

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 972 684**

51 Int. Cl.:

C12Q 1/6827 (2008.01)

C12Q 1/6834 (2008.01)

C12Q 1/6886 (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **11.04.2019 PCT/EP2019/059274**

87 Fecha y número de publicación internacional: **17.10.2019 WO19197541**

96 Fecha de presentación y número de la solicitud europea: **11.04.2019 E 19717477 (4)**

97 Fecha y número de publicación de la concesión europea: **20.12.2023 EP 3775274**

54 Título: **Procedimiento de detección de anomalías genéticas somáticas, combinación de sondas de captura y kit de detección**

30 Prioridad:

11.04.2018 EP 18305440

45 Fecha de publicación y mención en BOPI de la traducción de la patente:

14.06.2024

73 Titular/es:

**UNIVERSITÉ DE BOURGOGNE (50.0%)
Maison de l'Université, Esplanade Erasme
21000 Dijon, FR y
CGFL (CENTRE GEORGES FRANÇOIS LECLERC)
(50.0%)**

72 Inventor/es:

**BIDOT, ROMAIN;
VEGRAN, FRÉDÉRIQUE y
GHIRINGHELLI, FRANÇOIS**

74 Agente/Representante:

LINAGE GONZÁLEZ, Rafael

Observaciones:

**Véase nota informativa (Remarks, Remarques o
Bemerkungen) en el folleto original publicado por
la Oficina Europea de Patentes**

ES 2 972 684 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

DESCRIPCIÓN

Procedimiento de detección de anomalías genéticas somáticas, combinación de sondas de captura y kit de detección

5

Campo de la invención

La presente invención se refiere al campo de la detección del cáncer.

10 En particular, la presente invención se refiere a un procedimiento de detección de anomalías genéticas potencialmente presentes en un tumor utilizando sondas de captura específicas, a un kit de detección de estas anomalías genéticas y a una combinación de sondas de captura.

Descripción de la técnica relacionada

15

El cáncer es una de las principales causas de muerte en el mundo. Por ejemplo, en 2010, casi 1,5 millones de personas (esencialmente mujeres) desarrollaron cáncer de mama en todo el mundo. Se estima que entre el 5 y el 10% de estos casos eran hereditarios. Además, sólo en Estados Unidos se notificaron más de 1,5 millones de casos nuevos y más de 0,5 millones de muertes durante 2010.

20

El tratamiento que se realiza actualmente para el cáncer es principalmente una terapia sintomática que consiste principalmente en una terapia quirúrgica con una combinación de radioterapia y quimioterapia. Gracias a los avances en la tecnología médica, el cáncer es ahora una enfermedad casi curable si se puede detectar a tiempo.

25

De hecho, es deseable no sólo detectar el cáncer, sino también diagnosticar: un cáncer desarrollado en un lugar invisible a simple vista, una extensión del cáncer, una neoplasia maligna o un curso post-operatorio del cáncer, una recurrencia, una metástasis, etc.

30

El uso generalizado de la Secuenciación de Próxima Generación (NGS), también conocida como secuenciación masiva paralela, ha abierto nuevas vías para la investigación y el diagnóstico del cáncer. La NGS aportará enormes cantidades de datos nuevos sobre el cáncer y, especialmente, sobre su genética.

35

La NGS representa una forma eficaz de capturar una gran cantidad de información genómica sobre un cáncer. La mayoría de las tecnologías NGS giran en torno a la secuenciación por síntesis. Cada fragmento de ADN a secuenciar está unido a una matriz y luego la ADN polimerasa agrega secuencialmente los nucleótidos marcados. Una cámara de alta resolución captura la señal de cada nucleótido que se integra y anota las coordenadas espaciales y el tiempo. Luego, un programa informático puede inferir la secuencia en cada punto para generar una secuencia de ADN contigua, denominada lectura.

40

Múltiples mejoras tecnológicas han permitido que la NGS sea implementada con mayor facilidad en un flujo de trabajo clínico. Ahora ya no es necesario manipular las muestras de manera diferente a las muestras de diagnóstico estándar, y los avances recientes incluso han permitido derivar datos genómicos cada vez más complejos a partir de la sangre periférica de un paciente. El concepto de medicina de precisión va de la mano de la comprensión del genoma del cáncer determinado por la NGS.

45

Antes del desarrollo de la NGS, el genotipado de tumores se realizaba sólo en lugares/loci genómicos específicos que se sabía que mutaban con frecuencia en el cáncer, que se conocen como "puntos calientes", en general mediante la secuenciación de Sanger. Estos enfoques eran más adecuados para mutaciones activadoras recurrentes en oncogenes, como en el gen *KRAS* en el cáncer de colon y de pulmón.

50

Sin embargo, estos enfoques fueron insuficientes para identificar alteraciones en los supresores de tumores (en las que una alteración en cualquier parte del gen puede afectar su función) o el área cada vez más compleja de las alteraciones de los puntos críticos de "cola larga" en los oncogenes. Además, estos enfoques sólo permiten una exploración muy limitada del genoma humano.

55

Por lo tanto, las opciones de ensayo actuales implican enfoques que pueden capturar genes cancerosos conocidos ("paneles de genes"), enfoques de exoma completo, genoma completo y/o de transcriptoma completo. Existen varias compensaciones a la hora de aumentar la porción del genoma que está siendo secuenciado. La primera es una pérdida de cobertura por la misma cantidad de secuenciación. La cobertura, o profundidad, se define como el número promedio de lecturas mapeables en un locus determinado en su panel. Una cobertura más baja limita la capacidad de considerar con seguridad que una variante de fracción alélica baja es biológicamente real y no un artefacto técnico. La segunda es que la secuenciación del genoma completo y del exoma completo requiere la secuenciación de la línea germinal para mejorar la identificación de variantes somáticas verdaderas, que pueden descubrir trastornos hereditarios incidentales clínicamente relevantes.

60

65 Para lograr estos últimos enfoques, la gran mayoría de los kits de detección de cáncer existentes en el mercado

se basan en la secuenciación de amplicones.

La secuenciación de amplicones enriquece los genes diana mediante PCR con un conjunto de cebadores para exones de genes seleccionados antes del análisis por NGS. Especialmente, el conjunto de cebadores amplificará fragmentos de ADN grandes y diferentes. Estos protocolos tienen la ventaja de requerir menos ADN de entrada y menos tiempo de respuesta que los procedimientos de captura híbridos, lo cual es fundamental para la aplicación clínica, pero potencialmente la amplificación por PCR puede sesgar la fracción alélica observada. También extrae información de un porcentaje menor del material inicial, lo que aumenta aún más la posibilidad de sesgo al indicar variaciones en el número de copias. El análisis informático es relativamente fácil, ya que cualquier lectura que no se asigne a un locus entre los cebadores puede ser ignorada.

Una desventaja de esta simplicidad es que el ensayo es inherentemente incapaz de detectar fusiones inesperadas, porque el cebador 5' o 3' no lograría unirse al ADN translocado. Además, este procedimiento de secuenciación requiere que el ADN se encuentre en perfecto estado. Además, los cebadores específicos de estos diferentes fragmentos requieren condiciones de hibridación y elongación similares. Al ser difíciles de cumplir estas condiciones, la secuenciación de amplicones genera una gran cantidad de artefactos.

DAVID H. SPENCER ET AL, THE JOURNAL OF MOLECULAR DIAGNOSTICS (*LA REVISTA DE DIAGNÓSTICO MOLECULAR*), vol. 15, núm. 5, 1 de septiembre de 2013 (2013-09-01), páginas 623-633, se relaciona con la comparación de datos clínicos de Secuencia de Próxima Generación específicos de muestras de tejido fijadas con formalina y recién congeladas. El documento de patente WO2012/092426A1 se refiere a la optimización del análisis multigénico de muestras tumorales.

Por lo tanto, existe la necesidad de desarrollar un nuevo procedimiento de detección de anomalías genéticas somáticas presentes en un tumor que permita detectar varios tipos de cáncer con alta sensibilidad al mismo tiempo y preferentemente para diferentes pacientes.

En particular, existe la necesidad de desarrollar un nuevo procedimiento de detección que permita proporcionar un medicamento personalizado para cada paciente.

Sumario de la invención

La presente invención satisface estas necesidades ya que los inventores, sorprendente e inesperadamente, descubrieron que una combinación específica de sondas de captura integradas en captura híbrida permite obtener un procedimiento de medición de alta sensibilidad para detectar anomalías genéticas somáticas y esto, para varios tipos de cáncer y en un solo momento.

Además, los inventores han descubierto que la combinación específica de sondas de captura integradas en la captura híbrida es capaz no sólo de detectar el cáncer, sino también de diagnosticar el cáncer que se ha desarrollado en un lugar invisible a simple vista, la extensión del cáncer, la neoplasia maligna o el curso post-operatorio del cáncer, recurrencia, metástasis y similares.

De este modo, la presente invención proporciona un procedimiento de detección de anomalías genéticas somáticas potencialmente presentes en una muestra de tejido sólido (potencialmente canceroso), tomada de un sujeto, en el que dicho procedimiento de detección comprende las siguientes etapas:

- (a) corte de un ADN genómico extraído de dicha muestra de tejido, para obtener ADN genómico cortado;
- (b) preparar una biblioteca mediante captura por hibridación a partir del ADN genómico cortado obtenido en la etapa (a), mediante la implementación de una combinación de al menos 51 sondas de captura que tienen al menos 110 nucleótidos de longitud y al menos un 75% de homología de secuencia respectivamente con la SEQ ID NO: 1 a la SEQ ID NO: 51, pudiendo dichas sondas de captura hibridarse en fragmentos de las siguientes regiones diana de: ALK, BRAF, EGFR, EIF1AY, ERBB2, ERBB4, KDR, KIT, KRAS, MAP2K1, MET, NRAS, PDGFRA, SLC28A1, TP53;
- (c) secuenciación del ADN genómico cortado contenido en la biblioteca emitida en la etapa (b), y
- (d) analizar las secuencias de dicho ADN genómico cortado obtenido en la etapa (c), para detectar anomalías genéticas somáticas en dicha muestra de tejido.

La presente invención permite una alta sensibilidad del análisis de 15 genes de secuencia codificante completa (y secuencias de intrones cerradas) con una alta confiabilidad. Esto es la consecuencia de una señal de fondo muy baja, un diseño específico dirigido a los genes estudiados con teranósticos y el uso de controles internos. Finalmente, este diseño se puede utilizar en muestras sólidas (tumores) y líquidas (sangre).

La presente invención se refiere a una combinación de sondas de captura, que comprende al menos 51 sondas de captura que comprenden al menos 110 nucleótidos y al menos un 75% de homología de secuencia con, respectivamente, las sondas de la SEQ ID NO: 1 a la SEQ ID NO: 51.

Preferentemente, la combinación de sondas de captura además comprende al menos 100 sondas de captura adicionales, preferentemente al menos 200, en particular al menos 300, e idealmente al menos 1178 sondas de captura adicionales que comprenden al menos 110 nucleótidos y al menos un 75% de homología de secuencia con, respectivamente, las sondas de la SEQ ID NO: 52 a la SEQ ID NO: 1229.

5 La presente invención también se refiere a un kit de detección de anomalías genéticas somáticas potencialmente presentes en una muestra de tejido potencialmente canceroso tomada de un sujeto, que comprende la combinación de sondas de captura tal como se define anteriormente.

10 Las diversas realizaciones de la presente invención se describen especialmente en la memoria descriptiva detallada a continuación. Estas realizaciones pueden considerarse por separado o combinarse entre sí.

Descripción detallada de la invención y de las realizaciones preferentes

15 La práctica de la presente invención empleará, a menos que se indique lo contrario, técnicas convencionales de biología molecular, microbiología y técnicas de ADN recombinante, que están dentro de los conocimientos de la técnica. Estas técnicas se explican detalladamente en la literatura. Véase, por ejemplo, Sambrook, Fritsch & Maniatis, *Molecular Cloning: A Laboratory Manual (Clonación molecular: Un manual de laboratorio)*, cuarta edición (2012); *Oligonucleotide Synthesis (Síntesis de oligonucleótidos)* (M. J. Gait, ed., 1984); *Nucleic Acid Hybridization (Hibridación de ácidos nucleicos)* (B. D. Hames y S. J. Higgins, eds., 1984); *A Practical Guide to Molecular Cloning (Una guía práctica para la clonación molecular)* (B. Perbal, 1984); y una serie, *Methods in Enzymology (Procedimientos para la enzimología)* (Academic Press, Inc.).

A. Definiciones

25 Los términos “comprender” (y cualquier variación gramatical de los mismos, tales como “comprende” y “comprendiendo”), “tener” (y cualquier variación gramatical de los mismos, tales como “tiene” y “teniendo”), “contener” (y cualquier variación gramatical de los mismos, como “contiene” y “conteniendo”), e “incluir” (y cualquier variación gramatical de los mismos, como “incluye” e “incluyendo”) son verbos de enlace abiertos. Se utilizan para especificar la presencia de características, números enteros, etapas o componentes o grupos de los mismos indicados, pero no excluyen la presencia o adición de una o más características, números enteros, etapas o componentes o grupos de los mismos. Como resultado, un procedimiento, o una etapa en un procedimiento, que “comprende”, “tiene”, “contiene” o “incluye” una o más etapas o elementos posee tales una o más etapas o elementos, pero no está limitado a poseer sólo tales una o más etapas o elementos.

35 A menos que se indique lo contrario, todos los números o expresiones que se refieren a cantidades de ingredientes, intervalos, condiciones de reacción, etc. utilizados en la presente memoria descriptiva deben entenderse modificados en todos los casos por el término “aproximadamente”.

40 Además, a menos que se indique lo contrario, la indicación de un intervalo de valores “de X a Y” o “entre X a Y”, de acuerdo con la presente invención, significa que incluye los valores de X e Y.

45 Tal como se utilizan en la memoria descriptiva y las reivindicaciones, las formas singulares “un”, “uno”, “una”, “el” y “la” pueden incluir referencias en plural a menos que el contexto indique claramente lo contrario. Por ejemplo, el término “una célula” puede incluir una pluralidad de células, incluidas mezclas de las mismas.

50 El cáncer es la clase de enfermedad caracterizada por la proliferación rápida e incontrolada de células dentro de un tejido de un eucariota multitejido. Generalmente se piensa que los cánceres son enfermedades genéticas de las células somáticas, que surgen a través de mutaciones secuenciales que crean oncogenes e inactivan genes supresores de tumores.

55 El término “sujeto”, tal como se utiliza en la presente memoria descriptiva, generalmente se refiere a una entidad biológica que contiene materiales genéticos expresados. La entidad biológica puede ser una planta, un animal o un microorganismo, incluyendo, por ejemplo, bacterias, virus, hongos y protozoos. El sujeto puede ser tejidos, células y su progenie de una entidad biológica obtenida *in vivo* o cultivada *in vitro*. El sujeto puede ser un mamífero. El mamífero puede ser un ser humano. Se puede diagnosticar o sospechar que el ser humano tiene un alto riesgo de padecer una enfermedad. La enfermedad puede ser cáncer. Es posible que el ser humano no sea diagnosticado ni que haya sospecha de que tenga un alto riesgo de padecer una enfermedad.

60 Como se usa en la presente memoria descriptiva, una “muestra” o “muestra de ácido nucleico” puede referirse a cualquier sustancia que contenga o se presuma que contiene ácido nucleico. La muestra puede ser una muestra biológica obtenida de un sujeto. Los ácidos nucleicos pueden ser ARN, ADN, por ejemplo, ADN genómico, ADN mitocondrial, ADN viral, ADN sintético o ADNc con transcripción inversa a partir de ARN.

65 El término “ADN genómico” consiste de manera ventajosa en ADN del genoma que contiene todas las secuencias codificantes (exones) y no codificantes (intrones y otras).

Los ácidos nucleicos en una muestra de ácido nucleico generalmente sirven como plantillas para la extensión de un cebador hibridado. En algunas realizaciones, la muestra biológica es una muestra líquida. La muestra líquida puede ser sangre completa, plasma, suero, ascitis, líquido cefalorraquídeo, sudor, orina, lágrimas, saliva, muestra bucal, enjuague de cavidades o enjuague de órganos. La muestra líquida puede ser una muestra líquida esencialmente exenta de células (por ejemplo, plasma, suero, sudor, plasma, orina, sudor, lágrimas, saliva, esputo, líquido cefalorraquídeo). En otras realizaciones, la muestra biológica es una muestra biológica sólida, por ejemplo, heces o biopsia de tejido, por ejemplo, una biopsia de tumor. Una muestra también puede comprender constituyentes de cultivo celular *in vitro* (incluidos, entre otros, medio condicionado resultante del crecimiento de células en medio de cultivo celular, células recombinantes y componentes celulares). La muestra puede comprender una única célula, por ejemplo, una célula cancerosa, una célula tumoral circulante, una célula madre cancerosa y similares.

“Nucleótidos” y “nt” se usan indistintamente en la presente memoria descriptiva para referirse en general a moléculas biológicas que pueden formar ácidos nucleicos. Los nucleótidos pueden tener restos que contienen no sólo las conocidas bases púricas y pirimidínicas, sino también otras bases heterocíclicas que han sido modificadas. Tales modificaciones incluyen purinas o pirimidinas metiladas, purinas o pirimidinas aciladas, ribosas alquiladas u otros heterociclos. Además, el término “nucleótido” incluye aquellos restos que contienen hapteno, biotina o etiquetas fluorescentes y pueden contener no sólo azúcares ribosa y desoxirribosa convencionales, sino también otros azúcares.

Los nucleósidos o nucleótidos modificados también incluyen modificaciones en el resto de azúcar, por ejemplo, en las que uno o más de los grupos hidroxilo se reemplazan con átomos de halógeno o grupos alifáticos, se funcionalizan como éteres, aminas o similares. Los nucleósidos o nucleótidos modificados también pueden incluir ácido peptidonucleico (PNA). El ácido peptidonucleico generalmente se refiere a oligonucleótidos en los que la cadena principal de desoxirribosa ha sido reemplazada por una cadena principal que tiene enlaces peptídicos. Cada subunidad generalmente tiene adherida una base de origen natural o no natural. Una estructura fundamental de PNA ejemplar está construida a partir de unidades de repetición de N-(2-aminoetil)glicina unidas a través de enlaces amida. El PNA puede unirse tanto al ADN como al ARN para formar dúplex de PNA/ADN o PNA/ARN. Los dúplex de PNA/ADN o PNA/ARN resultantes se pueden unir con mayor afinidad que los dúplex de ADN/ADN o ADN/ARN correspondientes, como lo demuestra sus temperaturas de fusión más altas (T_m). La estructura fundamental neutra del PNA también puede hacer que la T_m de los dúplex de PNA/ADN (ARN) sea en gran medida independiente de la concentración de sal en una mezcla de reacción. Por lo tanto, el dúplex PNA/ADN puede ofrecer una ventaja sobre las interacciones dúplex de ADN/ADN que dependen en gran medida de la fuerza iónica. Las realizaciones ejemplares de PNA se describen en las Patentes de EE. UU. Nos. 7.223.833 y 5.539.083.

Los “nucleótidos” también pueden incluir nucleótidos que comprenden una base potenciadora de la T_m (por ejemplo, un nucleótido potenciador de la base T_m). Los nucleótidos base potenciadores de la T_m a modo de ejemplo incluyen, entre otros, nucleótidos con Superbases™, ácidos nucleicos bloqueados (LNA) o ácidos nucleicos puenteados (BNA). BNA y LNA generalmente se refieren a ribonucleótidos modificados en los que el resto de ribosa está modificado con un puente que conecta el oxígeno 2' y el carbono 4'. Generalmente, el puente “bloquea” la ribosa en la conformación endo 3' (Norte), que a menudo se encuentra en los dúplex en forma A. El término “ácido nucleico bloqueado” (LNA) generalmente se refiere a una clase de BNA, en la que el anillo de ribosa está “bloqueado” con un puente de metileno que conecta el átomo 2'-O con el átomo 4'-C. Los nucleósidos de LNA que contienen las seis nucleobases comunes (T, C, G, A, U y mC) que aparecen en el ADN y el ARN pueden formar pares de bases con sus nucleósidos complementarios de acuerdo con las reglas estándar de emparejamiento de bases de Watson-Crick. Por consiguiente, los nucleótidos base potenciadores de la T_m tales como los nucleótidos de BNA y LNA se pueden mezclar con bases de ADN o ARN en un oligonucleótido cuando se desee. La conformación de ribosa bloqueada mejora el apilamiento de la base y la organización previa de la estructura fundamental. El apilamiento de bases y la organización previa de la estructura fundamental pueden dar lugar a una mayor estabilidad térmica (por ejemplo, mayor T_m) y al poder discriminativo de los dúplex. El LNA puede discriminar desajustes de bases individuales en condiciones que no son posibles con otros ácidos nucleicos. El ácido nucleico bloqueado se describe, por ejemplo, en el documento de patente WO 99/14226. Los nucleótidos también pueden incluir nucleótidos modificados como se describe en la Solicitud de Patente Europea No. EP1995330.

Otros nucleótidos modificados pueden incluir fosforamidita 5-Me-dC-CE, fosforamidita 5-Me-dC-CPG, fosforamidita 2-Amino-dA-CE, fosforamidita N4-Et-dC-CE, fosforamidita N4-Ac-N4-Et-dC-CE, fosforamidita N6-Me-dA-CE, fosforamidita N6-Ac-N6-Me-dA-CE, ácidos nucleicos Zip (ZNA®, descritos en la literatura), fosforamidita con tapa de 5'-trimetoxistilbeno, fosforamidita con tapa de 5'-pireno, CPG con tapa de 3'-Uaq (Glen Research).

Aún otros nucleótidos modificados pueden incluir nucleótidos con bases de nucleósidos modificados tales como, por ejemplo, 2-aminopurina, 2,6-diaminopurina, 5-bromo-desoxiuridina, desoxiuridina, dT invertida, ddT invertida, ddC, 5-metildesoxicitidina, desoxinosina, 5-Nitroindol, bases de 2'-O-metil ARN, hidroximetil dC, Iso-dG e Iso-dC (Eragen Biosciences, Inc), bases de 2'-fluro que tienen una ribosa modificada con flúor.

Los términos “polinucleótidos”, “ácido nucleico”, “nucleótidos” y “oligonucleótidos” se pueden usar indistintamente. Pueden referirse a una forma polimérica de nucleótidos de cualquier longitud, ya sean desoxirribonucleótidos o ribonucleótidos, o análogos de los mismos. Los polinucleótidos pueden tener cualquier estructura tridimensional y pueden realizar cualquier función, conocida o desconocida. Los siguientes son ejemplos no limitantes de polinucleótidos: regiones codificantes o no codificantes de un gen o fragmento de gen, loci (locus) [lugares {lugar}] definido a partir de análisis de ligamiento, exones, intrones, ARN mensajero (ARNm), ARN de transferencia, ARN ribosómico, ribozimas, ADNc, polinucleótidos recombinantes, polinucleótidos ramificados, plásmidos, vectores, ADN aislado de cualquier secuencia, ARN aislado de cualquier secuencia, sondas de ácido nucleico y cebadores.

Un polinucleótido puede comprender nucleótidos modificados, tales como nucleótidos metilados y análogos de nucleótidos. Si están presentes, se pueden impartir modificaciones a la estructura de nucleótidos antes o después del ensamblaje del polímero. La secuencia de nucleótidos puede verse interrumpida por componentes no nucleotídicos. Un polinucleótido puede modificarse adicionalmente después de la polimerización, tal como mediante conjugación con un componente de etiquetado.

El término “polinucleótido diana”, “región diana” o “diana”, como se usa en la presente memoria descriptiva, generalmente se refiere a un polinucleótido de interés en estudio. En ciertas realizaciones, un polinucleótido diana contiene una o más secuencias que son de interés y están en estudio. Un polinucleótido diana puede comprender, por ejemplo, una secuencia genómica. El polinucleótido diana puede comprender una secuencia diana cuya presencia, cantidad y/o secuencia de nucleótidos, o cambios en estas, se desea determinar.

El polinucleótido diana puede ser una región de un gen asociado con una enfermedad. En general, la región es un exón. En algunas realizaciones, el gen es una diana farmacológica. El término “diana farmacológica”, como se usa en la presente memoria descriptiva, generalmente se refiere a un gen o vía celular que está modulada por una terapia de enfermedad. La enfermedad puede ser cáncer. Por consiguiente, el gen puede ser un gen conocido relacionado con el cáncer. Especialmente, el gen relacionado con el cáncer se selecciona del grupo que consiste en ALK, BRAF, EGFR, EIF1AY, ERBB2, ERBB4, KDR, KIT, KRAS, MAP2K1, MET, NRAS, PDGFRA, SLC28A1, TP53.

El término “secuencia genómica”, como se usa en la presente memoria descriptiva, generalmente se refiere a una secuencia que ocurre en un genoma. Debido a que los ARN se transcriben a partir de un genoma, este término abarca secuencias que existen en el genoma nuclear de un organismo, así como secuencias que están presentes en una copia de ADNc de un ARN (por ejemplo, un ARNm) transcrito a partir de dicho genoma.

Los términos “recocer”, “hibridar” o “unir” pueden referirse a dos secuencias, segmentos o hebras de polinucleótidos, se pueden usar indistintamente y tienen el significado habitual en la técnica. Dos secuencias complementarias (por ejemplo, ADN y/o ARN) pueden hibridarse o hibridarse formando enlaces de hidrógeno con bases complementarias para producir un polinucleótido bicatenario o una región bicatenaria de un polinucleótido.

Como se usa en la presente memoria descriptiva, el término “complementario” generalmente se refiere a una relación entre dos secuencias de ácido nucleico antiparalelas en las que las secuencias están relacionadas mediante las reglas de emparejamiento de bases: A se empareja con T o U y C se empareja con G. Una primera secuencia o segmento que es “perfectamente complementario” a una segunda secuencia o segmento es complementario en toda su longitud y no tiene desajustes. Una primera secuencia o segmento es “sustancialmente complementario” a una segunda secuencia de segmento cuando un polinucleótido que consiste en la primera secuencia es suficientemente complementario para hibridar específicamente con un polinucleótido que consiste en la segunda secuencia.

El término “dúplex” o “duplexado”, como se usa en la presente memoria descriptiva, puede describir dos polinucleótidos complementarios que tienen pares de bases, por ejemplo, hibridados entre sí.

Como se utiliza en la presente memoria descriptiva, el término “Tm” generalmente se refiere a la temperatura de fusión de un dúplex de oligonucleótidos en la que la mitad de los dúplex permanece hibridada y la mitad de los dúplex se disocia en hebras individuales. Véase Sambrook y Russell (2001; Molecular Cloning: A Laboratory Manual, 3^a ed., Cold Spring Harbor Press, Cold Spring Harbor N.Y., cap. 10).

Como se usa en la presente memoria descriptiva, “amplificación” de una secuencia de ácido nucleico generalmente se refiere a técnicas *in vitro* para aumentar enzimáticamente el número de copias de una secuencia diana. Los procedimientos de amplificación incluyen tanto procedimientos asimétricos (en los que el producto predominante es monocatenario) como procedimientos convencionales (en los que el producto predominante es bicatenario). Una “ronda” o “ciclo” de amplificación puede referirse a un ciclo de PCR en el que una molécula de ADN de plantilla bicatenaria se desnaturaliza en plantillas monocatenarias, los cebadores directos e inversos se hibridan con las plantillas monocatenarias para formar dúplex cebador/plantilla, Los

cebadores se extienden mediante una ADN polimerasa a partir de los dúplex cebador/plantilla para formar productos de extensión. En rondas de amplificación posteriores, los productos de extensión se desnaturalizan en plantillas monocatenarias y se repite el ciclo.

5 Los términos “plantilla”, “hebra plantilla”, “ADN plantilla” y “ácido nucleico plantilla” se pueden usar indistintamente en la presente memoria descriptiva para referirse a una hebra de ADN que se copia mediante un ciclo de amplificación.

10 El término “desnaturalizante”, tal como se utiliza en la presente memoria descriptiva, se refiere generalmente a la separación de un dúplex de ácido nucleico en dos hebras sencillas.

15 El término “extender”, como se usa en la presente memoria descriptiva, generalmente se refiere a la extensión de un cebador hibridado con un ácido nucleico plantilla mediante la adición de nucleótidos usando una enzima, por ejemplo, una polimerasa.

20 Un “cebador” es generalmente una secuencia de nucleótidos (por ejemplo, un oligonucleótido), generalmente con un grupo 3'-OH libre, que se hibrida con una secuencia plantilla (tal como un polinucleótido diana, o un producto de extensión del cebador) y es capaz de promover polimerización de un polinucleótido complementario a la plantilla. Un cebador puede ser, por ejemplo, una secuencia de la plantilla (tal como un producto de extensión del cebador o un fragmento de la plantilla creado después de la escisión de Rase de un complejo plantilla-ADN) que se hibrida con una secuencia en la propia plantilla (por ejemplo, como un bucle en horquilla), y que es capaz de promover la polimerización de nucleótidos. Por lo tanto, un cebador puede ser un cebador exógeno (por ejemplo, añadido) o un cebador endógeno (por ejemplo, fragmento de plantilla).

25 Los términos “determinar”, “medir”, “evaluar”, “valorar”, “ensayar” y “analizar” se pueden usar indistintamente en la presente memoria descriptiva para referirse a cualquier forma de medición, e incluyen determinar si un elemento está presente o no. Estos términos pueden incluir determinaciones tanto cuantitativas como cualitativas. La evaluación puede ser relativa o absoluta. “Evaluar la presencia de” puede incluir determinar la cantidad de algo presente, así como determinar si está presente o ausente.

30 El término “libre en solución”, como se usa en este caso, puede describir una molécula, tal como un polinucleótido, que no está unida ni unida a un soporte sólido.

35 El término “fragmento genómico”, como se usa en la presente memoria descriptiva, puede referirse a una región de un genoma, por ejemplo, un genoma animal o vegetal tal como el genoma de un ser humano, mono, rata, pez o insecto o planta. Un fragmento genómico puede estar o no ligado con un adaptador. Un fragmento genómico puede estar ligado con un adaptador (en cuyo caso tiene un adaptador ligado a uno o ambos extremos del fragmento, al menos al extremo 5' de una molécula), o puede estar ligado sin un adaptador.

40 “Preamplificación”, como se usa en la presente memoria descriptiva, generalmente se refiere a amplificación no clonal de ácidos nucleicos. Por ejemplo, la preamplificación de una biblioteca de ácidos nucleicos generalmente se realiza antes de la amplificación clonal de la biblioteca y/o la carga en un secuenciador.

45 El término “ligasa”, como se usa en la presente memoria descriptiva, generalmente se refiere a una enzima que se usa comúnmente para unir polinucleótidos entre sí o para unir los extremos de un único polinucleótido.

50 El término “ligación”, como se usa en la presente memoria descriptiva, generalmente se refiere a la unión de dos extremos de polinucleótidos o la unión de extremos de un único polinucleótido mediante la formación de un enlace covalente entre los extremos a unir. El enlace covalente puede ser un enlace fosfodiéster.

El término “ligación dependiente de ATP”, como se usa en la presente memoria descriptiva, generalmente se refiere a la ligación mediante una ligasa dependiente de ATP. En la presente memoria descriptiva se describe un mecanismo ejemplar de ligación dependiente de ATP.

55 Las especies de ácidos nucleicos “donantes” y “aceptores” generalmente se refieren a dos poblaciones distintas de moléculas de ácido nucleico que se unirán en una reacción de ligación. La especie “donante” generalmente se refiere a una población de moléculas de ácido nucleico que pueden aceptar un nucleósido monofosfato (NMP) en cualquier extremo 5' o 3'. La especie “aceptora” generalmente se refiere a una segunda población de moléculas de ácido nucleico que contienen un grupo OH 3' o 5' que puede ligarse a la especie “donante” a través del NMP en cualquier extremo 5' o 3' de la especie donante.

60 Las especies donantes yceptoras pueden ser cualquier especie de ácido nucleico. Estas pueden ser, por ejemplo, polinucleótidos aislados de una fuente biológica. La fuente biológica puede ser un sujeto. En la presente memoria descriptiva se describen sujetos y fuentes biológicas ejemplares. Pueden ser oligonucleótidos. Los procedimientos para preparar oligonucleótidos de secuencia específica son conocidos en la técnica e incluyen, por ejemplo, clonación y restricción de secuencias apropiadas y síntesis química directa. Los procedimientos de síntesis química pueden incluir, por ejemplo, el procedimiento de fosfotriéster descrito por Narang et al, 1979,

Methods in Enzymology 68:90, el procedimiento de fosfodiéster descrito por Brown et al, 1979, Methods in Enzymology 68: 109, el procedimiento de dietilfosforamidato divulgado en Beaucage et al, 1981, Tetrahedron Letters 22: 1859, y el procedimiento de soporte sólido divulgado en la Patente de EE. UU. No. 4.458.066. Pueden ser ARN o ADN. El ADN puede ser ADN parcial o totalmente desnaturalizado. El ADN puede ser ADN monocatenario (ss). Parcialmente desnaturalizado puede estar "deshilachado" en los extremos de manera que un extremo "deshilachado" puede comprender 1, 2, 3, 4, 5 o más de 5 nucleótidos no hibridados.

La especie de ácido nucleico donante y/o aceptor puede ser de cualquier tamaño, variando desde, por ejemplo, 1-50 nt, 10-100 nt, 50-200 nt, 100-400 nt, 200-600 nt, 500-1000 nt, 800-2000 nt, o más de 2000 nt. En algunas realizaciones, la especie de ácido nucleico donante y/o aceptor tiene más de 120 nt de longitud.

Las especies de ácido nucleico donante o aceptor pueden incluir, por ejemplo, ácidos nucleicos genómicos, secuencias adaptadoras y/o secuencias de códigos de barras. Las especies de ácido nucleico donante o aceptor pueden incluir oligonucleótidos. La especie de ácido nucleico donante o aceptor puede comprender una etiqueta detectable o una etiqueta de afinidad.

La etiqueta detectable puede ser cualquier molécula que permita detectar una molécula. Los ejemplos no limitantes de etiquetas detectables incluyen, por ejemplo, quelantes, agentes fotoactivos, restos radiactivos (por ejemplo, emisores alfa, beta y gamma), agentes fluorescentes, agentes luminiscentes, iones paramagnéticos o enzimas que producen una señal detectable en presencia de ciertos reactivos (p. ej., peroxidasa de rábano picante, fosfatasa alcalina, glucosa oxidasa).

Los compuestos fluorescentes ejemplares incluyen, por ejemplo, isotiocianato de fluoresceína, rodamina, ficoeriterina, ficocianina, alofocianina, o-ftaldehído, fluorescamina y fluoróforos disponibles comercialmente tales como Alexa Fluor 350, Alexa Fluor 488, Alexa Fluor 532, Alexa Fluor 546, Alexa Fluor 568, Alexa Fluor 594, Alexa Fluor 647, tintes DyLight como DyLight 488, DyLight 594, DyLight 647 y tintes BODIPY como BODIPY 493/503, BODIPY FL, BODIPY R6G, BODIPY 530/550, BODIPY TMR, BODIPY 558/568, BODIPY 558/568, BODIPY 564/570, BODIPY 576/589, BODIPY 581/591, BODIPY TR, BODIPY 630/650, BODIPY 650/665, Azul Cascada, Amarillo Cascada, Dansyl, lisamina rodamina B, Azul Marina, Verde Oregón 488, Verde Oregón 514, Azul Pacífico, rodamina 6G, verde rodamina, rojo rodamina, tetrametilrodamina y Rojo Texas. Dichos compuestos están disponibles comercialmente (véase, por ejemplo, Molecular Probes, Inc.).

La etiqueta de afinidad se puede seleccionar para que tenga afinidad con un resto de captura. La etiqueta de afinidad puede comprender, sólo a modo de ejemplo no limitante, biotina, destiobiotina, histidina, polihistidina, myc, hemaglutinina (HA), FLAG, una etiqueta fluorescente, una etiqueta de purificación por afinidad en tándem (TAP), una glutatión S transferasa (GST), o derivados de la misma. El resto de captura puede comprender, por ejemplo, avidina, estreptavidina, neutravidina™, níquel o glutatión u otra molécula capaz de unirse a la etiqueta de afinidad.

En algunas realizaciones, la especie aceptora y la especie donante pueden ser la misma especie. Por ejemplo, en algunas realizaciones un usuario puede desear circularizar un ácido nucleico lineal o formar concatémeros de una única especie de ácido nucleico.

El término "mezcla de reacción" como se usa en la presente memoria descriptiva generalmente se refiere a una mezcla de componentes necesarios para efectuar una reacción deseada. La mezcla puede comprender además un tampón (por ejemplo, un tampón Tris). La mezcla de reacción puede comprender además una sal monovalente. La mezcla de reacción puede comprender además un catión, por ejemplo, Mg y/o Mn. La concentración de cada componente es bien conocida en la técnica y puede optimizarse aún más por un experto habitual en la técnica. En algunas realizaciones, la mezcla de reacción también comprende aditivos que incluyen, entre otros, ácidos nucleicos de fondo/bloqueantes no específicos (por ejemplo, ADN de esperma de salmón), proteínas de fondo/bloqueantes no específicas (por ejemplo, albúmina de suero bovino, leche en polvo), bioconservantes (p. ej., azida sódica), potenciadores de la PCR (p. ej., betaina, trehalosa, etc.) e inhibidores (p. ej., inhibidores de la RNasa). En algunas realizaciones, se mezcla una muestra de ácido nucleico con la mezcla de reacción.

Un "sitio de unión de cebador" puede referirse a un sitio con el que se hibrida un cebador en un oligonucleótido o una cadena complementaria del mismo.

El término "separar", tal como se utiliza en la presente memoria descriptiva, puede referirse a la separación física de dos elementos (por ejemplo, por tamaño, afinidad, degradación de un elemento, etc.).

El término "secuenciación", como se usa en la presente memoria descriptiva, puede referirse a un procedimiento mediante el cual se obtiene la identidad de al menos 10 nucleótidos consecutivos (por ejemplo, la identidad de al menos 20, al menos 50, al menos 100, al menos 200 o al menos 500 o más nucleótidos consecutivos) de un polinucleótido.

El término "ligado a adaptador", como se usa en la presente memoria descriptiva, puede referirse a un ácido nucleico que ha sido ligado a un adaptador. El adaptador puede ligarse a un extremo 5' o un extremo 3' de una molécula de ácido nucleico, o puede agregarse a una región interna de una molécula de ácido nucleico.

5 El término "PCR puente" puede referirse a una reacción en cadena de la polimerasa en fase sólida en la que los cebadores que se extienden en la reacción están unidos a un sustrato por sus extremos 5'. Durante la amplificación, los amplicones forman un puente entre los cebadores unidos. La PCR en Puente (que también puede denominarse "PCR en grupo") se utiliza en la plataforma Solexa de Illumina. La PCR en Puente y la
10 plataforma Solexa de Illumina se describen generalmente en una variedad de publicaciones, por ejemplo, Gudmundsson et al (Nat. Genet. 2009 41: 1122-6), Out et al (Hum. Mutat. 2009 30: 1703-12) y Turner. (Nat. Methods 2009 6:315-6), Patente de EE. UU. No. 7.115.400, y las publicaciones de Solicitudes de Patente Nos. US20080160580 y US20080286795.

15 El término "secuencia de código de barras", tal como se utiliza en la presente memoria descriptiva, generalmente se refiere a una secuencia única de nucleótidos que puede codificar información sobre un ensayo. Una secuencia de código de barras puede codificar información relacionada con la identidad de un alelo interrogado, la identidad de un polinucleótido diana o lugar/locus genómico, la identidad de una muestra, un sujeto o cualquier combinación de los mismos. Una secuencia de código de barras puede ser una porción de un cebador, una sonda informadora o ambas. Una secuencia de código de barras puede estar en el extremo 5' o 3' de un
20 oligonucleótido, o puede estar ubicada en cualquier región del oligonucleótido. Una secuencia de código de barras puede o no ser parte de una secuencia de plantilla. Las secuencias de códigos de barras pueden variar ampliamente en tamaño y composición; las siguientes referencias proporcionan orientación para seleccionar conjuntos de secuencias de códigos de barras apropiados para realizaciones particulares: Brenner, Patente de EE. UU. No. 5.635.400; Brenner et al., Proc. Nacional. Acad. Sci., 97: 1665-1670 (2000); Shoemaker et al.,
25 Nature Genetics, 14: 450-456 (1996); Morris et al, Publicación de Patente Europea 0799897A1; Wallace, Patente de EE. UU. No. 5.981.179. Una secuencia de código de barras puede tener una longitud de aproximadamente 4 a 36 nucleótidos, aproximadamente de 6 a 30 nucleótidos o aproximadamente de 8 a 20 nucleótidos.

30 Los términos "anomalías genéticas somáticas" se relacionan con una alteración o mutación en el ADN que ocurre después de la concepción. Las mutaciones somáticas pueden ocurrir en cualquiera de las células del cuerpo excepto en las células germinales (esperma y óvulo) y, por lo tanto, no se transmiten a los niños. Estas alteraciones pueden provocar cáncer u otras enfermedades.

35 El término "mutación", tal como se utiliza en la presente memoria descriptiva, generalmente se refiere a un cambio de la secuencia de nucleótidos de un genoma. Las mutaciones pueden involucrar grandes secciones de ADN (p. ej., variación del número de copias). Las mutaciones pueden afectar a cromosomas completos (p. ej., aneuploidía). Las mutaciones pueden involucrar pequeñas secciones de ADN. Ejemplos de mutaciones que involucran pequeñas secciones de ADN incluyen, por ejemplo, mutaciones puntuales o polimorfismos de un solo nucleótido, polimorfismos de múltiples nucleótidos, inserciones (por ejemplo, inserción de uno o más nucleótidos
40 en un locus), cambios de múltiples nucleótidos, deleciones (por ejemplo, deleción de uno o más nucleótidos en un locus) e inversiones (por ejemplo, inversión de una secuencia de uno o más nucleótidos).

45 El término "locus", tal como se utiliza en la presente memoria descriptiva, puede referirse a un lugar o una ubicación de un gen, nucleótido o secuencia en un cromosoma. Un "alelo" de un locus, como se usa en la presente memoria descriptiva, puede referirse a una forma alternativa de un nucleótido o secuencia en el locus. Un "alelo de tipo salvaje" generalmente se refiere a un alelo que tiene la frecuencia más alta en una población de sujetos. Un alelo "de tipo salvaje" generalmente no está asociado con una enfermedad. Un "alelo mutante" generalmente se refiere a un alelo que tiene una frecuencia menor que un "alelo de tipo salvaje" en la población general y puede estar asociado con una enfermedad. Es posible que un "alelo mutante" no tenga por qué estar
50 asociado con una enfermedad. El término "alelo interrogado" generalmente se refiere al alelo que un ensayo está diseñado para detectar.

55 El término "polimorfismo de un solo nucleótido" o "SNP", como se usa en la presente memoria descriptiva, generalmente se refiere a un tipo de variación de secuencia genómica que resulta de una sustitución de un solo nucleótido dentro de una secuencia. "Alelos de SNP" o "alelos de un SNP" generalmente se refieren a formas alternativas del SNP en un locus particular. El término "alelo SNP interrogado" generalmente se refiere al alelo SNP para el que está diseñado un ensayo.

60 El término "variación del número de copias" o "CNV" se refiere a diferencias en el número de copias de la información genética. En muchos aspectos se refiere a diferencias en el número de copias por genoma de una región genómica. Por ejemplo, en un organismo diploide, el número de copias esperado para las regiones genómicas autosómicas es de 2 copias por genoma. Estas regiones genómicas deberían estar presentes en 2 copias por célula. Para una revisión reciente, véase Zhang et al. Annu. Rev. Genomics Hum. Genet. 2009. 10:451-81. La CNV es una fuente de diversidad genética en los seres humanos y puede asociarse con trastornos
65 y enfermedades complejos, por ejemplo, mediante la alteración de la dosis de genes, la alteración genética o la fusión genética. También pueden representar variantes polimórficas benignas. Las CNV pueden ser grandes, por

ejemplo, de más de 1 Mb, pero muchas son más pequeñas, por ejemplo, entre 100 bases y 1 Mb. Se han informado en humanos más de 38.000 CNV de más de 100 bases (y menos de 3 Mb). Junto con los SNP, estas CNV representan una cantidad significativa de variación fenotípica entre individuos. Además de tener impactos nocivos, por ejemplo, que causan enfermedades, también pueden resultar en variaciones ventajosas.

En ciertos casos, un oligonucleótido usado en el procedimiento descrito en la presente memoria descriptiva puede diseñarse usando una región genómica de referencia, es decir, una región genómica de secuencia de nucleótidos conocida, por ejemplo, una región cromosómica cuya secuencia está depositada en la base de datos Genbank del NCBI u otra base de datos, por ejemplo.

El término "genotipado", como se usa en la presente memoria descriptiva, generalmente se refiere a un proceso de determinación de diferencias en la composición genética (genotipo) de un individuo examinando la secuencia de ADN del individuo usando ensayos biológicos y comparándola con la secuencia de otro individuo o una secuencia de referencia.

Una "pluralidad" generalmente contiene al menos 2 miembros. En ciertos casos, una pluralidad puede tener al menos 10, al menos 100, al menos 1.000, al menos 10.000, al menos 100.000, al menos 1.000.000, al menos 10.000.000, al menos 100.000.000 o al menos 1.000.000.000 o más miembros.

El término "separar", como se usa en la presente memoria descriptiva, generalmente se refiere a la separación física de dos elementos (por ejemplo, mediante escisión, hidrólisis o degradación de uno de los dos elementos).

Los términos "etiqueta" y "resto detectable" se pueden usar indistintamente en la presente memoria descriptiva para referirse a cualquier átomo o molécula que pueda usarse para proporcionar una señal detectable y que pueda unirse a un ácido nucleico o proteína. Las etiquetas pueden proporcionar señales detectables por fluorescencia, radioactividad, colorimetría, gravimetría, difracción o absorción de rayos X, magnetismo, actividad enzimática y similares.

B. Procedimiento de detección

Aspectos de la invención se refieren a un procedimiento, una combinación de sondas de captura y kits que comprenden dicha combinación de sondas de captura que mejoran el seguimiento y tratamiento de un sujeto que padece una enfermedad, especialmente cáncer.

El cáncer puede ser un tumor, una leucemia tal como leucemia aguda, leucemia aguda de células T, leucemia linfocítica aguda, leucemia mielocítica aguda, leucemia mieloblástica, leucemia promielocítica, leucemia mielomonocítica, leucemia monocítica, eritroleucemia, leucemia crónica, leucemia mielocítica crónica (granulocítica) o leucemia linfocítica crónica, policitemia vera, linfomas tales como linfoma de Hodgkin, linfoma folicular o linfoma no Hodgkin, mieloma múltiple, macroglobulinemia de Waldenstrom, enfermedad de las cadenas pesadas, tumores sólidos, sarcomas, carcinomas como, por ejemplo, fibrosarcoma, mixosarcoma, liposarcoma, condrosarcoma, sarcoma osteogénico, linfangiosarcoma, mesotelioma, tumor de Ewing, leiomiomasarcoma, rhabdomiomasarcoma, carcinoma de colon, cáncer colorrectal, cáncer de páncreas, cáncer de mama, cáncer de ovario, cáncer de próstata, carcinoma de células escamosas, carcinoma de células basales, adenocarcinoma, carcinoma de glándulas sudoríparas, carcinoma de glándulas sebáceas, carcinoma papilar, adenocarcinomas papilares, cistadenocarcinoma, carcinoma medular, carcinoma broncogénico, carcinoma de células renales, hepatoma, carcinoma de vías biliares, coriocarcinoma, seminoma, carcinoma embrionario, tumor de Wilms, cáncer de cuello uterino, cáncer de útero, tumor testicular, carcinoma de pulmón, carcinoma de pulmón de células pequeñas, carcinoma de vejiga, carcinoma epitelial, glioma, craneofaringioma, ependimoma, pinealoma, hemangioblastoma, neuroma acústico, oligodendroglioma, meningioma, melanoma, neuroblastoma, retinoblastoma, cáncer de endometrio, cáncer de pulmón de células no pequeñas.

Se puede sospechar o saber que el sujeto alberga un tumor sólido, o puede ser un sujeto que previamente albergaba un tumor sólido.

Especialmente, la presente invención se refiere a un procedimiento de detección de anomalías genéticas somáticas potencialmente presentes en una muestra de tejido, tomada de un sujeto, en el que dicho procedimiento de detección comprende las siguientes etapas:

- (a) corte de un ADN genómico extraído de dicha muestra de tejido, para obtener ADN genómico cortado;
- (b) preparar una biblioteca mediante captura por hibridación a partir del ADN genómico cortado obtenido en la etapa (a), mediante la implementación de al menos 51 sondas de captura que tienen al menos 110 nucleótidos de longitud y al menos un 75% de homología de secuencia con respectivamente la SEQ ID NO: 1 a la SEQ ID NO: 51, pudiendo dichas sondas de captura hibridarse en fragmentos de las siguientes regiones diana: ALK, BRAF, EGFR, EIF1AY, ERBB2, ERBB4, KDR, KIT, KRAS, MAP2K1, MET, NRAS, PDGFRA, SLC28A1, TP53 ;
- (c) secuenciación del ADN genómico cortado de la biblioteca emitida en la etapa (b), y
- (d) analizar las secuencias de dicho ADN genómico cortado obtenido en la etapa (c), para detectar

anomalías genéticas somáticas en dicha muestra de tejido que pueden ser cancerosas.

En la presente memoria descriptiva se describen varias realizaciones de los procedimientos de secuenciación y ensayos para determinar el estado mutacional de una muestra.

5

Muestra

De acuerdo con la invención, la muestra de partida puede ser una muestra biológica obtenida de un sujeto. Las muestras biológicas pueden ser una muestra de tumor y células normales del sujeto. En realizaciones particulares, la muestra es una muestra biológica sólida, por ejemplo, una muestra tumoral.

10

En general, el ácido nucleico se aísla de la muestra tumoral y de las células normales usando cualquier procedimiento conocido en la técnica. El ácido nucleico es el ADN, y en particular el ADN genómico.

15

El ADN genómico de la muestra tumoral y de las células normales se puede utilizar para preparar una biblioteca de ADN genómico tumoral específica del sujeto y/o una biblioteca de ADN genómico normal. Las bibliotecas de ADN (y en particular la biblioteca de ADN genómico) se pueden usar para la secuenciación mediante una plataforma de secuenciación.

20

En algunas realizaciones, la muestra biológica sólida se procesa antes del ensayo basado en sonda. El procesamiento puede comprender la fijación en una solución de formalina, seguida de la inclusión en parafina (por ejemplo, es una muestra de FFPE). Alternativamente, el procesamiento puede comprender la congelación de la muestra antes de realizar el ensayo basado en sonda. En algunas realizaciones, la muestra no se fija ni se congela. La muestra no fijada y descongelada puede almacenarse, sólo a modo de ejemplo, en una solución de almacenamiento configurada para la conservación de ácido nucleico. En la presente memoria descriptiva se describen soluciones de almacenamiento ejemplares. En algunas realizaciones, los materiales distintos de ácidos nucleicos se pueden eliminar del material de partida usando tratamientos enzimáticos (por ejemplo, con una proteasa).

25

30

Opcionalmente, la muestra puede someterse a homogeneización, sonicación, prensa francesa, inmersión, congelación/descongelación, a lo que puede seguir una centrifugación. La centrifugación puede separar fracciones que contienen ácidos nucleicos de fracciones que no contienen ácidos nucleicos.

35

En algunas realizaciones, la muestra es una muestra biológica líquida. En la presente memoria descriptiva se describen muestras biológicas líquidas ejemplares. En algunas realizaciones, la muestra biológica líquida es una muestra de sangre (por ejemplo, sangre completa, plasma o suero). En algunas realizaciones, una muestra de sangre completa se somete a componentes acelulares (por ejemplo, plasma, suero) y componentes celulares mediante el uso de un reactivo de Ficoll descrito en detalle Fuss et al, Curr Protoc Immunol (2009) Capítulo 7: Unidad 7.1.

40

El ácido nucleico se puede aislar de la muestra biológica usando cualquier medio conocido en la técnica. Por ejemplo, el ácido nucleico se puede extraer de la muestra biológica usando técnicas de extracción líquida (por ejemplo, Trizol, DNazol). El ácido nucleico también se puede extraer utilizando kits disponibles comercialmente (p. ej., kit Qiagen DNeasy, kit QIAamp, kit Qiagen Midi, kit QIAprep spin, kits Promega Maxwell...).

45

Plataformas de secuenciación de próxima generación

El procedimiento de detección de la invención comprende una etapa de secuenciación de ADN genómico cortado contenido en una biblioteca obtenida mediante captura por hibridación con una combinación específica de sondas de captura.

50

La plataforma de secuenciación puede ser una plataforma de secuenciación de próxima generación (NGS). En algunas realizaciones, el procedimiento además comprende secuenciar las bibliotecas de ácidos nucleicos usando tecnología NGS. La tecnología NGS puede implicar la secuenciación de plantillas de ADN amplificadas clonalmente o moléculas de ADN individuales de forma masivamente paralela (por ejemplo, como se describe en Volkerding et al. Clin Chem 55:641-658 [2009]; Metzker M Nature Rev 1 1:31-46 [2010]). Además de la información de secuencia de alto rendimiento, NGS proporciona información cuantitativa digital, en el sentido de que cada secuencia leída es una "etiqueta de secuencia" contable que representa una plantilla de ADN clonal individual o una única molécula de ADN.

60

La plataforma de secuenciación de próxima generación puede ser una plataforma disponible comercialmente. Las plataformas disponibles comercialmente incluyen, por ejemplo, plataformas para secuenciación por síntesis, secuenciación de semiconductores de iones, pirosecuenciación, secuenciación con terminador de colorante reversible, secuenciación por ligación, secuenciación de una sola molécula, secuenciación por hibridación y secuenciación de nanoporos. Las plataformas para secuenciación por síntesis están disponibles, por ejemplo, en Illumina, 454 Life Sciences, Helicos Biosciences y Qiagen. Las plataformas de Illumina pueden incluir, por

65

ejemplo, la plataforma Solexa de Illumina, el analizador de genoma de Illumina, MiSeq, NextSeq500, Novaseq, HiSeq y se describen en Gudmundsson et al (Nat. Genet. 2009 41: 1122-6), Out et al (Hum. Mutat. 2009 30: 1703-12) y Turner (Nat. Methods 2009 6:315-6), publicaciones de Solicitudes de Patente de EE. UU. Nos. US20080160580 y US20080286795, Patentes Estadounidenses Nos. 6306597, 7115400 y 7232656. Las plataformas de 454 Life Sciences incluyen, por ejemplo, GS Flex y GS Junior, y se describen en la Patente de EE. UU. No. 7,323,305. Las plataformas de Helicos Biosciences incluyen la plataforma True Single Molecule Sequencing. Las plataformas para la secuenciación de semiconductores de iones incluyen, por ejemplo, la Máquina de Genoma Personal Ion Torrent (PGM) y se describen en la Patente de EE. UU. No. 7948015. Las plataformas para priosecuenciación incluyen el sistema GS Flex 454 y se describen en las Patentes de EE. UU. Nos. 7211390; 7244559; 7264929. Las plataformas y procedimientos para secuenciar mediante ligación incluyen, por ejemplo, la plataforma de secuenciación SOLiD y se describen en la Patente de EE. UU. No. 5750341. Las plataformas para la secuenciación de una sola molécula incluyen el sistema SMRT de Pacific Bioscience y la plataforma Helicos True Single Molecule Sequencing.

Por ejemplo, la tecnología de secuenciación de ADN puede utilizar la plataforma de secuenciación Ion Torrent, que combina la tecnología de semiconductores con una química de secuenciación para traducir directamente información codificada químicamente (A, C, G, T) en información digital (0, 1) en un chip semiconductor. Sin querer ceñirnos a ninguna teoría, cuando una polimerasa incorpora un nucleótido a una cadena de ADN, se libera un ion hidrógeno como subproducto. La plataforma Ion Torrent detecta la liberación del átomo de hidrógeno como un cambio de pH. Se puede utilizar un cambio detectado en el pH para indicar la incorporación de nucleótidos. La plataforma Ion Torrent comprende una serie de pocillos micromecanizados de alta densidad para realizar este proceso bioquímico de forma masivamente paralela. Cada pocillo contiene un miembro de biblioteca diferente, que puede amplificarse clonalmente. Debajo de los pocillos hay una capa sensible a los iones y debajo de ella un sensor de iones. La plataforma inunda secuencialmente la matriz con un nucleótido tras otro. Cuando se agrega un nucleótido, por ejemplo, un C, a una plantilla de ADN y luego se incorpora a una cadena de ADN, se liberará un ion hidrógeno. La carga de ese ion cambiará el pH de la solución, que puede identificarse mediante el sensor de iones de Ion Torrent. Si no se incorpora el nucleótido, no se registrará ningún cambio de tensión y no se llamará ninguna base. Si hay dos bases idénticas en una cadena de ADN, la tensión será el doble y el chip registrará dos bases idénticas llamadas. La identificación directa permite registrar la incorporación de nucleótidos en segundos. La preparación de la biblioteca para la plataforma Ion Torrent generalmente implica la ligación de dos adaptadores distintos en ambos extremos de un fragmento de ADN.

La tecnología de secuenciación de ADN utiliza una plataforma de secuenciación Illumina, que generalmente emplea amplificación en grupo de miembros de la biblioteca en una célula de flujo y un enfoque de secuenciación por síntesis. Los miembros de la biblioteca amplificados en grupo se someten a ciclos repetidos de extensión de base única dirigida por polimerasa. La extensión de una sola base puede implicar la incorporación de dNTP de terminación reversible, cada dNTP marcado con un fluoróforo removible diferente. Los dNTP de terminación reversible generalmente están modificados en 3' para evitar una mayor extensión por parte de la polimerasa. Después de la incorporación, el nucleótido incorporado puede identificarse mediante imágenes de fluorescencia. Después de la obtención de imágenes de fluorescencia, se puede eliminar el fluoróforo y se puede eliminar la modificación 3', lo que da como resultado un grupo hidroxilo 3', permitiendo así otro ciclo de extensión de una sola base. La preparación de la biblioteca para la plataforma Illumina generalmente implica la ligación de dos adaptadores distintos en ambos extremos de un fragmento de ADN.

Por lo tanto, la presente invención presenta la ventaja de poder utilizarse con los NGS disponibles en el mercado. Estas plataformas de secuenciación de próxima generación son conocidas por los expertos en la técnica y no se describirán más en la presente memoria descriptiva.

Las diferentes etapas del presente procedimiento de detección se describirán a continuación.

Etapas (a): Corte del ADN genómico

Corte/Fragmentación (i):

De acuerdo con un aspecto de la invención, el ADN genómico extraído (ADN iniciado) de la muestra biológica se corta/fragmenta/se dimensiona hasta las secuencias diana que tienen una longitud deseada.

El ADN genómico extraído se puede cortar mediante cualquier medio conocido en la técnica, tal como, por ejemplo, mediante corte mecánico, nebulización o sonicación.

En particular, hay tres enfoques disponibles para cortar/fragmentar cadenas de ácidos nucleicos: físico, enzimático y químico. La fragmentación del ADN generalmente se realiza mediante procedimientos físicos (es decir, corte acústico y sonicación) o procedimientos enzimáticos (es decir, cócteles de endonucleasas no específicos y reacciones de tagmentación de transposasa).

Por ejemplo, el corte acústico con un instrumento Covaris (Covaris, Woburn, MA) usualmente se realiza para

obtener fragmentos de ADN en el intervalo de 100 a 5000 pb, mientras que los Covaris g-TUBE se emplean para el intervalo de 6 a 20 kbp necesario para bibliotecas de emparejamiento de pares. Los procedimientos enzimáticos incluyen la digestión mediante DNasa I o Fragmentasa, una mezcla de dos enzimas (New England Biolabs, Ipswich MA). Un procedimiento enzimático alternativo para fragmentar el ADN es la tecnología de etiquetación Nextera de Illumina (Illumina, San Diego, CA) en la que una enzima transposasa fragmenta e inserta simultáneamente secuencias adaptadoras en el ADNds. Este procedimiento tiene varias ventajas, incluida la reducción del tiempo de preparación y manipulación de muestras.

Etapas posteriores de la preparación previa a la biblioteca (ii) a (vi):

Con la excepción de la preparación de Nextera de Illumina, las etapas previas a la preparación de la biblioteca (b) generalmente implican: (i) corte/fragmentación mencionado anteriormente, (ii) reparación de extremos, (iii) fosforilación de los extremos primarios 5', (iv) prolongación-A de los extremos 3' para facilitar la ligación a adaptadores de secuenciación, (v) ligación de adaptadores, y (vi) cierto número de ciclos pre-PCR para enriquecer el producto que tiene adaptadores ligados a ambos extremos.

Por ejemplo, la pre-PCR (subetapa vi) se puede realizar utilizando las siguientes condiciones: 1 ciclo de 2 min a 98 °C, 10 ciclos de 30 s a 98 °C, seguido de 30 s a 60 °C y 1 min a 72 °C. El alargamiento final es de 5 min a 72 °C.

Las principales diferencias en un flujo de trabajo de Ion Torrent son el uso de ligación de extremos romos a diferentes secuencias de adaptadores.

En general, una vez que se ha cortado el ADN inicial, los extremos del fragmento se despuntan y se fosforilan en 5' usando una mezcla de tres enzimas: polinucleótido quinasa T4, ADN polimerasa T4 y fragmento grande de Klenow. A continuación, los extremos 3' tienen una prolongación-A utilizando Taq polimerasa o fragmento de Klenow (exo-). Taq es más eficiente en prolongación-A, pero Klenow (exo-) se puede usar para aplicaciones en las que no se desea calentar, como la preparación de bibliotecas de emparejamiento de pares. Durante la reacción de ligación del adaptador, la relación óptima entre adaptador y fragmento es -10:1, calculada en función del número de copias o la molaridad. Demasiado adaptador favorece la formación de dímeros adaptadores que pueden ser difíciles de separar y dominar en la amplificación por PCR posterior. Se pueden realizar limpiezas basadas en perlas o columnas después de la reparación del extremo y las reacciones de la prolongación-A, pero después de la ligación encontramos que las limpiezas a base de perlas son más efectivas para eliminar el exceso de dímeros adaptadores.

Para facilitar la multiplexación, se pueden utilizar diferentes adaptadores de códigos de barras con cada muestra. Alternativamente, se pueden introducir códigos de barras en la etapa de amplificación por PCR utilizando diferentes cebadores de PCR con códigos de barras para amplificar diferentes muestras. Muchos proveedores ofrecen fácilmente reactivos de alta calidad con adaptadores con código de barras y cebadores de PCR en kits.

Un procedimiento alternativo es el kit de preparación de muestras de ADN Nextera (Illumina), que prepara bibliotecas de ADN genómico mediante el uso de una enzima transposasa para fragmentar y etiquetar simultáneamente el ADN en una reacción de un solo tubo denominada "etiquetación". La enzima diseñada tiene doble actividad; fragmenta el ADN y simultáneamente agrega adaptadores específicos a ambos extremos de los fragmentos. Estas secuencias adaptadoras se utilizan para amplificar el ADN del inserto mediante PCR. La reacción de PCR también agrega secuencias de índice (código de barras). El procedimiento de preparación mejora los protocolos tradicionales al combinar la fragmentación del ADN, la reparación final y la ligación del adaptador en una sola etapa. Este protocolo es muy sensible a la cantidad de entrada de ADN en comparación con los procedimientos de fragmentación mecánica. Para obtener eventos de transposición separados por las distancias apropiadas, la proporción de complejos de transposasa con respecto a la muestra de ADN es fundamental. Debido a que el tamaño del fragmento también depende de la eficiencia de la reacción, todos los parámetros de la reacción, como la temperatura y el tiempo de reacción, son críticos y deben controlarse estrictamente.

Preferentemente, el ADN genómico cortado obtenido en la etapa (a) tiene una longitud mayor o igual que 300 pb, preferentemente que oscila entre 300 y 600 pb, e idealmente que oscila entre entre 300 y 350 pb.

Como se usa en la presente memoria descriptiva, una longitud mayor o igual a 300 pb incluye los siguientes valores o cualquier intervalo entre estos valores: 300; 310; 320; 330; 340; 350; 360; 370; 380; 390; 400; 410; 420; 430; 440; 450; 460; 470; 480; 490; 500; 510; 520; 530; 540; 550; 560; 570; 580; 590; 600; 610; 620; 630; etc.

Etapas (b) Preparación de la biblioteca

De acuerdo con una realización de la invención, la preparación de la biblioteca comprende:

(b1) la hibridación del ADN genómico cortado obtenido en la etapa (a) con un conjunto específico de

sondas de captura, para enriquecer el ADN genómico cortado contenido en la biblioteca, y
(b2) la amplificación de la biblioteca para aumentar el ADN genómico cortado contenido en la biblioteca.

5 Como se usa en la presente memoria descriptiva, el término "biblioteca" se usa en la presente memoria descriptiva y puede referirse a una pluralidad de ADN genómico cortado obtenido de una muestra biológica.

Etapa (b1) Hibridación del ADN genómico cortado obtenido en la etapa (a)

10 Esta etapa se realiza mediante el uso de tecnología de captura híbrida.

Especialmente, en la captura híbrida, las secuencias de ADN cortadas obtenidas en la etapa (a) se hibridan con la combinación de sondas de captura de acuerdo con la invención.

15 La combinación de sondas de captura comprende al menos 43 sondas de captura que tienen al menos 110 nucleótidos de longitud y al menos un 75% de homología de secuencia con la SEQ ID NO: 1 a la SEQ ID NO: 43, respectivamente, pudiendo dichas sondas de captura hibridarse en fragmentos de al menos al menos 5 regiones diana seleccionadas de ALK (NM_004304.4), BRAF (NM_004333.4), EGFR (NM_005228.3), EIF1AY (NM_004681.3), ERBB2 (NM_004448.2), ERBB4 (NM_005235.2), KDR, KIT (NM_000222.2), KRAS (NM_033360.2), MAP2K1 MAP2K1 (NM_002755.3), MET MET (NM_001127500.1), NRAS (NM_002524.4), PDGFRA PDGFRA (NM_006206.4), SLC28A1 (NM_004213.4), TP53 (NM_000546.5).

Cada una de las regiones diana se especifica en la presente memoria descriptiva mediante su símbolo genético de 3 letras y su secuencia de referencia NCBI.

25 En general, la combinación de sondas de captura además comprende al menos 4 sondas de captura adicionales que comprenden al menos 110 nucleótidos y al menos un 75% de homología de secuencia con, respectivamente, las sondas de la SEQ ID NO: 44 a la SEQ ID NO: 47 y/o las sondas de la SEQ ID NO: 48 a la SEQ ID NO: 51 o combinación de las mismas.

30 Preferentemente, la combinación de sondas de captura además comprende al menos 100 sondas de captura adicionales, preferentemente al menos 200, en particular al menos 300, e idealmente al menos 1178 sondas de captura adicionales que comprenden al menos 110 nucleótidos y al menos un 75% de homología de secuencia con, respectivamente, las sondas de la SEQ ID NO: 52 a la SEQ ID NO: 1229.

35 En particular, la preparación de la biblioteca en la etapa (b) comprende la implementación de:

- 43 sondas de captura que tienen al menos un 75%, especialmente un 80%, preferentemente un 85%, en particular un 90% e idealmente un 95% y posiblemente un 100% de homología de secuencia con las sondas de captura de la SEQ ID NO: 1 a la SEQ ID NO: 43, respectivamente y/o;
- 40 – sondas de captura adicionales que tienen al menos un 75%, especialmente un 80%, preferentemente un 85%, en particular un 90% e idealmente un 95% y posiblemente un 100% de homología de secuencia con las sondas de captura de la SEQ ID NO: 44 a la SEQ ID NO: 47 y/o;
- sondas de captura adicionales que tienen al menos un 75%, especialmente un 80%, preferentemente un 85%, en particular un 90% e idealmente un 95% y posiblemente un 100% de homología de secuencia con las sondas de captura con las sondas de captura de la SEQ ID NO: 48 a la SEQ ID NO: 51 y/o;
- 45 – sondas de captura adicionales que tienen al menos un 75%, especialmente un 80%, preferentemente un 85%, en particular un 90% e idealmente un 95% y posiblemente un 100% de homología de secuencia con las sondas de captura con las sondas de captura de la SEQ ID NO: 52 a la SEQ ID NO: 1229, respectivamente.

50 En general, la preparación de la biblioteca en la etapa (b) comprende la implementación de sondas de captura que comprenden de 110 a 150 nucleótidos de longitud, preferentemente de 120 a 140 nucleótidos de longitud e idealmente de 125 a 135 nucleótidos de longitud desde la SEQ ID NO: 1 hasta la SEQ ID NO: 1229.

55 De acuerdo con una realización de la invención, la combinación de sondas de captura de acuerdo con la invención está biotinilada. La biotina se une a perlas de estreptavidina y luego el ADN no unido se elimina por lavado. Esto tiene la ventaja de una detección más fiable de los cambios en el número de copias.

60 En general, la preparación de la biblioteca puede comprender cualquier procedimiento conocido en la técnica o como se describe en la presente memoria descriptiva que sea adecuado con la combinación de sondas de captura de acuerdo con la invención.

65 Preferentemente, la hibridación con dichas sondas de captura tiene lugar a una temperatura de al menos 55 °C, preferentemente al menos a 58 °C y preferentemente a una temperatura que varía de 60 °C a 70 °C o de 58 a 65 °C.

En general, el ADN genómico cortado capturado en la etapa (b) tiene una longitud mayor o igual que 300 pb,

preferentemente que oscila entre 300 y 600 pb, e idealmente que oscila entre 300 y 350 pb.

5 Después del corte y purificación del ADN con perlas magnéticas, el ADN genómico cortado se repara, por ejemplo, gracias a una enzima polimerasa específica a 20 °C durante 30 minutos. A continuación, la muestra de ADN puede purificarse con perlas magnéticas y adenilarse en el extremo 3' gracias a una enzima específica a 37 °C durante 30 minutos.

10 Posteriormente, una vez purificados los fragmentos de ADN genómico con perlas magnéticas, se pueden ligar con un oligo adaptador (específico de la tecnología de secuenciación elegida) gracias, por ejemplo, a una ADN ligasa a 20 °C durante 15 minutos. Luego, los fragmentos de ADN se purifican, por ejemplo, con perlas magnéticas y se amplifican mediante PCR con una polimerasa de muy alta fidelidad (como: 98 °C 2 min, 6 a 10 ciclos de 98 °C 30 s + 65 °C 30 s + 72° C 1 min y 72 °C 10 min).

15 Después de la purificación con perlas magnéticas, la totalidad del ADN genómico amplificado se puede hibridar con el cóctel de sondas (por ejemplo, 95 °C durante 5 minutos y luego 65 °C durante la noche).

20 Luego, los fragmentos de ADN genómico hibridados pueden capturarse gracias a perlas magnéticas recubiertas de estreptavidina. Finalmente, las bibliotecas capturadas pueden amplificarse con cebadores de indexación mediante PCR con una polimerasa de muy alta fidelidad (tales como: 98 °C 2 min, 10 a 16 ciclos de 98 °C 30 s + 57 °C 30 s + 72 °C 1 min y 72 °C 10 min). A continuación, la muestra de ADN se puede purificar con perlas magnéticas. La cantidad y calidad de las bibliotecas indexadas se evalúan mediante TapeStation, Qubit, Bioanalyzer o qPCR.

25 Etapa (b2) Amplificación de la biblioteca

La amplificación de la biblioteca se realiza de manera ventajosa mediante tecnología PCR.

30 Para facilitar la multiplexación, se pueden utilizar diferentes adaptadores de códigos de barras con cada muestra. Alternativamente, se pueden introducir códigos de barras en la etapa de amplificación por PCR utilizando diferentes cebadores de PCR con códigos de barras para amplificar diferentes muestras. Muchos proveedores ofrecen fácilmente reactivos de alta calidad con adaptadores con código de barras y cebadores de PCR en kits.

Etapa (c) Secuenciación

35 A continuación, se secuencia el ADN genómico cortado, contenido en la biblioteca obtenida en la etapa (b).

40 En general, las bibliotecas enriquecidas con dianas se secuencian utilizando cualquier procedimiento conocido en la técnica o como se describe en la presente memoria descriptiva en parte "Plataformas de secuenciación de próxima generación". La secuenciación puede revelar la presencia de mutaciones en uno o más genes relacionados con el cáncer del conjunto.

45 En algunas realizaciones, se selecciona un subconjunto de 5, 6, 7, 8, 9, 10, 15 genes que albergan las mutaciones para un seguimiento adicional mediante la evaluación del ADN libre de células en una muestra de fluido aislada del sujeto en puntos de tiempo posteriores.

En algunas realizaciones, se selecciona un subconjunto de no más de 5 genes que albergan las mutaciones para un seguimiento adicional mediante la evaluación del ADN libre de células en una muestra de líquido aislada del sujeto en puntos de tiempo posteriores.

50 Etapa (d) Análisis de las secuencias de dicho ADN genómico cortado obtenido en la etapa (c)

La etapa de análisis (d) tiene como objetivo principal detectar anomalías genéticas somáticas potencialmente presentes en dicha muestra de tejido canceroso, tomada del sujeto.

55 La longitud de la secuencia leída puede variar dependiendo de la tecnología de secuenciación particular utilizada. Las plataformas NGS pueden proporcionar lecturas de secuencias que varían en tamaño desde decenas hasta cientos o miles de pares de bases. En algunas realizaciones del procedimiento descrito en la presente memoria descriptiva, las lecturas de secuencia tienen una longitud de aproximadamente 20 bases, una longitud de aproximadamente 25 bases, una longitud de aproximadamente 30 bases, una longitud de aproximadamente 35 bases, una longitud de aproximadamente 40 bases, una longitud de aproximadamente 45 bases, una longitud de aproximadamente 50 bases, una longitud de aproximadamente 55 bases, aproximadamente 60 bases, una longitud de aproximadamente 65 bases, una longitud de aproximadamente 70 bases, una longitud de aproximadamente 75 bases, una longitud de aproximadamente 80 bases, una longitud de aproximadamente 85 bases, una longitud de aproximadamente 90 bases, una longitud de aproximadamente 95 bases, una longitud de aproximadamente 100 bases, una longitud de aproximadamente 110 bases, una longitud de aproximadamente 120 bases, una longitud de aproximadamente 130, una longitud de aproximadamente 140

5 bases, una longitud de aproximadamente 150 bases, una longitud de aproximadamente 200 bases, una longitud de aproximadamente 250 bases, una longitud de aproximadamente 300 bases, una longitud de aproximadamente 350 bases, una longitud de aproximadamente 400 bases, una longitud de aproximadamente 450 bases, una longitud de aproximadamente 500 bases, una longitud de aproximadamente 600 bases, una longitud de aproximadamente 700 bases, una longitud de aproximadamente 800 bases, una longitud de aproximadamente 900 bases, una longitud de aproximadamente 1000 bases o una longitud de más de 1000 bases.

10 Se puede realizar la secuenciación parcial de fragmentos de ADN presentes en la muestra y se pueden contar las etiquetas de secuencia que comprenden lecturas que se asignan a un genoma de referencia conocido. Sólo las lecturas de secuencia que se alinean de forma única con el genoma de referencia pueden contarse como etiquetas de secuencia. En una realización, el genoma de referencia es la secuencia del genoma de referencia humano NCBI38/hg19, que está disponible en la red mundial en genome.ucsc.edu/. Otras fuentes de información pública sobre secuencias incluyen GenBank, dbEST, dbSTS, EMBL (el Laboratorio Europeo de Biología Molecular) y DDBJ (el Banco de Datos de ADN de Japón). El genoma de referencia también puede comprender la secuencia del genoma de referencia humano NCBI38/hg19 y un genoma de secuencias diana artificiales, que incluye secuencias diana polimórficas. En otra realización adicional, el genoma de referencia es un genoma de secuencia diana artificial que comprende secuencias diana polimórficas.

20 El mapeo de las etiquetas de secuencia se puede lograr comparando la secuencia de la etiqueta con la secuencia del genoma de referencia para determinar el origen cromosómico de la molécula de ácido nucleico secuenciada (por ejemplo, ADN libre de células), y no se necesita información de secuencia genética específica. Se encuentran disponibles varios algoritmos informáticos para alinear secuencias, incluidos, entre otros, BLAST (Altschul et al., 1990), BLITZ (MPsrch) (Sturrock & Collins, 1993), FASTA (Person & Lipman, 1988), BOWTIE (Langmead et al., Genome Biology 10:R25.1-R25.10 [2009]), o ELAND (Illumina, Inc., San Diego, CA, EE. UU.). En una realización, un extremo de las copias expandidas clonalmente de la molécula de ADN se secuencia y procesa mediante análisis de alineación bioinformática para el analizador del genoma Illumina, que utiliza el software de alineación eficiente a gran escala de bases de datos de nucleótidos (ELAND). El software adicional incluye SAMtools {SAMtools, Bioinformatics, 2009, 25(16):2078-9} y el procedimiento de compresión de clasificación de bloques de Burroughs-Wheeler que implica la clasificación de bloques o el preprocesamiento para hacer la compresión más eficiente.

35 Las plataformas de secuenciación descritas en la presente memoria descriptiva generalmente comprenden un soporte sólido inmovilizado sobre oligonucleótidos unidos a la superficie que permiten la captura e inmovilización de miembros de la biblioteca de secuenciación en el soporte sólido. Los oligonucleótidos unidos a la superficie generalmente comprenden secuencias complementarias a las secuencias adaptadoras de la biblioteca de secuenciación.

40 El análisis de las secuencias, en la etapa (d), se puede definir mediante dos parámetros: profundidad y tasa de cobertura.

45 La profundidad es el número de lecturas que incluyen un nucleótido determinado en la secuencia reconstruida (en x). La secuenciación profunda se refiere al concepto general de apuntar a un gran número de lecturas replicadas de cada región de una secuencia.

La tasa de cobertura es la zona cubierta por al menos una lectura (en %).

50 Preferentemente, en el análisis de las secuencias (d), la profundidad promedio es de al menos 200x, preferentemente de 200 a 10000x, en particular de 1000x a 1500x, e idealmente de 1000x a 1200x.

De acuerdo con una realización, en el análisis de las secuencias (d), la tasa de cobertura es de al menos el 90% a 200x, preferentemente del 90% al 99,9% a 500x, en particular del 95% al 99,9% a 500x, e idealmente que oscila entre 98% y 99,9% a 500x.

55 Las secuencias de dicho ADN genómico cortado, obtenidas en la etapa (c), se analizan para detectar posibles anomalías genéticas somáticas en dicha muestra de tejido.

60 Este análisis se puede lograr mediante técnicas/canalizaciones bioinformáticas bien conocidas, para identificar anomalías genéticas somáticas (potencialmente tumorales/cancerosas) en dicho ADN genómico cortado de la biblioteca en comparación con anomalías genéticas somáticas conocidas o con genomas normales. Estas anomalías genéticas somáticas pueden consistir en mutaciones somáticas de base única (SSM), mutaciones somáticas de inserción/delección (SIM) y cambios estructurales más grandes (reordenamientos y cambios en el número de copias de segmentos cromosómicos).

65

Combinación de sondas

La presente invención se refiere también a una combinación de sondas de captura, como se especifica anteriormente, es decir, que comprende al menos 51 sondas de captura que comprenden al menos 110 nucleótidos de longitud y al menos un 75% de homología de secuencia con, respectivamente, las sondas de la SEQ ID NO: 1 a SEQ ID NO: 51.

5 Preferentemente, la combinación de sondas de captura además comprende al menos 100 sondas de captura adicionales, preferentemente al menos 200, en particular al menos 300, e idealmente al menos 1178 sondas de captura adicionales que comprenden al menos 110 nucleótidos y al menos un 75% de homología de secuencia con, respectivamente, las sondas de la SEQ ID NO: 52 a la SEQ ID NO: 1229.

10 Por supuesto, todas las características descritas anteriormente para el procedimiento de detección de acuerdo con la invención se incorporan en la presente memoria descriptiva para la descripción de la combinación de sondas.

15 Kit de detección

La presente invención también se refiere a un kit de detección de anomalías genéticas somáticas potencialmente presentes en una muestra de tejido potencialmente canceroso tomada de un sujeto, que comprende la combinación de sondas de captura tal como se define anteriormente.

20 Por supuesto, todas las características descritas anteriormente para el procedimiento de detección de acuerdo con la invención o para la combinación de sondas se incorporan en la presente memoria descriptiva para la descripción del kit de detección.

25 En una realización, el kit de acuerdo con la invención además comprende al menos: agua libre de nucleasa, tampón de reparación de extremos 10x, mezcla de dNTP, ADN polimerasa T4, ADN polimerasa Klenow, polinucleótido quinasa T4, tampón polimerasa Klenow 10x, dATP, Exo(-) Klenow, tampón de ADN ligasa T4 5x, Mezcla de Oligo Adaptador SureSelect, ADN ligasa T4, cebador SureSelect, cebador inverso de PCR de precaptura de indexación SureSelect Illumina, tampón de reacción Herculase II 5x, mezcla de dNTP 100 mM, ADN polimerasa de fusión Herculase II, SureSelect Hyb 1, SureSelect Hyb 2, SureSelect Hyb 3, SureSelect Hyb 4, Bloque de Indexación SureSelect 1, Bloque SureSelect 2, Bloque de Indexación SureSelect Illumina 3, solución de bloque de RNasa, biblioteca de captura, Dynabeads MyOne Estreptavidina T1, perlas Agencourt AMPure XP, placas de plástico de 96 pocillos, o combinación de las mismas.

35 En algunas realizaciones, los kits de la invención incluyen además un material de empaque. Tal como se utiliza en la presente memoria descriptiva, el término "material de empaque" puede referirse a una estructura física que alberga los componentes del kit. El material de empaque puede mantener la esterilidad de los componentes del kit y puede estar hecho de material comúnmente utilizado para dichos fines (por ejemplo, papel, fibra corrugada, vidrio, plástico, papel de aluminio, ampollas, etc.). Los kits también pueden incluir un agente tamponante, un conservante o un agente estabilizante de proteína/ácido nucleico.

Descripción de las figuras

45 La presente invención se ilustrará a continuación sin limitarse a la misma mediante los siguientes ejemplos. Se hará referencia a las siguientes figuras en los ejemplos:

Figura 1: La Figura 1A muestra el resultado obtenido con las condiciones de corte para el kit de detección de tumor FFPE usando las instrucciones del protocolo del fabricante y la Figura 1B muestra el resultado obtenido con las condiciones de corte para el kit de detección de ADN sanguíneo usando las instrucciones del protocolo del fabricante;

Figura 2: La Figura 2A muestra el resultado obtenido con las condiciones de corte para el kit de detección de tumor FFPE de acuerdo con la invención utilizando el protocolo descrito en el párrafo 1.c siguiente y la Figura 2B muestra el resultado obtenido con las condiciones de corte para el kit de detección a partir de ADN sanguíneo de acuerdo con la invención utilizando el protocolo descrito en el párrafo 1.c siguiente.

Ejemplos

1- Material y procedimiento

60 a- Muestras

Para este ensayo se analizaron 100 muestras (tumores, plasmas y muestras comerciales). Especialmente, las muestras B15 a B17 se obtuvieron de archivos de nuestro laboratorio, las muestras CHA, VIA, LOB, RIG, LEM, GAR, TOR, PEN, LUC, LED, POQ, 17BM00026 se obtuvieron de muestras de plasma obtenidas en nuestro hospital con tubos STRECK, y la muestra comercial (V600E, G719S y G13D) se obtuvo de Horizon Discovery.

El ADN se extrajo de tumores incluidos en parafina y fijados con formalina. Un patólogo caracterizó los tumores para determinar el contenido de células tumorales y los envió a la plataforma de biología molecular para la extracción de ADN. Se extrajeron siete cortes de tumor de 15 µm utilizando el kit de purificación de ADN Maxwell 16 FFPE Plus LEV (Promega, Madison, EE. UU.) de acuerdo con el protocolo del fabricante evaluado mediante espectrofotometría con absorbancia a 230, 260 y 280 nm. El ADN se cuantificó mediante un ensayo fluorimétrico con un dispositivo Qubit.

Para las muestras de plasma, se extrajeron 4 ml de plasma obtenido de tubos STRECK con el kit de ácido nucleico circulante QIAamp (Qiagen, Hilden, Alemania), siguiendo el protocolo del fabricante.

b- Materiales/Equipos

Para el corte del ADN, se diluyeron aproximadamente 200 ng de ADN en tampón TE y se sonicaron gracias a un sonicador Covaris S220 (Covaris, Woburn, Mass, EE. UU.). Las bibliotecas se realizaron con un kit SureSelect XT personalizado (Agilent). Las bibliotecas de captura previa se concentran con un concentrador de vacío Plus (Eppendorf, Hamburgo, Alemania). Para la captura también se utilizan un dispositivo MixMate (Eppendorf) y un baño de agua.

La calidad de las bibliotecas se evaluó gracias a un sistema TapeStation (Agilent, Santa Clara, CA, EE. UU.). Las bibliotecas se secuenciaron en un sistema MiSeq (Illumina, San Diego, CA, EE. UU.) o un sistema NextSeq500 (Illumina).

Los reactivos utilizados para la preparación de la biblioteca fueron agua libre de nucleasa, tampón de reparación de extremos 10x, mezcla de dNTP, ADN polimerasa T4, ADN polimerasa Klenow, polinucleótido quinasa T4, tampón polimerasa Klenow 10x, dATP, Exo(-) Klenow, tampón de ADN ligasa T4 5x, Mezcla de Oligo Adaptador SureSelect, ADN Ligasa T4, Cebador SureSelect, cebador inverso de PCR de precaptura de indexación SureSelect Illumina, tampón de reacción Herculanase II 5x, mezcla de dNTP 100 mM, ADN Polimerasa Herculanase II Fusion, SureSelect Hyb 1, SureSelect Hyb 2, SureSelect Hyb 3, SureSelect Hyb 4, Bloque de Indexación SureSelect 1, Bloque de Indexación SureSelect 2, Bloque de Indexación SureSelect Illumina 3, solución de bloque de RNasa. Todos estos reactivos se compraron en el kit SureSelectXT. El kit Custom Capture Library es el objeto de la presente invención. Para la preparación también se utilizaron Dynabeads MyOne Streptavidin T1, perlas Agencourt AMPure XP y placas de plástico de 96 pocillos.

Sondas ensayadas de acuerdo con la invención	SEQ ID NO: 1 - 1229
Sondas de acuerdo con la técnica anterior (ejemplos comparativos)	La caracterización previa de las muestras se realizó mediante un kit de discriminación alélica personalizado dedicado (Applied Biosystems, Foster City, CA, EE. UU.), análisis de fragmentos y secuenciación de Sanger (Applied Biosystems), kit TruSeq Custom Amplicon (Illumina), Tumor Hotspot MASTR Plus (Multiplicom, Niel, Bélgica) o SureSelect XT Human All exon v5 (Agilent)

c- Ensayo de amplificación por captura híbrida

Para estos ensayos, el protocolo utilizado se describe a continuación. Todos los volúmenes indicados son para 1 muestra (pocillo). Estos volúmenes deben ajustarse con el número de muestras manipuladas al mismo tiempo.

✓ Evaluación del rendimiento de ADNg (ADN genómico)

Cuantificar el ADNg recién extraído antes de la preparación de la biblioteca. Utilice el procedimiento de evaluación del rendimiento fluorométrico para cuantificar el ADNg intacto (evite utilizar un espectrofotómetro para la evaluación del rendimiento del ADNg). Mida la proporción (260/280 nm) y asegúrese de que esté entre 1,8 y 2. NOTA: se requieren alrededor de 200 ng de ADNg para el experimento, pero se puede usar una cantidad menor (por ejemplo, para ADN extraído del plasma)

✓ Fragmentación del ADNg

Diluir aproximadamente 200 ng de ADN en un volumen final de 50 µl de tampón TE. A continuación, sonique en el dispositivo Covaris con el siguiente programa: Factor de trabajo = 10%, potencia incidente máxima = 175, ciclos por ráfaga = 200, temperatura del baño = 4 °C a 8 °C, tiempo de tratamiento = 120 segundos (para muestras FFPE) o 180 segundos (para muestras de tumores frescos) para obtener un fragmento de ADN de 300 a 350 pb.

Es posible comprobar que tus fragmentos de ADN tienen de 300 a 350 pb gracias a un Bioanizador (Agilent) o una TapeStation (Agilent).

ES 2 972 684 T3

✓ Reparación de extremos

5 Mezcle bien 35,2 µl de agua libre de nucleasa, 10 µl de tampón de reparación de extremos 10x, 1,6 µl de dNTP, 1 µl de ADN polimerasa T4, 2 µl de ADN polimerasa Klenow, polinucleótido quinasa T4 y 50 µl de ADN cortado. Incubar 30 min a 20 °C y luego mantener a +4 °C.

✓ Purificación magnética

10 Una vez que las perlas magnéticas (AMPure XP) alcancen la temperatura ambiente durante 30 minutos, agregue 180 µl de perlas homogéneas para finalizar las muestras de ADN reparadas. Mezclar bien e incubar durante 5 min a temperatura ambiente.

15 Luego, coloque la placa en un soporte magnético durante 5 minutos y deseche con cuidado el sobrenadante sin mover la placa del soporte.

Aún sin mover la placa del soporte, lave las perlas dos veces con 200 µl de etanol al 70% recién preparado durante 1 min.

20 Seque completamente las perlas (posiblemente a 37 ° C durante 3 a 5 minutos) y agregue 32 µl de agua libre de nucleasa a cada pocillo de muestra. Después de sellar y mezclar, incubar 2 min a temperatura ambiente y volver a colocar la placa en el soporte magnético durante 3 min. Transfiera el sobrenadante (30 µl) que contiene el ADN purificado a un pocillo de placa nuevo.

25 El ADN purificado se puede almacenar a -20 °C.

✓ 3' Adenilación

30 Mezcle bien 11 µl de agua libre de nucleasa, 10 µl de tampón de polimerasa Klenow 10x, 1 µl de dATP, 3 µl de Exo(-) Klenow y 30 µl de ADN reparado en el extremo purificado. Incubar 30 min a 37 °C y luego mantener a +4 °C. (Si usa termociclador, no use tapa calentada).

✓ Purificación magnética

35 Una vez que las perlas magnéticas (AMPure XP) alcancen la temperatura ambiente durante 30 minutos, agregue 90 µl de perlas homogéneas para finalizar las muestras de ADN reparadas. Mezclar bien e incubar durante 5 min a temperatura ambiente.

40 Luego, coloque la placa en un soporte magnético durante 5 minutos y deseche con cuidado el sobrenadante sin mover la placa del soporte.

Aún sin mover la placa del soporte, lave las perlas dos veces con 200 µl de etanol al 70% recién preparado durante 1 min.

45 Seque completamente las perlas (posiblemente a 37 ° C durante 3 a 5 minutos) y agregue 15 µl de agua libre de nucleasa a cada pocillo de muestra. Después de sellar y mezclar, incubar 2 min a temperatura ambiente y volver a colocar la placa en el soporte magnético durante 3 min.

50 Transfiera el sobrenadante (13 µl) que contiene el ADN purificado a un pocillo de placa nuevo y proceda inmediatamente a la ligación del adaptador de extremos emparejados.

✓ Ligación del adaptador de extremos emparejados

55 Mezcle bien 15,5 µl de agua libre de nucleasa, 10 µl de tampón de ADN ligasa T4 5x, 10 µl de oligo adaptador SureSelect 1/10, 1,5 µl de ADN ligasa T4 y 13 µl de ADN adenilado 3' purificado. Incubar 15 min a 20 °C y luego mantener a +4 °C (si usa termociclador, no use tapa calentada).

La placa se puede almacenar a -20 °C.

60 ✓ Purificación magnética

Una vez que las perlas magnéticas (AMPure XP) alcancen la temperatura ambiente durante 30 minutos, agregue 90 µl de perlas homogéneas para finalizar las muestras de ADN reparadas. Mezclar bien e incubar durante 5 min a temperatura ambiente.

65

ES 2 972 684 T3

Luego, coloque la placa en un soporte magnético durante 5 minutos y deseche con cuidado el sobrenadante sin mover la placa del soporte.

5 Aún sin mover la placa del soporte, lave las perlas dos veces con 200 µl de etanol al 70% recién preparado durante 1 min.

Seque completamente las perlas (posiblemente a 37 ° C durante 3 a 5 minutos) y agregue 32 µl de agua libre de nucleasa a cada pocillo de muestra. Después de sellar y mezclar, incubar 2 min a temperatura ambiente y volver a colocar la placa en el soporte magnético durante 3 min.

10

Transfiera el sobrenadante (30 µl) que contiene el ADN purificado a un pocillo de placa nuevo.

El ADN purificado se puede almacenar a -20 °C.

15 ✓ Amplificación de la biblioteca ligada al adaptador

Mezcle bien 6 µl de agua libre de nucleasa, 1,25 µl de cebador SureSelect, 1,25 µl de cebador inverso de PCR de precaptura de indexación SureSelect Illumina, 10 µl de tampón de reacción 5x Herculase II, 0,5 µl de mezcla de dNTP 100 mM, 1 µl de ADN polimerasa de fusión Herculase II y 30 µl de ADN purificado ligado al adaptador.

20

Realice el siguiente programa de PCR: 98 °C durante 2 min, 10 ciclos de 98 °C 30 s + 65 °C 30 s + 72 °C 1 min, 72 °C 10 s y luego mantenga a +4 °C.

25 ✓ Purificación magnética

Una vez que las perlas magnéticas (AMPure XP) alcancen la temperatura ambiente durante 30 minutos, agregue 90 µl de perlas homogéneas para finalizar las muestras de ADN reparadas. Mezclar bien e incubar durante 5 min a temperatura ambiente. Luego, coloque la placa en un soporte magnético durante 5 minutos y deseche con cuidado el sobrenadante sin mover la placa del soporte.

30

Aún sin mover la placa del soporte, lave las perlas dos veces con 200 µl de etanol al 70% recién preparado durante 1 min.

35 Seque completamente las perlas (posiblemente a 37 ° C durante 3 a 5 minutos) y agregue 30 µl de agua libre de nucleasa a cada pocillo de muestra. Después de sellar y mezclar, incubar 2 min a temperatura ambiente y volver a colocar la placa en el soporte magnético durante 3 min. Transfiera el sobrenadante (30 µl) que contiene el ADN purificado a un pocillo de placa nuevo.

40 El ADN purificado se puede almacenar a -20 °C.

Es posible comprobar que tus fragmentos de ADN tienen de 400 a 450 pb gracias a un Bioanalizador (Agilent) o una TapeStation (Agilent).

45 ✓ Hibridación de ADN purificado con la biblioteca de captura

Coloque la placa que contiene las bibliotecas ligadas al adaptador amplificadas purificadas en un concentrador de vacío y deshidrate las bibliotecas durante 33 minutos a temperatura ambiente (debe ser visible un pequeño volumen de líquido, aproximadamente 3 µl).

50 Durante este tiempo, prepare el tampón de hibridación que contiene: 6,63 µl de SureSelect Hyb 1, 0,27 µl de SureSelect Hyb 2, 2,65 µl de SureSelect Hyb 3 y 3,45 µl de SureSelect Hyb 4, mezcle bien y mantenga a temperatura ambiente hasta su uso.

55 A continuación, prepare la mezcla de bloques SureSelect que contiene: 2,5 µl del bloque de indexación SureSelect 1, 2,5 µl del bloque de indexación SureSelect 2 y 0,6 µl del bloque de indexación SureSelect Illumina 3. Mezcle bien y mantenga en hielo hasta su uso.

60 Agregue 5,6 µl de mezcla en bloque SureSelect al pocillo que contiene ADN y mezcle pipeteando. Transferir la placa sellada en un termociclador e incubar a 95 °C durante 5 min y mantener a 65 °C durante al menos 5 min (con la tapa a 105 °C).

65 Durante este tiempo, prepare 5 µl de una dilución al 10% de solución de bloque de RNasa en hielo y prepare la mezcla de hibridación de captura mezclando bien 13 µl de la mezcla de tampón de hibridación preparada anteriormente, 5 µl de solución de bloque de RNasa al 10% y 2 µl de biblioteca de captura. Mantener a temperatura ambiente hasta su uso.

ES 2 972 684 T3

Una vez finalizados los al menos 5 min a 65 °C, mantener la placa en el termociclador a 65 °C y agregar 20 µl de la mezcla de hibridación de captura a cada pocillo. Mezclar bien pipeteando y sellar la placa.

5 Incubar la placa durante la noche a 65 °C con una tapa caliente.

✓ Captura del ADN hibridado

10 Mañana por la mañana, precaliente el tampón de lavado SureSelect 2 a 65 °C. A continuación, resuspenda completamente las perlas magnéticas Dynabeads MyOne Streptavidin T1 (en vórtice). Luego, transfiera 50 µl de estas perlas a los pocillos de una nueva placa de PCR.

15 Agregue 200 µl de tampón de unión SureSelect a las perlas de 50 µl y mezcle pipeteando. Luego, coloque la placa en un soporte magnético durante 5 minutos y deseche con cuidado el sobrenadante sin mover la placa del soporte. Lave nuevamente 2 veces más con el tampón de unión SureSelect.

Finalmente, resuspender las perlas en 200 µl de tampón de unión SureSelect.

20 Manteniendo la placa de hibridación a 65 °C, transfiera el volumen total de cada mezcla de hibridación a la placa que contiene 200 µl de perlas de estreptavidina lavadas y mezcle bien pipeteando.

Sellar la placa y mezclar vigorosamente en un dispositivo MixMate a 1800 rpm durante 30 min a temperatura ambiente.

25 Después de girar brevemente, coloque la placa en un soporte magnético durante 5 minutos y deseche con cuidado el sobrenadante sin mover la placa del soporte. Retire la placa del soporte y resuspenda las perlas en 200 µl de tampón de lavado SureSelect 1, mezcle con pipeta. Incubar 15 min a temperatura ambiente, girar brevemente y volver a colocar la placa en un soporte magnético durante 5 min y desechar con cuidado el sobrenadante sin mover la placa del soporte. Retire la placa del soporte, resuspenda las perlas en 200 µl de tampón de lavado 2 precalentado (65 °C) y mezcle con pipeta. Sellar e incubar la placa en baño de agua a 65 °C durante 10 min.

35 Luego, coloque la placa en un soporte magnético durante 5 minutos y deseche con cuidado el sobrenadante sin mover la placa del soporte. Repite este lavado 2 veces más.

Después del tercer lavado, retire todo el sobrenadante y agregue 30 µl de agua libre de nucleasas a cada muestra, mezcle pipeteando y mantenga la placa en hielo hasta su uso.

40 ✓ Amplificación de las bibliotecas capturadas

45 Mezcle bien 18,5 µl de agua libre de nucleasa, 5 µl de tampón de reacción Herculase II 5x, 0,5 µl de mezcla de dNTP 100 mM, 1 µl de ADN polimerasa de fusión Herculase II, 1 µl de cebador de PCR directo post-captura de indexación SureSelect Illumina y agregue 31 µl de esta preparación a cada muestra. A continuación, agregue 5 µl del cebador de indexación apropiado a cada pocillo de muestra. Finalmente, transfiera 14 µl del ADN capturado a la mezcla de reacción, mezcle pipeteando, selle la placa y transfiera la placa a un termociclador.

Ejecute el siguiente programa de PCR: 98 °C durante 2 min, 16 ciclos de 98 °C 30 s + 57 °C 30 s + 72 °C 1 min, 72 °C 10 min y luego mantenga a +4 °C.

50 ✓ Purificación magnética

55 Deje que las perlas magnéticas (AMPure XP) alcancen la temperatura ambiente durante 30 minutos. A continuación, agregue 90 µl de perlas homogéneas para finalizar las muestras de ADN reparadas. Mezclar bien e incubar durante 5 min a temperatura ambiente. Luego, coloque la placa en un soporte magnético durante 5 minutos y deseche con cuidado el sobrenadante sin mover la placa del soporte.

Aún sin mover la placa del soporte, lave las perlas dos veces con 200 µl de etanol al 70% recién preparado durante 1 min.

60 Seque completamente las perlas (posiblemente a 37 °C durante 3 a 5 minutos) y agregue 30 µl de agua libre de nucleasa a cada pocillo de muestra. Después de sellar y mezclar, incubar 2 min a temperatura ambiente y volver a colocar la placa en el soporte magnético durante 3 min. Transfiera el sobrenadante (30 µl) que contiene la biblioteca purificada a un pocillo de placa nuevo.

65 La biblioteca purificada se puede almacenar a -20 °C durante mucho tiempo.

✓ Evaluación de calidad y cantidad de la biblioteca.

5 Verifique la calidad de la biblioteca mediante un analizador de fragmentos. El tamaño de la biblioteca debería ser de alrededor de 500 pb.

Evalúe la cantidad de ADN mediante cuantificación por electroforesis o cuantificación fluorimétrica.

10 Una vez cuantificada la biblioteca, agrupe las bibliotecas y prepárese para la inyección siguiendo las instrucciones de su dispositivo de secuenciación.

✓ Secuenciación

15 La secuenciación se puede realizar con 120 a 150 bases para la lectura 1, 9 ciclos para lecturas de índice y 120 a 150 bases para la lectura 2.

2- Resultados

20 a- Condiciones de corte

Haciendo referencia a las Figuras 1A a 2B que muestran las condiciones de corte de acuerdo con la técnica anterior y de acuerdo con la invención, se puede apreciar que las condiciones de corte de acuerdo con la invención permiten una buena acumulación de fragmentos y una buena calidad de secuenciación en comparación.

25 b- Sensibilidad del procedimiento de detección de anomalías genéticas somáticas presentes en las muestras

Los resultados de la sensibilidad de la combinación de sondas de captura de acuerdo con la invención se ilustran en las tablas siguientes.

Origen	Número de muestra	Variaciones observadas con procedimientos clásicos	Variaciones observadas con nuestro protocolo de captura
Tumor	B08,36	EGFR: Exón de delección 19	EGFR: Exón de delección 19
Tumor	B08,227	EGFR: Exón de delección 19 EGFR: Exón de inserción 20	EGFR: Exón de delección 19 EGFR: Exón de inserción 20
Tumor	B09,74	EGFR: L858R	EGFR: L858R
Tumor	B10,165	EGFR: Exón de delección 19	EGFR: Exón de delección 19
Tumor	B11,437	EGFR: Exón de delección 19	EGFR: Exón de delección 19
Tumor	B12,621	EGFR: Exón de delección 19	EGFR: Exón de delección 19
Tumor	B12,648	EGFR: Exón de delección 19	EGFR: Exón de delección 19
Tumor	B13,425	EGFR: Exón de delección 19	EGFR: Exón de delección 19
Tumor	B13,560	EGFR: Exón de delección 19	EGFR: Exón de delección 19 EGFR: S752F
Tumor	B14,488	EGFR: Exón de delección 19	EGFR: Exón de delección 19
Tumor	B14,626	EGFR: L858R	EGFR: L858R
Tumor	B15,133	EGFR: Exón de delección 19	EGFR: Exón de delección 19
Tumor	B15,169	EGFR: L858R EGFR: T790M	EGFR: L858R EGFR: T790M
Tumor	B15,305	EGFR: Exón de delección 19	EGFR: Exón de delección 19
Tumor	B15,812	EGFR: Exón de inserción 20	EGFR: Exón de inserción 20 EGFR: H773Y
Tumor	B15,853	EGFR: Exón de delección 19	EGFR: Exón de delección 19
Tumor	B15,878	EGFR: Exón de delección 19	EGFR: Exón de delección 19

ES 2 972 684 T3

	Origen	Número de muestra	Variaciones observadas con procedimientos clásicos	Variaciones observadas con nuestro protocolo de captura
5	Tumor	B16,237	<i>KRAS</i> : G12V	<i>KRAS</i> : G12V
	Tumor	B16,298	<i>EGFR</i> : Exón de delección 19	<i>EGFR</i> : Exón de delección 19
	Tumor	B16,357	<i>EGFR</i> : L858R	<i>EGFR</i> : L858R
10	Tumor	B17,12	<i>KRAS</i> : G12S	<i>KRAS</i> : G12S
	Tumor	B16,1251	<i>KRAS</i> : G12V	<i>KRAS</i> : G12V
	Tumor	B17,16	<i>KRAS</i> : G12V	<i>KRAS</i> : G12V
15	Tumor	B17,14	<i>KRAS</i> : G12C	<i>KRAS</i> : G12C
	Tumor	B17,21	<i>BRAF</i> : V600E	<i>BRAF</i> : V600E
	Tumor	B17,22	<i>EGFR</i> : Exón de delección 19	<i>EGFR</i> : Exón de delección 19
20	Tumor	B17,27	<i>KIT</i> : L576P	<i>KIT</i> : L576P <i>KRAS</i>: G12D
	Tumor	B 17,32	<i>KRAS</i> : G12D	<i>KRAS</i> : G12D
25	Tumor	B17,36	<i>KRAS</i> : G12V	<i>KRAS</i> : G12V <i>EGFR</i>: H304Y
	Tumor	B12,55	<i>EGFR</i> : L858R	<i>EGFR</i> : L858R <i>EGFR</i>: D612G
30	Tumor	B14,415	<i>EGFR</i> : Exón de delección 19	<i>EGFR</i> : Exón de delección 19
	Tumor	B14,750	<i>EGFR</i> : L858R	<i>EGFR</i> : L858R
	Tumor	B15,346	<i>EGFR</i> : Exón de delección 19	<i>EGFR</i> : Exón de delección 19
35	Tumor	B15,849	<i>EGFR</i> : Exón de delección 19	<i>EGFR</i> : Exón de delección 19
	Tumor	B16,996	<i>KRAS</i> : G12V	<i>KRAS</i> : G12V
	Tumor	B16,1053	<i>KRAS</i> : Q61H	<i>KRAS</i> : Q61H
40	Tumor	B16,1056	<i>NRAS</i> : Q61H	<i>NRAS</i> : Q61H
	Tumor	B16,1078	<i>KIT</i> : V654A <i>KIT</i> : Y823D	<i>KIT</i> : V654A <i>KIT</i> : Y823D
45	Tumor	B16,1120	<i>BRAF</i> : V600E	<i>BRAF</i> : V600E
	Tumor	B16,1124	<i>KRAS</i> : G12C	<i>KRAS</i> : G12C
	Tumor	B16,1126	<i>EGFR</i> : L858R	<i>EGFR</i> : L858R
50	Tumor	B16,1127	<i>KRAS</i> : G12C	<i>KRAS</i> : G12C
	Tumor	B16,1129	<i>BRAF</i> : V600E	<i>BRAF</i> : V600E
	Tumor	B16,1130	<i>KRAS</i> : G13D	<i>KRAS</i> : G13D
	Tumor	B16,1137	<i>BRAF</i> : V600E	<i>BRAF</i> : V600E
55	Tumor	B16,1138	<i>KRAS</i> : G13D	<i>KRAS</i> : G13D
	Tumor	B16,1148	<i>NRAS</i> : Q61L	<i>NRAS</i> : Q61L
60	Tumor	B16,418	<i>EGFR</i> : G719A	<i>EGFR</i> : G719A
	Tumor	B16,430	<i>EGFR</i> : L858R	<i>EGFR</i> : L858R
	Tumor	B16,655	<i>EGFR</i> : G719A	<i>EGFR</i> : G719A
	Tumor	B16,990	<i>ERBB2</i> : Exón de inserción 20	<i>ERBB2</i> : Exón de inserción 20
65	Tumor	B16,501	<i>EGFR</i> : Exón de delección 19	<i>EGFR</i> : Exón de delección 19

ES 2 972 684 T3

	Origen	Número de muestra	Variaciones observadas con procedimientos clásicos	Variaciones observadas con nuestro protocolo de captura
5	Tumor	B16,1141	<i>EGFR</i> : Exón de inserción 20	<i>EGFR</i> : Exón de inserción 20
	Tumor	B16,1143	<i>KRAS</i> : G12C	<i>KRAS</i> : G12C
	Tumor	B16,1145	<i>KRAS</i> : G12V	<i>KRAS</i> : G12V
10	Tumor	B16,1149	<i>KRAS</i> : G13D	<i>KRAS</i> : G13D
	Tumor	B16,1153	<i>KRAS</i> : K117N	<i>KRAS</i> : K117N
	Tumor	B16,1155	<i>KRAS</i> : G13C	<i>KRAS</i> : G13C
15	Tumor	B16,1157	<i>KRAS</i> : G12C	<i>KRAS</i> : G12C
	Tumor	B17,41	<i>KRAS</i> : G12A	<i>KRAS</i> : G12A
	Tumor	B17,47	<i>EGFR</i> : L858R	<i>EGFR</i> : L858R
20	Tumor	B17,48,1BP	<i>KRAS</i> : G13C	<i>KRAS</i> : G13C
	Tumor	B17,48,4DP	<i>KRAS</i> : Q61L	<i>KRAS</i> : Q61L
	Tumor	B17,58	<i>KRAS</i> : G12S	<i>KRAS</i> : G12S
25	Tumor	B17,59	<i>BRAF</i> : V600E	<i>BRAF</i> : V600E
	Tumor	B17,64	<i>KRAS</i> : G12D	<i>KRAS</i> : G12D
	Tumor	B17, 65	<i>BRAF</i> : V600E	<i>BRAF</i> : V600E
30	Tumor	B17,60	<i>EGFR</i> : L858R	<i>EGFR</i> : L858R
	Tumor	B17,61	<i>KRAS</i> : A146P	<i>KRAS</i> : A146P
	Tumor	B17,67	<i>BRAF</i> : V600K	<i>BRAF</i> : V600K
35	Tumor	B17,05	Sin mutación	Sin mutación
	Tumor	B17,11	Sin mutación	Sin mutación
	Tumor	B17,13	Sin mutación	Sin mutación
40	Tumor	B17,15	Sin mutación	Sin mutación
	Tumor	B17,19	Sin mutación	Sin mutación
	Tumor	B17,20	Sin mutación	Sin mutación
45	Tumor	B17,23	Sin mutación	Sin mutación
	Tumor	B17,05	Sin mutación	Sin mutación
	Tumor	B17,11	Sin mutación	Sin mutación
50	Tumor	B16,1152	Sin mutación	Sin mutación
	Tumor	B15,342	Sin mutación	<i>EGFR</i>: E709G <i>EGFR</i>: G719C
55	Tumor	B17,18	Sin mutación	<i>BRAF</i>: G596R
	Tumor	B15,820	Sin mutación	<i>EGFR</i>: T751P
	Tumor	N16,296	Sin mutación	<i>EGFR</i>: S768I
				<i>EGFR</i>: V769L
60	Tumor	B14,632	Sin mutación	<i>EGFR</i>: L858R
	Plasma	CHA	<i>KRAS</i> : G12C	Sin mutación
	Plasma	VIA	<i>KRAS</i> : G12S	<i>KRAS</i> : G12S
65	Plasma	LOB	<i>EGFR</i> : G719A	<i>EGFR</i> : G719A

Origen	Número de muestra	Variaciones observadas con procedimientos clásicos	Variaciones observadas con nuestro protocolo de captura
Plasma	RIG	Sin mutación	Sin mutación
Plasma	LEM	Sin mutación	Sin mutación
Plasma	GAR	<i>KRAS</i> : A146T	<i>KRAS</i> : A146T
Plasma	TOR	Sin mutación	Sin mutación
Plasma	PEN	Sin mutación	Sin mutación
Plasma	LUC	Sin mutación	Sin mutación
Plasma	LED	<i>KRAS</i> : G12C	<i>KRAS</i> : G12C
Plasma	POQ	<i>EGFR</i> : Exón de delección 19	<i>EGFR</i> : Exón de delección 19 <i>EGFR</i>: T790M
Plasma	17BM00026	<i>EGFR</i> : Exón de delección 19	<i>EGFR</i> : Exón de delección 19 <i>EGFR</i>: T790M

Origen	Variaciones esperadas	Porcentaje observado en ejecución 1	Porcentaje observado en ejecución 2	Porcentaje observado en ejecución 3
Comercial	<i>BRAF</i> : V600E	9%	9%	7%
	<i>EGFR</i> : G719S	18%	20%	20%
	<i>KRAS</i> : G13D	23%	26%	26%

Se podría observar en las tablas anteriores que la combinación de sondas de captura de acuerdo con la invención permite detectar anomalías genéticas con una alta precisión.

REIVINDICACIONES

- 5 1. Procedimiento de detección de anomalías genéticas somáticas potencialmente presentes en una muestra de tejido sólido, potencialmente canceroso, tomada de un sujeto, en el que dicho procedimiento de detección comprende las siguientes etapas:
- 10 (a) corte del ADN genómico extraído de dicha muestra de tejido sólido, para obtener ADN genómico cortado;
- (b) preparar una biblioteca mediante captura por hibridación a partir del ADN genómico cortado obtenido en la etapa (a), mediante la implementación de una combinación de al menos 51 sondas de captura que tienen al menos 110 nucleótidos de longitud y al menos un 75% de homología de secuencia respectivamente con la SEQ ID NO: 1 a la SEQ ID NO: 51, siendo la combinación de dichas
- 15 al menos 51 sondas de captura capaz de hibridarse en fragmentos de las siguientes regiones diana: ALK, BRAF, EGFR, EIF1AY, ERBB2, ERBB4, KDR, KIT, KRAS, MAP2K1, MET, NRAS, PDGFRA, SLC28A1, TP53;
- (c) secuenciación del ADN genómico cortado contenido en la biblioteca emitida en la etapa (b), y
- 20 (d) analizar las secuencias de dicho ADN genómico cortado obtenido en la etapa (c), para detectar anomalías genéticas somáticas en dicha muestra de tejido.
2. Procedimiento de detección de acuerdo con la reivindicación 1, en el que la preparación de la biblioteca en la etapa (b) comprende la implementación de la combinación de sondas de captura que además comprende al menos 100 sondas de captura adicionales, preferentemente al menos 200, en particular al menos 300, e idealmente al menos 1178 sondas de captura adicionales que comprenden al menos 110 nucleótidos de longitud y al menos un 75% de homología de secuencia con la SEQ ID NO: 52 a la SEQ ID NO: 1229.
- 25 3. Procedimiento de detección de acuerdo con la reivindicación 1 o 2, en el que la preparación de la biblioteca en la etapa (b) comprende la implementación de la combinación de sondas de captura que comprende:
- 30 - 51 sondas de captura que tienen al menos un 75%, preferentemente un 85%, en particular un 90% e idealmente un 95%, de homología de secuencia con las sondas de captura de la SEQ ID NO: 1 a la SEQ ID NO: 51;
- 35 - sondas de captura adicionales teniendo al menos un 75%, preferentemente un 85%, en particular un 90% e idealmente un 95% y posiblemente un 100% de homología de secuencia con las sondas de captura con las sondas de captura de la SEQ ID NO: 52 a la SEQ ID NO: 1229, respectivamente.
4. Procedimiento de detección de acuerdo con una cualquiera de las reivindicaciones precedentes, en el que la preparación de la biblioteca en la etapa (b) comprende la implementación de dicha combinación de sondas de captura que comprenden de 110 a 150 nucleótidos de longitud, preferentemente de 120 a 140 nucleótidos de longitud e idealmente de 125 a 135 nucleótidos de longitud.
- 40 5. Procedimiento de detección de acuerdo con una cualquiera de las reivindicaciones precedentes, en el que la preparación de la biblioteca en la etapa (b) comprende:
- 45 (b1) la hibridación del ADN genómico cortado obtenido en la etapa (a) con la combinación de sondas de captura, para enriquecer el ADN genómico cortado contenido en la biblioteca,
- (b2) la amplificación de la biblioteca.
- 50 6. Procedimiento de detección de acuerdo con la reivindicación 5, en el que la etapa (b1) tiene lugar a una temperatura de al menos 55 °C, preferentemente al menos a 58 °C y preferentemente a una temperatura que oscila entre 60 °C y 70 °C o entre 58 °C a 65 °C.
- 55 7. Procedimiento de detección de acuerdo con una cualquiera de las reivindicaciones precedentes, en el que el ADN genómico cortado obtenido en la etapa (a) y capturado en la etapa (b) tiene una longitud mayor o igual que 300 pb, preferentemente que oscila entre 300 y 600 pb, e idealmente que oscila entre 300 y 350 pb.
- 60 8. Procedimiento de detección de acuerdo con una cualquiera de las reivindicaciones precedentes, en el que, en el análisis de las secuencias (d), la profundidad promedio es de al menos 200x, preferentemente que oscila entre 200 y 10000x, en particular que oscila entre 500x y 1500x, e idealmente que oscila entre 500x y 1200x.
- 65 9. Procedimiento de detección de acuerdo con una cualquiera de las reivindicaciones precedentes, en el que, en el análisis de las secuencias (d), la tasa de cobertura es de al menos el 90% a 200x, preferentemente que oscila entre el 90% y el 99,9% a 500x, en particular que oscila entre entre el 95% y el 99,9% a 500x, e

idealmente que oscila entre el 98% y el 99,9% a 500x.

- 5
10. Procedimiento de detección de acuerdo con una cualquiera de las reivindicaciones precedentes, en el que las anomalías genéticas somáticas están ligadas al cáncer seleccionado entre al menos uno de estos cánceres sólidos siguientes: tumor, policitemia vera, linfomas tales como el linfoma de Hodgkin, linfoma folicular o linfoma no Hodgkin, mieloma múltiple, macroglobulinemia de Waldenstrom, enfermedad de cadenas pesadas, tumores sólidos, sarcomas, carcinomas tales como, por ejemplo, fibrosarcoma, mixosarcoma, liposarcoma, condrosarcoma, sarcoma osteogénico, linfangiosarcoma, mesotelioma, tumor de Ewing, leiomiosarcoma, rhabdomiosarcoma, carcinoma de colon, cáncer colorrectal, cáncer de páncreas, 10
cáncer de mama, cáncer de ovario, cáncer de próstata, carcinoma de células escamosas, carcinoma de células basales, adenocarcinoma, carcinoma de glándulas sudoríparas, carcinoma de glándulas sebáceas, carcinoma papilar, adenocarcinomas papilares, cistadenocarcinoma, carcinoma medular, carcinoma broncogénico, carcinoma de células renales, hepatoma, carcinoma de vías biliares, coriocarcinoma, seminoma, carcinoma embrionario, tumor de Wilms, cáncer de cuello uterino, cáncer de útero, tumor testicular, carcinoma de pulmón, carcinoma de pulmón de células pequeñas, carcinoma de vejiga, carcinoma epitelial, glioma, craneofaringioma, ependimoma, pinealoma, hemangioblastoma, neuroma acústico, oligodendroglioma, meningioma, melanoma, neuroblastoma, retinoblastoma, cáncer de endometrio, cáncer de pulmón de células no pequeñas, glioblastoma o cáncer de pulmón.
- 20
11. Combinación de sondas de captura, que comprende al menos 51 sondas de captura que comprenden al menos 110 nucleótidos y al menos un 75% de homología de secuencia con, respectivamente, las sondas de la SEQ ID NO: 1 a la SEQ ID NO: 51.
- 25
12. Combinación de sondas de captura de acuerdo con la reivindicación 11, que además comprende al menos:
- 100 sondas de captura adicionales, preferentemente al menos 200, en particular al menos 300, e idealmente al menos 1178 sondas de captura adicionales que comprenden al menos 110 nucleótidos y al menos un 75% de homología de secuencia con, respectivamente, las sondas de la SEQ ID NO: 52 a la SEQ ID NO: 1229.
- 30
13. Kit de detección de anomalías genéticas somáticas potencialmente presentes en una muestra de tejido potencialmente canceroso tomada de un sujeto, que comprende la combinación de sondas de captura tal como se define en la reivindicación 11 o la reivindicación 12.
- 35

Fig.1A

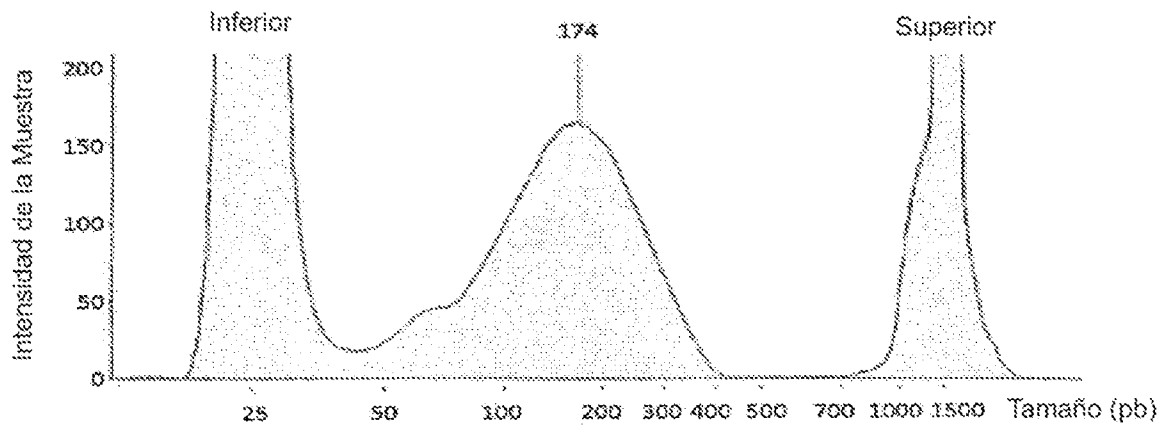


Fig.1B

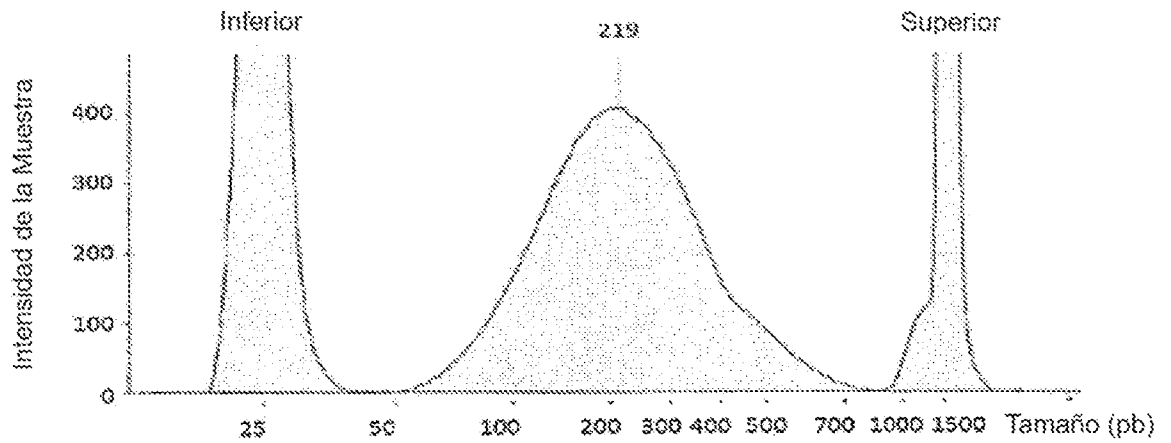


Fig.2A

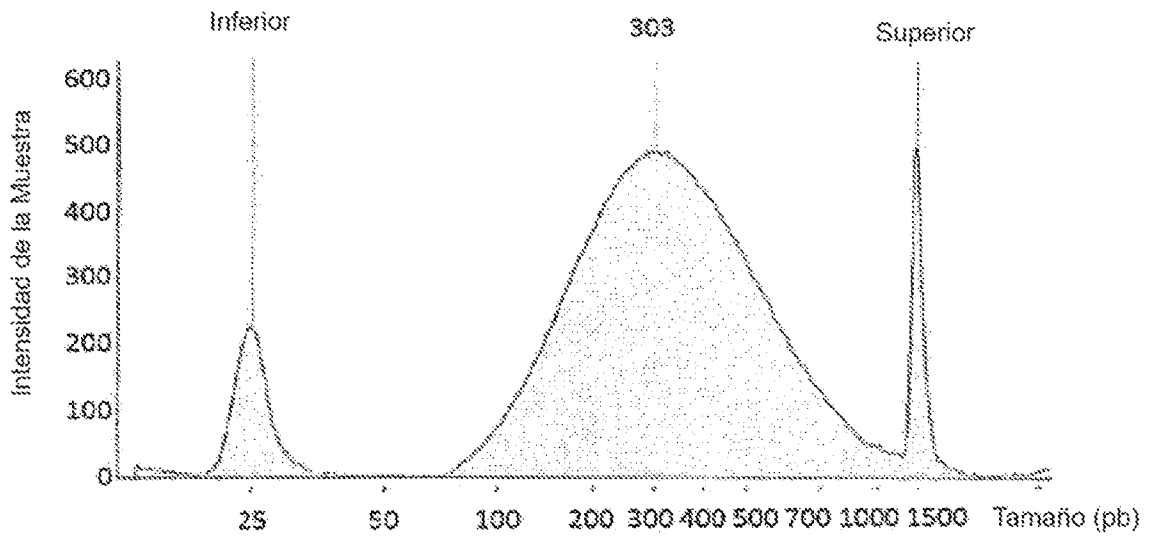


Fig.2B

