótesis de trabajo y objetivos

Las mutaciones *missense* somáticas de *PI3K* se encuentran en aproximadamente el 15% de los cánceres humanos, incluyendo cáncer de hígado, mama, colon, ovario, estómago, cerebro, pulmón y leucemia (de mayor a menor frecuencia respectivamente). Se sabe también que las mutaciones se concentran en algunos puntos "calientes" (*hotspots*): el exón 20 (H1047R), en el dominio quinasa, y el exón 9 (E542K y E545K), en el dominio helical. Desde que existe un inhibidor selectivo de p110 y de su isoforma α (alpelisib), ha cobrado mayor interés analizar la presencia de mutaciones en *PI3K* en cánceres de mama, particularmente en los tumores de tipo luminal¹⁹. Puesto que en teoría los carcinomas con *PI3K* mutado deben tener mayor actividad proliferativa y en el cáncer de mama luminal usamos el test EndoPredict (EP en adelante) para elegir la mejor estrategia terapéutica (separando dos grupos con riesgo de recurrencia bajo o alto, y en función de esto recomendando incluir quimioterapia para los de alto riesgo), nos hemos

planteado explorar la presencia de mutaciones en PI3K en una serie retrospectiva de casos de carcinoma de mama a los que se le había hecho el test EP previamente. El objetivo es determinar si los carcinomas de mama PI3K mutados están sobrerrepresentados entre los casos de alto riesgo y explorar si es plausible que la mutación contribuya a ese mayor riesgo de recurrencia de una manera significativa. Por tanto, la hipótesis alternativa (H_1) es hallar una correlación entre que las enfermas sean de alto riesgo y estén mutadas.

4.1 Objetivo primario

Aprendizaje de las técnicas para la secuenciación de los exones 9 y 20 del gen PI3K y realización del test EP a partir de las muestras histológicas a estudio.

4.2 Objetivos secundarios

Utilización de plataformas webs (Cbioportal, ncbi, UniProt, GraphPad...) en el minado de datos, su comprensión y empleo en la formulación de hipótesis. [Anexo 1](#).

Detección de errores en retrospectiva a partir de resultados confusos o discordantes y saber, tras su adecuada comprensión, la mejor manera de prevenirlos, abordarlos y categorizarlos según el grado de importancia.

Obtener una visión global del manejo de la paciente con cáncer de mama, desde que acude a consulta por clínica, se le realizan técnicas de imagen y se le extirpa, hasta la manipulación de la muestra quirúrgica por parte del departamento de Anatomía Patológica y su análisis molecular.

5. Material y métodos

Como esquema general, presentaremos los procedimientos a realizar dentro de cada variable estudiada: en *High/Low Risk* (HR/ LR en adelante), lo subdividiremos en la fabricación del excel y la obtención de resultados del test EndoPredict (purificación de ARN → quality control → realización de la prueba); en *PI3K mutated/wild type*, comenzaremos con la purificación del ADN (bien directamente a partir de la muestra de la paciente, o bien como paso conexo de la purificación del ARN), y continuaremos con la PCR del exón 9 y 20 y la

pirosecuenciación (para estos dos últimos pasos es necesaria la fabricación de los *primers*).
Figura 2.

5.1 Endopredict *High/Low Risk*

5.1.1 Obtención de datos

5.1.1.1 Purificación ARN

La purificación del ARN es un primer paso fundamental en muchos análisis (el test EP en nuestro caso) y plantea retos particulares debido a la fácil degradabilidad del ARN²⁰. Las RNAsas suelen ser muy ricas en enlaces disulfuro, lo que las convierte en enzimas increíblemente estables. De esta manera, es fácil contaminar un laboratorio de secuenciación de ARN con RNasa, pero es notoriamente difícil asegurar su completa erradicación. Esto se debe a la presencia de RNasa ambiental en fuentes microbianas en el aire, así como en la piel, el pelo o la saliva de los seres humanos²¹. Por tanto, es inexorable la utilización de guantes durante el procedimiento.

En principio, escogemos la muestra macroscópica de tejido mamario (que contiene el tumor luminal B de cuyas células queremos extraer ARN y purificarlo) y su corte histológico correspondiente. Así, si observamos minuciosamente el corte histológico podemos distinguir una distribución heterogénea conformada por: zonas eosinófilas (moradas), con células más agrupadas y más núcleos (de elección para el raspado que se realizará posteriormente); y zonas eosinófilas (rosadas), con células más dispersas y menos núcleos.

Seguidamente, superponemos el corte histológico en la muestra macroscópica en aras de poder dilucidar qué zona es la idónea para raspar con la cuchilla (es decir, las zonas de carcinoma, eosinófilas mencionadas anteriormente). Es beneficioso mirar previamente al microscopio el corte histológico con el objetivo de evitar raspar áreas que contengan carcinoma *in situ*.

Más tarde, cogemos un tubo Eppendorf (o tubo de microcentrifugación) y lo etiquetamos con el número de control de la paciente. Añadimos 150µL de FFPE Buffer, que disolverá la parafina, y el raspado. A continuación, lo introducimos en el ThermoMixer (**figura 8**) a 80°C

durante 30 minutos. A partir de este momento, podemos estructurar la purificación del ARN en 5 pasos: lisis, unión, lavar, eluir y destrucción del ADN genómico restante.

Lisar: una vez haya concluido el tiempo, retiramos el tubo de microcentrifugación del ThermoMixer y vertemos en él 50µL de proteinasa K, que se utiliza para la destrucción de proteínas en lisados celulares (tejidos) y para la liberación de ácidos nucleicos, ya que inactiva las DNasas y las RNasas con gran eficacia²². Después, reducimos la temperatura del ThermoMixer a 65°C y lo dejamos en el interior por 30 minutos más para que actúe la proteinasa K.

Unión: cuando retiremos el tubo Eppendorf, añadimos 850µL de solución binding (cloruro de guanidinio o tiocianato de guanidinio) y, mientras lo centrifugamos a 13.000 rpm. durante 1 minuto, preparamos un tubo nuevo de microcentrifugación y lo etiquetamos con el mismo número de control de la paciente. En este último, pipeteamos 50µL de las microesferas magnéticas de afinidad por ácidos nucleicos tras utilizar, en el envase que las contenía, el agitador tipo vórtex para que se resuspendan en una solución homogénea. Una vez finalice el centrifugado, pasamos 950µL de dicha solución de lisis al tubo con las esferas magnéticas. Para finalizar esta fase, bajamos el ThermoMixer a temperatura ambiente (27°C) e introducimos el nuevo tubo Eppendorf durante 15 minutos para que las esferas capturen el ARN.

Lavar: las microesferas magnéticas con grupos funcionales (oligo dT, y N6, entre otras... secreto comercial, kit FFPET RNA de Siemens) que se unen al ARN (secuencias que se unirán selectivamente a las moléculas de ARNm, ya que los ARNm suelen incluir una cola de poli-A²³). Por tanto, en el momento actual tendremos el ARNm que queremos purificar enlazado con las microesferas, pudiendo retirar el sobrenadante en un tubo nuevo de microcentrifugación (lo utilizaremos luego para la purificación del ADN).

En este último paso, y en los tres lavados siguientes, nos ayudaremos de un soporte magnético en el que sostendremos el tubo de microcentrifugación, para que las microesferas queden sujetas a su pared y poder, así, extraer el sobrenadante y los lavados sin dificultad respectivamente.

Utilizaremos tres soluciones, cada cual de mayor a menor concentración de sales. En ese mismo orden, realizamos tres lavados de 850 μ L en el tubo Eppendorf que contiene el ARNm unido a las microesferas. Cada una de las veces, entre cada lavado, y antes de retirar la solución, es necesario introducir el tubo de microcentrifugación en el agitador tipo vórtex y en la centrifugadora. Finalmente, centrifugamos las microesferas enlazadas con el ARNm a 13.000 rpm. durante 1 minuto, y con una pipeta desechamos el líquido que surja en el fondo del tubo Eppendorf.

Eluir: añadimos 100 μ L de elución (con el objetivo de desenlazar el ARNm de las microesferas magnéticas) en el tubo de microcentrifugación. Ajustamos el ThermoMixer a 70°C, lo introducimos, y lo dejamos por el tiempo de 10 minutos. En tanto que finalicen los 10 minutos, pipeteamos 90 μ L en un tubo Eppendorf nuevo (etiquetándolo con el mismo número control de la paciente). Ya podemos desechar el tubo de microcentrifugación que contiene las microesferas.

Destrucción del ADN genómico restante: cogemos DNasa y Buffer, añadiendo 2,5 μ L y 10 μ L al tubo Eppendorf respectivamente. Seguidamente, lo insertamos en el bloque Thermo a 37°C durante 30 minutos; y después, ya para concluir la purificación del ARN, lo introducimos en el ThermoMixer a 80°C durante 10 minutos, en aras de desnaturalizar la DNasa restante.

5.1.1.2 Quality Control

Una vez purificado el ARN, y previo a la realización del test EP, se procede a hacer el *quality control* (qc). Este consiste en un control positivo que ofrece la empresa, el cual cumple con dos funciones: corroborar la calidad de nuestra muestra y conocer el *Cycle of quantification* (Ct).

Para realizar el *quality control* se requiere de la RT-qPCR (retrotranscripción del ARN purificado a ADN complementario [ADNc] seguida por su análisis con PCR cuantitativa) por lo que, mientras se va preparando, aislamos 4,5 μ L de RNA (del tubo Eppendorf que estaba en el ThermoMixer en el paso de “destrucción del DNA genómico restante”) y 95 μ L de MasterMix (620 μ L de agua, 720 μ L de buffer y 29 μ L de retrotranscriptasa o *Reverse Transcriptase* [RT en

adelante] y *Taq DNA polymerase* [Taq en adelante], que es un tipo de ADN polimerasa termoestable) en otro tubo de microcentrifugación. Después, pipeteamos 20 µL en cada pocillo del qc (en total 4).

Una vez que finalice la RT-qPCR se obtiene el Ct, es decir, se cuantifica el número de ciclos que han sido necesarios para alcanzar una cantidad prefijada de ADN producto. Cuanto mayor sea la abundancia de ARN en la muestra, antes se alcanzará la cantidad prefijada y, por tanto, menor cantidad habrá que poner de ARN para hacer el test EP.

5.1.1.3 Test EndoPredict (EP)

El test EP es una prueba diseñada con el objeto de predecir el riesgo de recurrencia metastásica en aquellos cánceres de mama ER+ y HER2-. Esto se hace mediante el cálculo de un Score que se obtiene a partir de la cuantificación de la expresión de 11 genes (3 de referencia y 8 genes de interés). Según los resultados del EP, se puede clasificar a las pacientes en alto (predominio de proliferación) y bajo riesgo (predominio de diferenciación), lo cual determinará si se ven más beneficiadas de hormonoterapia y quimioterapia, o solo de hormonoterapia, respectivamente.

Los genes que tiene en cuenta son: 1) tres genes de referencia (*CALM2*, *OAZ1* y *RPL37A*) y 2) ocho genes de interés (*BIRC5*, *RBBP8*, *UBE2CA*, *IL6ST*, *AZGP1*, *DHCR7*, *MGP* y *STC2*). Dentro de estos últimos, los subdividimos en cinco de diferenciación (*AZGP1*, *IL6ST*, *MGP*, *RBBP8* y *STC2*) y tres de proliferación (*BIRC5*, *DHCR7* y *UBE2CA*).

Además se añade, para cada gen, un control positivo, o qREF, y un control negativo, o NTC. El qREF trata de comparar la expresión molecular de cada uno de estos genes alterados en la paciente con la expresión normal de los mismos. Un estudio de 2018 evaluó la calidad del test EP utilizando el seguimiento de los valores de los controles positivos (qREF) para 151 muestras independientes. Todas las mediciones resultaron ser altamente reproducibles²⁴.

En resumen, utilizamos una placa de EP con 12 columnas (correspondientes a los 12 genes) y 5 filas (las 3 primeras se corresponden con la muestra, la cuarta con el qREF, o control positivo, y la quinta con NTC, o control negativo).

Para comenzar, hemos de crear el *Mastermix* pipeteando en un tubo: 1) 620 μL de agua, 2) 720 μL de buffer (Mg, dNTP) y 3) 29 μL de un mix de enzimas (RT y Taq).

Luego, utilizamos tres tubos y vertemos en cada uno lo siguiente: 1) tubo para **NTC**, añadiéndose 124 μL de agua + 285 μL de *Mastermix*, 2) tubo para **qREF**, vertiéndose 285 μL de *Mastermix* y 3) tubo para la **muestra de ARN** pipeteando 742 μL de *Mastermix* + ARN muestra (25-50 μL , en función de los resultados del qc se añaden más o menos μL).

En la placa se ponen por cada pocillo 20 μL y se centrifuga antes de realizar la RT-qPCR:

- **RT:** se pone durante 30 minutos a 50°C para que se active la RT, que convierte el ARN en ADNc.
- **qPCR:** debido a que tenemos anticuerpos anti-Taq en la mezcla, se debe mantener durante 2 minutos a 95°C para destruir los anticuerpos y para que se libere la Taq. Pasados estos dos minutos, pasamos a la fase de desnaturalización (se abren las 2 hebras) que dura 15 segundos a 95°C. A continuación, para que se lleve a cabo la fase de hibridación y extensión, se requiere un tiempo de 30 segundos a 60°C. Finalmente, se deja 30 segundos a 40 °C.

Una vez finalice la RT-qPCR, obtenemos el EPclin oficial de la plataforma *web* de la empresa y cogemos los Ct de cada gen con el objetivo de introducirlos en nuestro Excel y sacar nuestro propio EPclin (más información en el apartado “5.1.1 Excel”) (**figura 9**).

5.1.2 Hoja de cálculo: Excel

Como base de datos para nuestro estudio, se seleccionaron pacientes diagnosticadas de cáncer de mama de tipo luminal B que se sometieron a un test EP, en total 100, registradas en el sistema informático del servicio de Anatomía Patológica del Hospital Universitario de Canarias.

Con el fin de contrastar los scores obtenidos en los informes de los test, proporcionados por la plataforma web de la empresa, así como entender el funcionamiento del test EP, tomamos la decisión de recalcular el score de las muestras nosotros mismos, empleando las fórmulas publicadas en 2011^{25,26,27} y datos clínicos de la paciente como la T y la N del tumor.

Para ello utilizamos el programa Excel, en el que preparamos en varias hojas los algoritmos para obtener tanto el *score* EP como el *score* EPclin a partir de los datos introducidos. Dichas ecuaciones fueron obtenidas a partir del artículo, ya citado, que detalla el funcionamiento del test EP. Para obtener los *scores*, partimos de las concentraciones obtenidas en una qPCR de 8 genes de interés (GOI; *gene of interest*) relacionados con el cáncer de mama (*BIRC5*, *UBE2CA*, *DHCR7*, *RBBP8*, *IL6ST*, *AZGP1*, *MGP*, y *STC2*), y de otros 3 genes denominados “de normalización” (*CALM2*, *OAZ1*, y *RPL37A*) que se emplearán como genes de referencia. La expresión relativa de cada GOI se denominó *Delta Cycle threshold* (DCt) y se obtienen empleando la siguiente fórmula que normaliza los valores en base a la expresión de los genes de referencia (**figura 3.1**).

Los DCt de cada uno de los genes se combinaron de la siguiente manera para obtener el *predictive unscaled risk score* (Su) (**figura 3.2**).

El EP *risk score* final oscila entre 0 y 15 por lo que se emplearon los siguientes criterios para obtener el EP *risk score* (S) a partir del Su (**figura 3.3**).

El EPclin (Sclin), por otra parte, es un *score* que combina el EP *risk score* y otros parámetros clínicos como el tamaño del tumor (T) y la afectación ganglionar (N). Se obtiene a partir de esta fórmula (**figura 3.4**). En ella, *t* hace referencia al tamaño del tumor (1: 1cm, 2: >1 a 2cm, 3: >2 a 5cm y 4: >5cm) y *n* al número de ganglios afectados (1: negativo, 2: 1–3 ganglios positivos, 3: 4–10 ganglios positivos, y 4: >10 ganglios positivos).

Finalmente el excel constó de cuatro hojas. En la primera, se introdujeron todos los parámetros de los 11 genes, que agrupamos en 3 de referencia (de color verde), y los 8 GOI los diferenciamos en 5 de predominio diferenciador (color rosa) y 3 de predominio proliferativo (color azul). En esta hoja se introdujeron los tres valores tomados de Ct para cada uno de los genes y se configuró para que el programa obtenga automáticamente la media aritmética. En la segunda y tercera hoja, el programa calcula los valores de DCt para los GOI de predominio diferenciador y proliferativo, respectivamente, empleando los valores medios de Ct para cada gen obtenidos en la primera hoja. En la última hoja, el programa obtiene, con los datos de las

anteriores páginas, los valores de EP *risk score* y, tras introducir los valores necesarios para la *T* (tamaño tumoral) y para la *N* (afectación ganglionar), el EPclin.

En un primer momento, tras introducir los valores de varios de nuestros casos en el excel comprobamos que el valor de EPclin obtenido no se asemejaba al indicado en el informe. Se planteó la posibilidad de que exista un error en la fórmula introducida en nuestro excel, y tras repasar uno a uno los pasos de cálculo observamos que el error se encontraba en que el programa no estaba aplicando la puntuación de los valores *T* y *N* a la fórmula. Tras solucionar este problema, vimos que la diferencia entre los valores de EPclin del informe y los obtenidos se reducía, pero no a cero. Seguía existiendo una diferencia en el cálculo de estos parámetros, sin embargo, en este caso la diferencia era siempre negativa, nuestro resultado era siempre más bajo tanto en el valor de EP como de EPclin, por lo que se revisó nuevamente el proceso de cálculo y, después de no encontrar fallos, decidimos calcular el error de las muestras. Nuestro valor de EP tenía un error medio de -0.62 con una desviación estándar de 0.084, lo que significa que en todas las muestras existe un error de cálculo negativo muy aproximado a la media. Lo mismo ocurrió con el EPclin, que presentaba un error medio de -0.19 con una desviación estándar de 0.071. Se decidió revisar el algoritmo una tercera vez y, sin volver a encontrar fallos, y dada la poca variabilidad y desviación estándar del error, se decidió asumir como error sistemático. La causa de este error debió ser o bien el empleo de una fórmula de cálculo desactualizada (dado que el único artículo donde aparece pública esta fórmula es del año 2011) o bien un error decimal debido a que sólo introducimos valores del orden de dos decimales en nuestra ecuación.

5.2 PI3K *mutated/wild type*

5.2.1 Purificación ADN

5.2.1.1 Cromatografía por afinidad

La purificación de ácidos nucleicos mediante columnas de centrifugación o purificación con minicolumnas es un método de extracción en fase sólida para purificar rápidamente los ácidos nucleicos. Este método se basa en el hecho de que los ácidos nucleicos se unirán a la fase sólida de sílice en determinadas condiciones. Recordemos que este ADN que queremos purificar

lo hemos obtenido del sobrenadante de la purificación del ARN (**figura 4.A**). Las etapas del método son lisar (se omite esta fase, pues ya se realizó en la purificación del ARN), unión, lavar y eluir²⁸.

Unión: preparamos en un portatubos tres hileras (**figura 4.B**): en la primera, colocamos el tubo que contiene el sobrenadante obtenido en la purificación del ARN; en la segunda, ponemos un recipiente con la fase sólida de sílice (o filtro de afinidad por el ADN); y en el tercero, un tubo Eppendorf en el que se depositará, más adelante, en la fase de eluir, el ADN purificado (por lo que este tubo no lo tocamos hasta el final). Etiquetamos todo con el número de control de la paciente.

Para finalizar esta fase, con una pipeta desechable trasparamos el sobrenadante (del tubo de la primera hilera) al recipiente con la matriz de afinidad por el ADN, e introducimos dicho tubo en la centrifugadora a 13.000 rpm. durante 1 minuto. Mientras transcurre el tiempo colocamos en la segunda hilera otro recipiente nuevo y, cuando este finalice, pasamos el filtro al mismo, que ya contiene el ADN a purificar (la condición indispensable para que se lleve a cabo el enlace entre la matriz y el ADN es la alta salinidad del sobrenadante, que traía ya de la purificación del ARN). El recipiente salido de la centrifugadora, que contiene líquido, se puede desechar.

Lavar: en esta fase realizaremos dos lavados: el primero, contiene 50mM de NaCl; y el segundo, 5mM de NaCl. En cada lavado añadiremos 0,5ml de cada solución, centrifugaremos a 13.000 rpm durante 1 minuto, y desecharemos el recipiente en cuestión, colocando el filtro en uno nuevo. Por último, hacemos una centrifugación en seco (sin soluciones), en aras de precipitar el líquido que pueda haber quedado durante los lavados.

Eluir: tras la centrifugación en seco, desechamos el recipiente y colocamos el filtro en el tubo Eppendorf de la tercera hilera. Añadimos 60µL de agua destilada (salinidad nula, con el objetivo de desenlazar el ADN de la matriz y que eluya sobre el tubo de microcentrifugación). Antes de centrifugar a 13.000 rpm. durante 1 minuto, esperamos 1 minuto a que la matriz absorba adecuadamente el agua destilada. Al acabar, ya tendremos en el tubo Eppendorf el ADN purificado.

5.2.1.2 Con fenol-cloroformo (PCI)

Otra técnica de purificación de ADN es la que está basada en fenol-cloroformo (PCI en adelante), en la que se extrae ADN a partir de muestras de tejido parafinado haciendo uso de técnicas físico-químicas. Este proceso puede dividirse en dos fases: 1) Lisis celular, 2) extracción y purificación de ADN.

Lisis celular. A partir de los portaobjetos con las muestras embebidas en parafina, se debe realizar un desparafinado sobre la zona con la mayor cantidad de células tumorales. Por ello, primero se delimita la zona haciendo uso del microscopio óptico, y luego se marca con un rotulador permanente. Aquí siempre se debe evitar marcar la zona no tumoral (tejido normal, necrótico o adiposo). Posteriormente, se hacen dos lavados de 5 minutos en xilol para desparafinar. Luego, para rehidratar, se hacen lavados de 5 segundos en concentraciones decrecientes de etanol (100° -96° - 90°), es decir, 5 segundos en 100°, 5 segundos en 96°, y otros 5 segundos en 90°. Finalmente, se lava con agua destilada.

Se raspa con una punta desechable solo el material tumoral de cada muestra que luego se añadirá en su tubo correspondiente que contiene 0,5 mL de TENS y 50µl de proteinasa K. A continuación, pasa a describirse el fundamento por el cual se digieren las proteínas y las membranas celulares para liberar el ADN de las muestras, proceso denominado lisis celular.

En esta fase se hace uso de dos soluciones: el TENS y la proteinasa K. El primero de ellos está compuesto por cuatro componentes: 1) Tris HCl: es un tamponador que mantiene el pH de la solución estable, a pH=8; 2) EDTA: es un agente quelante que inhibe a las nucleasas de DNA uniéndose a iones metálicos (Ca²⁺ y Mg²⁺); 3) NaCl, proporciona la salinidad necesaria para la actividad enzimática de la proteinasa K; 4) Sodio dodecilo sulfato (SDS): actúa como detergente, rompiendo la membrana lipídica. En segundo lugar se añade la proteinasa K, una exo y endoproteasa que se encarga de degradar proteínas. Luego se agitan los tubos con el vórtex y se incuban a 56°C durante toda la noche. Al día siguiente se incuban los tubos a 90°C durante 1 hora para romper los puentes (crosslinking) formados entre los grupos amino y el ADN por el

formaldehído. Una vez que se haya liberado el ADN de todos los elementos celulares, empieza la segunda fase, la denominada extracción y purificación de ADN con PCI).

Extracción y purificación de ADN. Primero se centrifugan los tubos a máxima velocidad (1300 r.p.m.) durante 1 minuto y luego se transfieren los sobrenadantes (lisados de la proteinasa K) a unos tubos nuevos con 0,5 mL de PCI. Posteriormente, se agitan bien hasta que ambas fases sean homogéneas y se centrifugan durante 5 minutos a máxima velocidad para separar la fase acuosa (que contiene los ácidos nucleicos) de la fase orgánica (que contiene el PCI y los restos celulares). Entre ambas se forma una interfase en la que quedan atrapadas proteínas de membrana y las nucleasas liberadas durante el proceso de lisis celular. Luego se pasa la fase acuosa con el ADN disuelto a un tubo nuevo de 2ml con 40 µl de acetato sódico 3M. Aquí es importante no aspirar la interfase ni la fase orgánica y usar pipetas de 1ml limpias para cada tubo. Luego se añaden 1500 µl de EtOH absoluto usando una pipeta de 3ml y se agitan antes de dejarlos 1 hora a -20°C para facilitar la precipitación del ADN.

Posteriormente, se centrifuga a máximas r.p.m. 15 minutos hasta que se forma un “pellet” y se tira el sobrenadante. Acto seguido se añaden 400µl de EtOH 70°C para lavar el “pellet” de ADN y se centrifuga durante 5 minutos. Se descarta el sobrenadante, se drenan los tubos en papel, y se dejan secar 5 min. Tras ello, se colocan los tubos con la tapa abierta en el termobloque a 56°C durante 5 min. para acelerar el proceso de evaporación del EtOH. Luego, se resuspende el “pellet” con 50 µL del tampón TE (10 mM de tris-clorhídrico y 1 mM de EDTA). Por último, se agita el tubo en el vórtex para que el pellet se disuelva en el tampón mencionado.

5.2.2 Cuantificación del ADN mediante el *NanoDrop*

Una vez purificado el ADN, es necesario cuantificar su concentración final resultante. Para ello, emplearemos el nanoespectrofotómetro, un dispositivo que emplea la espectrofotometría para calcular la concentración de ADN en cada muestra. Este dispositivo emite un haz de luz que atraviesa la muestra y alcanza un receptor. La muestra habrá absorbido parte de la luz en este proceso por lo que la cantidad de luz dispensada por el emisor y la

cantidad que detecta el receptor mostrarán disparidad, en base a la cual la máquina calculará la concentración de sustrato que hay en el medio (en nuestro caso, ADN disuelto en H₂O).

Para comenzar a cuantificar primero debemos calibrar la máquina asignándole un *blank* con la pipeta. Colocamos un microlitro en el medio donde se encuentra el sustrato que queremos cuantificar (H₂O) y le indicamos a la máquina que realice el blanqueo (esa será su línea base). Posteriormente, limpiamos el punto sensor y depositamos un microlitro de muestra de nuestro ADN purificado en él. Luego, le indicamos a la máquina que comience la medición, acto seguido obtendremos el resultado. Procedemos a realizar lo mismo con todas las muestras, limpiando bien el medidor entre cada una de ellas y anotando la concentración en el lateral del tubo de cada paciente (esto servirá más adelante para estimar cuánto ADN hemos de pipetear en los pocillos para la PCR).

5.2.3 Diseño de primers

La Reacción en Cadena de la Polimerasa (PCR) permite obtener de un fragmento de ADN particular un gran número de copias, denominadas amplicones, para su posterior análisis molecular. Para llevar a cabo la misma, es necesario que fabriquemos los *primers* o cebadores (tanto el *forward* como el *reverse*). Se tratan de oligonucleótidos que flanquean los extremos del fragmento de genoma que queremos amplificar y que sirven como punto de partida para la replicación. Sin lugar a dudas, son los *primers*, más que otra cosa, los que determinan el éxito o el fracaso de la reacción de amplificación. Es, por tanto, que debemos considerar una serie de puntos en su diseño: seleccionar cebadores con una distribución de base aleatoria, desechar secuencias con estructuras secundarias significativas, y enfrentarlos el uno con el otro para comprobar la complementariedad y evitar el fenómeno *primer dimer*. Otro aspecto importante es la Temperatura de *melting* (T_m) que se calcula de la siguiente manera:

$$T_m = 2 \times (A + T) + 4 \times (G + C)$$



Temperatura a la cual la mitad de las moléculas del primer están apareadas con el ADN molde.

Nos vamos a centrar en amplificar los exones en donde se localizan las mutaciones más frecuentes del gen *PIK3CA*, concretamente el exón 9 y el exón 20 para, más adelante, pirosecuenciarlos. En el diseño de los *primers* hemos utilizado las siguientes plataformas digitales: *Ensembl*, la base de datos del genoma, para obtener la secuencia normal de nucleótidos de dichos exones; *Primer3Plus*, el programa *open-source*, que fabrica los cebadores más adecuados; *oligoevaluator*, proporciona el número de cada nucleótido que conforma el cebador, la T_m estimada, y el acceso directo al *Basic Local Alignment Search Tool* (BLAST) que comprueba que sean específicos a nuestra región de interés y no se unan inespecíficamente a otras partes del genoma; y *reverse complementary*, que ayuda a localizar el *primer reverse* en nuestra secuencia normal de nucleótidos para los exones de interés. El diseño de los *primers* se adjunta en el [anexo 2](#).

5.2.4 PCR e9, e20

Como ya se ha mencionado anteriormente, la PCR trata de obtener múltiples copias de un fragmento de ADN para su posterior análisis molecular. Para ello, nos valdremos de los cebadores ya sintetizados, que flanquean la región diana que se va a secuenciar mediante la unión a las cadenas complementarias del ADN. La amplificación se produce al calentar el ADN para separar las dobles cadenas, tras lo que se enfría la reacción para permitir que los cebadores se unan a las dos cadenas de ADN y después se elongan las secuencias a partir de los cebadores con la ADN polimerasa. Este ciclo de calentamiento, enfriamiento y polimerización se desarrolla a través de varios ciclos, en cada uno de los cuales aumenta exponencialmente el número de copias del ADN diana²⁹.

A continuación, preparamos la placa de trabajo con el número de pocillos necesarios para el número de muestras que queramos amplificar (en nuestro caso 15, que repartiremos en 3 filas con 5 pozales en cada una de ellas). Paralelamente, trazamos una tabla, en un folio que tengamos a mano, que contenga el mismo número de cuadrículas y en disposición idéntica a nuestra placa de trabajo. Dentro de cada cuadrícula anotaremos el número de control de una paciente, cerciorándonos, así, de cuál ADN estamos trabajando en cada pocillo. En una gradilla, colocaremos tres tubos:

- Un tubo Eppendorf nuevo y rotulado como **MasterMix** (MM), donde pipeteamos DNA polimerasa, dNTPs, MgCl₂ (generalmente 1,5mM, pero la concentración varía según los *primers* a usar y la región a amplificar), Buffer de Reacción y estabilizantes.
- El tubo que contiene los *primers* que hemos diseñado para amplificar la región de *PI3K*.
- Los tubos de microcentrifugación que albergan el ADN de nuestras pacientes, siguiendo el orden de nuestra tabla.

Seguidamente, hemos de coger un nuevo tubo Eppendorf donde vertemos 240μL de MM (16μL por muestra) y 7,5μL de *primers* (0,5μL por muestra), lo agitamos, y traspasamos 15μL a cada pocillo de nuestra placa de trabajo. Una vez los rellenemos todos, es hora de introducir el ADN de cada paciente en su respectivo pocillo, y lo hacemos en la cantidad necesaria, basándonos en la concentración de material preexistente medida por el NanoDrop, y con el siguiente algoritmo:

- **<20 ng/μL:** añadiremos 2μL.
- **>20 y <200 ng/μL:** añadiremos 1μL.
- **>200 ng/μL:** abrimos el tubo, depositamos en la tapa 3μL de agua y disolvemos en él 1μL de ADN. Del resultado, vertemos 1μL en el pocillo correspondiente.

Debemos tener en cuenta que, cada vez que pipeteemos ADN en los pocillos, es beneficioso llegar hasta el fondo (es decir, el MM con los *primers*) y mezclarlo bien. Finalmente, colocamos la placa de trabajo en la PCR, la sellamos bien con cinta adhesiva con el objetivo de evitar que se evapore el contenido, y dejamos que el termociclador haga su trabajo. Acabada esta fase, los amplicones resultantes estarán listos para pirosecuenciarlos y observar cuáles de nuestras pacientes son *wild type* o *mutated*.

5.2.5 Pirosecuenciación

En el proceso de secuenciación se utiliza una cadena molde de ADN, donde se va a hibridar un cebador de oligonucleótidos y donde el ADN polimerasa extiende la cadena complementaria. En la secuenciación clásica (Sanger), el paso de la polimerización incorpora nucleótidos marcados con un fluorocromo (de hasta 4 colores, uno para cada nucleótido) que son

detectados por sistemas láser. Por otro lado, la pirosecuenciación es otra técnica en la que se detectan las moléculas de pirofosfato generadas con la incorporación de cada nucleótido y en la que se secuencian fragmentos más pequeños de ADN.

La pirosecuenciación utiliza la luminiscencia para detectar la incorporación de nucleótidos complementarios a los de la cadena molde de ADN (por ejemplo, productos de PCR desnaturalizados) a medida que se va sintetizando una nueva hebra. El objetivo final de este procedimiento es secuenciar fragmentos de ADN que contengan las regiones a analizar para luego cuantificar la frecuencia alélica de las diferentes mutaciones. En nuestro proyecto secuenciamos los productos de la PCR de las muestras de las 100 pacientes diagnosticadas de cáncer de mama. Y como ya se dijo previamente, dicho ADN se amplificó gracias al uso de cebadores específicamente diseñados para las regiones diana, es decir, aquellas donde se buscará cuantificar las mutaciones. Uno de estos cebadores viene con una molécula de biotina ya preincorporada de fábrica para la PCR, que servirá para poder capturar una sola hebra de ADN desnaturalizado y secuenciarla mediante esta técnica.

Antes de empezar la reacción de pirosecuenciación, se prepara una solución que está compuesta por tres componentes: 1) 300 μ l de agua; 2) 380 μ l de tampón de unión (facilita la interacción entre la biotina del cebador y la estreptavidina); 3) 10 μ l de estreptavidina (captura una hebra de ADN, la que lleva el primer biotinilado). Luego se debe echar 70 μ l de esta solución a cada pocillo de la placa de PCR y agitarla usando un Eppendorf Thermomixer durante, como mínimo, 5 minutos con el objetivo de capturar los productos de la PCR. Tras la agitación, estos se deben llevar a 3 recipientes ya previamente preparados en la mesa de trabajo de la estación de vacío PyroMark Q24 MDx. Para ello, se deben transferir las muestras mediante una herramienta de vacío (para eliminar los desechos y quedarnos solo con las cadenas de ADN desnaturalizado) a los diferentes recipientes en el siguiente orden: 1) etanol al 70 % (EtOH); 2) solución desnaturalizante compuesta por NaOH (al desnaturalizar el ADN, facilitará la eliminación de la hebra complementaria); 3) tampón de lavado para retornar el pH óptimo (**figura 10**).

Para la preparación de la placa de pirosecuenciación, se añaden 25 μ l de *Annealing Buffer* (facilita la unión de los cebadores de secuenciación a la hebra de ADN) y 0,7 μ l de cebadores

reversos por pocillo. Con el ManiFold, se añaden los productos de la PCR en cada uno de los pocillos. Luego, para desnaturalizar el ADN, se coloca la placa encima del *ThermoBlock* durante 2 minutos a 80°C. A continuación, se deja la placa fuera del mismo durante 8-10 minutos para dejar que el primer de secuenciación se hibride a la hebra a secuenciar. Mientras tanto se prepara un cartucho en el que se añaden 100 µl de Enzima en el pocillo superior izquierdo, 100µl de Sustrato en el superior derecho, y 60µl de cada nucleótido en orden alfabético (A-C-G-T) y en el mismo sentido de las agujas del reloj como se muestra en **figura 10**. Finalmente, se introducen tanto la placa como el cartucho en el equipo donde se llevará a cabo la reacción.

El fundamento por el cual se obtiene la secuenciación de las regiones diana de los pequeños fragmentos de ADN es gracias a una polimerasa que cataliza la incorporación de desoxinucleótidos trifosfato (dNTP) complementarios al ADN que se quiere analizar a partir de los cebadores que flanquean dichas regiones. Cuando esto ocurre, se libera una molécula de pirofosfato (PP) como producto de desecho. El PP actúa como sustrato por la ATP sulfurilasa, que cataliza la conversión de 5'-fosfosulfato a ATP. Este último es empleado por la luciferasa para reaccionar con la luciferina para generar, finalmente, luz. El pirosecuenciador, que funciona como una cámara oscura, detecta esta señal lumínica y la registra como picos de tamaño proporcional al número de nucleótidos añadidos a la nueva cadena creciente.

6. Resultados

Nuestro TFG se centra en la hipótesis que los cánceres de mama mutados en *PI3K* tienen características biológicas de proliferación / diferenciación diferentes a los no mutados. Consecuentemente, deberían observarse diferencias en los parámetros biológicos recogidos en el test de evaluación de riesgo de recurrencia conocido como Endopredict. Nos interesa pues un test de contraste de hipótesis entre dos variables cualitativas (χ^2 , o chi-cuadrado), que son, en nuestro caso, el resultado del test EP en *High Risk* o *Low Risk*, y la clasificación *PI3K mutated* o *wild type*. Además, nos interesa comparar también si hay variables biológicas integradas en el test EP que se asocien específicamente a la mutación en *PI3K*. En paralelo, nos interesa saber si las mutaciones en *PI3K* se asocian a valores de proliferación medidos por inmunohistoquímica para Ki67.

Al finalizar la parte de laboratorio, obtuvimos los datos del test EP de un total de 100 pacientes, de los cuales excluimos 8 por imposibilidad de obtención del Ki67 y 4 de la frecuencia alélica del *PI3K*, quedándose una N=89 (ya que hubo una paciente de la que no obtuvimos ni el Ki67 ni la frecuencia alélica). De esta manera, el siguiente paso fue definir cuál de nuestras variables sería la independiente (*PI3K* y Ki67) y cuál la dependiente (test EP) para, así mismo, establecer una jerarquía que ordenase de mayor a menor importancia el orden de las preguntas que realizaríamos en la estadística (**figura 11**):

- Variable independiente: *PI3K*.
 - Cualitativa dicotómica: *mutated / wild-type*.
 - Cualitativa no dicotómica (en base a la frecuencia alélica): grupo 0 (*wild-type*), grupo 1 (<26%) y grupo 2 (>25%).
- Variable independiente: Ki67.
 - Cuantitativa.
 - Cualitativa dicotómica: grupo 0 (<13%) y grupo 1 (>12%). Punto de corte establecido por la mediana de nuestra muestra. Además, este punto de corte es muy próximo al 14% que se ha considerado durante muchos años en la práctica médica habitual para separar los cánceres de mama luminales B de los A (de menor actividad proliferativa).
 - Cualitativa no dicotómica: consideramos grupo bajo ($\leq 15\%$), grupo medio (16-30%), y grupo alto ($\geq 31\%$). Es decir, usamos como puntos de corte 15% y 30%.
- Variable dependiente: test EP.
 - Cualitativa dicotómica: *high risk / low risk*.
 - Cuantitativas: EPclin, EPbio, ScoreDif, ScoreProl, *k*AZGP1*, *k*IL6ST*, *k*MGP*, *k*RBBP8*, *k*STC2*, *k*BIRC5*, *k*DHCR7* y *k*UBE2C*.

Después, introdujimos la hoja Excel con los datos de las 89 pacientes en el SPSS y empezamos a aplicar los test estadísticos no paramétricos según las variables y la jerarquía que hemos definido en el apartado anterior. Utilizamos cuatro test estadísticos principalmente:

- Chi-cuadrado: entre variables cualitativas.

- Rho de Spearman: entre variables cuantitativas.
- U de Mann de Mann-Whitney: entre variable cuantitativa y cualitativa dicotómica.
- Kruskal-Wallis: entre variable cuantitativa y cualitativa no dicotómica.

Tanto los test aplicados según los tipos de variables a estudio como los resultados de cada uno se pueden observar en la **tabla 1**, representados en dos tablas:

- **Tabla 1:** tipos de tests estadísticos utilizados según los tipos de variables tenidos en cuenta.
- **Tabla 2:** se visualiza de color amarillo aquellos tests que dieron significancia (es decir, $p < 0,05$).

Los resultados resumidos son los siguientes:

- **Ki67:** existe una correlación débil (0,202 de rho de Spearman, $p=0,054$) entre el Ki67 y el EPclin, y una correlación moderada (0,414 de rho de Spearman, $p=0,01$) entre el Ki67 y el EPbio. Además, dichas correlaciones son directamente proporcionales al ScoreProl (0,435 de rho de Spearman, $p=0,01$) e inversamente proporcionales al ScoreDif. Esto tiene sentido, ya que el Ki67, que se tiene en cuenta para clasificar el cáncer de mama en Luminal B, es un marcador de proliferación.

Cuando tratamos Ki67 como una variable cualitativa dicotómica o no dicotómica se obtiene significancia con los genes *UBE2C*, *BIRC5*, *MGP* y *AZGP1*, con el ScoreProl y el EPbio, añadiéndose el EPclin y el *High/Low Risk* cuando tratamos la variable Ki67 como dicotómica.

- **PI3K:** cuando tratamos *PI3K* como variable dicotómica (*mutated/wild-type*) o no dicotómica sale relación con el gen *UBE2C*, con un $p=0,003$ y $p=0,006$ respectivamente, añadiéndose el gen *RBBP8* ($p=0,033$) y el *MGP* ($p=0,02$), y el ScoreProl ($p=0,021$) cuando la manejamos de manera dicotómica.

A continuación, se nombran algunas limitaciones a destacar en el estudio: 1) tenemos una N pequeña de 89 pacientes; 2) al transformar la variable cuantitativa de frecuencia alélica de *PI3K* en cualitativa no dicotómica el punto de corte para definir los grupos lo establecimos arbitrariamente porque no hay información en la literatura que nos permita correlacionar la

frecuencia alélica con variables biológicas contempladas en este trabajo; 3) no se trata de 100 casos consecutivos, que era la intención original del trabajo. Debido a que el porcentaje de casos mutados para *PI3K* es de alrededor del 30% en nuestra serie, tuvimos que incluir casos mutados no consecutivos hasta obtener un número similar de mutados y no mutados.

7. Discusión

Finalmente hay cinco puntos a tratar según los resultados expuesto en el apartado anterior:

- **Ki67 se relaciona con un mayor HR en el test EP cuando se estratifica dicho marcador celular proliferativo de manera dicotómica según la mediana (12).** Esto puede plantear utilizarlo como predictor de un mayor riesgo de obtención de un resultado de HR en el test EP y, en consecuencia, un mayor riesgo de recidiva en los próximos 10 años. Teniendo en cuenta que el Ki67 se calcula en la clasificación molecular del cáncer de mama, y como los tumores a los que le realizamos el test EP eran Luminales B, cabe esperar la obtención de dicho resultado.
 - **Hay mayor correlación con el EPbio (sin los valores clínicos del T y la N del tumor) que con el EPclin (con los valores clínicos).** Esto se traduce en que este marcador de proliferación celular parece relacionarse más con la expresión de los genes del test EP que con el fenotipo clínico que manifieste la paciente.
 - **Existe una relación directamente proporcional con los genes de proliferación e inversamente proporcional con los genes de diferenciación.** Como cabría pensar, una célula que contenga mayor expresión de genes diferenciadores debería obtener un menor Ki67 en comparación con una que tenga más activados los de proliferación.
- ***PI3K* dicotómico se asocia a un mayor ScoreProl, pero no es suficiente para que haya significancia con el High/Low Risk o el EPclin/bio.** La obtención de una mutación de *PI3K* infiere en un mayor carácter proliferativo del cáncer de mama, no obstante, esto no es bastante para afirmar que el mismo tenga un mayor riesgo de recidiva en los próximos 10 años.

- ***PI3K* se relaciona a una mayor proliferación celular debido a la gran significancia que presenta con la expresión del gen *UBE2C* que mide el test EP.** Este gen codifica un miembro de la familia E2 de enzimas conjugadoras de ubiquitina. La modificación de proteínas mediante ubiquitinación es un mecanismo importante para dirigir al proteasoma para su degradación a proteínas mal plegadas o de corta vida media. Entre estas proteínas de corta vida media reguladas por ubiquitinación se encuentran las proteínas reguladoras de los puntos de transición del ciclo celular³. La ubiquitinación de proteínas implica al menos tres clases de enzimas: 1) activadores de ubiquitina, 2) conjugadores de ubiquitina, y 3) ligasas de ubiquitina-proteína. *UBE2CA* es esencial para la destrucción de ciclinas mitóticas y la progresión del ciclo celular, por lo que parece lógica su implicación en la progresión del cáncer. En particular, la sobreexpresión de *UBE2CA* se ha visto asociada a mal pronóstico en cáncer de próstata, riñón y esófago. Respecto al cáncer de mama, se ha publicado que la sobreexpresión de *UBE2C* juega un papel importante en la progresión del tumor y este gen podría ser un biomarcador de mal pronóstico¹¹⁻¹⁵.

Análisis bioinformático de repositorios de datos de expresión génica de cáncer de mama (como GEO y TCGA, <https://www.cbioportal.org>) ha encontrado un vínculo significativo entre la expresión de *UBE2C* y la proliferación celular en cáncer de mama, mediada por activación de la ruta AKT/mTOR, en la que *PI3K* es un componente esencial.

La invasión linfovascular (LVI) es uno de los indicadores histopatológicos de mal pronóstico, y se considera un prerrequisito de la metástasis⁸⁻¹⁰ pero muchos detalles moleculares del proceso de metástasis se desconocen hasta la fecha. De hecho, la sobreexpresión de *UBE2C* parece jugar un papel importante en la progresión del cáncer de mama, influyendo sobre la invasión linfovascular¹¹⁻¹⁵ y el estado de metástasis ganglionares linfáticas (N en el estadiaje TNM), uno de los factores pronósticos más robustos en cáncer de mama. Los niveles altos de *UBE2C* podrían facilitar el proceso de migración de células tumorales e invasión linfovascular mediante la activación de rutas de señalización como Wnt y *PI3K*. Esto podría explicar nuestros hallazgos, en los que hemos observado una

correlación entre la expresión de *UBE2C* y la presencia de mutaciones oncogénicas de *PI3K*. En conjunto, podrían potenciar el silenciamiento de la expresión de E-cadherina y aumentar la expresión de N-cadherina y EGFR, fenómenos implicados en la invasión linfovascular y metástasis del cáncer de mama.

- ***PI3K* se relaciona con dos genes de diferenciación incluidos en el testEP (*MGP* y *RBBP8*).**

- ***MGP***: este gen codifica un miembro de la familia de osteocalcina/Gla matrix. Para su activación, requiere vitamina K y actúa como un inhibidor de la formación de hueso y de la calcificación ectópica. Mutaciones en este gen se han visto asociadas al síndrome de Keutel (calcificación anormal de cartílagos, hipoplasia facial y estenosis pulmonar periférica). No se conoce el motivo de su sobreexpresión en carcinomas de mama de mayor riesgo de recurrencia, y su inclusión entre los genes de proliferación en Endopredict se hizo a base de estudios de asociación sin hipótesis previa.

- ***RBBP8***: este gen se expresa de manera ubicua en células en estados proliferativos. Codifica una proteína nuclear que interacciona con RB1, una de las proteínas clave reguladoras del ciclo celular. Estos complejos multiproteicos interaccionan con el correpresor transcripcional CTBP. *RBBP8* también se asocia con *BRCA1* y podría modular su función en aspectos tan importantes como la reparación del DNA, el control transcripcional y del checkpoint del ciclo celular. Podría considerarse un gen de supresión tumoral de la misma ruta que *BRCA1*.

- Este estudio allana el terreno para futuras posibles investigaciones que permitan estudiar con mayor exactitud el papel de *UBE2C* en el cáncer de mama.

- **Destacar que hallamos una mutación infrecuente de *PI3K* localizada también en el exón 20 (M1043I).** De esta se desconoce el nivel de evidencia para predecir un posible beneficio del alpelisib por parte de los pacientes que la presenten³⁰. Esta mutación se ha

relacionado con mutaciones en *PI3K* en cáncer de orofaringe y a una mayor respuesta al everolimus en el cáncer urotelial de vejiga.

8. Conclusiones

- En conclusión, las muestras de cáncer de mama con PI3K mutado presentan valores del Score de proliferación (a expensas del gen UBE2C) superiores a los no mutados, pero no es suficiente para alterar el resultado de alto/bajo riesgo de EP.
- También presentan mayores índices de proliferación medidas por Ki-67. Hay una correlación inmunohistoquímica entre un marcador como el Ki-67 y un parámetro de expresión génica (resultado de alto/bajo riesgo de EP).
- Además, las muestras con PI3K mutado expresan genes de diferenciación (a expensas de MGP y RBBP8), aunque con menor significancia estadística que los de proliferación.
- Hemos encontrado un caso con la mutación M1043I de PI3K, relativamente infrecuente.

El objetivo principal de este trabajo se ha cumplido ya que hemos sido capaces de identificar las mutaciones *missense* en los exones 9 y 20 de las pacientes a estudio mediante la pirosecuenciación, hallando incluso una mutación que es más infrecuente (M1034I), y realizado su test EndoPredict a partir de las muestras histológicas a estudio. Todas las mutaciones detectadas ya estaban descritas en la bibliografía consultada. En cuanto a la relación PI3K *mutated / wild-type* y el Ki67 con el test EndoPredict, puede concluirse que a mayor Ki67 puede esperarse un mayor riesgo de recidiva del tumor en los próximos diez años y que mutaciones en PI3K predisponen a un mayor fenotipo proliferativo del cáncer Luminal B, no siendo suficiente para mostrar relación ni con el EPclin/bio ni con el resultado de *High/Low Risk*.

También cumplimos los objetivos secundarios. Fuimos capaces de estudiar la estructura molecular del gen (y proteína a la que codifica) realizar el diseño de primers y obtener, amplificar, y secuenciar el material genético de las muestras histológicas de cáncer de mama Luminal B. Para la comprensión de nuestros hallazgos nos apoyamos en bases de datos científicas como: Cbioportal, ncbi, UniProt, GraphPad... Así mismo, detectamos errores que cometimos durante el procesamiento de las muestras para obtener las mutaciones en PI3K y el

EPclin de las mismas, los comprendimos, y hallamos la mejor manera de prevenirlos y abordarlos. Por último, conseguimos un profundo conocimiento de todas las etapas por la que pasa la muestra de la mama desde que se detecta la sospecha de cáncer hasta que llega al departamento de Anatomía Patológica y se realiza su análisis molecular.

9. ¿Qué hemos aprendido durante este TFG?

Este TFG ha significado para nosotros un gran aprendizaje, ya que hemos estudiado las bases moleculares de una cascada concreta implicada en la génesis del cáncer de mama (PI3K-AKT-mTOR). Para comenzar, tuvimos que estudiar y leer un montón sobre: qué es un gen, qué es un exón, qué tipos de mutaciones existen, qué es el test EndoPredict, qué algoritmo utiliza para llegar a su resultado, que es el cáncer de mama, cómo se clasifica, qué implicaciones tiene con PI3K, cómo se purifica ARN y ADN, cómo se utiliza la PCR, cómo funciona el pirosecuenciador, etc.

Todo esto ha influido en la perspectiva que tenemos ahora en el estudio de la medicina, pues muchas veces olvidamos la trascendencia que tiene conocer y comprender que las manifestaciones clínicas de los pacientes muchas veces son fruto de alteraciones genómicas detectadas en el laboratorio. También nos quedamos con la importancia del trabajo en equipo, de heredar el conocimiento de aquellos que ya tienen la experiencia, y de que un buen ambiente y armonía en el laboratorio hace mucho más ameno las horas de esfuerzo. Todo esto nos ha hecho, sin duda alguna, adquirir gran interés por el campo de la investigación que, hasta ahora, quizás no le dábamos la importancia que merecía. A pesar de la frustración, de los errores y de la repetición de algunos de los pasos que dábamos, hemos crecido como estudiantes y profesionales, demostrándonos, una vez más, que con estudio, diligencia y dedicación todo sale.

10. Bibliografía

1. PIK3CA phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha [Homo sapiens (human)] - Gene - NCBI [Internet]. [citado 29 de agosto de 2022]. Disponible en: <https://www.ncbi.nlm.nih.gov/gene/5290>
<https://www.cancer.net/es/tipos-de-cancer/cancer-de-mama/tipos-de-tratamiento>

2. Gene: PIK3CA (ENSG00000121879) - Summary - Homo_sapiens - Ensembl genome browser 108 [Internet]. [citado 29 de agosto de 2022]. Disponible en: http://www.ensembl.org/Homo_sapiens/Gene/Summary?db=core:g=ENSG00000121879;r=3:179148114-179240093;t=ENST00000263967
3. Di Donato N, Rump A, Mirzaa GM, Alcantara D, Oliver A, Schrock E, et al. Identification and Characterization of a Novel Constitutional PIK3CA Mutation in a Child Lacking the Typical Segmental Overgrowth of «PIK3CA-Related Overgrowth Spectrum». Hum Mutat. marzo de 2016;37(3):242-5.
4. Makki J. Diversity of Breast Carcinoma: Histological Subtypes and Clinical Relevance. Clin Med Insights Pathol. 2015;8:23-31.
5. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. mayo de 2021;71(3):209-49.
6. Alkabban FM, Ferguson T. Breast Cancer. En: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2022 [citado 27 de agosto de 2022]. Disponible en: <http://www.ncbi.nlm.nih.gov/books/NBK482286/>
7. Stemke-Hale K, Gonzalez-Angulo AM, Lluch A, Neve RM, Kuo WL, Davies M, et al. An integrative genomic and proteomic analysis of PIK3CA, PTEN, and AKT mutations in breast cancer. Cancer Res. 1 de agosto de 2008;68(15):6084-91.
8. PIK3CA-Related Overgrowth Spectrum - PubMed [Internet]. [citado 29 de agosto de 2022]. Disponible en: <https://pubmed.ncbi.nlm.nih.gov/23946963/>
9. Arenas JMB. Fundamentos de Ginecología. Ed. Médica Panamericana; 2009. 564 p. [Internet]. [citado 29 de agosto de 2022]. Disponible en: <https://books.google.es/books?hl=es&lr=&id=lqqaTcTQRLEC&oi=fnd&pg=PA13&dq=fundamentos+de+ginecolog%C3%ADa+sego&ots=zZ214ZA8Hp&sig=zWhkqjmib6p4A90fZbFXIqZ-sV4#v=onepage&q=fundamentos%20de%20ginecolog%C3%ADa%20sego&f=false>
10. Experiencia de las mujeres con cáncer de mama en España | AECC Observatorio [Internet]. [citado 29 de agosto de 2022]. Disponible en: <https://observatorio.contraelcancer.es/informes/experiencia-de-las-mujeres-con-cancer-de-mama-en-espana>
11. Carrera Salazar D. Utilidad de la técnica ROLL (radioguided occult lesion localization) en la exéresis de lesiones no palpables de mama [Internet] [Ph.D. Thesis]. TDX (Tesis Doctorals en Xarxa). Universitat Rovira i Virgili; 2016 [citado 29 de septiembre de 2022]. Disponible en: <https://www.tdx.cat/handle/10803/364780>
12. Cáncer de mama - Tipos de tratamiento [Internet]. Cancer.Net. 2012 [citado 29 de septiembre de 2022]. Disponible en: <https://www.cancer.net/es/tipos-de-c%C3%A1ncer/c%C3%A1ncer-de-mama/tipos-de-tratamiento>
13. Oncoguía SEGO. Cáncer infiltrante de mama. Guías de práctica clínica en cáncer ginecológico y mamario. Vol. 12. Publicaciones SEGO, junio 2017. [citado 27 de septiembre de 2022] Disponible en: <https://www.cuidamosdelamujer.es/es/profesionales/docencia-investigacion/revision-6-noviembre-2017.ficheros/759019-Oncoguia%20SEGO%20Cancer%20infiltrante%20de%20mama%20PDF.pdf>
14. Guía práctica quirúrgica de patología mamaria. Sociedad valenciana de cirugía. 2001. [citado 27 de septiembre de 2022]. Disponible en: http://www.sociedadvalencianadecirurgia.com/wp-content/uploads/2014/09/Guia_Patologia_Mamaria.pdf
15. Shen CH. Chapter 11 - Techniques in Sequencing. En: Shen CH, editor. Diagnostic

- Molecular Biology [Internet]. Academic Press; 2019. p. 277-302. [citado 27 de septiembre de 2022]. Disponible en:
<https://www.sciencedirect.com/science/article/pii/B9780128028230000110>
16. Emery. Elementos de genética médica y genómica - Peter D Turnpenny, Sian Ellard, Ruth Cleaver - Google Libros [Internet]. [citado 29 de septiembre de 2022]. Disponible en:
<https://books.google.es/books?id=yY1hEAAAQBAJ&printsec=frontcover&dq=Emery.+Elementos+de+gen%C3%A9tica+m%C3%A9dica+y+gen%C3%B3mica,+16.%C2%AA+Edici%C3%B3n&hl=es&sa=X&ved=2ahUKewjnnNvjt4j6AhVN6RoKHbp5DRAQ6AF6BAgGEAI#v=onepage&q&f=false>
 17. Madsen RR, Erickson EC, Rueda OM, Robin X, Caldas C, Toker A, et al. Positive correlation between transcriptomic stemness and PI3K/AKT/mTOR signaling scores in breast cancer, and a counterintuitive relationship with PIK3CA genotype. PLoS Genet. noviembre de 2021;17(11):e1009876. <https://pubmed.ncbi.nlm.nih.gov/34762647/>
 18. Askeland C, Wik E, Finne K, Birkeland E, Arnes JB, Collett K, et al. Stathmin expression associates with vascular and immune responses in aggressive breast cancer subgroups. Sci Rep. 19 de febrero de 2020;10(1):2914. <https://pubmed.ncbi.nlm.nih.gov/32076022/>
 19. Karakas B, Bachman KE, Park BH. Mutation of the PIK3CA oncogene in human cancers. Br J Cancer. 27 de febrero de 2006;94(4):455-9. <https://pubmed.ncbi.nlm.nih.gov/16449998/>
 20. Purificación de ARN [Internet]. [citado 26 de julio de 2022]. Disponible en:
<https://www.sigmaaldrich.com/ES/es/products/molecular-biology-and-functional-genomics/nucleic-acid-purification/rna-purification>
 21. Resolver el caso de la contaminación por RNasa - Phoseon Technology [Internet]. Tecnología Phoseon. [citado 26 de julio de 2022]. Disponible en:
<https://phoseon.com/es/in-the-news/the-usual-suspect-solving-the-case-of-rnase-contamination-in-rna-analysis/>
 22. Proteínasa K solución 20 mg/ml [Internet]. [citado 29 de junio de 2022]. Disponible en:
<https://www.itwreagents.com/iberia/es/product/proteinasa+k+soluci%C3%B3n+20+mg/ml/A4392>
 23. Purificación de ARNm con microesferas magnéticas - Labster Theory [Internet]. [citado 28 de julio de 2022]. Disponible en:
<https://theory.labster.com/mrna-purification-magnetic-beads-es/>
 24. Lehmann-Che J, Miquel C, Wong J, Callens C, Rouleau E, Quillien V, et al. First French Pilot Quality Assessment of the EndoPredict Test for Early Luminal Breast Carcinoma. Anticancer Res. mayo de 2018;38(5):2909-14. <https://pubmed.ncbi.nlm.nih.gov/29715116/>
 25. Filipits M, Rudas M, Jakesz R, et al. A new molecular predictor of distant recurrence in ER-positive, HER2-negative breast cancer adds independent information to conventional clinical risk factors. Clin Cancer Res. 2011;17(18):6012-6020. Disponible en:
<https://pubmed.ncbi.nlm.nih.gov/21807638/>
 26. Dubsky PC, Singer CF, Egle D, et al. The EndoPredict score predicts response to neoadjuvant chemotherapy and neoendocrine therapy in hormone receptor-positive, human epidermal growth factor receptor 2-negative breast cancer patients from the ABCSG-34 trial. Eur J Cancer. 2020;134:99-106. Disponible en:
<https://pubmed.ncbi.nlm.nih.gov/32502940/#:~:text=ABCSG%2D34%20trial-.The%20EndoPredict%20score%20predicts%20response%20to%20neoadjuvant%20chemotherapy%20and%20neoendocrine,doi%3A%2010.1016%2Fj>
 27. Dubsky P, Brase JC, Jakesz R, et al. Austrian Breast and Colorectal Cancer Study Group (ABCSG). The EndoPredict score provides prognostic information on late distant metastases in ER+/HER2- breast cancer patients. Br J Cancer. 2013 Dec 10;109(12):2959-64. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3859949/>
 28. Purificación de ácidos nucleicos mediante columnas de centrifugado. En: Wikipedia, la enciclopedia libre [Internet]. 2022 [citado 29 de julio de 2022]. Disponible en:

https://es.wikipedia.org/w/index.php?title=Purificaci%C3%B3n_de_%C3%A1cidos_nucleicos_mediante_columnas_de_centrifugado&oldid=143540045

29. Clonación acelular : reacción en cadena de la polimerasa (PCR) - Texto ilustrado e interactivo de biología molecular e ingeniería genética - ClinicalKey Student [Internet]. [citado 4 de julio de 2022]. Disponible en: <https://www-clinicalkey-com.accedys2.bbt.ull.es/student/content/book/3-s2.0-B9788480866477000144#hi0000358>
30. Martínez-Sáez O, Chic N, Pascual T, Adamo B, Vidal M, González-Farré B, et al. Frequency and spectrum of PIK3CA somatic mutations in breast cancer. Breast Cancer Res BCR. 13 de mayo de 2020;22(1):45. <https://pubmed.ncbi.nlm.nih.gov/32404150/>

11. Figuras

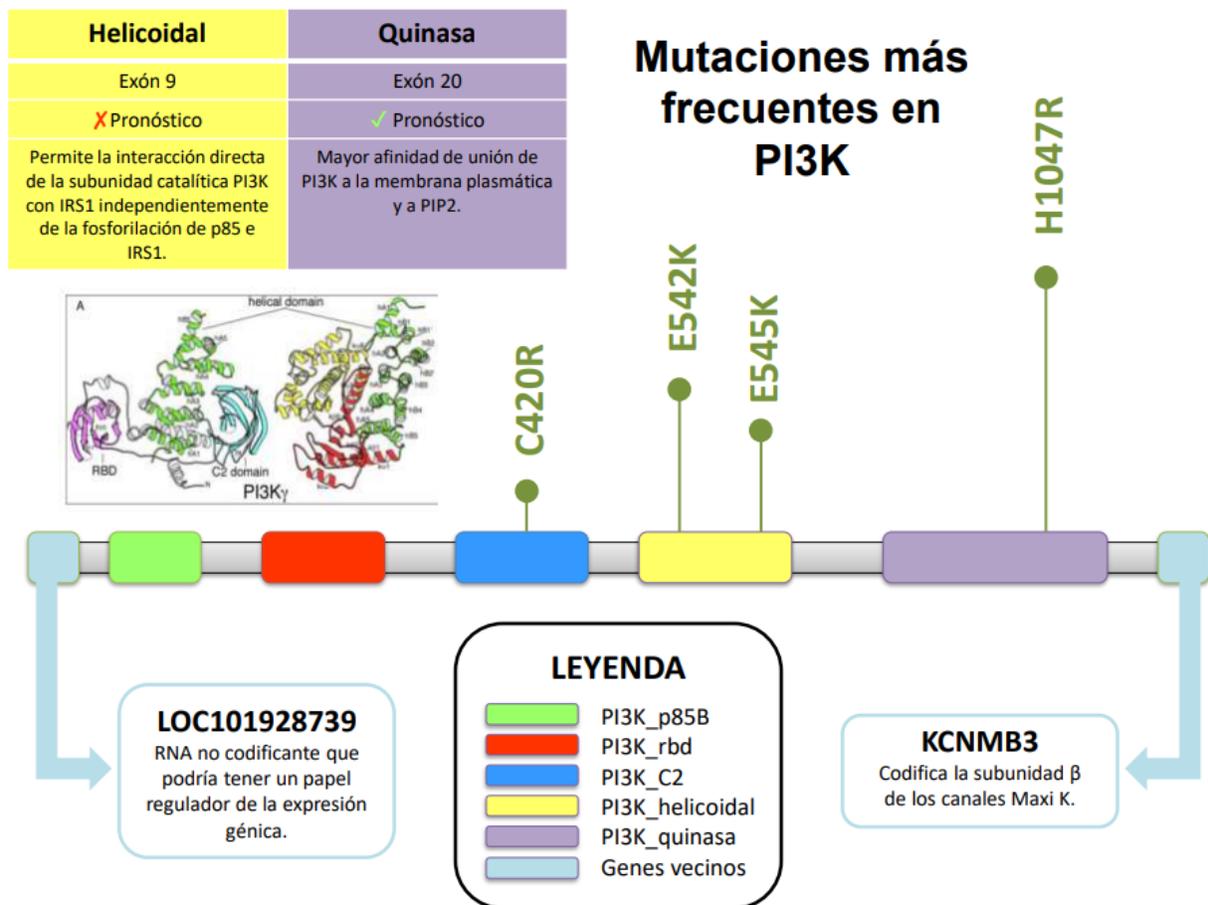


Figura 1. Se representa PI3K en el genoma humano y las mutaciones más frecuentes según el dominio al que nos referimos (p85B, rbd [*ras binding domain*], C2, helicoidal y quinasa). Además, se añaden los genes vecinos y sus funciones: LOC101928739 (ARN no codificante) y KCNMB3 (codifica la subunidad β de los canales Maxi K). Se detallan las diferencias principales entre las dos mutaciones más importantes que afectan a PI3K: el dominio helicoidal localizado

en el exón 9 y de peor pronóstico; y el dominio quinasa localizado en el exón 20 y de mejor pronóstico. La última fila de la tabla refleja cómo sería la ganancia de función a nivel molecular en cada caso. Por último, las cuatro mutaciones especificadas son *missense* (H1047R, E542K, E545K y C420R). H: Histidina, R: Arginina, E: Ácido glutámico, K: lisina, C: Cisteína.

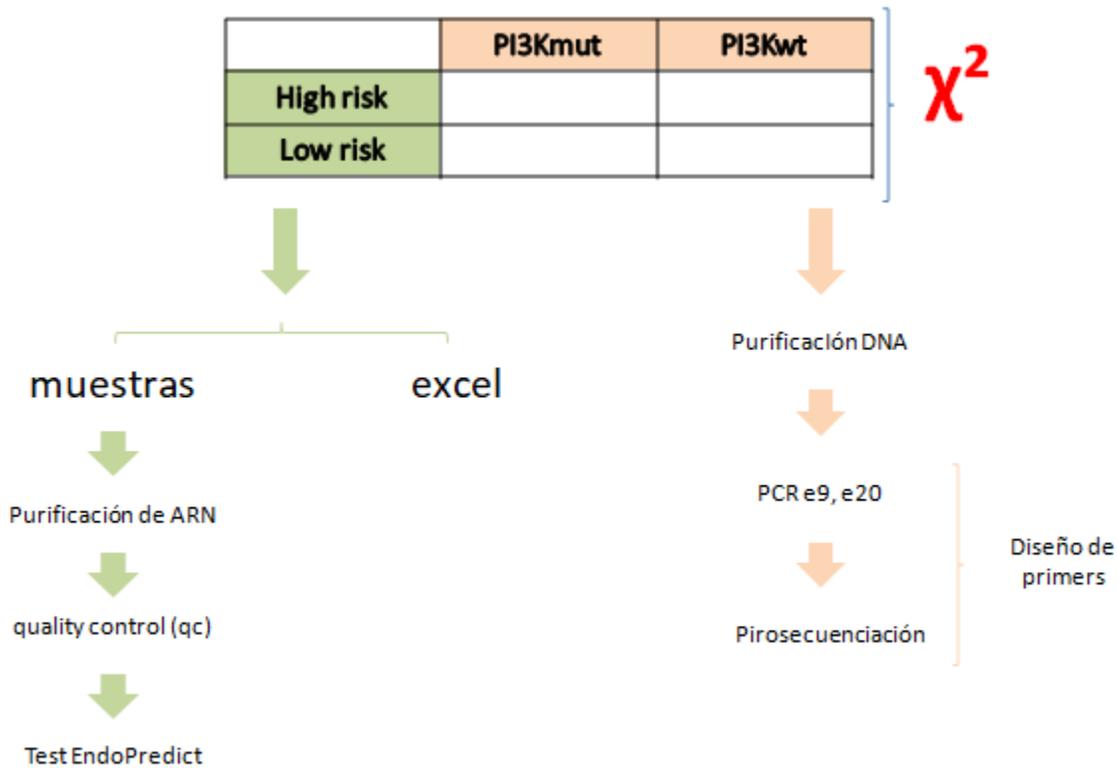


Figura 2. Esquema general del apartado “5. Material y métodos”. La hipótesis alternativa del TFG sería encontrar un mayor número de cánceres luminales B PI3K mutados en el grupo de alto riesgo del test EndoPredict. Para ello, y al tratarse de dos variables cualitativas, utilizamos una chi-cuadrado con el objetivo de encontrar una correlación. *Grosso modo*, se divide el proyecto en dos grandes grupos que difieren en cuanto a técnicas a utilizar: *High/Low Risk* y *PI3K mutated/wild type*.

$$\begin{aligned}
 & \text{3.1} \quad \Delta C_t(\text{GOI}) = 20 - C_t(\text{GOI}) + [C_t(\text{CALM2}) \\
 & \quad \quad \quad + C_t(\text{OAZ1}) + C_t(\text{RPL37A})]/3 \\
 & \quad \quad \quad \downarrow \\
 & \text{3.2} \quad s_u = 0.41 \cdot \Delta C_t(\text{BIRC5}) - 0.35 \cdot \Delta C_t(\text{RBBP8}) \\
 & \quad \quad \quad + 0.39 \cdot \Delta C_t(\text{UBE2C}) - 0.31 \cdot \Delta C_t(\text{IL6ST}) \\
 & \quad \quad \quad - 0.26 \cdot \Delta C_t(\text{AZGP1}) + 0.39 \cdot \Delta C_t(\text{DHCR7}) \\
 & \quad \quad \quad - 0.18 \cdot \Delta C_t(\text{MGP}) - 0.15 \cdot \Delta C_t(\text{STC2}) - 2.63 \\
 & \quad \quad \quad \downarrow \\
 & \text{3.3} \quad s = 0, \quad \text{if } 1.5 \cdot s_u + 18.95 < 0 \\
 & \quad \quad \quad s = 15, \quad \text{if } 1.5 \cdot s_u + 18.95 > 15 \\
 & \quad \quad \quad s = 1.5 \cdot s_u + 18.95, \quad \text{otherwise} \\
 & \quad \quad \quad \downarrow \\
 & \text{3.4} \quad s_{\text{clin}} = 0.35 \cdot t + 0.64 \cdot n + 0.28 \cdot s
 \end{aligned}$$

Figura 3. Fórmula del artículo publicado en 2011 para la fabricación del Excel que simula el algoritmo usado en el test EndoPredict.

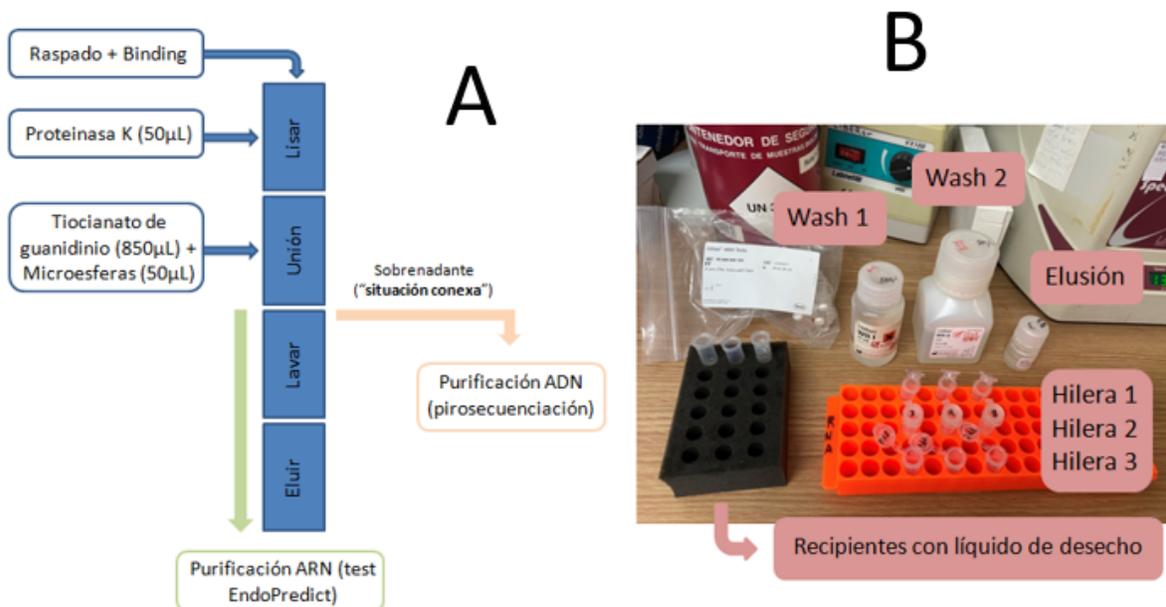
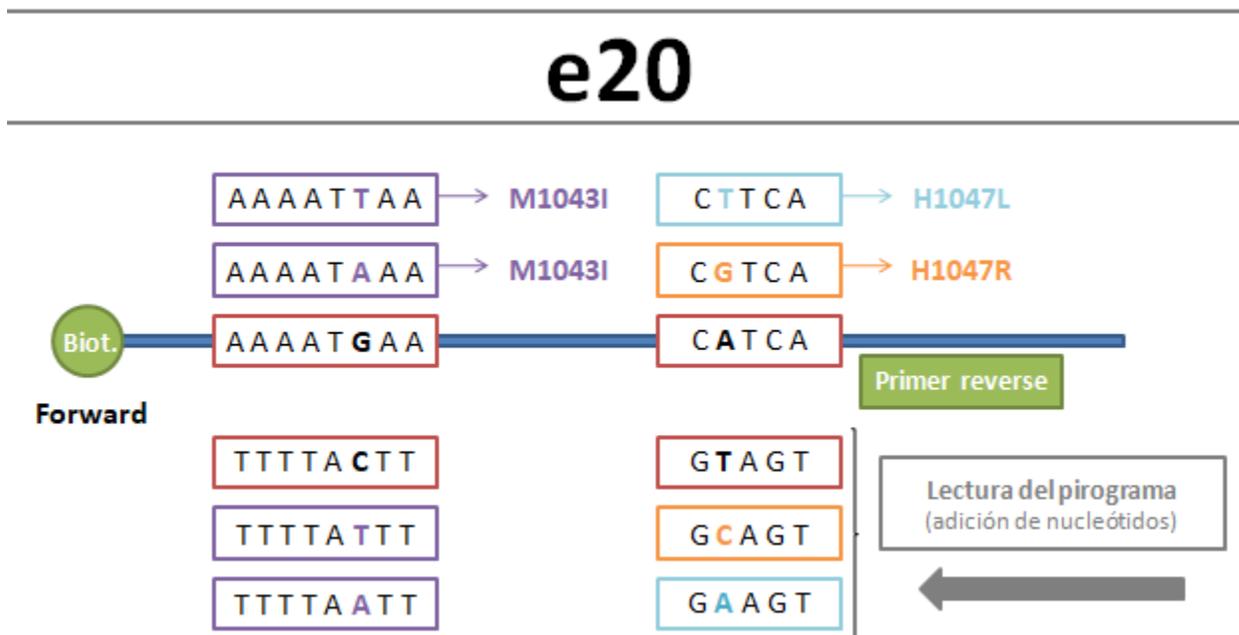


Figura 4.A Obtención de ADN purificado (para la pirosecuenciación) a partir del sobrenadante como punto conexo de la purificación del ARN (con el objetivo de la realización del test EndoPredict). Para ello, se utilizará la cromatografía por afinidad mediante unas columnas que atraparán el ADN en condiciones altas de salinidad. Una vez la disminuycamos hasta hacerla nula, usaremos la elusión (agua destilada) en aras de desenlazar el ADN de la matriz de afinidad y lograr que este precipite en el tubo Eppendorf de la tercera hilera. **Figura 4.B** Representación gráfica de los elementos utilizados en “purificación del ADN mediante cromatografía por afinidad”.



e9

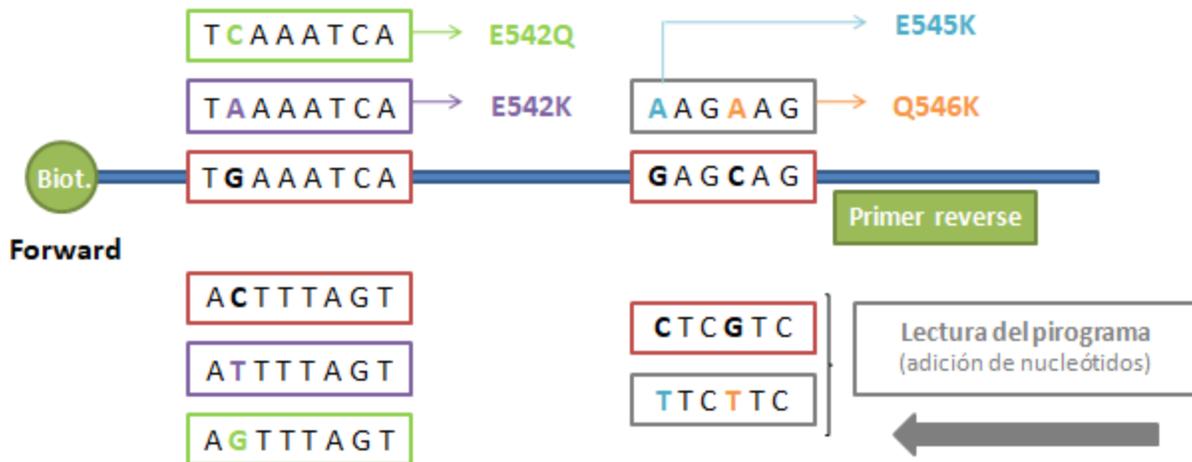


Figura 5. Lectura del pirograma de la secuenciación del exón 20 de PI3K en la posición 1047 y 1043. Las hebras que tienen los *primers* biotinilados son las *forward* y utilizamos cebadores reversos para la pirosecuenciación. La secuencia *wild-type* en el genoma es “CATCA”, leyéndose en el pirograma como “TGATG” (debido a que el primer es reverso y, por tanto, el pirosecuenciador empieza a introducir nucleótidos en sentido contrario), para la posición 1047. Si se cambiase la A por una G, el pirograma mostraría que ha incluido una C, tratándose de la mutación **H1047R**; en cambio, si se cambiase la A por una T, el pirograma mostraría que ha incluido una A, tratándose la mutación **H1047L**. Además, encontramos en una paciente una mutación rara “**M1043I**” donde, distintas mutaciones, en este caso una G por una A, o una G por una T, generan un cambio por el mismo aminoácido. **Figura 6. Lectura del pirograma de la secuenciación del exón 9 de PI3K en la posición 545 y 546, y 542.** La secuencia *wild-type* en el genoma es “GAGCAG”, leyéndose en el pirograma como “CTGCTC”, para la posición 545 y 546. Si se cambiase la G por una A, el pirograma mostraría que ha incluido una T, tratándose de la mutación **E545K**; en cambio, si se cambiase la C por una A, el pirograma mostraría que ha incluido una T, tratándose de la mutación **Q546K**. La secuencia *wild-type* en el genoma es “TGAAATCA”, leyéndose en el pirograma como “TGATTTCA”, para la posición 542. Si se cambiase la G por una A, el pirograma mostraría que ha incluido una T, tratándose de la

mutación **E542K**; en cambio, si se cambiase la **G** por una **C**, el pirograma mostraría que ha incluido una **G**, tratándose de la mutación **E542Q**. M: Metionina, I: Isoleucina.

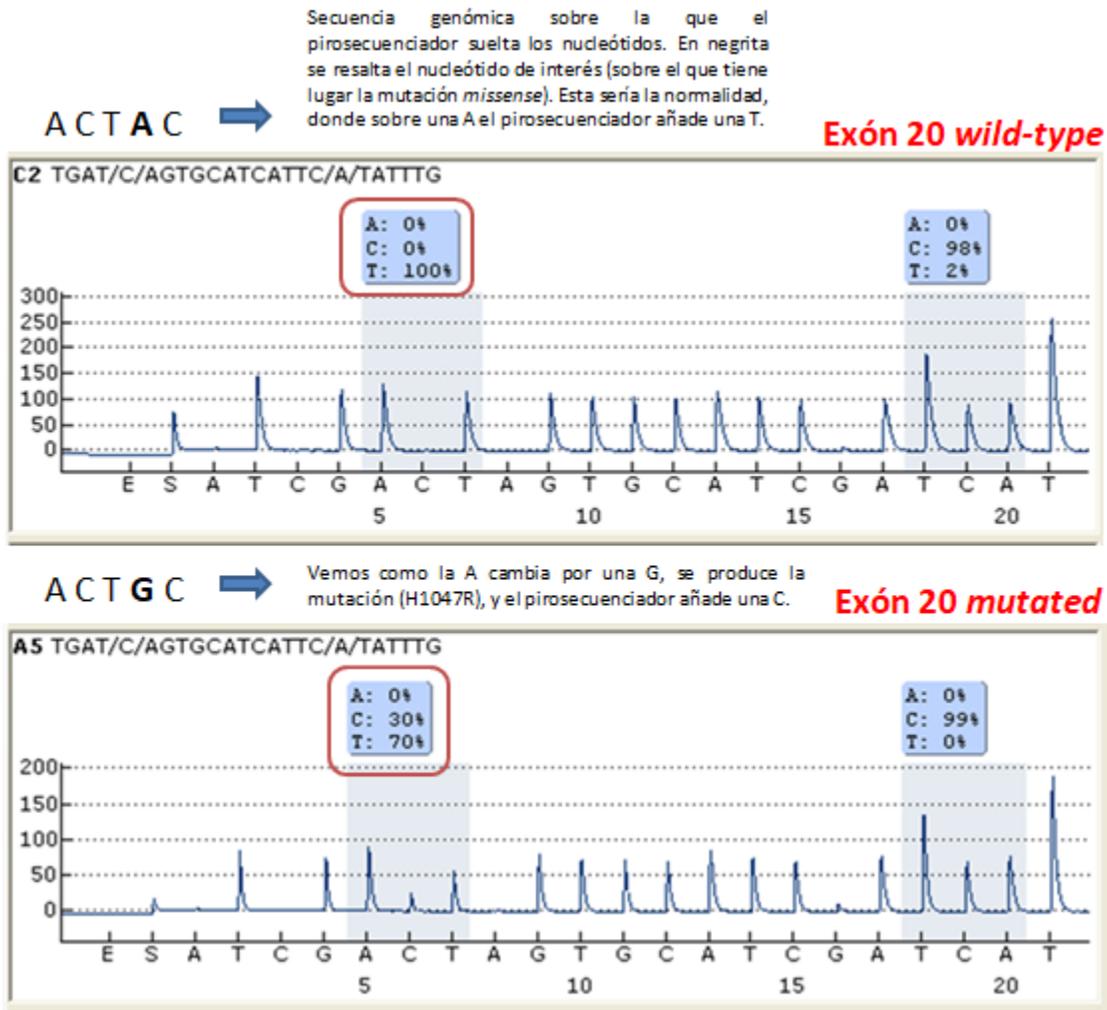


Figura 7. Comprensión del concepto de frecuencia alélica a partir del pirograma. Previo a la pirosecuenciación, realizamos una PCR, de la que obtenemos muchos fragmentos de ADN (o amplicones) de nuestra región de interés (localizada, en este caso, en el exón 20). El pirosecuenciador lo que va a realizar es introducir una serie de nucleótidos, uno por uno, en dicha región, con el objetivo de visualizar qué nucleótido añade donde esperamos ver la mutación (como se muestra en la imagen pasaría de añadir una T, donde había una A, en el *wild-type*, a soltar una C, donde había una G, en el *mutated*). En la imagen del *wild-type* observamos “T: 100%”, dando a entender que en todos los amplicones que secuenció añadió el nucleótido que se esperaba en condiciones normales (una T). No obstante, en la imagen del *mutated* pone “C: 30% y T: 70%”, queriendo decir que el 30% de las veces que secuenció los

amplicones añadió el nucleótido de la mutación (una C), y el 70% restante el que se espera encontrar fisiológicamente. Esto es lo que se conoce como frecuencia alélica. Una manera de entenderlo sería imaginar la muestra de la paciente y saber que, en el tejido que raspamos para purificar el ADN, nos llevaremos núcleos de células con ADN que portarán o no la mutación. En el primer caso, todas las células que raspamos no la portaban (puramente *wild-type*); en el segundo caso, el 30% de las células que raspamos la portaban (*mutated*). Se considera mutado a partir de un 10% de frecuencia alélica para ese nucleótido en el exón 20; y un 15% en el exón 9.



Figura 8. Herramientas utilizadas en la purificación del ARN.

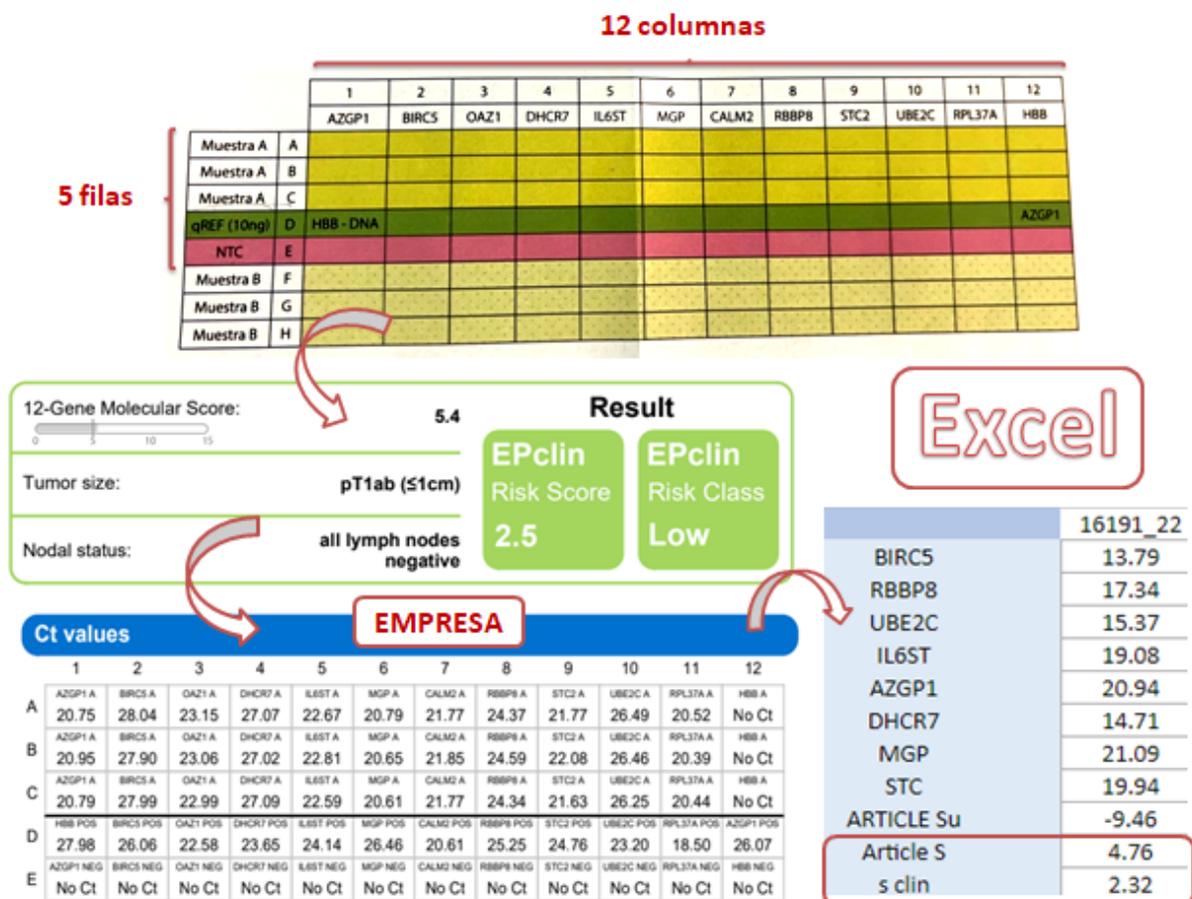


Figura 9. Placa de EndoPredict, informe de la empresa y traspaso de datos al Excel. “Article S” sería nuestro EP *risk score* de la empresa y el “s clin” nuestro EPclin. Estos dos datos obtenidos en el Excel provienen de la tabla “Ct values” que da el informe de la empresa, cuyos valores sometemos a un proceso integrado por diferentes fórmulas en el Excel (algoritmo obtenido por un estudio que publicó hace años la compañía).

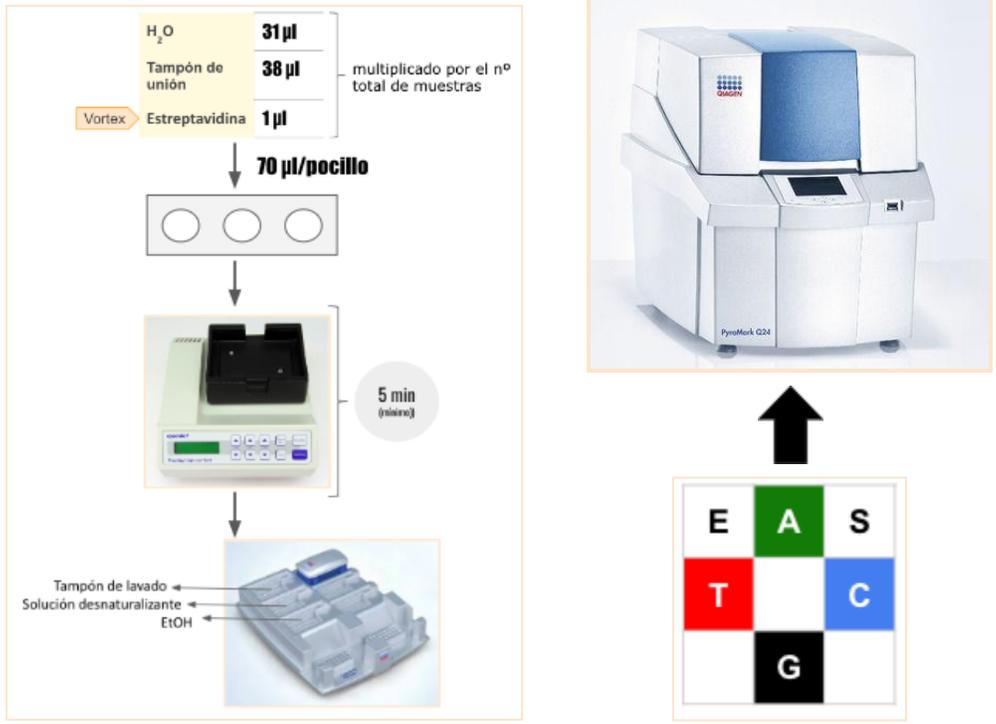


Figura 10. Pirosecuenciación (PyroMark Q24 MDx).

VARIABLES

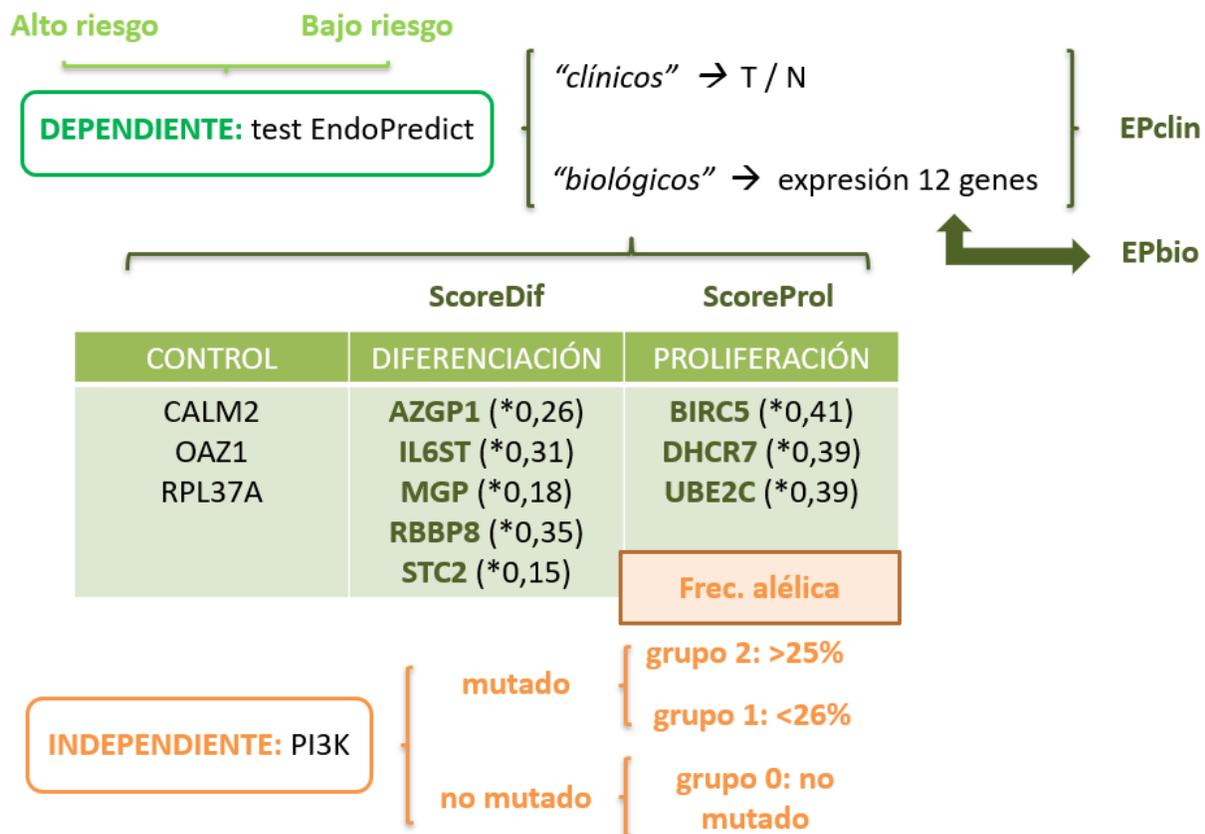


Figura 11. Definición de las variables y jerarquía.

12. Tablas

			Variable INDEPENDIENTE test EndoPredict													
			Ct H/L	Ct EPcln	Ct EPblo	Ct Sdif	Ct Sprol	Ct A2SP1	Ct IL6ST	Ct MGP	Ct RBBP8	Ct STC2	Ct BIRC5	Ct DHCR7	Ct UBE2C	
Variables D E P E N D I E N T E S	Ki67	Ct														
		Ct (2)														
		Ct (3)														
	PI3K	Ct (2) wt/mut														
		Ct (3) 0%, <26%, >25%														

			Variable INDEPENDIENTE test EndoPredict												
			Ct H/L	Ct EPcln	Ct EPblo	Ct Sdif	Ct Sprol	Ct A2SP1	Ct IL6ST	Ct MGP	Ct RBBP8	Ct STC2	Ct BIRC5	Ct DHCR7	Ct UBE2C
Variables D E P E N D I E N T E S	Ki67	Ct		0,202	0,414	-	+	-	-	-	-	-	+	+	+
		Ct (2)													
		Ct (3)													
	PI3K	Ct (2) wt/mut													
		Ct (3) 0%, <26%, >25%													

NO PARAMÉTRICOS

- p < 0,05
- U de Mann-Whitney
- Kruskal-Wallis
- Chi-cuadrado (χ^2)
- Rho de Spearman

Tabla 1. Realización de los test estadísticos no paramétricos según las variables y la jerarquía ya definidas.

13. Anexos

Anexo 1. Uno de los objetivos de nuestro Trabajo de Fin de Grado (TFG) es aprender a explotar plataformas de genes (cBioPortal for Cancer Genomics y Centro Nacional para la Información Biotecnológica), de proteínas (Uniprot) y de bioestadística (GraphPad), en aras de comprender la manera más efectiva de minar datos, ordenarlos, estudiarlos y compararlos, y de formular, con estos, hipótesis que se puedan contestar con muestras de pacientes y estudios accesibles para todo el mundo en la nube.

- Genes:
 - cBioPortal: entre otras cosas, hemos aprendido a utilizar CBioPortal, una base de datos genómica muy útil para visualizar, analizar y descargar información relativa al cáncer que estamos estudiando. Para ello la web dispone de varios vídeos explicativos que nos han resultado muy prácticos. Así, hemos sido capaces de formular nuestras propias hipótesis como, por ejemplo, que la expresión de genes implicados en la proliferación (MCM2, MCM5, MCM6, IRF2, PGK1, XRCC4, PARP3) es mayor en las muestras con mutación en PIK3CA que en las no mutadas (METABRIC, Nature 2012 & Nat Commun 2016). De esta forma, el programa nos permite recoger los datos de las medias poblacionales de expresión de ARNm de cada uno de los genes en relación a los dos grupos, de tal manera que podemos interpretar estadísticamente los datos en otros programas. En definitiva, se trata de una herramienta muy provechosa para nuestro futuro profesional.
 - Centro Nacional para la Información Biotecnológica (NCBI): es parte de la Biblioteca Nacional de Medicina de Estados Unidos, una rama de los Institutos Nacionales de Salud. Una vez entremos en la página web “<https://www.ncbi.nlm.nih.gov/>”, abrimos la pestaña situada a la izquierda del buscador y seleccionamos “Gene”. Luego, escribimos en el buscador el gen del cual queremos obtener información. En nuestro caso, PIK3CA (importante especificar la especie *Homo sapiens* a la derecha del gen). Seguidamente, clicamos en nuestro gen y se abre a nuestra disposición bastante información (el resumen del gen, la localización genómica, los genes vecinos, el número de

exones, la recopilación de artículos más relevantes en PubMed, etc.) Es importante destacar que cada gen tiene su propio código denominado “ID”.

- Proteínas:
 - UniProt: UniProt (de *universal protein*) es un repositorio central de datos gratuito sobre proteínas. En el buscador añades el nombre de la proteína y la especie de interés. Una vez cliques en ella, puedes acceder a multitud de información (la definición de la función de la proteína y de su actividad catalítica, los dominios por los que está compuesta, etc.) Es importante destacar que cada proteína tiene su propio código denominado “entry”.
- Bioestadística:
 - GraphPad: Se trata de un programa que facilita el análisis estadístico y nos permite crear una amplia gama de tipos de gráficos. Nosotros utilizamos especialmente el apartado de *QuickCalcs* que nos da la posibilidad de estimar la correlación existente entre variables cuantitativas o continuas, o cualitativas o categóricas. De esta manera, podemos complementar las hipótesis, formuladas, por ejemplo, con el cBioPortal, con la certeza que confiere el análisis estadístico.

Anexo 2.

Exón 10 (antes 9)	E21 (antes 20)
AGTAACAGACTAGCTAGAGACAATGAA TTAAGGGAAAATGACAAAGAACAGCT CAAAGCAATTTCTACACGAGATCCTCT CTCT[GA]AATCACT[GAGCAG]GAGAAA GATTTTCTATGGAGT CACAGgtaagtgctaaa atggagattctctgtttctttctttattacagaaaaataactgaat ttggctgatctcagcatgttttaccatactattggaa	GTTTCAGGAGATGTGTTACAAGGCTTAT CTAGCTATTCGACAGCATGCCAATCTCT TCATAAATCTTTTCTCAATGATGCTTGG CTCTGGAATGCCAGAACTACAATCTTTT GATGACATTGCATACATTCGAAAGACC CTAGCCTTAGATAAAACTGAGCAAGAG GCTTTGGAGTATTTTCATGAAACAAATG AATGATGCA[CAT]CATGGTGGCTGGAC AACA AAAATGGATTGGATCTTCCACAC AATTAAACAGCATGCATTGAACTGAAA AGATAACTGAGAAAATGAAAGCTCACT

	<p>CTGGATTCCACACTGCACTGTTAATAAC TCTCAGCAGGCAAAGACCGATTGCATA GGAATTGCACAATCCATGAACAGCATT AGAATTTACAGCAAGAACAGAAATAAA ATACTATATAATTTAAATAATGTAAACGC AAACAGGGTTTGATAGCACTTAAACTA GTTCAATTTCAAATTAAGCTTTAGAATA ATGCGCAATTTCATGTTATGCCTTAAGT CCAAAAGGTAAACTTTGA</p>
<p>Primers para la PCR (200 pb) de la mutación E542K y E545K:</p> <ul style="list-style-type: none"> • Left primer o <i>forward</i> (66°C): AGCTAGAGACAATGAATTAAGG GA. • Right primer o <i>reverse</i> (68°C): tctccatttagcacttacCTGTG. 	<p>Primers para la PCR (200 pb) de la mutación H1047R:</p> <ul style="list-style-type: none"> • Left primer o <i>forward</i> (54 °C): TCAGGAGATGTGTTACAAGGCT. • Right primer o <i>reverse</i> (66°C): TGAAATTGCGCATTATTCTAAAG CT.
<p>Primer para la pirosecuenciación (150 pb) de la mutación E542K y E545K:</p> <ul style="list-style-type: none"> • <i>Reverse primer:</i> CTCCATAGAAAATCTTT. 	<p>Primer para la pirosecuenciación (150 pb) de la mutación E542K y E545K:</p> <ul style="list-style-type: none"> • <i>Reverse primer:</i> TGTTGTCCAGCCACCA.