

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 905 448**

51 Int. Cl.:

**C12Q 1/6806** (2008.01)

**C12Q 1/6855** (2008.01)

**A61P 35/00** (2006.01)

12

## TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **11.03.2013 PCT/US2013/030201**

87 Fecha y número de publicación internacional: **14.11.2013 WO13169339**

96 Fecha de presentación y número de la solicitud europea: **11.03.2013 E 13787745 (2)**

97 Fecha y número de publicación de la concesión europea: **19.01.2022 EP 2847353**

54 Título: **Métodos para determinar una secuencia nucleotídica**

30 Prioridad:

**10.05.2012 US 201261645364 P**

**03.08.2012 US 201261679302 P**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:  
**08.04.2022**

73 Titular/es:

**THE GENERAL HOSPITAL CORPORATION  
(100.0%)**

**55 Fruit Street  
Boston, MA 02114, US**

72 Inventor/es:

**IAFRATE, ANTHONY JOHN;  
LE, LONG PHI y  
ZHENG, ZONGLI**

74 Agente/Representante:

**VALLEJO LÓPEZ, Juan Pedro**

ES 2 905 448 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

Métodos para determinar una secuencia nucleotídica

## 5 Referencia a solicitudes relacionadas

### Listado de secuencias

### 10 Apoyo gubernamental

La presente invención se realizó con financiación del gobierno con la subvención n. 5R21CA161590 otorgada por los "National Institutes of Health". El gobierno de los Estados Unidos posee determinados derechos sobre la invención.

### 15 Campo técnico

La tecnología descrita en el presente documento se refiere a métodos para determinar secuencias oligonucleotídicas.

### Antecedentes

20 El enriquecimiento dirigido antes de la secuenciación de última generación es más rentable que la secuenciación del genoma completo, el exoma completo y el transcriptoma completo y, por lo tanto, más práctico para la implementación general; tanto para la investigación científica como las aplicaciones clínicas. Por ejemplo, la elevada profundidad de cobertura proporcionada por los planteamientos de enriquecimiento dirigido posibilita un mayor rango dinámico para el recuento de alelo (en la evaluación de la expresión génica y el número de copias) y la detección de mutaciones de  
25 baja frecuencia, un rasgo crucial para la evaluación de las mutaciones somáticas en cáncer. Ejemplos de protocolos de enriquecimiento actuales para la secuenciación de última generación incluyen ensayos de captura basados en hibridación (TruSeq Capture, Illumina; captura híbrida SureSelect, Agilent) y análisis basado en la reacción en cadena de la polimerasa (PCR) (HaloPlex, Agilent; AmpliSeq, Ion Torrent; TruSeq Amplicon, Illumina; PCR en emulsión/digital, Raindance). Los planteamientos basados en hibridación no capturan solo las secuencias dirigidas cubiertas por las sondas de captura sino también las bases cercanas inespecíficas (*off-target*) que consumen la capacidad de secuenciación. Adicionalmente, estos métodos requieren mucho tiempo, son laboriosos y experimentan un nivel relativamente bajo de especificidad. Un planteamiento basado en amplificación por PCR es más sencillo y rápido pero por el diseño convencional requiere el uso de tanto cebadores directos como inversos que flanqueen los loci diana. En particular, para la detección de reordenamientos genómicos con parejas de fusión desconocidas, la PCR no es  
35 aplicable.

### Sumario

La tecnología descrita en el presente documento se dirige a métodos para determinar secuencias oligonucleotídicas. En algunas realizaciones, los métodos descritos en el presente documento se refieren a enriquecer secuencias diana antes de la secuenciación de las secuencias.

En un aspecto, la tecnología descrita en el presente documento se refiere a un método para determinar la secuencia nucleotídica contigua a una secuencia nucleotídica diana conocida, comprendiendo el método; (a) ligar un ácido  
45 nucleico diana que comprende la secuencia nucleotídica diana conocida a un adaptador de cola de oligonucleótido universal; (b) amplificar una porción del ácido nucleico diana y la hebra de amplificación del adaptador de cola de oligonucleótido universal con un primer cebador adaptador y un primer cebador específico de diana; (c) amplificar una porción del amplicón resultante de la etapa (b) con un segundo cebador adaptador y un segundo cebador específico de diana; (d) secuenciar la porción amplificada de la etapa (c) utilizando un primer y un segundo cebador de secuenciación; en donde el adaptador de cola de oligonucleótido universal comprende un primer extremo dúplex que se puede ligar y un segundo extremo deshibridado; en donde el adaptador de cola de oligonucleótido universal comprende una hebra de bloqueo y una hebra de amplificación; en donde la hebra de bloqueo comprende una porción dúplex 5'; en donde la hebra de amplificación comprende una porción 5' deshibridada, una porción dúplex 3' y un saliente T 3'; en donde la hebra de amplificación comprende secuencias de ácido nucleico idénticas a un primer y un  
55 segundo cebador de secuenciación; en donde las porciones dúplex de la hebra de bloqueo y la hebra de amplificación son sustancialmente complementarias y forman el primer extremo dúplex que se puede ligar que comprende un saliente T 3'; en donde la porción dúplex es de suficiente longitud para permanecer en forma dúplex a la temperatura de ligamiento; en donde el primer cebador específico de diana comprende una secuencia de ácido nucleico que puede hibridarse de forma específica con la secuencia nucleotídica diana conocida del ácido nucleico diana a la temperatura de hibridación; en donde el segundo cebador específico de diana comprende una porción 3' que comprende una  
60 secuencia de ácido nucleico que puede hibridarse de forma específica con una porción de la secuencia nucleotídica diana conocida comprendida por el amplicón resultante de la etapa (b), y una porción 5' que comprende una secuencia de ácido nucleico que es idéntica a un segundo cebador de secuenciación y el segundo cebador específico de diana es anidado con respecto al primer cebador específico de diana; en donde el primer cebador adaptador comprende una secuencia de ácido nucleico idéntica a una porción 5' del primer cebador de secuenciación; y en donde el segundo cebador adaptador comprende una secuencia de ácido nucleico idéntica a una porción del primer cebador de

secuencia y es anidado con respecto al primer cebador adaptador.

En algunas realizaciones, la hebra de bloqueo del adaptador de cola de oligonucleótido universal además puede comprender una porción deshibridada 3' que no es sustancialmente complementaria a la porción deshibridada 5' de la hebra de amplificación; y en donde la porción deshibridada 3' de la hebra de bloqueo no es sustancialmente complementaria a o sustancialmente idéntica a cualquiera de los cebadores. En algunas realizaciones, el segundo cebador adaptador puede ser anidado con respecto al primer cebador adaptador por al menos 3 nucleótidos. En algunas realizaciones, la porción de la hebra de amplificación que comprende una secuencia de ácido nucleico idéntica a un primer y un segundo cebador de secuenciación puede estar comprendida, al menos en parte, por la porción deshibridada 5' de la hebra de amplificación.

En algunas realizaciones, el primer cebador específico de diana además puede comprender una porción de la secuencia de la etiqueta 5' que comprende una secuencia de ácido nucleico de elevado contenido de GC que no es sustancialmente complementaria o sustancialmente idéntica a otra porción cualquiera de algunos de los cebadores. En algunas realizaciones, el primer cebador específico de diana además puede comprender una porción de la secuencia de la etiqueta 5' que comprende una secuencia de ácido nucleico de elevado contenido de GC que no se hibridará de forma específica con ninguna otra porción de ninguno de los cebadores o sus complementos a la temperatura de hibridación. En algunas realizaciones, el segundo cebador adaptador puede ser idéntico al primer cebador de secuenciación completo. En algunas realizaciones, las porciones de los cebadores específicos a diana que se hibridan de forma específica con la diana conocida pueden hibridarse de forma específica a una temperatura de aproximadamente 65 °C en un tampón de PCR.

En algunas realizaciones, el método puede incluir además; antes de la etapa (a), las etapas de: romper mecánicamente el ácido nucleico; someter el ácido nucleico a reparación del extremo; someter el ácido nucleico a fosforilación, y someter el ácido nucleico a adenilación. En algunas realizaciones, la muestra puede comprender ADN genómico. En algunas realizaciones, la muestra puede comprender ARN y el método puede comprender además una primera etapa para someter la muestra a un régimen de transcriptasa inversa. En algunas realizaciones, el régimen de transcriptasa inversa puede comprender la utilización de hexámeros aleatorios.

En algunas realizaciones, la secuencia diana conocida puede estar comprendida por un reordenamiento de genes. En algunas realizaciones, el reordenamiento de genes puede estar presente en un ácido nucleico seleccionado del grupo que consiste en: ADN genómico; ARN; y ADNc. En algunas realizaciones, el reordenamiento de genes puede comprender un oncogén. En algunas realizaciones, el reordenamiento de genes puede comprender un oncogén de fusión.

En algunas realizaciones, el producto de ácido nucleico puede secuenciarse mediante un método de secuenciación de última generación. En algunas realizaciones, el método de secuenciación de última generación puede comprender un método seleccionado del grupo que consiste en: Ion Torrent, Illumina, SOLiD, 454; secuenciación de firma masivamente paralela fase sólida, secuenciación de terminador con colorante reversible; y secuenciación de nanoesferas de ADN. En algunas realizaciones, los primeros y segundos cebadores de secuenciación son compatibles con el método de secuenciación de última generación seleccionado.

En algunas realizaciones, el método puede comprender poner en contacto la muestra, o separar porciones de la muestra, con una pluralidad de conjuntos de primeros y segundos cebadores específicos a diana. En algunas realizaciones, el método puede comprender poner en contacto una única mezcla de reacción que comprende la muestra con una pluralidad de conjuntos de primeros y segundos cebadores específicos a diana. En algunas realizaciones, la pluralidad de conjuntos de primeros y segundos cebadores específicos a diana pueden hibridarse de forma específica con secuencias nucleotídicas diana conocidas comprendidas por distintos genes. En algunas realizaciones, al menos dos conjuntos de primeros y segundos cebadores específicos a diana pueden hibridarse de forma específica con diferentes porciones de una secuencia nucleotídica diana conocida. En algunas realizaciones, al menos dos conjuntos de primeros y segundos cebadores específicos a diana pueden hibridarse de forma específica con diferentes porciones de un único gen que comprende una secuencia nucleotídica diana conocida. En algunas realizaciones, al menos dos conjuntos de primeros y segundos cebadores específicos a diana pueden hibridarse de forma específica a diferentes exones de un gen que comprende una secuencia nucleotídica diana conocida. En algunas realizaciones, la pluralidad de primeros cebadores específicos a diana puede comprender idénticas porciones de la secuencia de la etiqueta 5'. En algunas realizaciones, el adaptador de cola de oligonucleótido universal además puede comprender una porción de código de barras. En algunas realizaciones, múltiples muestras pueden conectarse cada una con un adaptador de cola de oligonucleótido universal con una porción de código de barras única y en donde las muestras se agrupan después de la etapa (a).

En algunas realizaciones, cada etapa de amplificación puede comprender un conjunto de ciclos de un régimen de amplificación por PCR de 5 ciclos a 20 ciclos de largo. En algunas realizaciones, los cebadores específicos a diana y los cebadores adaptador pueden diseñarse de modo que se hibridarán de forma específica con sus secuencias complementarias a una temperatura de hibridación de aproximadamente 61 a 72 °C. En algunas realizaciones, los cebadores específicos a diana y los cebadores adaptador pueden diseñarse de modo que se hibridarán de forma específica con sus secuencias complementarias a una temperatura de hibridación de aproximadamente 65 °C.

- En algunas realizaciones, la muestra puede comprender una muestra biológica obtenida de un sujeto. En algunas realizaciones, la muestra puede obtenerse de un sujeto que necesita tratamiento para una enfermedad asociada a una alteración genética. En algunas realizaciones, la enfermedad puede ser cáncer. En algunas realizaciones, la muestra puede comprender una población de células tumorales. En algunas realizaciones, la muestra puede comprender una biopsia tumoral. En algunas realizaciones, el cáncer puede ser cáncer de pulmón.
- En algunas realizaciones, la secuencia diana conocida puede estar comprendida por un gen asociado a la enfermedad. En algunas realizaciones, la secuencia diana conocida puede estar comprendida por un producto de reordenamiento de genes en la muestra. En algunas realizaciones, el producto de reordenamiento de genes puede ser un oncogén.
- En algunas realizaciones, una secuencia diana conocida puede comprender una secuencia de un gen seleccionado del grupo de: ALK; ROS1; y RET. En algunas realizaciones, al menos un conjunto de un primer cebador específico de diana y un segundo cebador específico de diana pueden seleccionarse del grupo que consiste en: SEQ ID NO: 5 y 6; SEQ ID NO: 7 y 8; SEQ ID NO: 9 y 10; SEQ ID NO: 11 y 12; SEQ ID NO: 13 y 14; SEQ ID NO: 15 y 16; SEQ ID NO: 17 y 18; SEQ ID NO: 19 y 20; SEQ ID NO: 21 y 22; SEQ ID NO: 23 y 24; SEQ ID NO: 25 y 26; SEQ ID NO: 27 y 28; SEQ ID NO: 29 y 30; SEQ ID NO: 31 y 32; SEQ ID NO: 33 y 34; SEQ ID NO: 35 y 36; y las SEQ ID NO: 37 y 38.
- En algunas realizaciones, la presencia de un reordenamiento génico de ALK en una muestra obtenida de un tumor en un sujeto puede indicar que el tumor es susceptible a tratamiento con un tratamiento seleccionado del grupo que consiste en: un inhibidor de ALK; crizotinib (PF-02341066); AP26113; LDK378; 3-39; AF802; IPI-504; ASP3026; AP-26113; X-396; GSK-1838705A; CH5424802; y NVP-TAE684.
- En algunas realizaciones, la presencia de un reordenamiento génico de ROS1 en una muestra obtenida de un tumor en un sujeto puede indicar que el tumor es susceptible a tratamiento con un tratamiento seleccionado del grupo que consiste en: un inhibidor de ROS; un inhibidor de ALK; crizotinib (PF-02341066); AP26113; LDK378; 3-39; AF802; IPI-504; ASP3026; AP-26113; X-396; GSK-1838705A; CH5424802; y NVP-TAE684.
- En algunas realizaciones, la presencia de un reordenamiento génico de RET en una muestra obtenida de un tumor en un sujeto puede indicar que el tumor es susceptible a tratamiento con un tratamiento seleccionado del grupo que consiste en: un inhibidor de RET; DP-2490; DP-3636; SU5416; BAY 43-9006; BAY 73-4506 (regorafenib); ZD6474; NVP-AST487; sorafenib; RPI-1; XL184; vandetanib; sunitinib; imatinib; pazopanib; axitinib; motesanib; gefitinib; y withaferin A.
- En un aspecto, la tecnología descrita en el presente documento se refiere a un método para tratar un cáncer, comprendiendo el método; detectar, en una muestra de tumor obtenida de un sujeto que necesita tratamiento para el cáncer, la presencia de uno o más reordenamientos de oncogenes de acuerdo con el método descrito en el presente documento; administrar un tratamiento para el cáncer que es efectivo contra los tumores que tienen cualquiera de los reordenamientos de oncogenes detectados. En algunas realizaciones, un tratamiento seleccionado del grupo que consiste en: un inhibidor de ALK; crizotinib (PF-02341066); AP26113; LDK378; 3-39; AF802; IPI-504; ASP3026; AP-26113; X-396; GSK-1838705A; CH5424802; y NVP-TAE684; puede ser efectivo contra tumores que tienen un reordenamiento de oncogén ALK. En algunas realizaciones, un tratamiento seleccionado del grupo que consiste en: un inhibidor de ROS1; un inhibidor de ALK; crizotinib (PF-02341066); AP26113; LDK378; 3-39; AF802; IPI-504; ASP3026; AP-26113; X-396; GSK-1838705A; CH5424802; y NVP-TAE684; puede ser efectivo contra tumores que tienen un reordenamiento de oncogén ROS1. En algunas realizaciones, un tratamiento seleccionado del grupo que consiste en: un inhibidor de RET; DP-2490; DP-3636; SU5416; BAY 43-9006; BAY 73-4506 (regorafenib); ZD6474; NVP-AST487; sorafenib; RPI-1; XL184; vandetanib; sunitinib; imatinib; pazopanib; axitinib; motesanib; gefitinib; y withaferin A; puede ser efectivo contra tumores que tienen un reordenamiento de oncogén RET.
- En un aspecto, la tecnología descrita en el presente documento se refiere a un método para determinar si un sujeto que necesita tratamiento para el cáncer será sensible a un tratamiento dado, comprendiendo el método; detectar, en una muestra de tumor obtenida del sujeto, la presencia de un reordenamiento de oncogenes de acuerdo con el método descrito en el presente documento; en donde se determina que un sujeto es sensible a un tratamiento que se dirige a un producto de reordenamiento de oncogenes si se detecta la presencia del reordenamiento del oncogén.
- En algunas realizaciones, cuando se detecta la presencia de un reordenamiento de oncogén ALK, el sujeto puede ser sensible a un tratamiento seleccionado del grupo que consiste en: un inhibidor de ALK; crizotinib (PF-02341066); AP26113; LDK378; 3-39; AF802; IPI-504; ASP3026; AP-26113; X-396; GSK-1838705A; CH5424802; y NVP-TAE684. En algunas realizaciones, cuando se detecta la presencia de un reordenamiento de oncogén ROS1, el sujeto puede ser sensible a un tratamiento seleccionado del grupo que consiste en: un inhibidor de ALK; crizotinib (PF-02341066); AP26113; LDK378; 3-39; AF802; IPI-504; ASP3026; AP-26113; X-396; GSK-1838705A; CH5424802; y NVP-TAE684.
- En algunas realizaciones, cuando se detecta la presencia de un reordenamiento de oncogén RET, el sujeto será sensible a un tratamiento seleccionado del grupo que consiste en: un inhibidor de RET; DP-2490; DP-3636; SU5416; BAY 43-9006; BAY 73-4506 (regorafenib); ZD6474; NVP-AST487; sorafenib; RPI-1; XL184; vandetanib; sunitinib; imatinib; pazopanib; axitinib; motesanib; gefitinib; y withaferin A.

En algunas realizaciones, el cáncer puede ser cáncer de pulmón.

### Breve descripción de los dibujos

La Figura 1 representa una ilustración esquemática de un ejemplo de construcción de genoteca para la secuenciación dirigida de ADN y ARN. 1) Se aplica un procedimiento estándar de síntesis de ADN bicatenario utilizando ácido nucleico total del espécimen FFPE como material de partida. Alternativamente, el material de partida también puede ser ADN<sub>g</sub> roto. 2) Después de limpiar, el ADN<sub>c</sub> o ADN<sub>g</sub> bicatenario se somete a reparación del extremo y adición de cola dA (*dA tailing*), directamente tras el ligamiento de un adaptador Y medio truncado sin limpieza intermedia. 3) Después de la limpieza por SPRI, la muestra ligada se somete a 14 ciclos de amplificación por PCR utilizando múltiples cebadores específicos a gen (GSP1) y cebador directo de longitud corta de Ion Torrent A 5' 20-mero (A20), con temperatura de hibridación de 65 °C. 4) Después de una segunda limpieza por SPRI, la muestra se somete a 14 ciclos adicionales de amplificación por PCR utilizando múltiples cebadores específicos a gen anidados (en dirección 3' de GSP1 y en la misma dirección) marcados con cebador inverso de Ion Torrent (P1\_GSP2s) y cebador directo A (30-mero completo), con temperatura de hibridación de 65 °C. 5) Después de una tercera limpieza por SPRI, el producto es la genoteca de Ion Torrent lista para la posterior PCR de emulsión y secuenciación. El agrupamiento de muestras puede llevarse a cabo después de la Etapa 2.

La Figura 2 representa los resultados del mapeo, que demuestran los diferentes productos de extensión del cebador (manchas) en una carrera de secuenciación de las secuencias de ROS1, incluyendo las lecturas que corresponden a ADN genómico (que abarca los límites de intrón-exón), ADN<sub>c</sub> (que abarca los exones) y ADN<sub>c</sub> de fusión (en el exón 34, mapeo sobre la pareja de fusión SLC34A2 no mostrado en la presente). Se consiguió 91,8 % de especificidad (127.446/138.787).

La Figura 3A representa una presentación esquemática de la estrategia de modificación dirigida con cebador anidado utilizando ROS1 como ejemplo. El panel objetivo de ensayo incluye un total de 17 pares de GSP1 y GSP2 para ROS1, ALK y RET. La Figura 3B representa una representación de los posibles tipos de productos de extensión utilizando moldes de ADN<sub>g</sub> y ADN<sub>c</sub>.

La Figura 4A representa la visualización del mapeo de las lecturas utilizando el Integrative Genomics Viewer. La Figura 4B representa la secuencia de dos secuencias de fusión sometidas a corte y empalme alternativos. La Figura 4C representa una tabla sumario que informa de los transcritos en los genes implicados comentada con detalles de exón, marco y cobertura de lectura de fusión.

La Figura 5 representa los resultados de un ejemplo de carrera de secuenciación.

Las Figuras 6 y 7 representan ejemplos de los resultados de las carreras de secuenciación para las secuencias de ALK y RET, respectivamente.

La Figura 8 representa una presentación esquemática del planteamiento de secuenciación dirigida descrita en el presente documento.

La Figura 9 representa esquemas de adaptadores de cola de oligonucleótidos universales ilustrativos.

### Descripción detallada

Las realizaciones de la tecnología descrita en el presente documento se refieren a métodos para determinar (es decir, secuenciar) secuencias oligonucleotídicas. En algunas realizaciones, los métodos descritos en el presente documento se refieren a métodos para enriquecimiento de secuencias diana antes de una etapa de secuenciación. En algunas realizaciones, la secuencia de un extremo de la secuencia diana que hay que enriquecer no es conocida antes de la etapa de secuenciación.

Por comodidad, a continuación, se proporciona el significado de algunos términos y expresiones utilizados en la memoria descriptiva, los ejemplos y las reivindicaciones adjuntas. A menos que se indique lo contrario, o esté implícito en el contexto, los siguientes términos y expresiones incluyen los significados proporcionados a continuación. Las definiciones se proporcionan para ayudar en la descripción de realizaciones particulares, y no pretenden limitar la invención reivindicada, debido a que el alcance de la invención está limitado únicamente por las reivindicaciones. A menos que se defina de otra forma, todos los términos técnicos y científicos utilizados en el presente documento tienen el mismo significado que el entendido comúnmente por un experto en la materia a la cual pertenece la presente invención. Si hay una aparente discrepancia entre la utilización de un término en la técnica y su definición proporcionada en el presente documento, prevalecerá la definición proporcionada en la memoria descriptiva.

Por comodidad, determinados términos empleados en el presente documento, en la memoria descriptiva, los ejemplos y las reivindicaciones adjuntas se recogen aquí.

Las expresiones "aumentar", "reducido", "reducción" o "inhibir" todas se utilizan en el presente documento, en general, para significar una disminución en una cantidad estadísticamente significativa. Sin embargo, para disipar cualquier duda, "reducido", "reducción", "disminuir" o "inhibir" significa una disminución de al menos un 10 % en comparación con un nivel de referencia, por ejemplo, una disminución de al menos aproximadamente un 20 %, o al menos aproximadamente un 30 %, o al menos aproximadamente un 40 %, o al menos aproximadamente un 50 %, o al menos aproximadamente un 60 %, o al menos aproximadamente un 70 %, o al menos aproximadamente un 80 %, o al menos aproximadamente un 90 % o hasta e incluido una disminución de un 100 % (por ejemplo, nivel ausente o nivel no

detectable en comparación con un nivel de referencia), o cualquier disminución entre 10-100 % en comparación con un nivel de referencia. En el contexto de un marcador o síntoma significa una disminución estadísticamente significativa en tal nivel. La disminución puede ser de, por ejemplo, al menos un 10 %, al menos un 20 %, al menos un 30 %, al menos un 40 % o más, y preferentemente se baja a un nivel aceptado como dentro del intervalo de normal para un individuo sin tal trastorno.

Las expresiones "aumentado", "aumentar", "potenciar" o "activar" todas se utilizan en el presente documento, en general, para significar un aumento en una cantidad estadísticamente significativa; para disipar cualquier duda, las expresiones "aumentado", "aumentar", "potenciar" o "activar" significan un aumento de al menos un 10 % en comparación con un nivel de referencia, por ejemplo, un aumento de al menos aproximadamente un 20 %, o al menos aproximadamente un 30 %, o al menos aproximadamente un 40 %, o al menos aproximadamente un 50 %, o al menos aproximadamente un 60 %, o al menos aproximadamente un 70 %, o al menos aproximadamente un 80 %, o al menos aproximadamente un 90 % o hasta e incluido un aumento de un 100 % o cualquier aumento entre 10-100 % en comparación con un nivel de referencia, o al menos un aumento de aproximadamente 2 veces, o al menos aproximadamente 3 veces, o al menos aproximadamente 4 veces, o al menos aproximadamente 5 veces o al menos aproximadamente 10 veces, o cualquier aumento entre 2 veces y 10 veces o superior en comparación con un nivel de referencia.

Como se utiliza en el presente documento, un "sujeto" significa un ser humano o un animal. Por lo general, el animal es un vertebrado, tal como un primate, roedor, animal doméstico o animal de caza. Los primates incluyen chimpancés, monos cynomolgous, monos araña y macacos, por ejemplo, Rhesus. Los roedores incluyen ratones, ratas, marmotas, hurones, conejos y hámsteres. Los animales domésticos y de caza incluyen vacas, caballos, cerdos, ciervos, bisontes, búfalos, especies felinas, por ejemplo, gato doméstico, especies caninas, por ejemplo, perro, zorro, lobo, especies avícolas, por ejemplo, pollo, emú, avestruz, y peces, por ejemplo, trucha, bagre y salmón. En algunas realizaciones, el sujeto es un mamífero, por ejemplo, un primate, por ejemplo, un ser humano. Los términos, "individuo", "paciente" y "sujeto" se utilizan en el presente documento de manera intercambiable.

Preferentemente, el sujeto es un mamífero. El mamífero puede ser un ser humano, primate no humano, ratón, rata, perro, gato, caballo o vaca, pero no se limita a estos ejemplos. Los mamíferos distintos de los seres humanos se utilizan de manera ventajosa como sujetos que representan modelos animales de, por ejemplo, cáncer de pulmón. Un sujeto puede ser macho o hembra.

Un sujeto puede ser uno al que previamente se le ha diagnosticado o se ha identificado como que padece o que tiene una afección que necesita tratamiento (por ejemplo, cáncer) o una o más complicaciones relacionadas con tal afección, y opcionalmente, ya se ha sometido a tratamiento para la afección o la complicación o complicaciones relacionadas con la afección. Alternativamente, un sujeto también puede ser uno al que previamente no se le ha diagnosticado que tenga la afección (por ejemplo, cáncer) o una o más complicaciones relacionadas con tal afección. Por ejemplo, un sujeto puede ser uno que presenta uno o más factores de riesgo para la afección o una o más complicaciones relacionadas con la afección o un sujeto que no presenta factores de riesgo.

Un "sujeto que necesita" tratamiento para una afección particular puede ser un sujeto que tiene esa afección, ha sido diagnosticado como que tiene esa afección, o está en riesgo de desarrollar esa afección.

Como se utiliza en el presente documento, una "enfermedad asociada a una alteración genética" se refiere a cualquier enfermedad que está causada por, al menos en parte, una alteración en el material genético del sujeto en comparación con un sujeto sano tipo silvestre, por ejemplo, una delección, una inserción, un SNP, un reordenamiento de genes. Una enfermedad puede estar causada por, al menos en parte, una alteración en el material genético del sujeto si la alteración aumenta el riesgo del sujeto de desarrollar la enfermedad, aumenta la susceptibilidad del sujeto a una enfermedad (incluidas enfermedades infecciosas o enfermedades con un componente infeccioso), causa la producción de una molécula asociada a enfermedad, o causa que las células lleguen a estar enfermas o ser anormales (por ejemplo, pérdida de la regulación del ciclo celular en células cancerosas). Las enfermedades pueden estar asociadas a múltiples alteraciones genéticas, por ejemplo, cánceres.

Como se utiliza en el presente documento, la expresión "ácido nucleico" o "secuencia de ácido nucleico" se refiere a cualquier molécula, preferentemente, una molécula polimérica, que incorpora unidades de ácido ribonucleico, ácido desoxirribonucleico o un análogo de los mismos. El ácido nucleico puede ser monocatenario o bicatenario. Un ácido nucleico monocatenario puede ser un ácido nucleico de una hebra de un ADN bicatenario desnaturalizado. Alternativamente, puede ser un ácido nucleico monocatenario no derivado de ningún ADN bicatenario. En un aspecto, el ácido nucleico molde es ADN. En otro aspecto, el molde es ARN. Las moléculas de ácido nucleico adecuadas son ADN, incluido ADN genómico o ADNc. Otras moléculas de ácido nucleico adecuadas son ARN, incluido ARNm.

La expresión "aislado" o "parcialmente purificado" como se utiliza en el presente documento se refiere, en el caso de un ácido nucleico, a un ácido nucleico separado de al menos otro componente (por ejemplo, ácido nucleico o polipéptido) que está presente con el ácido nucleico como se encuentra en su fuente natural y/o que estaría presente con el ácido nucleico cuando es expresado por la célula. Un ácido nucleico químicamente sintetizado o uno sintetizado utilizando transcripción/traducción *in vitro* se considera "aislado".

El término "gen" significa una secuencia de ácido nucleico que se transcribe (ADN) a ARN *in vitro* o *in vivo* cuando se operativamente liga a secuencias reguladoras apropiadas. El gen pueden incluir regiones reguladoras que preceden y siguen la región codificadora, por ejemplo, secuencias no traducidas 5' (5' UTR) o "líder" y secuencias 3' UTR o "remolque", así como secuencias intermedias (intrones) entre los segmentos codificadores individuales (exones).

Como se utiliza en el presente documento, el término "complementario" se refiere a la jerarquía de las preferencias de formación de pares de bases unidos por hidrógeno entre las bases nucleotídicas G, A, T, C y U, de modo que cuando dos polinucleótidos o secuencias de polinucleótidos dados se hibridan entre sí, A se empareja con T y G se empareja con C en el ADN, y G se empareja con C y A se empareja con U en el ARN. Como se utiliza en el presente documento, "sustancialmente complementaria" se refiere a una molécula de ácido nucleico o porción de la misma (por ejemplo, un cebador) que tiene al menos un 90 % de complementariedad a lo largo de la longitud completa de la molécula o porción de la misma con una segunda secuencia nucleotídica, por ejemplo, 90 % complementaria, 95 % complementaria, 98 % complementaria, 99 % complementaria o 100 % complementaria. Como se utiliza en el presente documento, "sustancialmente idéntica" se refiere a una molécula de ácido nucleico o porción de la misma que tiene al menos un 90 % de identidad a lo largo de la longitud completa de la molécula o porción de la misma con una segunda secuencia nucleotídica, por ejemplo, 90 % de identidad, 95 % de identidad, 98 % de identidad, 99 % de identidad o 100 % de identidad.

Como se utiliza en el presente documento, "específico" cuando se utiliza en el contexto de un cebador específico de un ácido nucleico diana se refiere a un nivel de complementariedad entre el cebador y la diana de modo que existe una temperatura de hibridación en la que el cebador se hibridará y mediará la amplificación del ácido nucleico diana y no se hibridará o mediará la amplificación de secuencias que no son diana presentes en una muestra.

Como se utiliza en el presente documento, "producto amplificado", "producto de amplificación" o "amplicón" se refiere a oligonucleótidos resultantes de una reacción de PCR que son copias de una porción de una hebra molde particular de ácido nucleico diana y/o su secuencia complementaria, que corresponde en secuencia nucleotídica a la secuencia oligonucleotídica molde y/o su secuencia complementaria. Un producto de amplificación además puede comprender secuencia específica a los cebadores y que flanquea la secuencia que es una porción del ácido nucleico diana y/o su complemento. Un "producto amplificado", como se describe en el presente documento en general será ADN bicatenario, aunque la referencia puede hacerse a sus hebras individuales.

Como se utiliza en el presente documento, una "porción" de una molécula de ácido nucleico se refiere al conjunto contiguo de los nucleótidos comprendidos por esa molécula. Una porción puede comprender todos o solo un subconjunto de los nucleótidos comprendidos por la molécula. Una porción puede ser bicatenaria o monocatenaria.

Como se utiliza en el presente documento, las expresiones "tratar", "tratamiento", "que trata", o "mejora" se refieren a tratamientos terapéuticos, en donde el objetivo es revertir, aliviar, mejorar, inhibir, ralentizar o parar la progresión o la gravedad de una afección asociada a una enfermedad o trastorno, por ejemplo, cáncer de pulmón. El término "tratar" incluye reducir o aliviar al menos un efecto adverso o síntoma de una afección, enfermedad o trastorno asociado a una dolencia. El tratamiento es en general "efectivo" si se reducen uno o más síntomas o marcadores clínicos. Alternativamente, el tratamiento es "efectivo" si se reduce o para la progresión de una enfermedad. Es decir, "tratamiento" incluye no solo la mejora de los síntomas o los marcadores, sino también un cese de, o al menos una ralentización de, el progreso o empeoramiento de los síntomas en comparación con lo que se esperaría en ausencia de tratamiento. Los resultados clínicos beneficiosos o deseados incluyen, pero sin limitación, alivio de uno o más síntomas, disminución del grado de la enfermedad, una patología estabilizada (es decir, que no empeora), retraso o ralentización de la progresión de la enfermedad, mejora o paliación de la patología, remisión (ya sea parcial o total), y/o mortalidad disminuida, ya sea detectable o no detectable. El término "tratamiento" de una enfermedad también incluye proporcionar alivio de los síntomas o los efectos secundarios de la enfermedad (incluido el tratamiento paliativo).

La expresión "estadísticamente significativo" o "significativamente" se refiere a la significancia estadística y en general significa dos desviaciones típicas (2SD) por debajo de la concentración normal, o inferior, del catalizador.

Excepto en los ejemplos operativos, o en donde se indique lo contrario, todos los números que expresan cantidades de ingredientes o condiciones de reacción utilizados en el presente documento deben entenderse como modificados en todos los ejemplos por el término "aproximadamente". La expresión "aproximadamente" cuando se utiliza junto con porcentajes puede significar  $\pm 1$  %.

Como se utiliza en el presente documento, la expresión "que comprende" o "comprende" se utilizan en referencia a composiciones, métodos y respectivos componentes de los mismos, que son esenciales para el método o a la composición, pero abiertos a la inclusión de elementos no especificados, ya sean esenciales o no.

La expresión "que consiste en" se refiere a composiciones, métodos y componentes correspondientes de los mismos como se describe en el presente documento, que son exclusivos de cualquier elemento no mencionado en esa descripción de la realización.

Como se utiliza en el presente documento, la expresión "que consiste esencialmente en" se refiere a los elementos requeridos para una realización dada. La expresión permite la presencia de elementos que no afectan materialmente a las características básicas y novedosas o funcionales de esa realización.

Los términos en singular "un", "uno/una", y "el/la" incluyen los referentes plurales a menos que el contexto indique claramente lo contrario. De manera similar, se pretende que la palabra "o" incluya "y" a menos que el contexto indique claramente lo contrario. Aunque pueden utilizarse métodos y materiales similares o equivalentes a los descritos en el presente documento a la hora de poner en práctica o probar la presente divulgación, a continuación, se describen los métodos y materiales adecuados. La abreviatura, "e.g" procede del Latín *exempli gratia* y se utiliza en el presente documento para indicar un ejemplo no limitante. Por tanto, la abreviatura "e.g" es sinónimo de la expresión "por ejemplo".

Definiciones de términos comunes en biología celular y biología molecular pueden encontrarse en "The Merck Manual of Diagnosis and Therapy", 19ª edición, publicado por Merck Research Laboratories, 2006 (ISBN 0-911910-19-0); Robert S. Porter et al. (eds.), "The Encyclopedia of Molecular Biology", publicado por Blackwell Science Ltd., 1994 (ISBN 0-632-02182-9). También pueden encontrarse definiciones de términos comunes en biología molecular en Benjamin Lewin, "Genes X", publicado por Jones & Bartlett Publishing, 2009 (ISBN-10: 0763766321); Kendrew et al. (eds.), "Molecular Biology and Biotechnology: a Comprehensive Desk Reference", publicado por VCH Publishers, Inc., 1995 (ISBN 1-56081-569-8) y "Current Protocols in Protein Sciences" 2009, Wiley Intersciences, Coligan et al., eds.

A menos que se indique lo contrario, la presente invención se realizó utilizando procedimientos estándares, como se describe, por ejemplo, en Sambrook et al., "Molecular Cloning: A Laboratory Manual" (3 ed.), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., EE. UU. (2001); y Davis et al., "Basic Methods in Molecular Biology", Elsevier Science Publishing, Inc., Nueva York, EE. UU. (1995).

Otros términos se definen en el presente documento dentro de la descripción de los diferentes aspectos de la invención.

En el presente documento se describen métodos para determinar la secuencia nucleotídica contigua a una secuencia nucleotídica diana conocida. Métodos de secuenciación tradicionales generan información de secuencia aleatoria (por ejemplo, secuenciación escopeta (*shotgun*) o entre dos secuencias conocidas que se utilizan para diseñar cebadores. Por el contrario, los métodos descritos en el presente documento, en algunas realizaciones, permiten determinar la secuencia nucleotídica (por ejemplo, secuenciar) en dirección 5' o en dirección 3' de una región única de la secuencia conocida con un elevado nivel de especificidad y sensibilidad.

En algunas realizaciones, los métodos descritos en el presente documento se refieren a un método de enriquecimiento de secuencias nucleotídicas específicas antes de determinar la secuencia nucleotídica utilizando una tecnología de secuenciación de última generación. En algunas realizaciones, los métodos de enriquecimiento de secuencias nucleotídicas específicas no comprenden enriquecimiento por hibridación.

En algunas realizaciones, la tecnología descrita en el presente documento puede referirse a un método para determinar la secuencia nucleotídica contigua a una secuencia nucleotídica diana conocida, comprendiendo el método; (a) ligar un ácido nucleico diana que comprende la secuencia nucleotídica diana conocida a un adaptador de cola de oligonucleótido universal; (b) amplificar una porción del ácido nucleico diana y la hebra de amplificación del adaptador de cola de oligonucleótido universal con un primer cebador adaptador y un primer cebador específico de diana; (c) amplificar una porción del amplicón resultante de la etapa (b) con un segundo cebador adaptador y un segundo cebador específico de diana; y (d) secuenciar la porción amplificada de la etapa (c) utilizando un primer y un segundo cebador de secuenciación. Como se utiliza en el presente documento, la expresión "ácido nucleico diana" se refiere a una molécula de ácido nucleico que comprende tanto la secuencia de ácido nucleico que hay que determinar como la secuencia nucleotídica diana conocida. El ácido nucleico diana puede ser de cualquier longitud y puede ser bicatenario o monocatenario. Como se utiliza en el presente documento, la expresión "secuencia nucleotídica diana conocida" se refiere a una porción de un ácido nucleico diana para la cual se conoce la secuencia (por ejemplo, la identidad y el orden de las bases nucleotídicas que comprenden el ácido nucleico. La secuencia nucleotídica diana conocida puede ser de cualquier longitud de 10 o más nucleótidos, preferentemente 30 o más nucleótidos (por ejemplo, 30 nucleótidos, 40 nucleótidos, 50 nucleótidos o más). Como se utiliza en el presente documento, la expresión "secuencia nucleotídica contigua a" se refiere a una secuencia nucleotídica que está localizada en la misma molécula de ácido nucleico (es decir, el ácido nucleico diana) como la secuencia nucleotídica diana conocida y o bien en dirección 5' o en dirección 3' de la secuencia nucleotídica diana conocida. La secuencia nucleotídica contigua puede comprender cualquier longitud de la secuencia nucleotídica. En algunas realizaciones, la secuencia nucleotídica contigua a la secuencia nucleotídica diana conocida comprende 1 kb o menos de la secuencia nucleotídica, por ejemplo, 1 kb o menos de la secuencia nucleotídica, 750 pb o menos de la secuencia nucleotídica, 500 pb o menos de la secuencia nucleotídica, 400 pb o menos de la secuencia nucleotídica, 300 pb o menos de la secuencia nucleotídica, 200 pb o menos de la secuencia nucleotídica, 100 pb o menos de la secuencia nucleotídica. Cuando una muestra comprende diferentes ácidos nucleicos diana que comprenden la secuencia nucleotídica diana conocida (por ejemplo, una célula donde la

secuencia nucleotídica diana conocida se da múltiples veces en el genoma, o por separado, cromosomas no idénticos), puede haber múltiples secuencias que comprendan la "secuencia nucleotídica contigua a" la secuencia nucleotídica diana conocida. Como se utiliza en el presente documento, la expresión "determinar la secuencia nucleotídica", se refiere a determinar la identidad y las posiciones relativas de las bases nucleotídicas que comprenden un ácido nucleico.

En la etapa (a) del método descrito en el presente documento, el adaptador de cola de oligonucleótido universal se puede ligar al ácido nucleico diana. En algunas realizaciones, el ácido nucleico diana puede estar comprendido por una muestra que comprende una pluralidad de ácidos nucleicos, algunos de los cuales no comprenden el ácido nucleico diana. En algunas realizaciones, el adaptador de cola de oligonucleótido universal se puede ligar sustancialmente a todos los ácidos nucleicos en una muestra. En algunas realizaciones, el adaptador de cola de oligonucleótido universal se puede ligar a los ácidos nucleicos que comprenden la secuencia de ácido nucleico diana como a los ácidos nucleicos que no comprenden la secuencia de ácido nucleico diana.

Como se utiliza en el presente documento, la expresión "adaptador de cola de oligonucleótido universal" se refiere a una molécula de ácido nucleico comprendida por dos hebras (una hebra de bloqueo y una hebra de amplificación) y que comprende un primer extremo dúplex que se puede ligar y un segundo extremo deshibridado. La hebra de bloqueo del adaptador de cola de oligonucleótido universal comprende una porción dúplex 5'. La hebra de amplificación comprende una porción 5' deshibridada, una porción dúplex 3' y un saliente T 3' y secuencias de ácidos nucleicos idénticas a un primer y un segundo cebador de secuenciación. Las porciones dúplex de la hebra de bloqueo y la hebra de amplificación son sustancialmente complementarias y forman el primer extremo dúplex que se puede ligar que comprende un saliente T 3' y la porción dúplex es de suficiente longitud para permanecer en forma dúplex a la temperatura de ligamiento.

En algunas realizaciones, la porción de la hebra de amplificación que comprende una secuencia de ácido nucleico idéntica a un primer y un segundo cebador de secuenciación puede estar comprendida, al menos en parte, por la porción deshibridada 5' de la hebra de amplificación.

En algunas realizaciones, el adaptador de cola de oligonucleótido universal puede comprender una porción dúplex y una porción deshibridada, en donde la porción deshibridada comprende solo la porción 5' de la hebra de amplificación, es decir, la hebra de bloqueo en su totalidad es una porción dúplex.

En algunas realizaciones, el adaptador de cola de oligonucleótido universal puede tener una forma "Y", es decir, la porción deshibridada puede comprender porciones de la hebra de bloqueo como de la hebra de amplificación que están deshibridadas. La porción deshibridada de la hebra de bloqueo puede ser más corta, más larga, o igual a la longitud de la porción deshibridada de la hebra de amplificación. En algunas realizaciones, la porción deshibridada de la hebra de bloqueo puede ser más corta que la porción deshibridada de la hebra de amplificación. Los adaptadores de cola de oligonucleótidos universales en forma Y tienen la ventaja de que la porción deshibridada de la hebra de bloqueo no se someterá a extensión 3' durante un régimen de PCR.

En algunas realizaciones, la hebra de bloqueo del adaptador de cola de oligonucleótido universal además puede comprender una porción deshibridada 3' que no es sustancialmente complementaria a la porción deshibridada 5' de la hebra de amplificación; y en donde la porción deshibridada 3' de la hebra de bloqueo no es sustancialmente complementaria a o sustancialmente idéntica a cualquiera de los cebadores. En algunas realizaciones, la hebra de bloqueo del adaptador de cola de oligonucleótido universal además puede comprender una porción deshibridada 3' que no se hibridará de forma específica con la porción deshibridada 5' de la hebra de amplificación a la temperatura de hibridación; y en donde la porción deshibridada 3' de la hebra de bloqueo no se hibridará de forma específica con ninguno de los cebadores o sus complementos a la temperatura de hibridación.

En algunas realizaciones, la porción dúplex del adaptador de cola de oligonucleótido universal (por ejemplo, las porciones dúplex de una cualquiera o ambas hebras) es de al menos 7 pares de bases de longitud, por ejemplo, 7 pb o más, 8 pb o más, 9 pb o más, 10 pb o más, 11 pb o más, 12 pb o más, 13 pb o más o 14 pb o más de longitud. En algunas realizaciones, la porción dúplex de un adaptador de cola de oligonucleótido universal puede ser de al menos 30 pb o más, por ejemplo, 30 pb o más, 31 pb o más, 32 pb o más, 33 pb o más, 34 pb o más, 35 pb o más, 40 pb o más o 50 pb o más de longitud. La porción dúplex de un adaptador de cola de oligonucleótido universal no debe ser tan largo como para suprimir la amplificación por PCR de los amplicones deseados en el régimen de amplificación por PCR para usar. Algunos métodos de secuenciación de última generación utilizan las moléculas de adaptador en forma de Y. Estas moléculas de adaptador en forma de Y requieren porciones dúplex que son de longitud limitada, (por ejemplo, 17 pb o menos) para evitar la formación de horquillas intramoleculares durante varias etapas de la PCR (por ejemplo, PCR con enriquecimiento de genoteca, PCR puente, o PCR en emulsión). Los adaptadores de cola de oligonucleótidos universales en forma de Y de los métodos descritos en el presente documento no se someten a esta limitación del extremo dúplex como los dos Este efecto de supresión de la PCR resultante del extremo dúplex no es aplicable en esta invención puesto que los dos cebadores diana pueden acceder al fragmento diana de una manera interna. En algunas realizaciones, la porción dúplex de un adaptador de cola de oligonucleótido universal puede ser de al menos 18 pb o más, por ejemplo, 18 pb o más, 19 pb o más, 20 pb o más, 21 pb o más, 22 pb o más, 23 pb o más, 24 pb o más o 25 pb o más de longitud.

Ejemplos ilustrativos de adaptadores de cola de oligonucleótidos universales se muestran en la Figura 9.

El ligamiento del adaptador de cola de oligonucleótido universal se puede conseguir mediante cualquier método conocido en la técnica, por ejemplo, ligamiento de extremo romo o ligamiento TA. En algunas realizaciones, antes del ligamiento del adaptador de cola de oligonucleótido universal, los ácidos nucleicos en una muestra pueden someterse a reparación de extremo del ácido nucleico para dejar romos los extremos del ácido nucleico. La reparación de extremo es bien conocida en la técnica y en el mercado hay disponibles kits y/o enzimas relevantes, (por ejemplo, la molécula de reparación de extremo NEBNEXT™ (n.º de cat E6050L; New England Biolabs; Ipswich, MA).

En algunas realizaciones, antes del ligamiento del adaptador de cola de oligonucleótido universal, los ácidos nucleicos en una muestra pueden someterse a fosforilación y/o adenilación. La adenilación puede proporcionar un saliente de adenosina en el extremo 3' de un ácido nucleico. A continuación, se puede ligar un segundo ácido nucleico con un saliente 3' de tionina al primer ácido nucleico por ligamiento TA. Los métodos de ligamiento son bien conocidos en la técnica y en el mercado hay disponibles kits y/o enzimas relevantes, por ejemplo, la molécula de adición de cola dA NEBNEXT™ (n.º de cat E6053L; New England Biolabs; Ipswich, MA) puede utilizarse para la adenilación de un extremo romo de un ácido nucleico. En algunas realizaciones, los adaptadores de cola de oligonucleótidos universales pueden proporcionarse con un saliente 3' de tionina.

Cada una de las etapas (b) y (c) de los métodos descritos en el presente documento puede comprender un régimen de amplificación por PCR, es decir, un conjunto de ciclos de amplificación por la reacción en cadena de la polimerasa (PCR). Como se utiliza en el presente documento, la expresión "régimen de amplificación" se refiere a un proceso de amplificación de forma específica, es decir, de aumento de la abundancia de, una secuencia de ácido nucleico de interés, y más particularmente, la amplificación exponencial que se da cuando los productos de una extensión por polimerasa previa sirven como moldes para las sucesivas rondas de extensión. Un régimen de amplificación por PCR de acuerdo con la invención comprende al menos uno, y preferentemente al menos 5 o más ciclos iterativos, donde cada ciclo comprende las etapas de: 1) separación de hebra (por ejemplo, desnaturalización térmica); 2) hibridación del cebador oligonucleotídico con moléculas modelo; y 3) extensión por polimerasa de ácido nucleico de los cebadores hibridados. Las condiciones y los tiempos necesarios para cada una de estas etapas son planeados por el experto en la materia. Un régimen de amplificación de acuerdo con los métodos descritos en el presente documento preferentemente se realiza en un termociclador, muchos de los cuales están disponibles en el mercado.

La PCR requiere la utilización de una polimerasa de ácido nucleico. Como se utiliza en el presente documento, la expresión "polimerasa de ácido nucleico" se refiere a una enzima que cataliza la polimerización dependiente de molde de nucleósido trifosfatos para formar productos de extensión de cebador que son complementarios a la secuencia del ácido nucleico molde. Una enzima polimerasa de ácido nucleico inicia la síntesis en el extremo 3' de un cebador hibridado y avanza en dirección hacia el extremo 5' del molde. Numerosas polimerasas de ácido nucleico son conocidas en la técnica y están disponibles en el mercado. Un grupo de polimerasas de ácido nucleico preferidas son termoestables, es decir, conservan la función después de ser sometidas a temperaturas suficientes para desnaturalizar las hebras hibridadas de los ácidos nucleicos complementarios, por ejemplo, 94 °C o a veces superior.

Como se entiende en la técnica, la PCR requiere ciclos que incluyen una etapa de separación de hebra que en general implica calentar la mezcla de reacción. Como se utiliza en el presente documento, la expresión "separación de hebra" o "separar las hebras" significa tratamiento de una muestra de ácido nucleico de modo que las moléculas bicatenarias complementarias se separen en dos hebras sencillas disponibles para la hibridación con un cebador de oligonucleótido. Más específicamente, la separación de hebra de acuerdo con los métodos descritos en el presente documento, se consigue calentando la muestra de ácido nucleico por encima de su  $T_m$ . En general, para una muestra que contiene moléculas de ácido nucleico en tampón adecuado para una polimerasa de ácido nucleico, el calentamiento a 94 °C es suficiente para conseguir la separación de hebra. Un tampón ilustrativo contiene KCl 50 mM, Tris-HCl 10 mM (pH 8,8 a 25 °C),  $MgCl_2$  0,5 a 3 mM y BSA al 0,1 %.

Como también se entiende en la técnica, la PCR requiere de cebadores de hibridación con ácidos nucleicos molde. Cualquier hebra de un ácido nucleico diana puede ser un ácido nucleico molde, ya que el ácido nucleico molde se define como un ácido nucleico monocatenario con el cual se hibridará de forma específica un cebador dado. Como se utiliza en el presente documento, "hibridar" se refiere a permitir que dos hebras de ácido nucleico complementarias o sustancialmente complementarias hibriden, y más particularmente, cuando se usa en el contexto de la PCR, hibriden de modo que se forme un sustrato de extensión de cebador para una enzima polimerasa dependiente de molde. Las condiciones para la hibridación del ácido nucleico diana y cebador varían con la longitud y la secuencia del cebador y se basan en la  $T_m$  calculada para el cebador. En general, una etapa de hibridación en un régimen de amplificación implica reducir la temperatura tras la etapa de separación de hebra hasta una temperatura en función de la  $T_m$  calculada para la secuencia del cebador, durante un tiempo suficiente para permitir tal hibridación. La  $T_m$  puede ser prevista fácilmente por un experto en la materia utilizando cualquiera de una cantidad de algoritmos ampliamente disponibles (por ejemplo, programa informático de diseño de cebador QLIGO™ (Molecular Biology Insights Inc. Colorado) y programa informático de diseño de cebador VENTRO NTI™ (Invitrogen, Inc. California) y programas disponibles en internet, incluidos Primer3, Oligo Calculator, and NetPrimer (Premier Biosoft; Palo Alto, CA; y gratuitos en la página <http://www.premierbio-soft.com/netprimer/netprlaunch/Help/xnetprlaunch.html>). Por ejemplo, la  $T_m$  de un

cebador puede calcularse utilizando la siguiente fórmula, que es utilizada por el programa informático NetPrimer y está descrita con más detalle en Frieir et al. *PNAS* 1986 83:9373-9377.

$$T_m = \Delta H / (\Delta S + R \cdot \ln(C/4)) + 16,6 \log ([K^+] / (1 + 0,7 [K^+])) - 273,15$$

en donde,  $\Delta H$  es entalpía para la formación de la hélice;  $\Delta S$  es entropía para la formación de la hélice;  $R$  es la constante molar de un gas (1,987 cal/°C \* mol);  $C$  es la concentración del ácido nucleico; y  $[K^+]$  es la concentración de sal. Para la mayoría de los regímenes de amplificación, la temperatura de hibridación se selecciona para ser de aproximadamente 5 °C por debajo de la  $T_m$  prevista, aunque pueden usarse temperaturas más cercanas a y por encima de la  $T_m$  (por ejemplo, entre 1 °C y 5 °C por debajo de la  $T_m$  prevista o entre 1 °C y 5 °C por encima de la  $T_m$  prevista), así como, por ejemplo, temperaturas de más de 5 °C por debajo de la  $T_m$  prevista (por ejemplo, 6 °C por debajo, 8 °C por debajo, 10 °C por debajo o inferior). En general, cuanto más cercana es la temperatura de hibridación a la  $T_m$ , más específico es la hibridación. El tiempo permitido para hibridar el cebador durante un régimen de amplificación por PCR depende mucho del volumen de la reacción, por lo que mayores volúmenes requieren mayores tiempos, pero también depende de las concentraciones de cebador y molde, por lo que concentraciones relativamente superiores de cebador a molde requieren menos tiempo que las concentraciones relativamente inferiores. Dependiendo del volumen y la concentración relativa de cebador/molde, las etapas de hibridación en un régimen de amplificación pueden estar en el orden de 1 segundo a 5 minutos, pero en general estarán entre 10 segundos y 2 minutos, preferentemente en el orden de 30 segundos a 2 minutos. Como se utiliza en el presente documento, "hibridar sustancialmente" se refiere a un grado de hibridación durante un régimen de amplificación por PCR que es suficiente para producir un nivel detectable de un producto amplificado de forma específica.

La PCR también depende de la extensión por polimerasa de los cebadores hibridados en cada ciclo. Como se utiliza en el presente documento, la expresión "extensión por polimerasa" significa la incorporación dependiente del molde de al menos un nucleótido complementario, mediante una polimerasa de ácido nucleico, sobre el extremo 3' de un cebador hibridado. La extensión por polimerasa preferentemente agrega más de un nucleótido, preferentemente hasta e incluido nucleótidos que corresponden a la longitud completa del molde. Las condiciones para la extensión por polimerasa varían con la identidad de la polimerasa. La temperatura utilizada para la extensión por polimerasa en general se basa en las propiedades de actividad conocidas de la enzima. No obstante, cuando se requiere que las temperatura de hibridación estén, por ejemplo, por debajo de las temperaturas óptimas para la enzima, con frecuencia será aceptable utilizar una temperatura de extensión inferior. En general, aunque las enzimas conserven al menos actividad parcial por debajo de sus temperaturas óptimas de extensión, la extensión por polimerasa por las polimerasas termoestables más comúnmente utilizadas (por ejemplo, polimerasa Taq y sus variantes) se realiza a 65 °C a 75 °C, preferentemente aproximadamente 68-72 °C.

La extensión del cebador se realiza en condiciones que permiten la extensión de los cebadores oligonucleotídicos hibridados. Como se utiliza en el presente documento, la expresión "condiciones que permiten la extensión de un oligonucleótido hibridado de modo que se generen productos de extensión" se refiere al conjunto de condiciones que incluyen, por ejemplo, la temperatura, las concentraciones de sal y el cofactor, pH y la concentración de enzima en las cuales una polimerasa de ácido nucleico cataliza la extensión del cebador. Tales condiciones variarán con la identidad de la polimerasa de ácido nucleico que se utilice, pero las condiciones para un gran número de enzimas polimerasa útiles son bien conocidas por el experto en la materia. Un conjunto ilustrativo de las condiciones es KCl 50 mM, Tris-HCl 10 mM (pH 8,8 a 25 °C), MgCl<sub>2</sub> 0,5 a 3 mM, 200 uM de cada dNTP, y BSA al 0,1 % a 72 °C, en las cuales la polimerasa Taq cataliza la extensión del cebador. Las condiciones para la iniciación y la extensión por lo general incluyen la presencia de al menos uno, pero más preferentemente los cuatro desoxinucleósido trifosfatos diferentes y un agente inductor de la polimerización tal como la ADN polimerasa o la transcriptasa inversa, en un tampón adecuado (en este contexto "tampón" incluye disolventes (en general acuosos) más cofactores necesarios y reactivos que afectan al pH, la fuerza iónica, etc.) y a una temperatura adecuada.

En algunas realizaciones, cada etapa de amplificación puede comprender un conjunto de ciclos de un régimen de amplificación por PCR de 5 ciclos a 20 ciclos de largo. En algunas realizaciones, cada etapa de amplificación puede comprender un conjunto de ciclos de un régimen de amplificación por PCR de 10 ciclos a 20 ciclos de largo. En algunas realizaciones, cada etapa de amplificación puede comprender un conjunto de ciclos de un régimen de amplificación por PCR de 12 ciclos a 16 ciclos de largo. En algunas realizaciones, la temperatura de hibridación puede ser de menos de 70 °C. En algunas realizaciones, la temperatura de hibridación puede ser de menos de 72 °C.

En diversas realizaciones, los métodos y composiciones descritos en el presente documento se refieren a la realización de un régimen de amplificación por PCR con uno o más de los tipos de cebadores descritos en el presente documento. Como se utiliza en el presente documento, "cebador" se refiere a una molécula de polinucleótido ADN o ARN o un análogo de la misma capaz de hibridarse de forma específica con un molde de polinucleótido y proporcionar un extremo 3' que sirva como sustrato para una polimerasa dependiente de molde para producir un producto de extensión que es complementario al molde de polinucleótido. Un cebador útil en los métodos descritos en el presente documento generalmente es monocatenario, y un cebador y su complemento pueden hibridarse para formar un polinucleótido bicatenario. Los cebadores de acuerdo con los métodos y composiciones descritos en el presente documento pueden ser de menos de o igual a 300 nucleótidos de longitud, por ejemplo, menos de o igual a 300 o 250, o 200, o 150, o 100, o 90, o 80, o 70, o 60, o 50, o 40, o 30 o inferior, o 20 o inferior, o 15 o inferior, pero al menos 10 nucleótidos de

longitud. Los métodos de fabricación de cebadores son bien conocidos en la técnica, y numerosas fuentes comerciales ofrecen servicios de síntesis de oligonucleótidos adecuados para proporcionar cebadores de acuerdo con los métodos y composiciones descritos en el presente documento, por ejemplo, INVITROGEN™ Custom DNA Oligos; Life Technologies; Grand Island, NY o custom DNA Oligos de IDT; Coralville, Iowa).

En algunas realizaciones, después de que el adaptador de cola de oligonucleótido universal se ligue a los ácidos nucleicos en una muestra (por ejemplo, los ácidos nucleicos diana), el ácido nucleico diana puede amplificarse en una primera etapa de amplificación (es decir, etapa (b)). La primera etapa de amplificación puede ser un conjunto de ciclos de amplificación por PCR que utilizan un primer cebador específico de diana y un primer cebador adaptador de cola.

Como se utiliza en el presente documento, la expresión "primer cebador específico de diana" se refiere a un oligonucleótido monocatenario que comprende una secuencia de ácido nucleico que puede hibridarse de forma específica con el ácido nucleico diana a la temperatura de hibridación.

En algunas realizaciones, el primer cebador específico de diana puede comprender una porción de la secuencia de la etiqueta 5'. En algunas realizaciones, todos los primeros cebadores específicos a diana presentes en una reacción pueden comprender porciones idénticas a la secuencia de la etiqueta 5'. En una reacción de PCR múltiple, diferentes especies de cebador pueden interactuar entre sí de una manera inespecífica no deseada, conduciendo a la extensión del cebador y posteriormente a la amplificación por ADN polimerasa. Estos dímeros de cebador tienden a ser cortos, y su amplificación eficaz puede sobrepasar la reacción y dominar dando como resultado una escasa amplificación de la secuencia diana deseada. La inclusión de una secuencia de la etiqueta 5' sobre el primer o los primeros cebadores específicos a diana causa algún dímero de cebador potencial que puede conservar las mismas colas complementarias sobre ambos extremos. En ciclos de PCR posteriores, los dímeros de cebadores se desnaturalizarían en dímeros de cebadores de ADN monocatenario, que comprenden cada uno secuencias complementarias sobre sus dos extremos que se introducen por la etiqueta 5'. En lugar de hibridación de cebador con estos dímeros de cebadores de ADN monocatenario, preferentemente se daría una horquilla intramolecular (una estructura tipo mango) debido a la accesibilidad próxima de las etiquetas complementarias en la misma molécula de dímero de cebador en lugar de una interacción intramolecular con nuevos cebadores en distintas moléculas. Como resultado, estos dímeros de cebadores se amplifican de forma muy ineficaz, de modo que los cebadores no se consumen de manera exponencial por los dímeros no deseados para la amplificación. En su lugar los cebadores marcados pueden permanecer en una concentración elevada y suficiente para la amplificación específica deseada de las secuencias diana. La acumulación de los dímeros de cebadores puede ser también un detrimento para la PCR múltiple debido a que compiten por y consumen otros reactivos en la reacción. En algunas realizaciones, la secuencia de la etiqueta 5' puede ser una secuencia rica en GC, es decir, la secuencia de la etiqueta puede comprender al menos un 50 % de contenido de GC, al menos un 55 % de contenido de GC, al menos un 60 % de contenido de GC, al menos un 65 % de contenido de GC, al menos un 70 % de contenido de GC, al menos un 75 % de contenido de GC, al menos un 80 % de contenido de GC, o contenido de GC superior. En algunas realizaciones, la secuencia de la etiqueta puede comprender al menos un 60 % de contenido de GC. En algunas realizaciones, la secuencia de la etiqueta puede comprender al menos un 65 % de contenido de GC.

Como se utiliza en el presente documento, la expresión "segundo cebador específico de diana" se refiere a un oligonucleótido monocatenario que comprende una porción 3' que comprende una secuencia de ácido nucleico que puede hibridarse de forma específica con una porción de la secuencia nucleotídica diana conocida comprendida por el amplicón resultante de la etapa (b), y una porción 5' que comprende una secuencia de ácido nucleico que es idéntica a un segundo cebador de secuenciación. El segundo cebador específico de diana es anidado con respecto al primer cebador específico de diana. En algunas realizaciones, el segundo cebador específico de diana es anidado con respecto al primer cebador específico de diana por al menos 3 nucleótidos, por ejemplo, por 3 o más, 4 o más, 5 o más, 6 o más, 7 o más, 8 o más, 9 o más, 10 o más o 15 o más nucleótidos.

En algunas realizaciones, todos los segundos cebadores específicos a diana presentes en una reacción comprenden la misma porción 5'. En algunas realizaciones, la porción 5' la porción 5' puede servir para suprimir los dímeros de cebadores como se describe para la etiqueta 5' del primer cebador específico de diana descrito anteriormente en el presente documento.

En algunas realizaciones, los primeros y segundos cebadores específicos a diana son sustancialmente complementarios a la misma hebra del ácido nucleico diana. En algunas realizaciones, las porciones de los primeros y segundos cebadores específicos a diana que se hibridan de forma específica con la secuencia diana conocida pueden comprender un total de al menos 20 bases únicas de la secuencia nucleotídica diana conocida, por ejemplo, 20 o más bases únicas, 25 o más bases únicas, 30 o más bases únicas, 35 o más bases únicas, 40 o más bases únicas o 50 o más bases únicas. En algunas realizaciones, las porciones de los primeros y segundos cebadores específicos a diana que se hibridan de forma específica con la secuencia diana conocida pueden comprender un total de al menos 30 bases únicas de la secuencia nucleotídica diana conocida,

Como se utiliza en el presente documento, la expresión "primer cebador adaptador" se refiere a una molécula de ácido nucleico que comprende una secuencia de ácido nucleico idéntica a una porción 5' del primer cebador de secuenciación. Como el primer cebador adaptador de cola es por ello idéntico a al menos una porción de la secuencia

de la hebra de amplificación (en contraposición al complementario), no será capaz de hibridarse de forma específica con ninguna porción del propio adaptador de cola de oligonucleótido universal.

Como se utiliza en el presente documento, la expresión "segundo cebador adaptador" se refiere a una molécula de ácido nucleico que comprende una secuencia de ácido nucleico idéntica a una porción del primer cebador de secuenciación y es anidado con respecto al primer cebador adaptador. Como el segundo cebador adaptador de cola es por ello idéntico a al menos una porción de la secuencia de la hebra de amplificación (en contraposición al complementario), no será capaz de hibridarse de forma específica con ninguna porción del propio adaptador de cola de oligonucleótido universal. En algunas realizaciones, el segundo cebador adaptador es idéntico al primer cebador de secuenciación.

El segundo cebador adaptador debe ser anidado con respecto al primer cebador adaptador, es decir, el primer cebador adaptador comprende una secuencia de ácido nucleico idéntica a la hebra de amplificación que no está comprendida por el segundo cebador adaptador y que está ubicada más cerca del extremo 5' del cebador de amplificación que cualquiera de las secuencias idénticas a la hebra de amplificación que está comprendida por el segundo cebador adaptador. En algunas realizaciones, el segundo cebador adaptador es anidado por al menos 3 nucleótidos, por ejemplo, por 3 nucleótidos, por 4 nucleótidos, por 5 nucleótidos, por 6 nucleótidos, por 7 nucleótidos, por 8 nucleótidos, por 9 nucleótidos, por 10 nucleótidos o más.

En algunas realizaciones, el primer cebador adaptador puede comprender una secuencia de ácido nucleico idéntica a aproximadamente las 20 bases más 5' de la hebra de amplificación del adaptador de cola de oligonucleótido universal y el segundo cebador adaptador puede comprender una secuencia de ácido nucleico idéntica a aproximadamente 30 bases de la hebra de amplificación del adaptador de cola de oligonucleótido universal, con una base 5' que es al menos 3 nucleótidos 3' del extremo 5' de la hebra de amplificación.

La utilización de los cebadores adaptadores anidados elimina la posibilidad de producir amplicones finales que sean amplificables (por ejemplo, durante una PCR puente o PCR en emulsión) pero que no se puedan secuenciar, una situación que puede surgir durante los métodos semianidados. En otras situaciones, los planteamientos semianidados que utilizan un cebador idéntico a un cebador de secuenciación pueden dar como resultado el paso de los productos de amplificación no deseados de la primera etapa de PCR a la segunda etapa de PCR y producirían por último lecturas de secuenciación artificiales. La utilización de dos cebadores adaptadores, como se describe en el presente documento puede reducir, y en algunas realizaciones eliminar, estos problemas.

En el primer ciclo de amplificación por PCR de la primera etapa de amplificación, el primer cebador específico de diana puede hibridarse de forma específica con una hebra molde de cualquier ácido nucleico que comprenda la secuencia nucleotídica diana conocida. Dependiendo de la orientación con la cual se diseñó el primer cebador específico de diana, se sintetizará la secuencia en dirección 5' o en dirección 3' de la secuencia nucleotídica diana conocida, y el complementario para la hebra molde. Si, durante la fase de extensión de la PCR, el extremo 5' de la hebra molde termina en un adaptador de cola de oligonucleótido universal, el extremo 3' del producto recién sintetizado comprenderá la secuencia complementaria al primer cebador adaptador de cola. En ciclos de amplificación por PCR posteriores, tanto el primer cebador específico de diana como el primer cebador adaptador de cola serán capaces de hibridarse de forma específica con las hebras apropiadas de la secuencia de ácido nucleico diana y pueden amplificarse la secuencia entre la secuencia de ácido nucleico diana conocida y el adaptador de cola de oligonucleótido universal, (es decir, copiada).

En la próxima etapa (es decir, etapa (c)) del método descrito en el presente documento, una porción de la porción amplificada resultante de la etapa (b) se amplifica en una segunda etapa de amplificación. La segunda etapa de amplificación puede ser un conjunto de ciclos de amplificación por PCR que utilizan un segundo cebador específico de diana y un primer cebador de secuenciación. El segundo conjunto de ciclos de amplificación por PCR pueden tener parámetros de PCR idénticos a, o que difieren de, aquellos del primer conjunto de ciclos de amplificación por PCR. Por ejemplo, los regímenes de amplificación por PCR de las etapas (b) y (c) pueden tener las mismas temperaturas de hibridación o diferentes o las mismas duraciones de la etapa de extensión o diferentes.

Los métodos descritos en el presente documento permiten determinar la secuencia nucleotídica contigua a una secuencia nucleotídica diana conocida sobre cualquiera de los flancos o ambos de la secuencia nucleotídica diana conocida. Independientemente de si el ácido nucleico diana existe normalmente como un ácido nucleico monocatenario o bicatenario, la información de secuencia normalmente se representa en un formato monocatenario (Hebra A), de 5' a 3'. Si hay que determinar la secuencia 5' de la secuencia nucleotídica diana conocida de la Hebra A, los cebadores específicos a gen pueden ser complementarios (es decir, hibridarse con) a la Hebra A. Si hay que determinar la secuencia 3' de la secuencia nucleotídica diana conocida de la Hebra A, los cebadores específicos a gen pueden ser idénticos a la Hebra A, de modo que se hibridarán con la hebra complementaria de un ácido nucleico diana bicatenario. Tales consideraciones del diseño del cebador son bien conocidas por los expertos en la materia.

En algunas realizaciones, los métodos descritos en el presente documento, que se refieren a la utilización de un primer y un segundo cebador específico de gen pueden dar como resultado ensayos con una tasa específica superior, por ejemplo, 70-90 %. En algunas realizaciones, los ensayos y los métodos descritos en el presente documento pueden

tener una tasa de especificidad de diana de al menos un 85 %.

En algunas realizaciones, los cuatro tipos de cebadores se diseñan de modo que se hibridarán de forma específica con sus secuencias complementarias a una temperatura de hibridación de aproximadamente 61 a 72 °C, por ejemplo, de aproximadamente 61 a 69 °C, de aproximadamente 63 a 69 °C, de aproximadamente 63 a 67 °C, de aproximadamente 64 a 66 °C. En algunas realizaciones, los cuatro tipos de cebadores se diseñan de modo que se hibridarán de forma específica con sus secuencias complementarias a una temperatura de hibridación de menos de 72 °C. En algunas realizaciones, los cuatro tipos de cebadores se diseñan de modo que se hibridarán de forma específica con sus secuencias complementarias a una temperatura de hibridación de menos de 70 °C. En algunas realizaciones, los cuatro tipos de cebadores se diseñan de modo que se hibridarán de forma específica con sus secuencias complementarias a una temperatura de hibridación de menos de 68 °C. En algunas realizaciones, los cuatro tipos de cebadores se diseñan de modo que se hibridarán de forma específica con sus secuencias complementarias a una temperatura de hibridación de aproximadamente 65 °C.

En algunas realizaciones, las porciones de los cebadores específicos a diana que se hibridan de forma específica con la secuencia nucleotídica diana conocida se hibridarán de forma específica a una temperatura de aproximadamente 61 a 72 °C, por ejemplo, de aproximadamente 61 a 69 °C, de aproximadamente 63 a 69 °C, de aproximadamente 63 a 67 °C, de aproximadamente 64 a 66 °C. En algunas realizaciones, las porciones de los cebadores específicos a diana que se hibridarán de forma específica con la secuencia nucleotídica diana conocida se hibridarán de forma específica a una temperatura de aproximadamente 65 °C en un tampón de PCR.

En algunas realizaciones, los cebadores y/o adaptadores descritos en el presente documento no pueden comprender bases modificadas (por ejemplo, los cebadores y/o adaptadores no pueden comprender una amina 3' de bloqueo).

En la próxima etapa (es decir, etapa (d)) de los métodos descritos en el presente documento, se puede secuenciar la porción amplificada resultante de la etapa (c). En algunas realizaciones, se puede secuenciar mediante un método de secuenciación de última generación. Como se utiliza en el presente documento, "secuenciación de última generación" se refiere a las tecnologías de secuenciación de oligonucleótidos que tienen la capacidad de secuenciar oligonucleótidos a velocidades por encima de aquellas posibles con los métodos de secuenciación convencionales (por ejemplo, secuenciación Sanger), debido a la realización y lectura de miles a millones de reacciones de secuenciación en paralelo. Ejemplos no limitantes de métodos/plataformas de secuenciación de última generación incluyen secuenciación de firma masivamente paralela (Lyns Therapeutics); pirosecuenciación 454 (454 Life Sciences/Roche Diagnostics); secuenciación de terminador con colorante reversible en fase sólida (Solexa/Illumina); tecnología SOLID (Applied Biosystems); secuenciación por semiconductores iónicos (Ion Torrent); secuenciación de nanoesferas de ADN (Complete Genomics); y tecnologías disponibles de Pacific biosciences, Intelligen Biosystems, Oxford Nanopore Technologies y Helicos Biosciences. En algunas realizaciones, los cebadores de secuenciación pueden comprender porciones compatibles con el método de secuenciación de última generación seleccionado. Las tecnologías de secuenciación de última generación y las restricciones y los parámetros de diseño de cebadores de secuenciación asociados son bien conocidos en la técnica (véase, por ejemplo, Shendure, et al., "Next-generation DNA sequencing", *Nature*, 2008, vol. 26, n.º 10, 1135-1145; Mardis, "The impact of next-generation sequencing technology on genetics", *Trends in Genetics*, 2007, vol. 24, n.º 3, págs. 133-141; Su, et al., "Next-generation sequencing and its applications in molecular diagnostics" *Expert. Rev. Mol. Diagn.*, 2011, 11(3):333-43; Zhang et al., "The impact of next-generation sequencing on genomics", *J. Genet. Genomics*, 2011, 38(3):95-109; Nyren, P. et al. *Anal Biochem* 208: 17175 (1993); Bentley, D. R. *Curr. Opin. Genet. Dev.* 16:545-52 (2006); Strausberg, R. L., et al. *Drug Disc Today* 13:569-77 (2008); patente de Estados Unidos n.º 7.282.337; patente de Estados Unidos n.º 7.279.563; patente de Estados Unidos n.º 7.226.720; patente de Estados Unidos n.º 7.220.549; patente de Estados Unidos n.º 7.169.560; patente de Estados Unidos n.º 6.818.395; patente de Estados Unidos n.º 6.911.345; las publicaciones de Estados Unidos n.º 2006/0252077; 2007/0070349; y 2007/0070349).

En algunas realizaciones, la etapa de secuenciación depende de la utilización de un primer y un segundo cebador de secuenciación. En algunas realizaciones, el primer y segundo cebador de secuenciación se selecciona por ser compatible con un método de secuenciación de última generación como se describe en el presente documento.

Los métodos para alinear las lecturas de secuenciación con las bases de datos de secuencia conocidas de secuencias de ADNc y/o genómico son bien conocidos en la técnica y el programa informático para este proceso está disponible en el mercado. En algunas realizaciones, las lecturas (menos la secuenciación de la secuencia nucleotídica del adaptador y/o cebador) que no se mapean, en su totalidad, en las bases de datos de secuencia tipo silvestre pueden ser reordenamientos genómicos o mutaciones indel grandes. En algunas realizaciones, las lecturas (menos la secuenciación de la secuencia nucleotídica de adaptador y/o cebador) que comprenden las secuencias que se mapean en para localizaciones múltiples en el genoma pueden ser reordenamientos genómicos.

En algunas realizaciones, los ácidos nucleicos y/o sus productos de amplificación pueden aislarse de enzimas, cebadores o componentes tampón antes y/o después de alguna de las etapas a-d. Los métodos para aislar los ácidos nucleicos son bien conocidos en la técnica. En algunas realizaciones, el aislamiento puede comprender limpieza por inmovilización reversible en fase sólida (SPRI). Los métodos para la limpieza por SPRI son bien conocidos en la técnica y se encuentran disponibles en el mercado, por ejemplo, purificación de PCR por Agencourt AMPure XP (n.º

de cat A63880, Beckman Coulter; Brea, CA). En algunas realizaciones, las enzimas pueden inactivarse mediante tratamiento térmico.

En algunas realizaciones, el ácido nucleico diana puede estar comprendido por ADN genómico. En algunas realizaciones, el ácido nucleico diana puede estar comprendido por ácido ribonucleico (ARN), por ejemplo, ARNm. En algunas realizaciones, el ácido nucleico diana puede estar comprendido por ADNc. Muchos de los métodos de secuenciación adecuados para su uso en los métodos descritos en el presente documento proporcionan carreras de secuenciación con longitudes óptimas de lectura de decenas a cientos de bases nucleotídicas (por ejemplo, la tecnología Ion Torrent puede producir longitudes de lectura de 200-400 pb). Los ácidos nucleicos diana comprendidos, por ejemplo, por ADN genómico o ARNm, pueden comprender moléculas de ácido nucleico que son sustancialmente mayores que esta longitud óptima de lectura. Para que la porción de ácido nucleico amplificada resultante de la etapa (c) sea de una longitud adecuada para su uso en una tecnología de secuenciación particular, la distancia media entre la secuencia nucleotídica diana conocida y un extremo del ácido nucleico diana al que se puede ligar el adaptador de cola de oligonucleótido universal debe ser tan próxima a la longitud óptima de lectura de la tecnología seleccionada como sea posible. Por ejemplo, si la longitud óptima de lectura de una tecnología de secuenciación dada es de 200 pb, entonces las moléculas de ácido nucleico amplificadas de acuerdo con los métodos descritos en el presente documento deben tener una longitud media de aproximadamente 400 pb o menos. Los ácidos nucleicos diana comprendidos por, por ejemplo, ADN genómico o ARNm, se pueden romper, por ejemplo, de forma mecánica o por enzimas, para generar fragmentos de cualquier tamaño deseado antes de la etapa (a). Ejemplos no limitantes de los procesos de ruptura mecánica incluyen sonicación, nebulización y la tecnología de ruptura AFA™ disponible de Covaris (Woburn, MA). En algunas realizaciones, un ácido nucleico diana comprendido por ADN genómico se puede romper de forma mecánica mediante sonicación. En algunas realizaciones, cuando el ácido nucleico diana está comprendido por ARN, la muestra puede someterse a un régimen de transcriptasa inversa para generar molde de ADN y, a continuación, el molde de ADN puede romperse. En algunas realizaciones, el ARN diana se puede romper antes de la realización del régimen de transcriptasa inversa. En algunas realizaciones, una muestra que comprende el ARN diana puede utilizarse en los métodos descritos en el presente documento utilizando ácidos nucleicos totales extraídos de especímenes frescos o degradados; sin la necesidad de eliminación de ADN genómico para la secuenciación de ADNc; sin la necesidad de agotamiento de ARN ribosómico para la secuenciación de ADNc; sin la necesidad de rotura mecánica o enzimática en cualquiera de las etapas; sometiendo el ARN a síntesis de ADNc bicatenario utilizando hexámeros aleatorios; y sometiendo al ácido nucleico a reparación del extremo, fosforilación y adenilación en un tubo único.

En algunas realizaciones, la secuencia nucleotídica diana conocida puede estar comprendida por un reordenamiento de genes. Los métodos descritos en el presente documento son adecuados para determinar la presencia y/o la identidad de un reordenamiento de genes ya que previamente se debe conocer la identidad de solo una mitad del reordenamiento de genes (es decir, la mitad del reordenamiento de genes que es identificada por los cebadores específicos a diana). En algunas realizaciones, el reordenamiento de genes puede comprender un oncogén. En algunas realizaciones, el reordenamiento de genes puede comprender un oncogén de fusión.

En algunas realizaciones, el ácido nucleico diana puede estar comprendido por una muestra. En algunas realizaciones, el ácido nucleico diana puede estar comprendido por una muestra obtenida de un sujeto. En algunas realizaciones una muestra puede ser una muestra diagnóstica obtenida de un sujeto. En algunas realizaciones, una muestra además puede comprender proteínas, células, fluidos, fluidos biológicos, conservantes y/u otras sustancias. A modo de ejemplo no limitante, una muestra puede ser un frotis de mejilla, sangre, suero, plasma, esputo, fluido cerebrospinal, orina, lágrimas, aislados alveolares, líquido pleural, líquido pericárdico, líquido quístico, tejido tumoral, tejido, una biopsia, saliva, una aspiración o combinaciones de los mismos. En algunas realizaciones, una muestra puede obtenerse por extirpación o biopsia.

En algunas realizaciones, la muestra puede obtenerse de un sujeto que necesita tratamiento para el cáncer. En algunas realizaciones, la muestra puede comprender una población de células tumorales, por ejemplo, al menos una célula tumoral. En algunas realizaciones, la muestra puede comprender una biopsia de tumor, incluyendo, pero sin limitación, tejido de biopsia no tratado o tejido de biopsia tratado (por ejemplo, tejido de biopsia fijado en formol y/o embebido en parafina).

En algunas realizaciones, la muestra puede ser recién recogida. En algunas realizaciones, la muestra puede almacenarse antes de utilizarse en los métodos y composiciones descritos en el presente documento. En algunas realizaciones, la muestra es una muestra no tratada. Como se utiliza en el presente documento, "muestra no tratada" se refiere a una muestra biológica que no ha tenido ningún pretratamiento de muestra previo excepto la dilución y/o suspensión en una solución. En algunas realizaciones, una muestra puede obtenerse de un sujeto y conservarse o procesarse antes de ser utilizada en los métodos y composiciones descritos en el presente documento. A modo de ejemplo no limitante, una muestra puede ser incrustada en cera de parafina, refrigerada o congelada. Una muestra congelada puede descongelarse antes de determinar la presencia de un ácido nucleico de acuerdo con los métodos y composiciones descritos en el presente documento. En algunas realizaciones, la muestra puede ser una muestra procesada o tratada. Métodos ilustrativos para tratar o procesar una muestra incluyen, pero sin limitación, centrifugación, filtración, sonicación, homogeneización, calentamiento, congelación y descongelación, puesta en contacto con un conservante (por ejemplo, anticoagulante o inhibidor de la nucleasa) y cualquier combinación de los

5 mismos. En algunas realizaciones, la muestra puede tratarse con un reactivo químico y/o biológico. Los reactivos químicos y/o biológicos pueden emplearse para proteger y/o conservar la estabilidad de la muestra o el ácido nucleico comprendida por la muestra durante el procesamiento y/o almacenamiento. Además, o como alternativa, los reactivos químicos y/o biológicos pueden emplearse para liberar los ácidos nucleicos de otros componentes de la muestra. A modo de ejemplo no limitante, una muestra sanguínea puede tratarse con un anticoagulante antes de utilizarse en los métodos y composiciones descritos en el presente documento. El experto en la técnica es bien consciente de los métodos y procesos para el procesamiento, la conservación o el tratamiento de las muestras para el análisis de ácido nucleico. En algunas realizaciones, la muestra puede ser una muestra fluida aclarada, por ejemplo, mediante centrifugación. En algunas realizaciones, la muestra puede aclararse mediante centrifugación a baja velocidad (por ejemplo, 3.000 x g o menos) y la recogida del sobrenadante que comprende la muestra fluida aclarada.

15 En algunas realizaciones, el ácido nucleico presente en una muestra puede aislarse, enriquecerse o purificarse antes de utilizarse en los métodos y composiciones descritos en el presente documento. Los métodos de aislamiento, enriquecimiento o purificación de ácidos nucleicos a partir de una muestra son bien conocidos por los expertos en la materia. A modo de ejemplo no limitante, los kits para el aislamiento del ADN genómico de diferentes tipos de muestras están disponibles en el mercado (por ejemplo n.º del catálogo 51104, 51304, 56504 y 56404; Qiagen; Germantown, MD).

20 Los métodos descritos en el presente documento pueden utilizarse en múltiples técnicas. En realizaciones de los métodos descritos en el presente documento, las múltiples aplicaciones pueden incluir determinar la secuencia nucleotídica contigua a una o más secuencias nucleotídicas diana conocidas. Como se utiliza en el presente documento, la "PCR múltiple" se refiere a una variante de PCR donde la amplificación simultánea de más de un ácido nucleico diana en un recipiente de reacción y posterior determinación de la secuencia de los productos de amplificación utilizando más de un conjunto de primeros y segundos cebadores específicos a diana. Múltiple puede referirse a la detección de entre aproximadamente 2-1.000 secuencias diana diferentes en una única reacción. Como se utiliza en el presente documento, múltiple se refiere a la detección de cualquier intervalo entre 2-1.000, por ejemplo, entre 5-500, 25-1.000 o 10-100 secuencias diana diferentes en una reacción única, etc. El término "múltiple" como se aplica a la PCR implica que hay cebadores específicos para al menos dos secuencias diana diferentes en la misma reacción de PCR.

30 En algunas realizaciones, los ácidos nucleicos diana en una muestra, o distintas porciones de una muestra, pueden amplificarse con una pluralidad de primeros y segundos cebadores específicos a diana. En algunas realizaciones, la pluralidad de primeros y segundos cebadores específicos a diana puede estar presente en una mezcla de reacción única, por ejemplo, pueden producirse múltiples productos de amplificación en la misma mezcla de reacción. En algunas realizaciones, la pluralidad de conjuntos de primeros y segundos cebadores específicos a diana puede hibridarse de forma específica con secuencias diana conocidas comprendidas por distintos genes. En algunas realizaciones, al menos dos conjuntos de primeros y segundos cebadores específicos a diana pueden hibridarse de forma específica con diferentes porciones de una secuencia diana conocida. En algunas realizaciones, al menos dos conjuntos de primeros y segundos cebadores específicos a diana pueden hibridarse de forma específica con diferentes porciones de una secuencia diana conocida comprendida por un único gen. En algunas realizaciones, al menos dos conjuntos de primeros y segundos cebadores específicos a diana pueden hibridarse de forma específica con diferentes exones de un gen que comprende una secuencia diana conocida. En algunas realizaciones, la pluralidad de primeros cebadores específicos a diana puede comprender idénticas porciones de la secuencia de la etiqueta 5'.

45 En realizaciones de los métodos descritos en el presente documento, las múltiples aplicaciones pueden incluir determinar la secuencia nucleotídica contigua a una o más secuencias nucleotídicas diana conocidas en múltiples muestras en una reacción de secuenciación o carrera de secuenciación. Las múltiples muestras pueden ser de diferentes orígenes, por ejemplo, de diferentes tejidos y/o diferentes sujetos. En dichas realizaciones, el adaptador de cola de oligonucleótido universal además puede comprender una porción de código de barras. En algunas realizaciones, puede añadirse un adaptador de cola de oligonucleótido universal con una única porción de código de barras a cada muestra y se liga a los ácidos nucleicos en la misma; a continuación, las muestras se agrupan después de la etapa (a). Cada lectura de secuenciación resultante de un producto de amplificación por ello comprenderá un código de barras que identifique qué muestra comprendía el ácido nucleico molde original del cual se deriva el producto de amplificación. El uso de porciones de código de barras en las aplicaciones de secuenciación de última generación es bien conocido en la técnica y está descrito, por ejemplo, en Margulies, M. et al. "Genome Sequencing in Microfabricated High-Density Picolitre Reactors", *Nature*, 437, 376-80 (2005); Mikkelsen, T. et al. "Genome-Wide Maps of Chromatin State in Pluripotent and Lineage-Committed Cells", *Nature*, 448, 553-60 (2007); McLaughlin, S. et al. "Whole-Genome Resequencing With Short Reads: Accurate Mutation Discovery With Mate Pairs and Quality Values", ASHG Annual Meeting (2007); Shendure I. et al. "Accurate Multiplex Polony Sequencing of an Evolved Bacterial Genome", *Science*, 309, 1728-32 (2005); Harris, T. et al. "Single-Molecule DNA Sequencing of a Viral Genome", *Science*, 320, 106-9 (2008); Simen, B. et al. "Prevalence of LoW Abundance Drug Resistant Variants by Ultra Deep Sequencing in Chronically HIV-infected Antiretroviral (ARV) Naive Patients and the Impact on Virologic Outcomes", 16th International HIV Drug Resistance Workshop, Barbados (2007); Thomas, R. et al. "Sensitive Mutation Detection in Heterogeneous Cancer Specimens by Massively Parallel Picoliter Reactor Sequencing", *Nature Med.*, 12, 852-855 (2006); Mitsuya, Y et al. "Minority Human Immunodeficiency Virus Type 1 Variants in Antiretroviral-Naive Persons With Reverse Transcriptase Codon 215 Revertant Mutations", *J. Vir.*, 82, 10747-10755 (2008); Binladen, J. et al. "The Use

of Coded PCR Primers Enables High-Throughput Sequencing of Multiple Homolog Amplification Products by 454 Parallel Sequencing", *PLoS ONE*, 2, e197 (2007); y Hoffmann, C. et al. "DNA Bar Coding and PyROS1equencing to Identify Rare HIV Drug Resistance Mutations", *Nuc. Acids Res.*, 35, e91 (2007).

- 5 En algunas realizaciones de la tecnología descrita en el presente documento, la determinación de la secuencia contigua a una secuencia diana de oligonucleótido conocida puede proporcionar información relevante al tratamiento de la enfermedad, y/o puede comprender un método para tratar la enfermedad. En algunas realizaciones, la muestra puede ser de un sujeto que necesita tratamiento para una enfermedad asociada a una alteración genética. En algunas realizaciones, la secuencia diana de oligonucleótido conocida puede estar comprendida por un gen asociado a la enfermedad, por ejemplo, un oncogén. En algunas realizaciones, la secuencia contigua a una secuencia diana oligonucleotídica conocida y/o la secuencia diana oligonucleotídica conocida pueden comprender una mutación o anomalía genética que está asociada a la enfermedad, por ejemplo, un SNP, una inserción, una delección y/o un reordenamiento de genes. En algunas realizaciones, la secuencia contigua a una secuencia diana oligonucleotídica conocida y/o la secuencia diana oligonucleotídica conocida presentes en una muestra pueden comprender un producto de reordenamiento de genes. En algunas realizaciones, el reordenamiento de genes puede ser un oncogén, por ejemplo, un oncogén de fusión.

Determinados tratamientos contra el cáncer son en particular efectivos contra los tumores que comprenden determinados oncogenes, por ejemplo, un agente de tratamiento que se dirige a la acción o la expresión de un oncogén de fusión dado puede ser efectivo contra tumores que comprenden ese oncogén de fusión pero no contra los tumores que carecen del oncogén de fusión. Los métodos descritos en el presente documento pueden permitir la determinación de secuencias específicas que revelan el estado del oncogén (por ejemplo, mutaciones, SNP y/o reordenamientos). Como se describe en el presente documento, los métodos descritos en el presente documento además pueden permitir la determinación de secuencias específicas cuando se conoce solo la secuencia de un solo flanco, por ejemplo, los métodos descritos en el presente documento pueden determinar la presencia y la identidad de los reordenamientos de genes que implican oncogenes conocidos donde no se conocen la localización precisa y/o la pareja de reordenamiento antes de que se realicen los métodos descritos en el presente documento.

En algunas realizaciones, la tecnología descrita en el presente documento se refiere a un método para tratar un cáncer, comprendiendo el método; detectar, en una muestra de tumor obtenida de un sujeto que necesita tratamiento para el cáncer, la presencia de uno o más reordenamientos de oncogenes de acuerdo con el método descrito en el presente documento; administrar un tratamiento para el cáncer que es efectivo contra los tumores que tienen cualquiera de los reordenamientos de oncogenes detectados. En algunas realizaciones, la tecnología descrita en el presente documento se refiere a un método para determinar si un sujeto que necesita tratamiento para el cáncer será sensible a un tratamiento dado, comprendiendo el método; detectar, en una muestra de tumor obtenida del sujeto, la presencia de un reordenamiento de oncogenes de acuerdo con el método descrito en el presente documento; en donde se determina que un sujeto es sensible a un tratamiento que se dirige a un producto de reordenamiento de oncogenes si se detecta la presencia del reordenamiento del oncogén.

En algunas realizaciones, por ejemplo, la muestra es obtenida de un sujeto que necesita tratamiento para el cáncer de pulmón, la secuencia diana oligonucleotídica conocida puede comprender una secuencia de un gen seleccionado del grupo de ALK; ROS1; y RET. Los reordenamientos de genes que implican los genes ALK, ROS1 y RET y que dan como resultado oncogenes de fusión son bien conocidos en la técnica (véase, por ejemplo, Soda et al. *Nature* 2007 448:561-6; Rikova et al. *Cell* 2007 131:1190-1203; Kohno et al. *Nature Medicine* 2012 18:375-7; Takouchi et al. *Nature Medicine* 2012 18:378-81). Sin embargo, la localización precisa del reordenamiento de genes (por ejemplo, dónde se ha dado el reordenamiento en el gen ALK, ROS1, y/o RET), y la identidad del segundo gen implicado en el reordenamiento puede variar. En los métodos descritos en el presente documento, la presencia y la identidad de tal reordenamiento puede detectarse sin tener que conocer la localización del reordenamiento o la identidad del segundo gen implicado en el reordenamiento de genes.

En algunas realizaciones, la secuencia diana conocida puede comprender la secuencia de un gen seleccionado del grupo de: ALK; ROS1; y RET. En algunas realizaciones, al menos un conjunto de un primer cebador específico de diana y un segundo cebador específico de diana pueden seleccionarse del grupo que consiste en; SEQ ID NO: 5 y 6; SEQ ID NO: 7 y 8; SEQ ID NO: 9 y 10; SEQ ID NO: 11 y 12; SEQ ID NO: 13 y 14; SEQ ID NO: 15 y 16; SEQ ID NO: 17 y 18; SEQ ID NO: 19 y 20; SEQ ID NO: 21 y 22; SEQ ID NO: 23 y 24; SEQ ID NO: 25 y 26; SEQ ID NO: 27 y 28; SEQ ID NO: 29 y 30; SEQ ID NO: 31 y 32; SEQ ID NO: 33 y 34; SEQ ID NO: 35 y 36; y las SEQ ID NO: 37 y 38.

En algunas realizaciones, la presencia de un reordenamiento génico de ALK en una muestra obtenida de un tumor en un sujeto puede indicar que el tumor es susceptible a tratamiento con un tratamiento seleccionado del grupo que consiste en: un inhibidor de ALK; crizotinib (PF-02341066); AP26113; LDK378; 3-39; AF802; IPI-504; ASP3026; AP-26113; X-396; GSK-1838705A; CH5424802; inhibidores de diamino y aminopirimidina de la actividad cinasa de ALK tal como NVP-TAE684 y PF-02341066 (véase, por ejemplo, Galkin et al. *Proc. Natl. Acad. Sci. USA*, 2007, 104:270-275; Zou et al. *Cancer Res*, 2007, 67:4408-4417; Hallberg y Palmer *F1000 Med Reports* 2011 3:21; y Sakamoto et al. *Cancer Cell* 2011 19:679-690) y moléculas desveladas en el documento WO 04/079326. Un inhibidor de ALK puede incluir cualquier agente que reduzca la expresión y/o la actividad cinasa de ALK o una porción del mismo, incluidos, por ejemplo, oligonucleótidos, moléculas pequeñas y/o péptidos que reduzcan la expresión y/o la actividad de ALK o

una porción de los mismos. Como se utiliza en el presente documento, "cinasa de linfoma anaplásico" o "ALK" se refiere a una tirosina cinasa transmembrana normalmente implicada en regulación neuronal en la forma tipo silvestre. La secuencia nucleotídica del gen ALK y el ARNm es conocida para un número de especies, incluida la humana (por ejemplo, SEQ ID NO: 2 (ARNm), ID gen de NCBI: 238).

En algunas realizaciones, la presencia de un reordenamiento génico de ROS1 en una muestra obtenida de un tumor en un sujeto puede indicar que el tumor es susceptible a tratamiento con un tratamiento seleccionado del grupo que consiste en: un inhibidor de ROS1 y un inhibidor de ALK como se ha descrito anteriormente en el presente documento (por ejemplo, crizotinib). Un inhibidor de ROS1 puede incluir cualquier agente que reduzca la expresión y/o la actividad cinasa de ROS1 o una porción del mismo, incluidos, por ejemplo, oligonucleótidos, moléculas pequeñas y/o péptidos que reduzcan la expresión y/o la actividad de ROS1 o una porción de los mismos. Como se utiliza en el presente documento, "oncogén 1 c-ROS1" o "ROS1" (también denominado en la técnica como ROS1-1) se refiere a una tirosina cinasa transmembrana de la subfamilia sevenless y que interactúa con PTPN6. La secuencia nucleotídica del gen ROS1 y el ARNm es conocida para un número de especies, incluida la humana (por ejemplo, SEQ ID NO: 1 (ARNm), ID gen de NCBI: 238).

En algunas realizaciones, la presencia de un reordenamiento génico de RET en una muestra obtenida de un tumor en un sujeto puede indicar que el tumor es susceptible a tratamiento con un tratamiento seleccionado del grupo que consiste en: un inhibidor de RET; DP-2490, DP-3636, SU5416; BAY 43-9006, BAY 73-4506 (regorafenib), ZD6474, NVP-AST487, sorafenib, RPI-1, XL184, vandetanib, sunitinib, imatinib, pazopanib, axitinib, motesanib, gefitinib y witaferin A (véase, por ejemplo, Samadi et al. *Surgery* 2010 148:1228-36; Cuccuru et al. *JNCI* 2004 13:1006-1014; Akeno-Stuart et al. *Cancer Research* 67:6956; Grazma et al. *J. Clin. Oncol.* 2010 28:15s 5559; Mologni et al. *J. Mol. Endocrinol* 2006 37:199-212; Calmomagno et al. *Journal NCI* 2006 98:326-334; Mologni. *Curr. Med. Chem.* 2011 18:162-175 y los compuestos desvelados en el documento WO 06/034833; la publicación de patente de Estados Unidos 2011/0201598 y la patente de Estados Unidos n.º 8.067.434. Un inhibidor de RET puede incluir cualquier agente que reduzca la expresión y/o la actividad cinasa de RET o una porción del mismo, incluidos, por ejemplo, oligonucleótidos, moléculas pequeñas y/o péptidos que reduzcan la expresión y/o la actividad de RET o una porción de los mismos. Como se utiliza en el presente documento, "reordenado durante la transfección" o "RET" se refiere a un receptor tirosina cinasa de la superfamilia cadherina que está implicado en el desarrollo de la cresta neural y reconoce las moléculas de señalización de la familia del factor neurotrófico derivado de la línea de células gliales. La secuencia nucleotídica del gen ROS1 y el ARNm es conocida para un número de especies, incluida la humana (por ejemplo, las SEQ ID NO: 3-4 (ARNm), ID gen de NCBI: 5979).

Ejemplos no limitantes adicionales de aplicaciones de los métodos descritos en el presente documento incluyen detección de marcadores de malignidad hematológica y paneles de los mismos (por ejemplo, incluidos aquellos para detectar los reordenamientos genómicos en linfomas y leucemias), detección de los reordenamientos genómicos relacionados con sarcoma y paneles de los mismos; y detección de los reordenamientos génicos de IGH/TCR y paneles de los mismos para ensayar el linfoma.

En algunas realizaciones, los métodos descritos en el presente documento se refieren a tratar un sujeto que tiene o ha sido diagnosticado como que tiene, por ejemplo, cáncer con un tratamiento para el cáncer. Los sujetos que tienen cáncer pueden ser identificados por un médico utilizando métodos actuales de diagnóstico de cáncer. Por ejemplo, los síntomas y/o complicaciones del cáncer de pulmón que caracterizan estas condiciones y ayudan en el diagnóstico son bien conocidos en la técnica e incluyen, pero sin limitación, respiración débil, ganglios linfáticos inflamados por encima de la clavícula, sonidos anormales en los pulmones, matidez cuando se da golpecitos en el pecho, y dolor de pecho. Las pruebas que pueden ayudar al diagnóstico de, por ejemplo, cáncer de pulmón incluyen, pero sin limitación, rayos x, análisis sanguíneo para elevados niveles de determinadas sustancias (por ejemplo, calcio), escaneos por TC, y biopsia tumoral. Un historial familiar de cáncer de pulmón, o la exposición a factores de riesgo para el cáncer de pulmón (por ejemplo, fumar o exposición al humo y/o a la contaminación del aire) pueden ayudar también en la determinación de si el sujeto es probable que tenga cáncer de pulmón o en la realización de un diagnóstico de cáncer de pulmón.

El cáncer puede incluir, pero sin limitación, carcinoma, incluido adenocarcinoma, linfoma, blastoma, melanoma, sarcoma, leucemia, cáncer de células escamosas, cáncer de pulmón microcítico, cáncer de pulmón no microcítico, cáncer gastrointestinal, linfoma de Hodgkin y no Hodgkin, cáncer pancreático, glioblastoma, carcinoma basocelular, cáncer de las vías biliares, cáncer de vejiga, cáncer cerebral, incluidos glioblastomas y meduloblastomas; cáncer de mama, cáncer de cuello uterino, coriocarcinoma; cáncer de colon, cáncer colorrectal, carcinoma de endometrio, cáncer endometrial; cáncer de esófago, cáncer gástrico; diversos tipos de cáncer de cabeza y cuello, neoplasias intraepiteliales, incluida la enfermedad de Bowen y la enfermedad de Paget; neoplasias hematológicas, incluida la leucemia linfocítica y mielógena aguda; sarcoma de Kaposi, tricoleucemia; leucemia mielógena crónica, leucemias relacionadas al SIDA y linfoma o leucemia de linfocitos T del adulto; cáncer renal tal como cáncer de células renales, leucemia/linfoma linfoblástica aguda de linfocitos T, linfomas, incluida la enfermedad de Hodgkin y linfomas linfocíticos; cáncer de hígado tal como carcinoma hepático y hepatoma, carcinoma de células de Merkel, melanoma, mieloma múltiple; neuroblastomas; cáncer oral, incluido carcinoma escamocelular; cáncer de ovario, incluidos aquellos que surgen de células epiteliales, sarcomas, incluido el leiomiomasarcoma, rhabdomyosarcoma, liposarcoma, fibrosarcoma y osteosarcoma; cáncer pancreático; cáncer de piel, incluido el melanoma, células estromales, células germinales y

células mesenquimales; cáncer de próstata, cáncer de recto; cáncer de vulva, cáncer renal, incluido el adenocarcinoma; cáncer testicular, incluidos tumores germinales tales como seminoma, no seminoma (teratomas, coriocarcinomas), tumores estromales y tumores de células germinales; cáncer de tiroides, incluido el adenocarcinoma de tiroides y el carcinoma medular; cáncer de esófago, carcinoma de las glándulas salivales y tumores de Wilms. En algunas realizaciones, el cáncer puede ser cáncer de pulmón.

En algunas realizaciones, los métodos descritos en el presente documento comprenden administrar una cantidad efectiva de las composiciones descritas en el presente documento, por ejemplo, un tratamiento para el cáncer a un sujeto para aliviar un síntoma de un cáncer. Como se utiliza en el presente documento, "aliviar un síntoma de un cáncer" es mejorar cualquier afección o síntoma asociado al cáncer. En comparación con un control no tratado equivalente, tal reducción es de al menos un 5 %, un 10 %, un 20 %, un 40 %, un 50 %, un 60 %, un 80 %, un 90 %, un 95 %, un 99 % o más medido mediante cualquier técnica estándar. Los expertos en la materia conocen diversos medios para administrar las composiciones descritas en el presente documento a sujetos. Dichos métodos pueden incluir, pero sin limitación, administración oral, parenteral, intravenosa, intramuscular, subcutánea, transdérmica, respiratoria (aerosol), pulmonar, cutánea, tópica, por inyección o intratumoral. La administración puede ser local o sistémica. La expresión "cantidad efectiva" como se utiliza en el presente documento se refiere a la cantidad de un tratamiento necesario para aliviar al menos uno o más síntomas de la enfermedad o el trastorno, y se refiere a una cantidad suficiente de la composición farmacológica para proporcionar el efecto deseado. Por lo tanto, la expresión "cantidad terapéuticamente efectiva" se refiere a una cantidad que es suficiente para efectuar un efecto anticanceroso particular cuando se administra a un sujeto típico. Una cantidad efectiva como se utiliza en el presente documento, en diversos contextos, también incluiría una cantidad suficiente para retrasar el desarrollo de un síntoma de la enfermedad, alterar el curso de un síntoma de la enfermedad (por ejemplo, pero sin limitación, ralentizar la progresión de un síntoma de la enfermedad) o invertir un síntoma de la enfermedad. Por tanto, en general no es viable para especificar una "cantidad efectiva" exacta. Sin embargo, en cualquier caso, un experto en la materia puede determinar una "cantidad efectiva" adecuada utilizando únicamente experimentación rutinaria. Se puede hacer un seguimiento de los efectos de cualquier dosificación particular mediante un bioensayo adecuado. La dosificación es determinada por un médico y se ajusta, según sea necesario, para adaptarse a los efectos observados del tratamiento.

Ejemplos no limitantes de un tratamiento para el cáncer pueden incluir terapia con radiación, cirugía, gemcitabina, cisplatino, paclitaxel, carboplatino, bortezomib, AMG479, vorinostat, rituximab, temozolomida, rapamicina, ABT-737, PI-103; agentes alquilantes tales como tiotepa y ciclofosfamida CYTOXAN®; alquilsulfonatos tales como busulfano, improsulfano y piposulfano; aziridinas tales como benzodopa, carbocuna, meturedopa y uredopa; etileniminas y metilmelaminas que incluyen altretamina, trietilenmelamina, trietilenfosforamida, trietilentiofosforamida y trimetilolmelamina; acetogeninas (especialmente bullatacina y bullatacinona); una camptotecina (incluido el análogo sintético topotecán); briostatina; calistatina; CC-1065 (incluidos sus análogos sintéticos adozelesina, carzelesina y bizelesina); criptoficinas (en particular, criptoficina 1 y criptoficina 8); dolastatina; duocarmicina (incluyendo los análogos sintéticos, KW-2189 y CB1-TM1); eleuterobina; pancratistatina; una sarcodictina; espongiatina; mostazas nitrogenadas tales como clorambucilo, clornafazina, colofosfamida, estramustina, ifosfamida, mecloretamina, clorhidrato de óxido de mecloretamina, melfalán, novembiquina, fenesterina, prednimustina, trofosfamida, mostaza de uracilo; nitroureas tales como carmustina, clorozotocina, fotemustina, lomustina, nimustina y ranimustina; antibióticos tales como los antibióticos de enediina (por ejemplo, caliqueamicina, especialmente caliqueamicina gamma 11 y caliqueamicina omega 11 (véase, por ejemplo, Agnew, *Chem. Intl. Ed. Engl.*, 33: 183-186 (1994)); dinemicina, incluida dinemicina A; bisfosfonatos, tal como clodronato; una esperamicina; así como cromóforo de neocarzinostatina y cromóforos de antibióticos de cromoproteína enediina relacionados), aclacinomisinas, actinomicina, autramicina, azaserina, bleomicinas, cactinomicina, carabicina, caminomicina, carcinofilina, cromomicinas, dactinomicina, daunorrubicina, detorrubicina, 6-diazo-5-oxo-L-norleucina, doxorubicina ADRIAMYCIN® (incluidas morfolino-doxorrubicina, cianomorfolino-doxorrubicina, 2-pirrolin-doxorrubicina y desoxidoxorrubicina), epirubicina, esorubicina, idarrubicina, marcelomicina, mitomicinas tales como mitomicina C, ácido micofenólico, nogalamicina, olivomicinas, peplomicina, porfiromicina, puromicina, quelamicina, rodorrubicina, estreptonigrina, estreptozocina, tubercidina, ubenimex, cinostatina, zorrubicina; antimetabolitos tales como metotrexato y 5-fluorouracilo (5-FU); análogos de ácido fólico tales como denopterina, metotrexato, pteropterina, trimetrexato; análogos de purina tales como fludarabina, 6-mercaptopurina, tiampirina, tioguanina; análogos de pirimidina tales como ancitabina, azacitidina, 6-azauridina, carmofur, citarabina, didesoxiuridina, doxifluridina, enocitabina, floxuridina; andrógenos tales como calusterona, propionato de dromostanolona, epitioestanol, mepitioestano, testolactona; antipararrenales tales como aminoglutetimida, mitotano, trilostano; reabastecedor de ácido fólico tal como ácido frolínico; aceglatona; glucósido de aldofosfamida; ácido aminolevulínico; eniluracilo; amsacrina; bestrabucilo; bisantreno; edatraxato; defofamina; demecolcina; diaziquna; elformitina; acetato de eliptinio; una epotilona; etoglúcido; nitrato de galio; hidroxiurea; lentinano; lonidainina; maitansinoides tales como maitansina y ansamitocinas; mitogazona; mitoxantrona; mopidanmol; nitraerina; pentostatina; fenamet; pirarrubicina; losoxantrona; ácido podofilínico; 2-etilhidrazida; procarbazona; complejo de polisacárido PSK® (JHS Natural Products, Eugene, Oreg.); razoxano; rizoxina; sizofurano; espirogermanio; ácido tenuazónico; triaziquona; 2,2',2"-triclorotrietilamina; tricotecenos (especialmente toxina T-2, verracurina A, roridina A y anguidina); uretano; vindesina; dacarbazina; manomustina; mitobronitol; mitolactol; pipobromano; gacitosina; arabinósido ("Ara-C"); ciclofosfamida; tiotepa; taxoides, por ejemplo, paclitaxel TAXOL® (Bristol-Myers Squibb Oncology, Princeton, N.J.), ABRAXANE®, exento de Cremofor, formulación de nanopartículas de paclitaxel diseñadas por ingeniería con albúmina (American Pharmaceutical Partners, Schaumburg, Ill.) y doxetaxel TAXOTERE® (Rhône-Poulenc Rorer, Antony, Francia); clorambucilo; gemcitabina GEMZAR®; 6-tioguanina;

mercaptapurina; metotrexato; análogos de platino tales como cisplatino, oxaliplatino y carboplatino; vinblastina; platino; etopósido (VP-16); ifosfamida; mitoxantrona; vincristina; NAVELBINE.RTM. vinorelbina; novantrona; tenipósido; edatrexato; daunomicina; aminopterina; xeloda; ibandronato; irinotecán (Camptosar, CPT-11) (incluido el régimen de tratamiento de irinotecán con 5-FU y leucovorina); inhibidor de la topoisomerasa RFS 2000; difluorometilornitina (DMFO); retinoides tales como ácido retinoico; capecitabina; combretastatina; leucovorina (LV); oxaliplatino, incluido el régimen de tratamiento con oxaliplatino (FOLFOX); lapatinib (Tykerb.RTM.); inhibidores de PKC-alfa, Raf, H-Ras, EGFR (por ejemplo, erlotinib (TARCEVA®)) y VEGF-A que reducen la proliferación celular y sales farmacéuticamente aceptables, ácidos o derivados de cualquiera de los anteriores. Adicionalmente, los métodos de tratamiento además pueden incluir la utilización de la radiación o la terapia con radiación. Además, los métodos de tratamiento además pueden incluir la utilización de tratamientos por cirugía.

En algunas realizaciones, los métodos descritos en el presente documento pueden ser aplicables para la resecuenciación, por ejemplo, para confirmar secuencias de baja calidad y/o complejas particularmente relevantes obtenidas mediante secuenciación no dirigida de una gran cantidad de ácidos nucleicos. A modo de ejemplos no limitantes, los métodos descritos en el presente documento pueden permitir la resecuenciación dirigida de paneles de genes de la enfermedad dirigidos (por ejemplo, 10-100 genes), la resecuenciación para confirmar variantes obtenidas en proyectos de secuenciación a gran escala, resecuenciación de exoma completo y/o resecuenciación dirigida para la detección de variantes nucleotídicas sencillas, variantes nucleotídicas múltiples, inserciones, deleciones, cambios del número de copias y estado de la metilación.

En algunas realizaciones, los métodos descritos en el presente documento pueden permitir la secuenciación de la microbiota, la secuenciación de muestra antigua y/o el genotipado de nueva variante de virus.

Las publicaciones citadas en el presente documento se proporcionan únicamente para su divulgación antes de la fecha de presentación de la presente solicitud. Nada en este sentido debe interpretarse como una admisión de que los inventores no tienen derecho a ser anterior a dicha divulgación en virtud de una invención anterior o por cualquier otra razón. Todas las declaraciones sobre la fecha o la representación sobre el contenido de estos documentos se basan en la información disponible para los solicitantes y no constituyen un reconocimiento en cuanto a la exactitud de las fechas o el contenido de estos documentos.

La descripción de las realizaciones de la divulgación no tiene la intención de ser exhaustiva o limitar la divulgación a las formas precisas desveladas. Aunque las realizaciones específicas de, y los ejemplos para, la divulgación se describen en el presente documento con fines ilustrativos, son posibles diferentes modificaciones dentro del alcance de la divulgación, como reconocerán los expertos en la técnica relevante. Por ejemplo, aunque las etapas o las funciones del método estén presentes en un orden dado, realizaciones alternativas pueden realizar funciones en un orden diferente, o las funciones pueden realizarse básicamente al mismo tiempo. Los descubrimientos de la divulgación proporcionada en el presente documento pueden aplicarse a otros procedimientos o métodos como sea apropiado. Las diferentes realizaciones descritas en el presente documento se pueden combinar para proporcionar realizaciones adicionales. Los aspectos de la divulgación pueden modificarse, si es necesario, para emplear las composiciones, las funciones y los conceptos de las anteriores referencias y la aplicación para proporcionar realizaciones más adicionales de la divulgación. Estos y otros cambios pueden realizarse a la divulgación considerando la descripción detallada anteriormente. Se pretende que todas estas modificaciones estén incluidas dentro del alcance de las reivindicaciones adjuntas.

Los elementos específicos de cualquiera de las anteriores realizaciones pueden combinarse o sustituirse para los elementos en otras realizaciones. Asimismo, aunque se han descrito las ventajas asociadas con determinadas realizaciones de la divulgación en el contexto de estas realizaciones, otras realizaciones también pueden presentar tales ventajas, y no todas las realizaciones necesitan presentar tales ventajas para estar dentro del alcance de la divulgación.

La tecnología descrita en el presente documento se ilustra además mediante los siguientes ejemplos que de ninguna manera deben interpretarse como limitantes adicionales.

## Ejemplos

### Ejemplo 1

En una realización, en el presente documento se describe un ensayo que utiliza un novedoso adaptador en forma de Y medio truncado para la construcción de genoteca de secuenciación de última generación seguido de dos reacciones en cadena de la polimerasa con extremo único anidadas capaces de enriquecimiento dirigido rápido y eficaz utilizando ARN o ADN genómico como molde a partir de especímenes frescos o fijados en formol y embebidos en parafina (FFPE). Este método de enriquecimiento posibilita la resecuenciación dirigida de ADNc o ADNg para la detección potencial de alteraciones genéticas (variantes nucleotídicas sencillas, inserciones/deleciones y número de copias), alteraciones epigenéticas (metilación), expresión génica y reordenamientos genómicos.

El enriquecimiento dirigido antes de la secuenciación de última generación es más rentable que la secuenciación del

genoma completo, el exoma completo y el transcriptoma completo y, por lo tanto, más práctico para la implementación general tanto para la investigación científica como las aplicaciones clínicas. La elevada profundidad de cobertura proporcionada por los planteamientos de enriquecimiento dirigido permite un mayor rango dinámico para el recuento de alelo (en evaluación de la expresión génica y el número de copias) y la detección de mutaciones de baja frecuencia, un rasgo crucial para la evaluación de las mutaciones somáticas en cáncer. Antes de que sea posible la general implementación de la secuenciación del genoma completo o el exoma completo, el pilar de la secuenciación de última generación clínica implicará seleccionar paneles de la enfermedad con un número distinto de dianas génicas. Igualmente, los estudios científicos para el análisis de grandes tamaños de muestras basándose en las dianas génicas definidas también requerirán un medio económico para el genotipado. Los ensayos descritos en el presente documento serían aplicables en ambos de estos escenarios.

Para la detección de reordenamientos inter e intracromosómicos, la secuenciación del genoma completo o transcriptoma es útil para nuevos descubrimientos de reordenamientos genómicos y no requiere conocimiento previo de las parejas génicas/cromosómicas implicadas. Sin embargo, estos planteamientos de transcriptoma y genoma completo actualmente no están en práctica en el marco clínico debido al elevado coste, la baja profundidad de secuenciación que da como resultado escasa sensibilidad y la alta demanda de análisis bioinformáticos. La hibridación fluorescente *in situ* (FISH) ha sido el método de referencia para la detección de los reordenamientos genómicos en las clínicas; sin embargo, el ensayo es de bajo rendimiento y su implementación requiere equipo especial y un conocimiento experto o experiencia. La RT-PCR es también efectiva para detectar tales reordenamientos pero requiere de conocimiento de tanto las parejas 5' como 3' y no se puede poner a escala para un gran número de dianas/muestras.

Ejemplos de ensayos de enriquecimiento preparatorios comúnmente utilizados para la secuenciación de última generación incluyen ensayos de captura basados en hibridación (TruSeq Capture, Illumina; captura híbrida SureSelect, Agilent) y análisis basado en la reacción en cadena de la polimerasa (PCR) (HaloPlex, Agilent; AmpliSeq, Ion Torrent; TruSeq Amplicon, Illumina; PCR en emulsión/digital, Raindance). Los planteamientos basados en hibridación no solo capturan las secuencias dirigidas cubiertas por las sondas de captura sino también las bases cercanas inespecíficas que consumen la capacidad de secuenciación adicional. Adicionalmente, estos métodos requieren mucho tiempo, son laboriosos y menos específicos. Un planteamiento basado en amplificación por PCR es más sencillo y rápido pero mediante el diseño convencional requiere de tanto cebadores directos como inversos que flanqueen los loci diana. En particular, para la detección de reordenamientos genómicos con parejas de fusión desconocidas, la PCR no es aplicable.

En el presente documento se describe un ensayo de enriquecimiento dirigido que utiliza un novedoso adaptador en forma de Y medio truncado para la construcción de la genoteca de secuenciación de última generación, que posibilita la rápida secuenciación de ADNc o ADNg de especímenes frescos o FFPE. La Figura 1 esboza la construcción de genoteca para el enriquecimiento dirigido utilizando un adaptador Y medio truncado para la secuenciación Ion Torrent como ejemplo. Notablemente, este método es adaptable para cualquier otra plataforma de secuenciación de última generación, incluidas, pero sin limitación, Illumina, SOLiD y 454. En resumen, un ADNc o ADNg bicatenario, partido aleatoriamente puede someterse a reparación en el extremo, adenilación y, a continuación, ligarse a uno o ambos extremos con el adaptador Y universal para crear una secuencia universal para iniciar la PCR y la posterior secuenciación. Una ronda inicial de PCR que utiliza un cebador específico de diana marcado con una cola de relleno (la cola de relleno ayuda en la multiplexación de un elevado número de dianas) y un cebador igual a las 20 bases en el saliente adaptador Y 5'. Se lleva a cabo una segunda ronda de PCR con un cebador igual a las 30 bases en el saliente adaptador Y 5' y un segundo cebador específico de diana anidado en tándem que se hibrida en dirección 3' del primer cebador específico de diana. El segundo cebador anidado en tándem está marcado en 5' con la secuencia del segundo cebador requerido para la posterior PCR en emulsión o agrupación dependiendo de la tecnología de secuenciación. La elevada especificidad en este sistema se consigue con los cebadores anidados en tándem unidireccionales que cubrirán de manera efectiva una secuencia diana de 36-40 bases o más (dependiendo del espacio entre los cebadores). Cabe destacar que, el número de ciclos de la PCR pueden optimizarse dependiendo de cuánto material de ácido nucleico de partida se usa, el número de muestras agrupadas, y el número de dianas.

En resumen, el método utiliza un adaptador Y medio truncado para marcar de forma universal todos los fragmentos de ADNbc con un extremo 3' común que se utiliza para dos rondas de PCR que utilizan dos cebadores específicos a diana anidados unidireccionales para la especificidad. La aplicación de cebadores anidados marcados también evita los efectos de la homodimerización y la heterodimerización del cebador cuando se dirigen a muchos sitios diferentes en el genoma o transcriptoma. Además de las dianas de multiplexación, los métodos descritos en el presente documento también permiten el agrupamiento de múltiples muestras después del ligamiento del adaptador Y (Etapa 2) durante el cual las muestras individuales se ligan a adaptadores de código de barras únicos. Una vez añadido el código de barras mediante el ligamiento del adaptador Y, se pueden agrupar múltiples muestras en un tubo de reacción para posterior enriquecimiento dirigido.

A diferencia de la detección de reordenamientos genómicos que es más tolerante a los errores de secuenciación de inserción y delección tales como los encontrados en la secuenciación por Ion Torrent y 454 como resultado de homopolímeros, la detección de mutaciones de base única y bases múltiples (incluidas las inserciones y las delecciones) requiere secuenciación más precisa alcanzable con la plataforma de secuenciación Illumina. Para este fin, la genoteca con adaptador Y medio truncado puede convertirse en una genoteca de Illumina, por ejemplo, mediante

la amplificación del brazo 3' truncado saliente del adaptador Y utilizando un cebador marcado con Illumina hacia la secuencia del adaptador (con código de barras o sin código de barras) y cebadores anidados específicos a gen marcados (GSP2 marcados con cebador invertido Illumina). De manera similar, la secuenciación bidireccional también se puede conseguir introduciendo un cebador de secuenciación mediante el brazo del adaptador Y 3' truncado saliente utilizando un cebador marcado durante la primera PCR, Etapa 3.

Los ensayos y métodos descritos en el presente documento se han aplicado sobre ADNc derivado de especímenes FFPE para la detección de las fusiones del gen *ROS1* utilizando siete cebadores específicos a gen que se dirigen a siete exones que comprenden el dominio cinasa de *ROS1*. Los reordenamientos genómicos de *ROS1* se detectaron con parejas de fusión conocidas y previamente desconocidas tales como las fusiones "*SLC34A2* Exón 12 - *ROS1* Exón 34" y "*EZR* Exón 9 - *ROS1* Exón 34". El ensayo alcanzó elevada especificidad de diana (aproximadamente 85-95 % cuando se mapea utilizando la referencia hg19 de genoma humano), que posibilita la secuenciación de múltiples muestras incluso dentro de plataformas de secuenciación a la escala más pequeña (y, por tanto, las menos caras) tal como el chip PGM 314 de Ion Torrent con resultados de cobertura elevados (7 loci dirigidos, 5 muestras, >1000 X cobertura para cada diana por muestra). La Figura 2 muestra el mapeo de las lecturas de secuenciación para los loci diana en el dominio cinasa del gen *ROS1*.

Las ventajas de los métodos descritos en el presente documento incluyen:

1. Los métodos descritos en el presente documento pueden utilizarse para enriquecimiento dirigido de ADNc o ADNg bicatenario de muestras frescas o FFPE, permitiendo el mapeo de los extremos 5' y 3' de los fragmentos diana de ADNc o ADNg, utilizando la secuenciación de última generación. Los planteamientos actuales basados en hibridación requieren de días de preparación e hibridación, y muestran inferior especificidad con captura de secuencias cercanas inespecíficas. Los planteamientos actuales basados en la amplificación requieren cebadores directos e inversos conocidos mediante diseño y no son susceptibles a multiplexación a gran escala de dianas.
2. Análisis bioinformático simple. Dependiendo del número de las dianas seleccionadas, el análisis de los datos generados a partir de este ensayo es simple y rápido. Adicionalmente, puede establecerse para ser compatible con las herramientas bioinformáticas existentes en la actualidad. Por tanto, pequeños laboratorios clínicos son capaces de realizar el análisis de datos sin la necesidad de inversión significativa en herramientas bioinformáticas, lo cual sería una limitación para la implementación clínica general de la secuenciación de última generación.
3. Elevada especificidad (aproximadamente 85-95 %). Una construcción de adaptador Y convencional que tiene ambos componentes de cebadores directos e inversos (salientes 5' y 3') para la construcción rutinaria de la genoteca introducirá un elevado nivel de secuenciación inespecífica de fondo debido al remanente del material de partida la de biblioteca genómica o de transcriptoma "secuenciable". Por ejemplo, si la diana es de un tamaño de 100 pb y está presente en dos copias en un genoma, la relación de diana y genoma es de aproximadamente  $1:3 \times 10^7$ . Los planteamientos basados en hibridación normalmente emplean cebos oligonucleotídicos biotinilados para extraer los fragmentos diana hibridados mediante perlas magnéticas recubiertas con estreptavidina. Justo un evento de unión no específica fuera de lo posible  $3 \times 10^7$  de posibilidades podría drásticamente dar como resultado una tasa de inespecificidad del 50 %. La utilización de un adaptador Y medio truncado evita de manera efectiva este asunto de remanente de fondo, puesto que el material de partida de la genoteca no puede amplificarse en las posteriores etapas preparatorias para la secuenciación. La especificidad adicional se consigue utilizando dos cebadores unidireccionales, con GSP2 en dirección 3' de GSP1. Efectivamente, la utilización de los dos cebadores en tándem produce un sitio de cebado de diana de 40 (asumir dos cebadores de 20 pares de bases de longitud) y mayor especificidad que con solo un cebador. La utilización de cebador 20-mero 5' en la primera PCR y el cebador 30-mero completo (sirve como un cebador anidado para el sitio de cebado universal) en la segunda PCR aumenta más la especificidad. Por último, la especificidad adicional se consigue con la utilización de cebadores marcados para ambas etapas de la PCR. Los cebadores marcados, a través de la formación de horquilla intramolecular, previenen la propagación de los homodímeros y los heterodímeros de cebador que podrían sobrepasar las reacciones de la PCR y conducir a productos artificiales no deseados; por lo tanto, los cebadores marcados permiten la capacidad de multiplexación de muchas dianas mientras que se evita los dímeros de cebadores.
4. Coste económico. Los componentes claves en los métodos descritos en el presente documento son cebadores marcados no modificados convencionales, reactivos de PCR estándar, y termociclación rutinaria. A diferencia de los métodos de captura basados en hibridación o equipos de PCR digitales microfluídicos, el protocolo de enriquecimiento dirigido descrito evita la utilización de oligonucleótidos biotinilados relativamente caros, perlas magnéticas recubiertas con estreptavidina, y equipo especial. Una vez fabricadas, las mezclas de cebador GSP1 y GSP2 agrupadas pueden utilizarse para miles de reacciones.
5. Automatización. Debido a que el método descrito en el presente documento depende de técnicas de PCR estándares y de limpieza por SPRI, existe un potencial para la automatización simple en aplicaciones de alto rendimiento. Los volúmenes pueden ajustarse para placas de 384 pocillos para la implementación de ultra-alto rendimiento.

Las aplicaciones de los métodos y los ensayos descritos en el presente documento incluyen, pero sin limitación: 1. Un panel de translocación del cáncer de pulmón que consiste en dianas terapéuticas conocidas que incluyen los genes *ALK*, *ROS1* y *RET*; 2. paneles de malignidad hematológica que incluye aquellos para detectar los reordenamientos genómicos en linfomas y leucemias; 3. Panel de reordenamiento genómico del sarcoma; 4. Panel de reordenamiento génico de *IGH/TCR* para el ensayo del linfoma; 5. Panel del gen de la enfermedad dirigido para la resecuenciación

(10-100 genes); 6. Secuenciación dirigida para confirmar variantes de proyectos de secuenciación a gran escala; 7. Potencial para la resecuenciación dirigida del exoma completo; 8. Resecuenciación dirigida para la detección de variantes nucleotídicas sencillas, variantes nucleotídicas múltiples, inserciones, deleciones, cambios del número de copias y estado de la metilación; 9. Secuenciación de microbiota; 10. Secuenciación de muestra antigua; 11. Genotipado de nueva variante de virus.

## **EJEMPLO 2: Ensayo de secuenciación de última generación dirigida para la detección simultánea de los reordenamientos de los genes *ALK*, *ROS1* y *RET***

El conocimiento del estado de reordenamiento cromosómico en cáncer es importante para la terapia dirigida individualizada. Recientemente, se han descrito tres receptores de tirosina cinasas principales implicados en los reordenamientos en cáncer de pulmón. El reordenamiento de genes que implica el gen *ALK* se ha establecido como una diana terapéutica. Los datos tempranos del ensayo clínico sugieren que un inhibidor de *ROS1* es efectivo en el tratamiento de pacientes que son positivos para el reordenamiento de *ROS1*. La evidencia *in vitro* ha mostrado que las células tumorales que albergan el reordenamiento de *RET* son sensibles a un inhibidor de *RET*. Por tanto, *ALK*, *ROS1*, y *RET* actualmente representan tres dianas terapéuticas importantes en el cáncer de pulmón.

Los ensayos clínicos actuales para la detección de los reordenamientos de genes incluyen hibridación fluorescente *in situ* (FISH), inmunohistoquímica (IHC) y reacción en cadena de la polimerasa con transcriptasa inversa (RT-PCR). FISH e IHC no tienen la capacidad para la creciente demanda de elevado volumen de ensayo/cribado debido a su bajo rendimiento, elevado coste e interpretación compleja. Los ensayos de RT-PCR requieren de completo conocimiento de ambas parejas de fusión que a veces no están disponibles e impactan de manera efectiva en la sensibilidad clínica. La RT-PCR múltiple se ha utilizado para detener fusiones variantes que implican diferentes exones entre dos parejas de fusión conocidas. En general, estos ensayos se limitan a una diana génica en un enfoque temporal que requiere equipos de reacción múltiple y análisis.

Los últimos ensayos basados en secuenciación de última generación para la secuenciación por captura de transcriptoma o diana se han aplicado mucho con el fin de investigación y descubrimiento en grandes instalaciones de secuenciación. Estos ensayos en el marco de la investigación en general consiguen baja profundidad de secuenciación debido al gran número de dianas, lo que produce baja sensibilidad analítica y posterior escasa sensibilidad clínica. Aunque son razonables para grandes instalaciones de secuenciación, estos ensayos aún no están dentro del alcance para muchos laboratorios clínicos.

Anteriormente se han descrito métodos y ensayos relacionados con un método de secuenciación dirigida que aplica un novedoso adaptador en forma de Y medio truncado para la construcción de la genoteca de secuenciación de última generación seguido de dos reacciones en cadena de la polimerasa anidadas con extremo único que posibilitan el enriquecimiento dirigido rápido y efectivo utilizando ARN o ADN genómico como molde de especímenes frescos o fijados en formol y embebidos en parafina (FFPE).

Según este método, en este Ejemplo se describe un ensayo específico para la detección de reordenamientos de los genes *ALK*, *ROS1* y *RET* (Figuras 3A-3B). Los cebadores específicos a gen (GSP1) se diseñan para cebar los exones cerca o en el dominio de la cinasa. Los cebadores anidados (GSP2) se diseñan para cebar en dirección 3' de GSP1 pero dentro de los mismos exones y próximo a sus GSP1 emparejados. El panel actualmente incluye siete pares de cebadores para identificar *ROS1* exones 31 a 37; cuatro pares de cebadores para identificar *ALK* exones 19 a 22; y seis pares de cebadores para identificar *RET* exones 8 a 13.

Este ensayo puede adaptarse a diferentes plataformas de secuenciación NGS utilizando la secuencia del adaptador específico de plataforma tras la configuración del adaptador Y medio funcional y los cebadores GSP2. Después de la secuenciación, se hizo un mapa de las lecturas en referencia al genoma humano tras la identificación de los genes parejas de fusión, el corte y empalme alternativo, así como el estado estructural de las fusiones utilizando un algoritmo bioinformático desarrollado por los inventores (por ejemplo, Figura 4A y 4B). El resultado es una simple tabla comentada para un informe rápido (Figura 4C). Utilizando este ensayo de panel de diana de tres genes con código de barras de multimuestra, las muestras positivas para el reordenamiento de *ROS1*, *ALK* y *RET* se han detectado utilizando una carrera de secuenciación Ion Torrent.

Este ensayo es aplicable a ARN degradado, tal como ARN extraído de un espécimen fijado en formol y embebido en parafina (FFPE), que es el material clínico más disponible para la prueba de diagnóstico molecular. Este ensayo tiene la ventaja de las plataformas NGS de laboratorio, un conjunto de programas de análisis informático simples y, por lo tanto, implementación relativamente fácil para los laboratorios clínicos. Cabe destacar que, este ensayo no requiere de conocimiento previo de las parejas de fusión y producirá elevada sensibilidad clínica en la detección de diferentes parejas génicas (y los correspondientes exones múltiples) asociadas al gen diana principal. Adicionalmente, el ensayo será funcional para ambas parejas de fusión conocidas en dirección 5' o en dirección 3' siempre que una de las parejas sea conocida. La multiplexación y la secuenciación profunda permite el ensayo de muestras con baja celularidad tumoral, fusiones raras y eventos de corte y empalme alternativos raros. La información detallada del reordenamiento específico de paciente proporcionada por este ensayo será útil para evaluar la respuesta terapéutica específica de genotipo y quizás un marcador tumoral específico de paciente para el mínimo seguimiento de la enfermedad residual.

Según los reactivos de síntesis oligo convencionales, las enzimas disponibles y la capacidad de multiplexación de muchas muestras en una carrera, este ensayo será un ensayo clínico rentable para detectar reordenamientos de genes.

- 5 Tabla 1: Cebadores adaptados para la plataforma Ion Torrent, v1. Los nombres de los cebadores indican el gen diana y el exón del gen diana que comprende la secuencia nucleotídica diana conocida. R1 designa un primer cebador específico de diana y R2 designa un segundo cebador específico de diana. En esta tabla se detalla un conjunto de un primer cebador específico de diana y un segundo específico de diana para cada exón enumerado de cada gen.

| Nombre del cebador | Secuencia                                           | SEQ ID NO |
|--------------------|-----------------------------------------------------|-----------|
| ALKex19_P1R2       | CCTCTCTATGGGCAGTCGGTGATGCGAGAGTGGCAGGTG TGG         | 5         |
| ALKex19_tag.R1     | GGATCTCGACGCTCTCCCTCAGAGGTCACCACAGAGAGG ATCAG       | 6         |
| ALKex20_P1R2       | CCTCTCTATGGGCAGTCGGTGATCATGGCTTGCAGCTCC TGGT        | 7         |
| ALKex20_tag.R1     | GGATCTCGACGCTCTCCCTGCAGCTCCATCTGCATGGCTT G          | 8         |
| ALKex21_P1R2       | CCTCTCTATGGGCAGTCGGTGATGGCCTTCATACACCTC CCCAAA      | 9         |
| ALKex21_tag.R1     | GGATCTCGACGCTCTCCCTTTGGGCATTCCGGACACC               | 10        |
| ALKex22_P1R2       | CCTCTCTATGGGCAGTCGGTGATAGGAAATCCAGTTCGT CCTGTT CAGA | 11        |
| ALKex22_tag.R1     | GGATCTCGACGCTCTCCCTGATCAGGGCTTCCATGAGGA AATC        | 12        |
| RETex10_P1R2       | CCTCTCTATGGGCAGTCGGTGATGGCTCCCCAGGCTCGT GT          | 13        |
| RETex10_tag.R1     | GGATCTCGACGCTCTCCCTAGGTGCCATAGCCAGCTTTA ATCC        | 14        |
| RETex11_P1R2       | CCTCTCTATGGGCAGTCGGTGATATCACCGTGCGGCACA GCTC        | 15        |
| RETex11_tag.R1     | GGATCTCGACGCTCTCCCTGAGGACAGCGGCTGCGATCA             | 16        |
| RETex12_P1R2       | CCTCTCTATGGGCAGTCGGTGATAGAACCAAGTTCTTCC GAGGGAAT    | 17        |
| RETex12_tag.R1     | GGATCTCGACGCTCTCCCTTCCAAATTGCGCTTCTCCTAG AGTT       | 18        |
| RETex13_P1R2       | CCTCTCTATGGGCAGTCGGTGATACAGCAGGTCTCGCAG CTCAC       | 19        |
| RETex13_tag.R1     | GGATCTCGACGCTCTCCCTTGACCTGCTTCAGGACGTTG AA          | 20        |
| RETex8_P1R2        | CCTCTCTATGGGCAGTCGGTGATCTTGCTGACTGCACAG GACAGG      | 21        |
| RETex8_tag.R1      | GGATCTCGACGCTCTCCCTTCCTCACACTCCAGCCGTCTC            | 22        |

(continuación)

| Nombre del cebador | Secuencia                                             | SEQ ID NO |
|--------------------|-------------------------------------------------------|-----------|
| RETex9_P1R2        | CCTCTCTATGGGCAGTCGGTGATTGGTGCTGGGAGAGCA GGT           | 23        |
| RETex9_tag.R1      | GGATCTCGACGCTCTCCCTCCGTCGGGGCAGGTCTTG                 | 24        |
| ROS1Ex31_P1R2      | CCTCTCTATGGGCAGTCGGTGATGGCTGCATGAAGTTTT AACATGG       | 25        |
| ROS1Ex31_tag.R1    | GGATCTCGACGCTCTCCCTTGATATTACAGACATAAGCA GGACCTTGG     | 26        |
| ROS1Ex32_P1R2      | CCTCTCTATGGGCAGTCGGTGATCTAGTAATTTGGGAAT GCCTGGTTT     | 27        |
| ROS1Ex32_tag.R1    | GGATCTCGACGCTCTCCCTTTCAGCTTTCTCCCACTGTAT TGAA         | 28        |
| ROS1Ex33_P1R2      | CCTCTCTATGGGCAGTCGGTGATCATCTTCCACCTTAAAT TCTGGTTCTGTA | 29        |
| ROS1Ex33_tag.R1    | GGATCTCGACGCTCTCCCTCAGGATCCATTAAATGTCAT CTTCC         | 30        |
| ROS1Ex34_P1R2      | CCTCTCTATGGGCAGTCGGTGATAGTAAGTATGAACTT GTTTCTGGTATCC  | 31        |
| ROS1Ex34_tag.R1    | GGATCTCGACGCTCTCCCTGGTCAGTGGGATTGTAACAA CCAG          | 32        |
| ROS1Ex35_P1R2      | CCTCTCTATGGGCAGTCGGTGATCACCCCTTCCTTGGCAC TTT          | 33        |
| ROS1Ex35_tag.R1    | GGATCTCGACGCTCTCCCTTCTTTGTCTTCGTTTATAAGC ACTGTC       | 34        |
| ROS1Ex36_P1R2      | CCTCTCTATGGGCAGTCGGTGATTCAATCTCCTCTTGGG TTGGA         | 35        |
| ROS1Ex36_tag.R1    | GGATCTCGACGCTCTCCCTCCGAGGGAAGGCAGGAAGA TT             | 36        |
| ROS1Ex37_P1R2      | CCTCTCTATGGGCAGTCGGTGATCAGGAATTCAATCTTC TCCTGGTC      | 37        |
| ROS1Ex37_tag.R1    | GGATCTCGACGCTCTCCCTCTCATCAGATGTGCCTCCTTC AG           | 38        |

**Listado de secuencias**

- 5 <110> THE GENERAL HOSPITAL CORPORATION
- <120> MÉTODOS PARA DETERMINAR UNA SECUENCIA NUCLEOTÍDICA
- <130> 030258-074962-PCT
- 10 <140>
- <141>
- <150> 61/679.302

<151> 03/08/2012  
 <150> 61/645.364  
 <151> 10/05/2012  
 5 <160> 48  
 <170> PatentIn versión 3.5  
 10 <210> 1  
 <211> 6267  
 <212> ADN  
 <213> *Homo sapiens*  
 15 <400> 1  
 agctgcaagt ggcgggcgcc caggcagatg cgatccagcg gctctggggg cggcagcgggt 60  
 ggtagcagct ggtacctccc gccgcctctg ttcggagggg cgcggggcac cgaggtgctt 120  
 tccggccgcc ctctggtcgg ccacccaaag ccgcgggcgc tgatgatggg tgaggagggg 180  
 gcggcaagat ttcgggcgcc cctgccctga acgccctcag ctgctgccgc cggggccgct 240  
 ccagtgcctg cgaactctga ggagccgagg cgccggtgag agcaaggacg ctgcaaactt 300  
 gcgcagcgcg ggggctggga ttcacgcca gaagttcagc aggcagacag tccgaagcct 360  
 tcccgcagcg gagagatagc ttgaggggtgc gcaagacggc agcctccgcc ctcggttccc 420  
 gccagaccg ggcagaagag cttggaggag ccaaaggaa cgcaaaaggc ggccaggaca 480  
 gcgtgcagca gctgggagcc gccgttctca gccttaaaag ttgcagagat tggaggctgc 540  
 cccgagaggg gacagacccc agctccgact gcggggggca ggagaggacg gtacccaact 600  
 gccacctccc ttcaaccata gtagttcttc tgtaccgagc gcagcgagct acagacgggg 660  
 gcgcggcact cggcgcgagg agcgggaggc tcaaggctcc agccagtgag cccagtgtgc 720  
 ttgagtgtct ctggactcgc ccctgagctt ccaggctctgt ttcatttaga ctctgtctcg 780  
 cctccgtgca gttgggggaa agcaagagac ttgcgcgcac gcacagtcct ctggagatca 840  
 ggtggaagga gccgctgggt accaaggact gttagagacc tcttcccatc tcggggagag 900  
 cgaagggtga ggctgggccc ggagagcagt gtaaacggcc tcctccggcg ggatgggagc 960  
 catcgggctc ctgtggctcc tgccgctgct gctttccacg gcagctgtgg gctccgggat 1020  
 ggggaccggc cagcgcgcg gctccccagc tgcggggccg ccgctgcagc cccgggagcc 1080

# ES 2 905 448 T3

|             |             |            |             |             |             |      |
|-------------|-------------|------------|-------------|-------------|-------------|------|
| actcagctac  | tcgcgccctgc | agaggaagag | tctggcagtt  | gacttcgtgg  | tgccctcgct  | 1140 |
| cttccgtgtc  | tacgcccggg  | acctactgct | gccaccatcc  | tcctcggagc  | tgaaggctgg  | 1200 |
| caggcccagag | gcccgcggct  | cgctagctct | ggactgcgcc  | ccgctgctca  | ggttgctggg  | 1260 |
| gccggcgccg  | ggggtctcct  | ggaccgccgg | ttcaccagcc  | ccggcagagg  | cccggacgct  | 1320 |
| gtccaggggtg | ctgaagggcg  | gctccgtgcg | caagctccgg  | cgtgccaaagc | agttggtgct  | 1380 |
| ggagctgggc  | gaggaggcga  | tcttgagggg | ttgcgtcggg  | ccccccgggg  | aggcggctgt  | 1440 |
| ggggctgctc  | cagttcaatc  | tcagcgagct | gttcagttgg  | tggattcgcc  | aaggcgaagg  | 1500 |
| gcgactgagg  | atccgcctga  | tgcccagaaa | gaaggcgtcg  | gaagtgggca  | gagaggggaag | 1560 |
| gctgtccgcg  | gcaattcgcg  | cctcccagcc | ccgccttctc  | ttccagatct  | tcgggactgg  | 1620 |
| tcatagctcc  | ttggaatcac  | caacaaacat | gccttctcct  | tctcctgatt  | attttacatg  | 1680 |
| gaatctcacc  | tggataatga  | aagactcctt | ccctttcctg  | tctcatcgca  | gccgatattg  | 1740 |
| tctggagtgc  | agctttgact  | tcccctgtga | gctggagtat  | tcccctccac  | tgcattgacct | 1800 |
| caggaaccag  | agctggtcct  | ggcgccgcct | cccctccgag  | gaggcctccc  | agatggactt  | 1860 |
| gctggatggg  | cctggggcag  | agcgttctaa | ggagatgccc  | agaggctcct  | ttctccttct  | 1920 |
| caacacctca  | gctgactcca  | agcacaccat | cctgagtcgg  | tggatgagga  | gcagcagtga  | 1980 |
| gcactgcaca  | ctggccgtct  | cggtgcacag | gcacctgcag  | ccctctggaa  | ggtacattgc  | 2040 |
| ccagctgctg  | ccccacaacg  | aggctgcaag | agagatcctc  | ctgatgccca  | ctccagggaa  | 2100 |
| gcatgggttg  | acagtgtctc  | agggaagaat | cgggcgctcca | gacaacccat  | ttcgagtggc  | 2160 |
| cctggaatac  | atctccagtg  | gaaaccgcag | cttgtctgca  | gtggacttct  | ttgccctgaa  | 2220 |
| gaactgcagt  | gaaggaacat  | cccaggtctc | caagatggcc  | ctgcagagct  | ccttcacttg  | 2280 |
| ttggaatggg  | acagtcctcc  | agcttgggca | ggcctgtgac  | ttccaccagg  | actgtgccc   | 2340 |
| gggagaagat  | gagagccaga  | tgtgccggaa | actgcctgtg  | ggtttttact  | gcaactttga  | 2400 |
| agatggcttc  | tgtggctgga  | ccaaggcac  | actgtcaccc  | cacactcctc  | aatggcaggt  | 2460 |
| caggacccta  | aaggatgccc  | ggttccagga | ccaccaagac  | catgctctat  | tgctcagtac  | 2520 |
| cactgatgtc  | cccgttcttg  | aaagtgtctc | agtgaccagt  | gctacgtttc  | ctgcaccgat  | 2580 |
| caagagctct  | ccatgtgagc  | tccgaatgtc | ctggctcatt  | cgtggagtct  | tgaggggaaa  | 2640 |
| cgtgtccttg  | gtgctagtgg  | agaacaaaac | cggaaggag   | caaggcagga  | tggctctggca | 2700 |
| tgtcgccgcc  | tatgaaggct  | tgagcctgtg | gcagtggatg  | gtgttgccctc | tcctcgatgt  | 2760 |
| gtctgacagg  | ttctggctgc  | agatggctgc | atgggtggga  | caaggatcca  | gagccatcgt  | 2820 |
| ggcttttgac  | aatatctcca  | tcagcctgga | ctgctacctc  | accattagcg  | gagaggacaa  | 2880 |
| gatcctgcag  | aatacagcac  | ccaaatcaag | aaacctgttt  | gagagaaacc  | caaacaagga  | 2940 |
| gctgaaaccc  | ggggaaaatt  | caccaagaca | gaccccatc   | tttgacccta  | cagttcattg  | 3000 |

# ES 2 905 448 T3

|            |             |            |             |            |            |      |
|------------|-------------|------------|-------------|------------|------------|------|
| gctgttcacc | acatgtgggg  | ccagcggggc | ccatggcccc  | acccaggcac | agtgcacaaa | 3060 |
| cgcctaccag | aactccaacc  | tgagcgtgga | ggtgggggagc | gagggccccc | tgaaaggcat | 3120 |
| ccagatctgg | aaggtgccag  | ccaccgacac | ctacagcatc  | tcgggctacg | gagctgctgg | 3180 |
| cgggaaaggc | gggaagaaca  | ccatgatgcg | gtcccacggc  | gtgtctgtgc | tgggcatctt | 3240 |
| caacctggag | aaggatgaca  | tgctgtacat | cctgggtggg  | cagcaggagg | aggacgcctg | 3300 |
| ccccagtaca | aaccagttaa  | tccagaaagt | ctgcattgga  | gagaacaatg | tgatagaaga | 3360 |
| agaaatccgt | gtgaacagaa  | gcgtgcatga | gtgggcagga  | ggcggaggag | gagggggtgg | 3420 |
| agccacctac | gtattttaaga | tgaaggatgg | agtgccgggtg | cccctgatca | ttgcagccgg | 3480 |
| aggtggtggc | agggcctacg  | gggccaagac | agacacgttc  | cacccagaga | gactggagaa | 3540 |
| taactcctcg | gttctagggc  | taaacggcaa | ttccggagcc  | gcaggtggtg | gaggtggctg | 3600 |
| gaatgataac | acttccttgc  | tctgggcccg | aaaatctttg  | caggagggtg | ccaccggagg | 3660 |
| acattcctgc | ccccaggcca  | tgaagaagtg | gggggtgggag | acaagagggg | gtttcggagg | 3720 |
| gggtggaggg | gggtgctcct  | caggtggagg | aggcggagga  | tatataggcg | gcaatgcagc | 3780 |
| ctcaaacaat | gaccccgaaa  | tgatggggga | agatgggggtt | tccttcatca | gtccactggg | 3840 |
| catcctgtac | accccagctt  | taaaagtgat | ggaaggccac  | ggggaagtga | atattaagca | 3900 |
| ttatctaaac | tgcagtcact  | gtgaggtaga | cgaatgtcac  | atggaccctg | aaagccacaa | 3960 |
| ggtcatctgc | ttctgtgacc  | acgggacggg | gctggctgag  | gatggcgtct | cctgcattgt | 4020 |
| gtcaccacc  | ccggagccac  | acctgccact | ctcgctgac   | ctctctgtgg | tgacctctgc | 4080 |
| cctcgtggcc | gccctggtcc  | tggctttctc | cggcatcatg  | attgtgtacc | gccggaagca | 4140 |
| ccaggagctg | caagccatgc  | agatggagct | gcagagccct  | gagtacaagc | tgagcaagct | 4200 |
| ccgcacctcg | accatcatga  | ccgactacaa | ccccaaactac | tgctttgctg | gcaagacctc | 4260 |
| ctccatcagt | gacctgaagg  | aggtgcccg  | gaaaaacatc  | accctcattc | ggggtctggg | 4320 |
| ccatggcgcc | tttggggagg  | tgtatgaagg | ccaggtgtcc  | ggaatgccca | acgacccaag | 4380 |
| ccccctgcaa | gtggctgtga  | agacgctgcc | tgaagtgtgc  | tctgaacagg | acgaactgga | 4440 |
| tttcctcatg | gaagccctga  | tcatcagcaa | attcaaccac  | cagaacattg | ttcgctgcat | 4500 |
| tggggtgagc | ctgcaatccc  | tgccccgggt | catcctgctg  | gagctcatgg | cggggggaga | 4560 |
| cctcaagtcc | ttcctccgag  | agacccgccc | tcgcccagagc | cagccctcct | ccctggccat | 4620 |
| gctggacctt | ctgcacgtgg  | ctcgggacat | tgctgtggc   | tgtcagtatt | tggaggaaaa | 4680 |
| ccacttcatc | caccgagaca  | ttgctgccag | aaactgcctc  | ttgacctgtc | caggccctgg | 4740 |
| aagagtggcc | aagattggag  | acttcgggat | ggcccagagac | atctacaggg | cgagctacta | 4800 |
| tagaaagggg | ggctgtgcca  | tgctgccagt | taagtggatg  | ccccagagg  | ccttcatgga | 4860 |

# ES 2 905 448 T3

```

aggaatattc acttctaaaa cagacacatg gtcctttgga gtgctgctat gggaaatctt 4920
ttctcttgga tatatgccat accccagcaa aagcaaccag gaagttctgg agtttgtcac 4980
cagtggaggc cggatggacc caccacaaga ctgccctggg cctgtatacc ggataatgac 5040
tcagtgctgg caacatcagc ctgaagacag gcccaacttt gccatcattt tggagaggat 5100
tgaatactgc acccaggacc cggatgtaat caacaccgct ttgccgatag aatatggtcc 5160
acttgtggaa gaggaagaga aagtgcctgt gagggccaag gaccctgagg gggttcctcc 5220
tctcctggtc tctcaacagg caaaacggga ggaggagcgc agcccagctg cccaccacc 5280
tctgcctacc acctcctctg gcaaggctgc aaagaaaccc acagctgcag agatctctgt 5340
tcgagtcctt agagggcccg ccgtggaagg gggacacgtg aatatggcat tctctcagtc 5400
caaccctcct tcggagttgc acaagggtcca cggatccaga aacaagccca ccagcttgtg 5460
gaaccaacg tacggctcct gggttacaga gaaaccacc aaaaagaata atcctatagc 5520
aaagaaggag ccacacgaca ggggtaacct ggggctggag ggaagctgta ctgtcccacc 5580
taacgttgca actgggagac ttccgggggc ctcactgtc ctagagccct ctctcgtgac 5640
tgccaatatg aaggaggtac ctctgttcag gctacgtcac ttcccttgtg ggaatgtcaa 5700
ttacggctac cagcaacagg gcttgccctt agaagccgct actgcccctg gagctggcca 5760
ttacgaggat accattctga aaagcaagaa tagcatgaac cagcctgggc cctgagctcg 5820
gtcgcacact cacttctctt ccttgggac cctaagaccg tggaggagag agaggcaatg 5880
gctccttcac aaaccagaga ccaaattgtca cgttttgttt tgtgccaacc tattttgaag 5940
taccaccaa aaagctgtat ttgaaaatg ctttagaaaag gttttgagca tgggttcac 6000
ctattctttc gaaagaagaa aatatcataa aatgagtga taaatacaag gccagatgt 6060
ggttgcataa ggtttttatg catgtttgtt gtatacttcc ttatgcttct ttcaaattgt 6120
gtgtgctctg cttcaatgta gtcagaatta gctgcttcta tgtttcatag ttgggggtcat 6180
agatgtttcc ttgccttgtt gatgtggaca tgagccattt gaggggagag ggaacggaaa 6240
taaaggagtt atttgtaatg actaaaa 6267

```

<210> 2  
 <211> 7368  
 <212> ADN  
 <213> *Homo sapiens*

<400> 2

```

caagctttca agcattcaaa ggtctaaatg aaaaaggcta agtattattt caaaaggcaa 60
gtatatccta atatagcaaa acaaacaaag caaatccat cagctactcc tccaattgaa 120
gtgatgaagc ccaaataatt catatagcaa aatggagaaa attagaccgg catctaaaa 180
atctgccatt ggtgaagtga tgaagaacat ttactgtctt attccgaagc ttgtcaattt 240

```

|            |             |             |            |             |             |      |
|------------|-------------|-------------|------------|-------------|-------------|------|
| tgcaactctt | ggctgcctat  | ggattttctgt | ggtgcagtgt | acagttttta  | atagctgcct  | 300  |
| aaagtcgtgt | gtaactaatc  | tgggccagca  | gcttgacctt | ggcacaccac  | ataatctgag  | 360  |
| tgaaccgtgt | atccaaggat  | gtcacttttg  | gaactctgta | gatcagaaaa  | actgtgcttt  | 420  |
| aaagtgtcgg | gagtcgtgtg  | aggttggctg  | tagcagcgcg | gaaggtgcat  | atgaagagga  | 480  |
| agtactggaa | aatgcagacc  | taccaactgc  | tccctttgct | tcttccattg  | gaagccacaa  | 540  |
| tatgacatta | cgatggaaat  | ctgcaaactt  | ctctggagta | aaatacatca  | ttcagtggaa  | 600  |
| atatgcacaa | cttctgggaa  | gctggactta  | tactaagact | gtgtccagac  | cgtcctatgt  | 660  |
| ggtcaagccc | ctgcaccctt  | tactgagta   | cattttccga | gtggtttgga  | tcttcacagc  | 720  |
| gcagctgcag | ctctactccc  | ctccaagtcc  | cagttacagg | actcatcctc  | atggagttcc  | 780  |
| tgaaactgca | cctttgatta  | ggaatattga  | gagctcaagt | cccgaactg   | tggaagtcag  | 840  |
| ctgggatcca | cctcaattcc  | caggtggacc  | tattttgggt | tataacttaa  | ggctgatcag  | 900  |
| caaaaatcaa | aaattagatg  | caggacaca   | gagaaccagt | ttccagtttt  | actccacttt  | 960  |
| accaaatact | atctacaggt  | tttctattgc  | agcagtaaat | gaagttgggtg | aggggtccaga | 1020 |
| agcagaatct | agtattacca  | cttcatcttc  | agcagttcaa | caagaggaac  | agtggtctctt | 1080 |
| tttatccaga | aaaacttctc  | taagaaagag  | atctttaaaa | catttagtag  | atgaagcaca  | 1140 |
| ttgccttcgg | ttggatgcta  | tataccataa  | tattacagga | atatctgttg  | atgtccacca  | 1200 |
| gcaaattggt | tattttctctg | aaggaactct  | catatgggcg | aagaaggctg  | ccaacatgtc  | 1260 |
| tgatgtatct | gacctgagaa  | ttttttacag  | aggttcagga | ttaatttctt  | ctatctccat  | 1320 |
| agattggctt | tatcaaagaa  | tgtatttcat  | catggatgaa | ctggatgtg   | tctgtgattt  | 1380 |
| agagaactgc | tcaaacatcg  | aggaaattac  | tccaccctct | attagtgcac  | ctcaaaaaat  | 1440 |
| tgtggctgat | tcatacaatg  | ggtatgtctt  | ttacctctg  | agagatggca  | tttatagagc  | 1500 |
| agaccttctt | gtaccatctg  | gccggtgtgc  | agaagctgtg | cgtattgtgg  | agagttgcac  | 1560 |
| gttaaaggac | tttgcaatca  | agccacaagc  | caagcgaatc | atttacttca  | atgacactgc  | 1620 |
| ccaagtcttc | atgtcaacat  | ttctggatgg  | ctctgcttcc | catctcatcc  | tacctcgcat  | 1680 |
| cccctttgct | gatgtgaaaa  | gttttgcttg  | tgaaaacaat | gactttcttg  | tcacagatgg  | 1740 |
| caaggtcatt | ttccaacagg  | atgctttgtc  | ttttaatgaa | ttcatcgtgg  | gatgtgacct  | 1800 |
| gagtcacata | gaagaatttg  | ggtttggtaa  | cttggtcatc | tttggctcat  | cctcccagct  | 1860 |
| gcaccctctg | ccaggccgcc  | cgcaggagct  | ttcggtgctg | tttggctctc  | accaggctct  | 1920 |
| tgttcaatgg | aagcctcctg  | cccttgccat  | aggagccaat | gtcatcctga  | tcagtgatat  | 1980 |
| tattgaactc | tttgaattag  | gcccttctgc  | ctggcagaa  | tggacctatg  | aggtgaaagt  | 2040 |
| atccacccaa | gaccctcctg  | aagtcactca  | tattttcttg | aacataagtg  | gaaccatgct  | 2100 |
| gaatgtacct | gagctgcaga  | gtgctatgaa  | atacaagggt | tctgtgagag  | caagttctcc  | 2160 |

# ES 2 905 448 T3

|            |             |             |             |            |             |      |
|------------|-------------|-------------|-------------|------------|-------------|------|
| aaagaggcca | ggccctggt   | cagagccctc  | agtgggtact  | accctggtgc | cagctagtga  | 2220 |
| accaccattt | atcatggctg  | tgaagaaga   | tggtcttgg   | agtaaaccat | taaatagctt  | 2280 |
| tggtccagga | gagttcttat  | cctctgatat  | aggaaatgtg  | tcagacatgg | attggtataa  | 2340 |
| caacagcctc | tactacagtg  | acacgaaagg  | cgacgttttt  | gtgtggctgc | tgaatgggac  | 2400 |
| ggatatctca | gagaattatc  | acctaccag   | cattgcagga  | gcaggggctt | tagcttttga  | 2460 |
| gtggctgggt | cactttctct  | actgggctgg  | aaagacatat  | gtgatacaaa | ggcagtctgt  | 2520 |
| gttgacggga | cacacagaca  | ttgttacctc  | cgtgaagcta  | ttggtgaatg | acatggtggt  | 2580 |
| ggattcagtt | ggtggatatc  | tctactggac  | cacactctat  | tcagtggaaa | gcaccagact  | 2640 |
| aaatggggaa | agttcccttg  | tactacagac  | acagccttgg  | ttttctggga | aaaaggtaat  | 2700 |
| tgctctaact | ttagacctca  | gtgatgggct  | cctgtattgg  | ttggttcaag | acagtcaatg  | 2760 |
| tattcacctg | tacacagctg  | ttcttcgggg  | acagagcact  | ggggatacca | ccatcacaga  | 2820 |
| atgtgcagcc | tggagtactt  | ctgaaatttc  | ccagaatgca  | ctgatgtact | atagtggctg  | 2880 |
| gctgttctgg | atcaatggct  | ttaggattat  | cacaactcaa  | gaaataggtc | agaaaaccag  | 2940 |
| tgtctctggt | ttggaaccag  | ccagatttaa  | tcagttcaca  | attattcaga | catcccttaa  | 3000 |
| gccctgccca | gggaactttt  | cctttacccc  | taagggttatt | ccagattctg | ttcaagagtc  | 3060 |
| ttcatttagg | attgaaggaa  | atgcttcaag  | ttttcaaata  | ctgtggaatg | gtccccctgc  | 3120 |
| ggtagactgg | ggtgtagttt  | tctacagtgt  | agaatttagt  | gctcattcta | agttcttggc  | 3180 |
| tagtgaacaa | cactctttac  | ctgtattttac | tgtggaagga  | ctggaacctt | atgccttatt  | 3240 |
| taatctttct | gtcactcctt  | atacctactg  | gggaaagggc  | cccaaaacat | ctctgtcact  | 3300 |
| tcgagcacct | gaaacagttc  | catcagcacc  | agagaacccc  | agaatattta | tattaccaag  | 3360 |
| tggaaaatgc | tgcaacaaga  | atgaagttgt  | ggtggaattt  | aggtggaaca | aacctaaagca | 3420 |
| tgaaaatggg | gtgttaacaa  | aatttgaaat  | tttctacaat  | atatccaatc | aaagtattac  | 3480 |
| aaacaaaaca | tgtgaagact  | ggattgctgt  | caatgtcact  | ccctcagtga | tgtcttttca  | 3540 |
| acttgaaggc | atgagtccca  | gatgctttat  | tgccttccag  | gttagggcct | ttacatctaa  | 3600 |
| ggggccagga | ccatatgctg  | acgttgtaaa  | gtctacaaca  | tcagaaatca | acctatttcc  | 3660 |
| tcacctcata | actcttcttg  | gtaacaagat  | agttttttta  | gatatggatc | aaaatcaagt  | 3720 |
| tgtgtggacg | ttttcagcag  | aaagagttat  | cagtgccgtt  | tgctacacag | ctgataatga  | 3780 |
| gatgggatat | tatgctgaag  | gggactcact  | ctttcttctg  | cacttgcaca | atcgctctag  | 3840 |
| ctctgagctt | ttccaagatt  | cactggtttt  | tgatatcaca  | gttattacaa | ttgactggat  | 3900 |
| ttcaaggcac | ctctactttg  | cactgaaaga  | atcacaaaat  | ggaatgcaag | tatttgatgt  | 3960 |
| tgatcttgaa | cacaagggtga | aatatcccag  | agaggtgaag  | attcacataa | ggaattcaac  | 4020 |

# ES 2 905 448 T3

aataatthtct ttttctgtat atcctctttt aagtcgcttg tattggacag aagtttccaa 4080  
 ttttggctac cagatgttct actacagtat tatcagtcac accttgcacc gaattctgca 4140  
 acccacagct acaaaccaac aaaacaaaag gaatcaatgt tcttgtaatg tgactgaatt 4200  
 tgagttaagt ggagcaatgg ctattgatac ctctaaccta gagaaacat tgatatactt 4260  
 tgccaaagca caagagatct gggcaatgga tctggaaggc tgtcagtggt ggagagttat 4320  
 cacagtacct gctatgctcg caggaaaaac ccttgtagc ttaactgtgg atggagatct 4380  
 tatatactgg atcatcacag caaaggacag cacacagatt tatcaggcaa agaaaggaaa 4440  
 tggggccatc gtttccaggg tgaaggccct aaggagtagg catatcttgg cttacagttc 4500  
 agttatgcag ccttttccag ataaagcgtt tctgtctcta gcttcagaca ctgtggaacc 4560  
 aactatactt aatgccacta acactagcct cacaatcaga ttacctctgg ccaagacaaa 4620  
 cctcacatgg tatggcatca ccagccctac tccaacatac ctgggtttatt atgcagaagt 4680  
 taatgacagg aaaaacagct ctgacttgaa atatagaatt ctggaatttc aggacagtat 4740  
 agctcttatt gaagatttac aaccattttc aacatacatg atacagatag ctgtaaaaaa 4800  
 ttattattca gatcctttgg aacatttacc accaggaaaa gagatttggg gaaaaactaa 4860  
 aaatggagta ccagaggcag tgcagctcat taatacaact gtgcggtcag acaccagcct 4920  
 cattatatct tggagagaat ctcaacaagg aaatggacct aaagaatcag tccgttatca 4980  
 gttggcaatc tcacacctgg ccctaattcc tgaaactcct ctaagacaaa gtgaatttcc 5040  
 aaatggaagg ctcaactctc ttgttactag actgtctggt ggaaatattt atgtgttaaa 5100  
 ggttcttgcc tgccactctg aggaaatgtg gtgtacagag agtcatcctg tcaactgtga 5160  
 aatgtttaac acaccagaga aaccttattc cttgggtcca gagaacacta gtttgcaatt 5220  
 taattggaag gctccattga atgttaacct catcagattt tgggttgagc tacagaagtg 5280  
 gaaatacaat gagttttacc atgttaaaac ttcatgcagc caaggctcctg cttatgtctg 5340  
 taatatcaca aatctacaac cttatacttc atataatgtc agagtagtgg tggtttataa 5400  
 gacgggagaa aatagcacct cacttccaga aagctttaag acaaaagctg gagtcccaa 5460  
 taaaccaggc attcccaa atactagaagg gagtaaaaat tcaatacagt gggagaaagc 5520  
 tgaagataat ggatgtagaa ttacatacta tatccttgag ataagaaaga gcacttcaaa 5580  
 taatttacag aaccagaatt taaggtgga gatgacattt aatggatcct gcagtagtgt 5640  
 ttgcacatgg aagtcacaaa acctgaaagg aatatttcag ttcagagtag tagctgcaaa 5700  
 taatctaggg tttggtgaat atagtggaat cagtgagaat attatattag ttggagatga 5760  
 tttttggata ccagaaacaa gtttcatact tactattata gttggaatat ttctggttgt 5820  
 tacaatccca ctgacctttg tctggcatag aagattaaag aatcaaaaaa gtgccaagga 5880  
 aggggtgaca gtgcttataa acgaagacaa agagttggct gagctgcgag gtctggcagc 5940

# ES 2 905 448 T3

|             |             |            |            |            |             |      |
|-------------|-------------|------------|------------|------------|-------------|------|
| cggagtaggc  | ctggctaata  | cctgctatgc | aatacatact | cttccaaccc | aagaggagat  | 6000 |
| tgaaaatctt  | cctgccttcc  | ctcgggaaaa | actgactctg | cgtctcttgc | tgggaagtgg  | 6060 |
| agcctttgga  | gaagtgtatg  | aaggaacagc | agtggacatc | ttaggagttg | gaagtggaga  | 6120 |
| aatcaaagta  | gcagtgaaga  | ctttgaagaa | gggttccaca | gaccaggaga | agattgaatt  | 6180 |
| cctgaaggag  | gcacatctga  | tgagcaaatt | taatcatccc | aacattctga | agcagcttgg  | 6240 |
| agtttgtctg  | ctgaatgaac  | ccaatacat  | tatcctggaa | ctgatggagg | gaggagacct  | 6300 |
| tcttacttat  | ttgcgtaaag  | cccggatggc | aacgttttat | ggtcctttac | tcaccttggg  | 6360 |
| tgaccttgta  | gacctgtgtg  | tagatatctc | aaaaggctgt | gtctacttgg | aacggatgca  | 6420 |
| tttcattcac  | agggatctgg  | cagctagaaa | ttgccttgtt | tccgtgaaag | actataccag  | 6480 |
| tccacggata  | gtgaagattg  | gagactttgg | actcgccaga | gacatctata | aaaatgatta  | 6540 |
| ctatagaaag  | agaggggaag  | gcctgctccc | agttcgggtg | atggctccag | aaagtttgat  | 6600 |
| ggatggaatc  | ttcactactc  | aatctgatgt | atggctcttt | ggaattctga | tttgggagat  | 6660 |
| tttaactctt  | ggatcatcag  | cttatccagc | tcattccaac | cttgatgtgt | taaactatgt  | 6720 |
| gcaaacagga  | gggagactgg  | agccaccaag | aaattgtcct | gatgatctgt | ggaatttaat  | 6780 |
| gaccagtgcc  | tgggctcaag  | aaccgacca  | aagacctact | tttcatagaa | ttcaggacca  | 6840 |
| acttcagtta  | ttcagaaatt  | ttttcttaaa | tagcatttat | aagtcagag  | atgaagcaaa  | 6900 |
| caacagtgga  | gtcataaatg  | aaagctttga | aggtgaagat | ggcgaatgta | tttggttgaa  | 6960 |
| ttcagatgac  | attatgccag  | ttgctttaat | ggaaacgaag | aaccgagaag | ggttaaaacta | 7020 |
| tatgggtactt | gctacagaat  | gtggccaagg | tgaagaaaag | tctgagggtc | ctctaggctc  | 7080 |
| ccaggaatct  | gaatcttgtg  | gtctgaggaa | agaagagaag | gaaccacatg | cagacaaaga  | 7140 |
| tttctgccaa  | gaaaaacaag  | tggcttactg | cccttctggc | aagcctgaag | gcctgaacta  | 7200 |
| tgcctgtctc  | actcacagtg  | gatatggaga | tgggtctgat | taatagcgtt | gtttgggaaa  | 7260 |
| tagagagttg  | agataaacac  | tctcattcag | tagttactga | aagaaaactc | tgctagaatg  | 7320 |
| ataaatgtca  | tgggtggtcta | taactccaaa | taaacaatgc | aacgttcc   |             | 7368 |

<210> 3  
 <211> 4174  
 <212> ADN  
 <213> *Homo sapiens*  
  
 <400> 3

## ES 2 905 448 T3

|                                                                    |     |
|--------------------------------------------------------------------|-----|
| agtcccgcgga ccgaagcagg gcgcgcagca gcgctgagtg ccccggaacg tgcgtcgcgc | 60  |
| ccccagtgtc cgtcgcgtcc gccgcgcccc gggcggggat ggggcggcca gactgagcgc  | 120 |
| cgcacccgcc atccagaccc gccggcccta gccgcagtcc ctccagccgt ggccccagcg  | 180 |
| cgcacggggcg atggcgaagg cgacgtccgg tgccgcgggg ctgcgtctgc tgttgctgct | 240 |

|             |            |            |             |             |            |      |
|-------------|------------|------------|-------------|-------------|------------|------|
| gctgctgccg  | ctgctaggca | aagtggcatt | gggcctctac  | ttctcgaggg  | atgcttactg | 300  |
| ggagaagctg  | tatgtggacc | aggcggccgg | cacgcccttg  | ctgtacgtcc  | atgccctgcg | 360  |
| ggacgcccct  | gaggaggtgc | ccagcttccg | cctgggccag  | catctctacg  | gcacgtaccg | 420  |
| cacacggctg  | catgagaaca | actggatctg | catccaggag  | gacaccggcc  | tcctctacct | 480  |
| taaccggagc  | ctggaccata | gctcctggga | gaagctcagt  | gtccgcaacc  | gcggctttcc | 540  |
| cctgctcacc  | gtctacctca | aggtcttcct | gtcaccacaca | tccttctgtg  | agggcgagtg | 600  |
| ccagtggcca  | ggctgtgccc | gcgtatactt | ctccttcttc  | aacacctcct  | ttccagcctg | 660  |
| cagctccctc  | aagccccggg | agctctgctt | cccagagaca  | aggccctcct  | tccgcattcg | 720  |
| ggagaaccga  | ccccaggca  | ccttcacca  | gttccgcctg  | ctgcctgtgc  | agttcttgtg | 780  |
| ccccaacatc  | agcgtggcct | acaggctcct | ggagggtgag  | ggtctgccct  | tccgctgcgc | 840  |
| cccggacagc  | ctggaggtga | gcacgcgctg | ggccctggac  | cgcgagcagc  | gggagaagta | 900  |
| cgagctggtg  | gccgtgtgca | ccgtgcacgc | cggcgcgcgc  | gaggaggtgg  | tgatggtgcc | 960  |
| cttcccgggtg | accgtgtacg | acgaggacga | ctcggcgccc  | accttccccg  | cgggcgtcga | 1020 |
| caccgccagc  | gccgtggtgg | agttcaagcg | gaaggaggac  | accgtggtgg  | ccacgctgcg | 1080 |
| tgtcttcgat  | gcagacgtgg | tacctgcac  | aggggagctg  | gtgaggcggt  | acacaagcac | 1140 |
| gctgctcccc  | ggggacacct | ggggccagca | gaccttccgg  | gtggaacact  | ggcccaacga | 1200 |
| gacctcggtc  | caggccaacg | gcagcttcgt | gcgggcgacc  | gtacatgact  | ataggctggt | 1260 |
| tctcaaccgg  | aacctctcca | tctcggagaa | ccgcaccatg  | cagctggcgg  | tgctggtcaa | 1320 |
| tgactcagac  | ttccagggcc | caggagcggg | cgctctcttg  | ctccacttca  | acgtgtcggg | 1380 |
| gctgccggtc  | agcctgcacc | tgcccagtac | ctactccctc  | tccgtgagca  | ggagggctcg | 1440 |
| ccgatttgcc  | cagatcggga | aagtctgtgt | ggaaaactgc  | caggcattca  | gtggcatcaa | 1500 |
| cgtccagtac  | aagctgcatt | cctctggtgc | caactgcagc  | acgctagggg  | tggtcacctc | 1560 |
| agccgaggac  | acctcgggga | tcctgtttgt | gaatgacacc  | aaggccctgc  | ggcggcccaa | 1620 |
| gtgtgccgaa  | cttcactaca | tggtggtggc | caccgaccag  | cagacctcta  | ggcaggccca | 1680 |
| ggcccagctg  | cttgtaacag | tgagggggtc | atatgtggcc  | gaggaggcgg  | gctgccccct | 1740 |
| gtcctgtgca  | gtcagcaaga | gacggctgga | gtgtgaggag  | tgtggcggcc  | tgggctcccc | 1800 |
| aacaggcagg  | tgtgagtgga | ggcaaggaga | tggcaaaggg  | atcaccagga  | acttctccac | 1860 |
| ctgctctccc  | agcaccaaga | cctgccccga | cggccactgc  | gatgttgtgg  | agaccaaga  | 1920 |
| catcaacatt  | tgccctcagg | actgcctccg | gggcagcatt  | gttggggggac | acgagcctgg | 1980 |
| ggagccccgg  | gggattaaag | ctggctatgg | cacctgcaac  | tgcttccctg  | aggaggagaa | 2040 |
| gtgcttctgc  | gagcccgaag | acatccagga | tccactgtgc  | gacgagctgt  | gccgcacggt | 2100 |

# ES 2 905 448 T3

|            |            |             |             |             |             |      |
|------------|------------|-------------|-------------|-------------|-------------|------|
| gatcgagcc  | gctgtcctct | tctccttcat  | cgtctcgggtg | ctgctgtctg  | ccttctgcat  | 2160 |
| ccactgctac | cacaagtttg | cccacaagcc  | acccatctcc  | tcagctgaga  | tgaccttccg  | 2220 |
| gaggcccgc  | caggccttcc | cggtcagcta  | ctcctcttcc  | ggtgcccgc   | ggccctcgct  | 2280 |
| ggactccatg | gagaaccagg | tctccgtgga  | tgcttcaag   | atcctggagg  | atccaaagtg  | 2340 |
| ggaattccct | cggaagaact | tggttcttgg  | aaaaactcta  | ggagaaggcg  | aatttgaaa   | 2400 |
| agtggtaag  | gcaacggcct | tccatctgaa  | aggcagagca  | gggtacacca  | cgggtggcgt  | 2460 |
| gaagatgctg | aaagagaacg | cctccccgag  | tgagcttca   | gacctgctgt  | cagagttcaa  | 2520 |
| cgtcctgaag | caggtcaacc | acccacatgt  | catcaaattg  | tatggggcct  | gcagccagga  | 2580 |
| tggcccgcct | ctcctcatcg | tggagtacgc  | caaatacggc  | tccctgcggg  | gcttctccg   | 2640 |
| cgagagccgc | aaagtggggc | ctggctacct  | gggcagtgga  | ggcagccgca  | actccagctc  | 2700 |
| cctggaccac | ccggatgagc | gggccctcac  | catgggcgac  | ctcatctcat  | ttgcctggca  | 2760 |
| gatctcacag | gggatgcagt | atctggccga  | gatgaagctc  | gttcatcggg  | acttggcagc  | 2820 |
| cagaaacatc | ctggtagctg | agggggcgaa  | gatgaagatt  | tcggatttcg  | gcttgtcccg  | 2880 |
| agatgtttat | gaagaggatt | cctacgtgaa  | gaggagccag  | ggtcggattc  | cagttaaattg | 2940 |
| gatggcaatt | gaatcccttt | ttgatcatat  | ctacaccacg  | caaagtgatg  | tatggtcttt  | 3000 |
| tgggtgcctg | ctgtgggaga | tcgtgaccct  | agggggaaac  | ccctatcctg  | ggattcctcc  | 3060 |
| tgagcggctc | ttcaaccttc | tgaagaccgg  | ccaccgatg   | gagaggccag  | acaactgcag  | 3120 |
| cgaggagatg | taccgcctga | tgctgcaatg  | ctggaagcag  | gagccggaca  | aaaggccggt  | 3180 |
| gtttgcggac | atcagcaaag | acctggagaa  | gatgatgggt  | aagaggagag  | actacttgga  | 3240 |
| ccttgcggcg | tccactccat | ctgactccct  | gatttatgac  | gacggcctct  | cagaggagga  | 3300 |
| gacaccgctg | gtggactgta | ataatgcccc  | cctccctcga  | gccctccctt  | ccacatggat  | 3360 |
| tgaaaacaaa | ctctatggta | gaatttccca  | tgcatctact  | agattctagc  | accgctgtcc  | 3420 |
| cctctgcact | atccttcctc | tctgtgatgc  | tttttaaaaa  | tgtttctggt  | ctgaacaaaa  | 3480 |
| ccaaagtctg | ctctgaacct | ttttatttgt  | aaatgtctga  | ctttgcatcc  | agtttacatt  | 3540 |
| taggcattat | tgcaactatg | tttttctaaa  | aggaagtgaa  | aataagtgta  | attaccacat  | 3600 |
| tgccagcaa  | cttaggatgg | tagaggaaaa  | aacagatcag  | ggcggaaactc | tcaggggaga  | 3660 |
| ccaagaacag | gttgaataag | gcgcttctgg  | ggtgggaatc  | aagtcatagt  | acttctactt  | 3720 |
| taactaagtg | gataaatata | caaactctggg | gaggtattca  | gttgagaaag  | gagccaccag  | 3780 |
| caccactcag | cctgcaactg | gagcacagcc  | aggttcccc   | agacccctcc  | tgggcaggca  | 3840 |
| ggtgcctctc | agaggccacc | cggcactggc  | gagcagccac  | tggccaagcc  | tcagccccag  | 3900 |
| tcccagccac | atgtcctcca | tcaggggtag  | cgagggttga  | ggagctggct  | ggccctggga  | 3960 |
| ggacgcaccc | ccactgctgt | tttcacatcc  | tttcccttac  | ccaccttcag  | gacggttgtc  | 4020 |

# ES 2 905 448 T3

|                                                                   |      |
|-------------------------------------------------------------------|------|
| acttatgaag tcagtgctaa agctggagca gttgcttttt gaaagaacat ggtctgtggt | 4080 |
| gctgtggtct tacaatggac agtaaatatg gttcttgcca aaactccttc ttttgtcttt | 4140 |
| gattaaatac tagaaattta aaaaaaaaaa aaaa                             | 4174 |

<210> 4  
 <211> 5629  
 <212> ADN  
 <213> *Homo sapiens*

<400> 4

|                                                                   |      |
|-------------------------------------------------------------------|------|
| agtccccgca ccgaagcagg gcgcgcagca gcgctgagtg ccccggaacg tgcgctcgcg | 60   |
| ccccagtgtc cgtcgcgtcc gccgcgcccc gggcggggat gggcgggcca gactgagcgc | 120  |
| cgcacccgcc atccagacct gccggcccta gccgcagtcc ctccagccgt ggccccagcg | 180  |
| cgcacgggcg atggcgaagg cgacgtccgg tgccgcgggg ctgcgtctgc tgttgctgct | 240  |
| gctgctgccg ctgctaggca aagtggcatt gggcctctac ttctcgaggg atgcttactg | 300  |
| ggagaagctg tatgtggacc aggcggccgg cagcccttg ctgtacgtcc atgccctgcg  | 360  |
| ggacgcccct gaggaggtgc ccagcttccg cctggggcag catctctacg gcacgtaccg | 420  |
| cacacggctg catgagaaca actggatctg catccaggag gacaccggcc tcctctacct | 480  |
| taaccggagc ctggaccata gtcctggga gaagctcagt gtccgcaacc ggggctttcc  | 540  |
| cctgctcacc gtctacctca aggtcttcc gtccaccaca tcccttcgtg agggcgagtg  | 600  |
| ccagtggcca ggctgtgcc gcgtatactt ctcttcttc aacacctcct ttccagcctg   | 660  |
| cagctccctc aagccccggg agctctgctt ccagagaca aggccctcct tccgcattcg  | 720  |
| ggagaaccga cccccaggca cttccacca gttccgcctg ctgcctgtgc agttcttgtg  | 780  |
| ccccaacatc agcgtggcct acaggctcct ggagggtgag ggtctgccct tccgctgcgc | 840  |
| cccggacagc ctggaggtga gcacgcgtg ggcctggac cgcgagcagc gggagaagta   | 900  |
| cgagctggtg gccgtgtgca ccgtgcacgc cggcgcgcg gaggaggtgg tgatggtgcc  | 960  |
| cttcccgggtg accgtgtacg acgaggacga ctggcgccc accttccccg cgggcgtcga | 1020 |
| caccgccagc gccgtggtgg agttcaagcg gaaggaggac accgtggtgg ccacgctgcg | 1080 |
| tgtcttcgat gcagacgtgg tacctgcac aggggagctg gtgaggcgg acacaagcac   | 1140 |
| gctgctcccc ggggacacct gggcccagca gaccttccgg gtggaacct ggcccaacga  | 1200 |
| gacctcggtc caggccaacg gcagcttcgt gcgggcgacc gtacatgact ataggctggt | 1260 |
| tctcaaccgg aacctctcca tctcgagaa ccgcaccatg cagctggcgg tgctggtcaa  | 1320 |
| tgactcagac ttccagggcc caggagcggg cgtcctcttg ctccacttca acgtgtcgg  | 1380 |
| gctgccggtc agcctgcacc tgcccagta ctactccctc tccgtgagca ggagggctcg  | 1440 |
| ccgatttgcc cagatcggga aagtctgtgt ggaaaactgc caggcattca gtggcatcaa | 1500 |

|                                                                    |      |
|--------------------------------------------------------------------|------|
| cgtccagtac aagctgcatt cctctggtgc caactgcagc acgctagggg tggtcacctc  | 1560 |
| agccgaggac acctcgggga tcctgtttgt gaatgacacc aaggccctgc ggcgcccaa   | 1620 |
| gtgtgccgaa cttcactaca tgggtggtggc caccgaccag cagacctcta ggaggccca  | 1680 |
| ggcccagctg cttgtaacag tggaggggtc atatgtggcc gaggaggcgg gctgccccct  | 1740 |
| gtcctgtgca gtcagcaaga gacggctgga gtgtgaggag tgtggcggcc tgggctcccc  | 1800 |
| aacaggcagg tgtgagtgga ggcaaggaga tggcaaaggg atcaccagga acttctccac  | 1860 |
| ctgctctccc agcaccaaga cctgccccga cggccactgc gatgttgttg agacccaaga  | 1920 |
| catcaacatt tgccctcagg actgcctccg gggcagcatt gttgggggac acgagcctgg  | 1980 |
| ggagccccgg gggattaaag ctggctatgg cacctgcaac tgcttccctg aggaggagaa  | 2040 |
| gtgcttctgc gagcccgaag acatccagga tccactgtgc gacgagctgt gccgcacggt  | 2100 |
| gatcgcagcc gctgtcctct tctccttcat cgtctcgggtg ctgctgtctg ccttctgcat | 2160 |
| ccactgctac cacaagtttg ccacaaagcc acccatctcc tcagctgaga tgaccttccg  | 2220 |
| gaggcccgcc caggccttcc cggtcagcta ctctcttcc ggtgcccgcc ggccctcgct   | 2280 |
| ggactccatg gagaaccagg tctccgtgga tgccttcaag atcctggagg atccaaagtg  | 2340 |
| ggaattccct cggaagaact tggttcttgg aaaaactcta ggagaaggcg aatttgga    | 2400 |
| agtggccaag gcaacggcct tccatctgaa aggagagca gggtaacca cggtggccgt    | 2460 |
| gaagatgctg aaagagaacg cctccccgag tgagcttcga gacctgctgt cagagttcaa  | 2520 |
| cgtcctgaag caggtcaacc acccacatgt catcaaattg tatggggcct gcagccagga  | 2580 |
| tggcccgtc ctctcatcg tggagtacgc caaatacggc tccctgcggg gcttctccg     | 2640 |
| cgagagccgc aaagtggggc ctggctacct gggcagtgga ggcagccga actccagctc   | 2700 |
| cctggaccac ccggatgagc gggccctcac catgggcgac ctcatctcat ttgcctggca  | 2760 |
| gatctcacag gggatgcagt atctggccga gatgaagctc gttcatcggg acttggcagc  | 2820 |
| cagaaacatc ctggtagctg aggggcgga gatgaagatt tcggatttcg gcttgtcccg   | 2880 |
| agatgtttat gaagaggatt cctacgtgaa gaggagccag ggtcggattc cagttaa     | 2940 |
| gatggcaatt gaatcccttt ttgatcatat ctacaccacg caaagtgatg tatggtcttt  | 3000 |
| tgggtgtcctg ctgtgggaga tcgtgaccct agggggaaac ccctatcctg ggattcctcc | 3060 |
| tgagcggctc ttcaaccttc tgaagaccgg ccaccggatg gagaggccag acaactgcag  | 3120 |
| cgaggagatg taccgcctga tgctgcaatg ctggaagcag gagccggaca aaaggccggt  | 3180 |
| gtttgcgga atcagcaaag acctggagaa gatgatggtt aagaggagag actacttga    | 3240 |
| ccttgccggc tccactccat ctgactccct gatttatgac gacggcctct cagaggagga  | 3300 |
| gacaccgctg gtggactgta ataatgcccc cctccctcga gccctccctt ccacatggat  | 3360 |

# ES 2 905 448 T3

|             |            |            |             |             |             |      |
|-------------|------------|------------|-------------|-------------|-------------|------|
| tgaaaacaaa  | ctctatggca | tgtcagaccc | gaactggcct  | ggagagagtc  | ctgtaccact  | 3420 |
| cacgagagct  | gatggcacta | acactgggtt | tccaagatat  | ccaaatgata  | gtgtatatgc  | 3480 |
| taactggatg  | ctttcaccct | cagcggcaaa | attaatggac  | acgtttgata  | gttaacattt  | 3540 |
| ctttgtgaaa  | ggtaatggac | tcacaagggg | aagaaacatg  | ctgagaatgg  | aaagtctacc  | 3600 |
| ggccctttct  | ttgtgaacgt | cacattggcc | gagccgtgtt  | cagttcccag  | gtggcagact  | 3660 |
| cgttttttgt  | agtttgtttt | aacttccaag | gtggtttttac | ttctgatagc  | cgggtgatttt | 3720 |
| ccctcctagc  | agacatgcca | caccgggtaa | gagctctgag  | tcttagtggt  | taagcattcc  | 3780 |
| tttctcttca  | gtgcccagca | gcacccagtg | ttggtctgtg  | tccatcagtg  | accaccaaca  | 3840 |
| ttctgtgttc  | acatgtgtgg | gtccaacact | tactacctgg  | tgtatgaaat  | tggacctgaa  | 3900 |
| ctgttggatt  | tttctagttg | ccgccaacaa | aggcaaaaaa  | atttaaacad  | gaagcacaca  | 3960 |
| cacaaaaaag  | gcagtaggaa | aaatgctggc | cctgatgacc  | tgtccttatt  | cagaatgaga  | 4020 |
| gactgcgggg  | ggggcctggg | ggtagtgtca | atgccctcc   | agggctggag  | gggaagaggg  | 4080 |
| gccccgagga  | tgggcctggg | ctcagcattc | gagatcttga  | gaatgatttt  | tttttaataca | 4140 |
| tgcaaccttt  | ccttaggaag | acatttggtt | ttcatcatga  | ttaagatgat  | tcctagattt  | 4200 |
| agcacaatgg  | agagattcca | tgccatcttt | actatgtgga  | tgggtggtatc | agggagaggg  | 4260 |
| gctcacaaag  | cacatttgtc | ccccggggcc | accacatcat  | cctcacgtgt  | tcggtagtga  | 4320 |
| gcagccacta  | cccctgatga | gaacagtatg | aagaaagggg  | gctgttggag  | tcccagaatt  | 4380 |
| gctgacagca  | gaggctttgc | tgctgtgaat | cccacctgcc  | accagcctgc  | agcacacccc  | 4440 |
| acagccaagt  | agaggcgaaa | gcagtggctc | atcctacctg  | ttaggagcag  | gtagggcttg  | 4500 |
| tactcacttt  | aatttgaatc | ttatcaactt | actcataaag  | ggacaggcta  | gctagctgtg  | 4560 |
| ttagaagtag  | caatgacaat | gaccaaggac | tgctacacct  | ctgattacaa  | ttctgatgtg  | 4620 |
| aaaaagatgg  | tgtttggttc | ttatagagcc | tgtgtgaaag  | gcccattggat | cagctcttcc  | 4680 |
| tgtgtttgta  | atttaatgct | gctacaagat | gtttctgttt  | cttagattct  | gaccatgact  | 4740 |
| cataagcttc  | ttgtcattct | tcattgcttg | tttgtggtca  | cagatgcaca  | acactcctcc  | 4800 |
| agtcttgtgg  | gggcagcttt | tgggaagtct | cagcagctct  | tctggctgtg  | ttgtcagcac  | 4860 |
| tgtaacttcg  | cagaaaagag | tcggattacc | aaaacactgc  | ctgctcttca  | gacttaaagc  | 4920 |
| actgatagga  | cttaaaatag | tctcattcaa | atactgtatt  | ttatataggc  | atttcacaaa  | 4980 |
| aacagcaaaa  | ttgtggcatt | ttgtgaggcc | aaggcttggg  | tgctgtgtga  | atagagcctt  | 5040 |
| gtggtgtgtg  | cgcacacacc | cagagggaga | gtttgaaaaa  | tgcttattgg  | acacgtaacc  | 5100 |
| tggctctaata | ttgggctgtt | tttcagatac | actgtgataa  | gttctttttac | aaatatctat  | 5160 |
| agacatggta  | aacttttggg | tttcagatat | gcttaatgat  | agtcttacta  | aatgcagaaa  | 5220 |
| taagaataaaa | ctttctcaaa | ttattaaaaa | tgccctacaca | gtaagtgtga  | attgctgcaa  | 5280 |

# ES 2 905 448 T3

|    |                                                                          |      |
|----|--------------------------------------------------------------------------|------|
|    | caggtttgtt ctcaggaggg taagaactcc aggtctaaac agctgaccca gtgatgggga        | 5340 |
|    | atttatcctt gaccaattta tccttgacca ataacctaatt tgtctattcc tgagttataa       | 5400 |
|    | aagtcccat ccttattagc tctactggaa ttttcataca cgtaaatgca gaagttacta         | 5460 |
|    | agtattaagt attactgagt attaagtagt aatctgtcag ttattaaaaat ttgtaaaatc       | 5520 |
|    | tatttatgaa aggtcattaa accagatcat gttccttttt ttgtaatcaa ggtgactaag        | 5580 |
|    | aaaatcagtt gtgtaaataa aatcatgtat cataaaaaaa aaaaaaaaaa                   | 5629 |
| 5  | <210> 5<br><211> 42<br><212> ADN<br><213> Secuencia artificial           |      |
| 10 | <220><br><223> Descripción de la secuencia artificial: Cebador sintético |      |
|    | <400> 5<br>cctctctatg ggcagtcggt gatgcgagag tggcaggtgt gg                | 42   |
| 15 | <210> 6<br><211> 44<br><212> ADN<br><213> Secuencia artificial           |      |
| 20 | <220><br><223> Descripción de la secuencia artificial: Cebador sintético |      |
|    | <400> 6<br>ggatctcgac gctctccctc agaggtcacc acagagagga tcag              | 44   |
| 25 | <210> 7<br><211> 43<br><212> ADN<br><213> Secuencia artificial           |      |
| 30 | <220><br><223> Descripción de la secuencia artificial: Cebador sintético |      |
| 35 | <400> 7<br>cctctctatg ggcagtcggt gatcatggct tgcagctcct ggt               | 43   |
| 40 | <210> 8<br><211> 41<br><212> ADN<br><213> Secuencia artificial           |      |
|    | <220><br><223> Descripción de la secuencia artificial: Cebador sintético |      |
| 45 | <400> 8<br>ggatctcgac gctctccctg cagctccatc tgcagtgctt g                 | 41   |
| 50 | <210> 9<br><211> 45<br><212> ADN<br><213> Secuencia artificial           |      |
|    | <220><br><223> Descripción de la secuencia artificial: Cebador sintético |      |
| 55 | <400> 9                                                                  |      |

|    |                                                                          |    |
|----|--------------------------------------------------------------------------|----|
|    | cctctctatg ggcagtcggt gatggccttc atacacctcc ccaaa                        | 45 |
| 5  | <210> 10<br><211> 37<br><212> ADN<br><213> Secuencia artificial          |    |
| 10 | <220><br><223> Descripción de la secuencia artificial: Cebador sintético |    |
|    | <400> 10<br>ggatctcgac gctctccctt tgggcattcc ggacacc                     | 37 |
| 15 | <210> 11<br><211> 49<br><212> ADN<br><213> Secuencia artificial          |    |
| 20 | <220><br><223> Descripción de la secuencia artificial: Cebador sintético |    |
|    | <400> 11<br>cctctctatg ggcagtcggt gataggaaat ccagttcgtc ctgttcaga        | 49 |
| 25 | <210> 12<br><211> 43<br><212> ADN<br><213> Secuencia artificial          |    |
| 30 | <220><br><223> Descripción de la secuencia artificial: Cebador sintético |    |
|    | <400> 12<br>ggatctcgac gctctccctg atcagggctt ccatgaggaa ate              | 43 |
| 35 | <210> 13<br><211> 41<br><212> ADN<br><213> Secuencia artificial          |    |
| 40 | <220><br><223> Descripción de la secuencia artificial: Cebador sintético |    |
| 45 | <400> 13<br>cctctctatg ggcagtcggt gatggctccc caggctcgtg t                | 41 |
| 50 | <210> 14<br><211> 43<br><212> ADN<br><213> Secuencia artificial          |    |
|    | <220><br><223> Descripción de la secuencia artificial: Cebador sintético |    |
| 55 | <400> 14<br>ggatctcgac gctctcccta ggtgccatag ccagctttaa tcc              | 43 |
| 60 | <210> 15<br><211> 43<br><212> ADN<br><213> Secuencia artificial          |    |
| 65 | <220><br><223> Descripción de la secuencia artificial: Cebador sintético |    |
|    | <400> 15                                                                 |    |

|    |                                                                          |    |
|----|--------------------------------------------------------------------------|----|
|    | cctctctatg ggcagtcggt gatatacccg tgcggcacag ctc                          | 43 |
| 5  | <210> 16<br><211> 39<br><212> ADN<br><213> Secuencia artificial          |    |
| 10 | <220><br><223> Descripción de la secuencia artificial: Cebador sintético |    |
|    | <400> 16<br>ggatctcgac gctctccctg aggacagcgg ctgcgatca                   | 39 |
| 15 | <210> 17<br><211> 47<br><212> ADN<br><213> Secuencia artificial          |    |
| 20 | <220><br><223> Descripción de la secuencia artificial: Cebador sintético |    |
| 25 | <400> 17<br>cctctctatg ggcagtcggt gatagaacca agttcttcg agggaaat          | 47 |
| 30 | <210> 18<br><211> 44<br><212> ADN<br><213> Secuencia artificial          |    |
| 35 | <220><br><223> Descripción de la secuencia artificial: Cebador sintético |    |
| 40 | <400> 18<br>ggatctcgac gctctccctt ccaaattgc ctctcctag agtt               | 44 |
| 45 | <210> 19<br><211> 44<br><212> ADN<br><213> Secuencia artificial          |    |
| 50 | <220><br><223> Descripción de la secuencia artificial: Cebador sintético |    |
| 55 | <400> 19<br>cctctctatg ggcagtcggt gatacagcag gtctcgcagc tcac             | 44 |
| 60 | <210> 20<br><211> 41<br><212> ADN<br><213> Secuencia artificial          |    |
| 65 | <220><br><223> Descripción de la secuencia artificial: Cebador sintético |    |
|    | <400> 20<br>ggatctcgac gctctccctt gacctgcttc aggacgttga a                | 41 |
|    | <210> 21<br><211> 45<br><212> ADN<br><213> Secuencia artificial          |    |
|    | <220><br><223> Descripción de la secuencia artificial: Cebador sintético |    |
|    | <400> 21                                                                 |    |

|    |                                                                          |    |
|----|--------------------------------------------------------------------------|----|
|    | cctctctatg ggcagtcggt gatcttgctg actgcacagg acagg                        | 45 |
| 5  | <210> 22<br><211> 40<br><212> ADN<br><213> Secuencia artificial          |    |
| 10 | <220><br><223> Descripción de la secuencia artificial: Cebador sintético |    |
|    | <400> 22<br>ggatctcgac gctctccctt cctcacactc cagccgtctc                  | 40 |
| 15 | <210> 23<br><211> 42<br><212> ADN<br><213> Secuencia artificial          |    |
| 20 | <220><br><223> Descripción de la secuencia artificial: Cebador sintético |    |
|    | <400> 23<br>cctctctatg ggcagtcggt gattggtgct gggagagcag gt               | 42 |
| 25 | <210> 24<br><211> 37<br><212> ADN<br><213> Secuencia artificial          |    |
| 30 | <220><br><223> Descripción de la secuencia artificial: Cebador sintético |    |
| 35 | <400> 24<br>ggatctcgac gctctccctc cgtcggggca ggtcttg                     | 37 |
| 40 | <210> 25<br><211> 46<br><212> ADN<br><213> Secuencia artificial          |    |
|    | <220><br><223> Descripción de la secuencia artificial: Cebador sintético |    |
| 45 | <400> 25<br>cctctctatg ggcagtcggt gatggctgca tgaagtttta acatgg           | 46 |
| 50 | <210> 26<br><211> 48<br><212> ADN<br><213> Secuencia artificial          |    |
|    | <220><br><223> Descripción de la secuencia artificial: Cebador sintético |    |
| 55 | <400> 26<br>ggatctcgac gctctccctt gatattacag acataagcag gaccttgg         | 48 |
| 60 | <210> 27<br><211> 48<br><212> ADN<br><213> Secuencia artificial          |    |
|    | <220><br><223> Descripción de la secuencia artificial: Cebador sintético |    |
| 65 | <400> 27                                                                 |    |

|    |                                                                          |    |
|----|--------------------------------------------------------------------------|----|
|    | cctctctatg ggcagtcggt gatctagtaa ttgggaatg cctgggtt                      | 48 |
| 5  | <210> 28<br><211> 44<br><212> ADN<br><213> Secuencia artificial          |    |
| 10 | <220><br><223> Descripción de la secuencia artificial: Cebador sintético |    |
|    | <400> 28<br>ggatctcgac gctctccctt tcagctttct cccactgtat tgaa             | 44 |
| 15 | <210> 29<br><211> 52<br><212> ADN<br><213> Secuencia artificial          |    |
| 20 | <220><br><223> Descripción de la secuencia artificial: Cebador sintético |    |
|    | <400> 29<br>cctctctatg ggcagtcggt gatcatcttc cacctaaat tctggtctg ta      | 52 |
| 25 | <210> 30<br><211> 44<br><212> ADN<br><213> Secuencia artificial          |    |
| 30 | <220><br><223> Descripción de la secuencia artificial: Cebador sintético |    |
|    | <400> 30<br>ggatctcgac gctctccctc aggatccatt aaatgtcatc ttcc             | 44 |
| 35 | <210> 31<br><211> 52<br><212> ADN<br><213> Secuencia artificial          |    |
| 40 | <220><br><223> Descripción de la secuencia artificial: Cebador sintético |    |
| 45 | <400> 31<br>cctctctatg ggcagtcggt gatagtaagt atgaaacttg ttctggtat cc     | 52 |
| 50 | <210> 32<br><211> 43<br><212> ADN<br><213> Secuencia artificial          |    |
|    | <220><br><223> Descripción de la secuencia artificial: Cebador sintético |    |
| 55 | <400> 32<br>ggatctcgac gctctccctg gtcagtggga ttgtaacaac cag              | 43 |
| 60 | <210> 33<br><211> 43<br><212> ADN<br><213> Secuencia artificial          |    |
| 65 | <220><br><223> Descripción de la secuencia artificial: Cebador sintético |    |
|    | <400> 33                                                                 |    |

|    |                                                                                 |    |
|----|---------------------------------------------------------------------------------|----|
|    | cctctctatg ggcagtcggt gatcacccct tccttggcac ttt                                 | 43 |
| 5  | <210> 34<br><211> 46<br><212> ADN<br><213> Secuencia artificial                 |    |
| 10 | <220><br><223> Descripción de la secuencia artificial: Cebador sintético        |    |
|    | <400> 34<br>ggatctcgac gctctccctt cttgtcttc gttataagc actgtc                    | 46 |
| 15 | <210> 35<br><211> 45<br><212> ADN<br><213> Secuencia artificial                 |    |
| 20 | <220><br><223> Descripción de la secuencia artificial: Cebador sintético        |    |
| 25 | <400> 35<br>cctctctatg ggcagtcggt gatttcaatc tccttctggg ttgga                   | 45 |
| 30 | <210> 36<br><211> 40<br><212> ADN<br><213> Secuencia artificial                 |    |
| 35 | <220><br><223> Descripción de la secuencia artificial: Cebador sintético        |    |
| 40 | <400> 36<br>ggatctcgac gctctccctc cgaggaagg caggaagatt                          | 40 |
| 45 | <210> 37<br><211> 47<br><212> ADN<br><213> Secuencia artificial                 |    |
| 50 | <220><br><223> Descripción de la secuencia artificial: Cebador sintético        |    |
| 55 | <400> 37<br>cctctctatg ggcagtcggt gatcaggaat tcaatcttct cctggtc                 | 47 |
| 60 | <210> 38<br><211> 42<br><212> ADN<br><213> Secuencia artificial                 |    |
| 65 | <220><br><223> Descripción de la secuencia artificial: Polinucleótido sintético |    |
|    | <400> 39                                                                        |    |

# ES 2 905 448 T3

|    |                                                                                  |     |
|----|----------------------------------------------------------------------------------|-----|
|    | aggatgtccc tgtcaaggct cctgagacct ttgataacat aaccattagc agagaggctc                | 60  |
|    | aggctggagt cccaaataaa ccaggcattc ccaaattact aga                                  | 103 |
| 5  | <210> 40<br><211> 88<br><212> ADN<br><213> Secuencia artificial                  |     |
| 10 | <220><br><223> Descripción de la secuencia artificial: Oligonucleótido sintético |     |
|    | <400> 40                                                                         |     |
|    | tcaaggctct gagacctttg ataacataac cattagcaga gaggctcagg ctggagtcca                | 60  |
| 15 | aatcaaacag gcattcccaa atactaga                                                   | 88  |
|    | <210> 41<br><211> 81<br><212> ADN<br><213> Secuencia artificial                  |     |
| 20 | <220><br><223> Descripción de la secuencia artificial: Oligonucleótido sintético |     |
|    | <400> 41                                                                         |     |
| 25 | gagacctttg ataacataac cattagcaga gaggctcagg ctggagtccc aaataaacca                | 60  |
|    | ggcattccca aaattactag a                                                          | 81  |
| 30 | <210> 42<br><211> 96<br><212> ADN<br><213> Secuencia artificial                  |     |
| 35 | <220><br><223> Descripción de la secuencia artificial: Oligonucleótido sintético |     |
|    | <400> 42                                                                         |     |
|    | gtcctgtcaa ggctctgaga cttttgataa cataaccatt agcagagagg ctcaggctgg                | 60  |
|    | agtcccaa at aaaccaggca ttcccaaatt actaga                                         | 96  |
| 40 | <210> 43<br><211> 103<br><212> ADN<br><213> Secuencia artificial                 |     |
| 45 | <220><br><223> Descripción de la secuencia artificial: Polinucleótido sintético  |     |
|    | <400> 43                                                                         |     |
|    | gcaggatgtc cctgtcaagg ctctgagacc ttgataaca taaccattag cagagaggct                 | 60  |
| 50 | caggatgatt ttgggatacc agaaacaagt ttcatactta cta                                  | 103 |
|    | <210> 44<br><211> 115                                                            |     |

|    |                                                                          |     |  |
|----|--------------------------------------------------------------------------|-----|--|
|    | <212> ADN                                                                |     |  |
|    | <213> Secuencia artificial                                               |     |  |
|    | <220>                                                                    |     |  |
| 5  | <223> Descripción de la secuencia artificial: Polinucleótido sintético   |     |  |
|    | <400> 44                                                                 |     |  |
|    | <b>ggcgcaggag gggcaggatg tccctgtcaa ggctctgaga cctttgataa cataaccatt</b> | 60  |  |
|    | <b>agcagagagg ctccaggatga tttttggata ccagaaacaa gtttcatact tacta</b>     | 115 |  |
| 10 | <210> 45                                                                 |     |  |
|    | <211> 46                                                                 |     |  |
|    | <212> ADN                                                                |     |  |
|    | <213> Secuencia artificial                                               |     |  |
| 15 | <220>                                                                    |     |  |
|    | <223> Descripción de la secuencia artificial: Oligonucleótido sintético  |     |  |
|    | <400> 45                                                                 |     |  |
| 20 | ccatctcatc cctgcgtgc tccgactcag ctaaggaac tggcgt                         | 46  |  |
|    | <210> 46                                                                 |     |  |
|    | <211> 20                                                                 |     |  |
|    | <212> ADN                                                                |     |  |
| 25 | <213> Secuencia artificial                                               |     |  |
|    | <220>                                                                    |     |  |
|    | <223> Descripción de la secuencia artificial: Oligonucleótido sintético  |     |  |
| 30 | <400> 46                                                                 |     |  |
|    | cgccagggtc cagtcacgac                                                    | 20  |  |
|    | <210> 47                                                                 |     |  |
|    | <211> 70                                                                 |     |  |
| 35 | <212> ADN                                                                |     |  |
|    | <213> Secuencia artificial                                               |     |  |
|    | <220>                                                                    |     |  |
|    | <223> Descripción de la secuencia artificial: Oligonucleótido sintético  |     |  |
| 40 | <400> 47                                                                 |     |  |
|    | <b>aatgatacgg cgaccaccga gatctacact agatcgcaca ctctttccct acacgacgct</b> | 60  |  |
|    | <b>cttccgatct</b>                                                        | 70  |  |
| 45 | <210> 48                                                                 |     |  |
|    | <211> 54                                                                 |     |  |
|    | <212> ADN                                                                |     |  |
|    | <213> Secuencia artificial                                               |     |  |
| 50 | <220>                                                                    |     |  |
|    | <223> Descripción de la secuencia artificial: Oligonucleótido sintético  |     |  |
|    | <400> 48                                                                 |     |  |
|    | gatcggaaga gccagttacc ttagctgagt cggagacacg cagggatgag atgg              | 54  |  |

## REIVINDICACIONES

1. Un método para determinar la secuencia nucleotídica contigua a una secuencia nucleotídica diana conocida, comprendiendo el método:

- (i) romper una preparación de ácido nucleico que comprende ADN genómico o ARNm para generar un fragmento de ácido nucleico que comprende una secuencia nucleotídica diana conocida, en donde la secuencia nucleotídica diana conocida comprende una secuencia asociada a un reordenamiento de genes, opcionalmente en donde el reordenamiento de genes da como resultado un oncogén de fusión;
- (ii) ligar un adaptador en forma Y al fragmento de ácido nucleico que comprende la secuencia nucleotídica diana conocida para producir un producto de ligamiento;
- (iii) amplificar el producto de ligamiento utilizando un primer cebador adaptador y un primer cebador específico de diana;
- (iv) amplificar un producto de amplificación de (iii) con un segundo cebador adaptador y un segundo cebador específico de diana, en donde el segundo cebador adaptador y el segundo cebador específico de diana son anidados en relación con el primer cebador adaptador y el primer cebador específico de diana, respectivamente, en donde la amplificación del producto de amplificación de (iii) se realiza utilizando una enzima polimerasa de ácido nucleico capaz de iniciar la síntesis en un extremo 3' de un cebador hibridado con un molde y procediendo hacia un extremo 5' del molde, y en donde el molde es el producto de amplificación de (iii), por lo que se produce de este modo un producto de amplificación adicional que tiene un extremo 3'; y
- (v) secuenciar el producto de amplificación adicional que tiene el extremo 3' de (iv) utilizando un primer cebador de secuenciación y un segundo cebador de secuenciación.

2. El método de la reivindicación 1, que comprende además, después de la etapa (i) y antes de la etapa (ii):

- (a) reparación del extremo del fragmento de ácido nucleico para dejar romos los extremos del fragmento de ácido nucleico;
- (b) fosforilación del fragmento del ácido nucleico; y/o
- (c) adenilación del fragmento de ácido nucleico para producir un saliente de adenosina 3' en el fragmento de ácido nucleico.

3. El método de la reivindicación 1 o de la reivindicación 2, en donde la preparación de ácido nucleico se prepara a partir de un espécimen biológico fresco o de un espécimen biológico fijado en formol y embebido en parafina.

4. El método de la reivindicación 1, en donde, después de romper mecánicamente la preparación de ácido nucleico, el fragmento de ácido nucleico resultante se somete a un procedimiento de limpieza, en donde opcionalmente el procedimiento de limpieza es un procedimiento de limpieza por inmovilización reversible en fase sólida (SPRI).

5. El método de la reivindicación 2 parte (a), (b) o (c), en donde el fragmento de ácido nucleico no se somete a un procedimiento de limpieza después de la reparación de extremo, la fosforilación y/o la adenilación y antes del ligamiento del adaptador en forma Y al fragmento de ácido nucleico en la etapa (ii).

6. El método de la reivindicación 1, en donde el adaptador en forma Y es un adaptador en forma Y medio truncado, o en donde el adaptador en forma Y tiene un brazo 3' truncado.

7. El método de la reivindicación 1, en donde el adaptador en forma Y comprende una porción monocatenaria y una porción bicatenaria, en donde la porción bicatenaria comprende hasta 17 pares de bases (pb), y/o en donde la porción bicatenaria comprende al menos 7 pb.

8. El método de la reivindicación 1, en donde el ligamiento en la etapa (ii) comprende realizar: una reacción de ligamiento del extremo romo; o una reacción de ligamiento saliente; o una reacción de ligamiento TA.

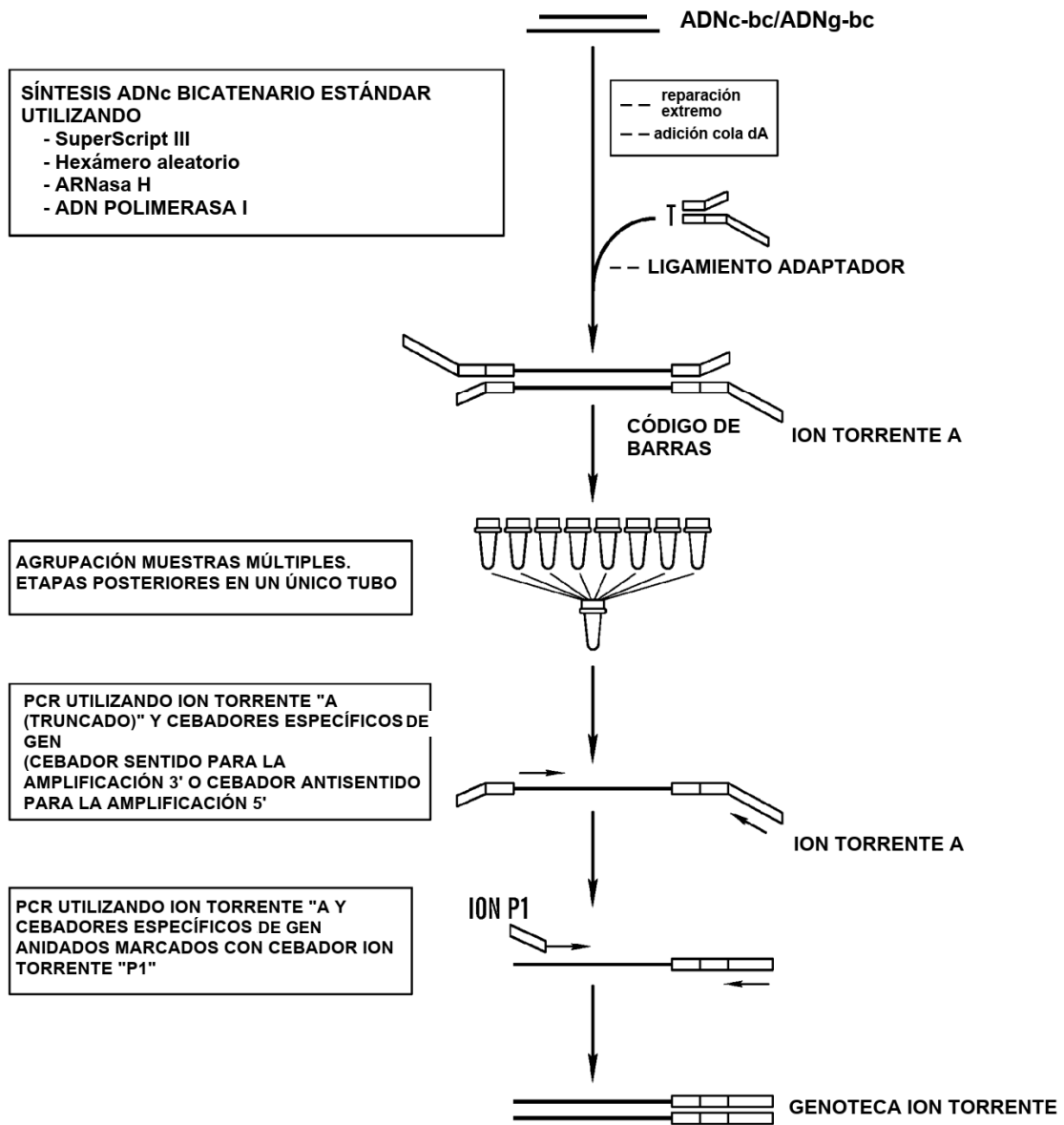
9. El método de la reivindicación 1, en donde:

- (a) el primer cebador adaptador comprende una secuencia de ácido nucleico idéntica a una porción 5' del primer cebador de secuenciación, en donde la porción 5' comprende al menos 20 nucleótidos;
- (b) el primer cebador específico de diana comprende una secuencia de ácido nucleico que puede hibridarse de forma específica con la secuencia nucleotídica diana conocida a una temperatura de hibridación en un intervalo de 61 °C a 72 °C, opcionalmente en donde la temperatura de hibridación es de aproximadamente 65 °C;
- (c) la amplificación en la etapa (iii) comprende realizar una reacción en cadena de la polimerasa (PCR) utilizando el primer cebador adaptador y el primer cebador específico de diana opcionalmente en donde la PCR comprende de 5 ciclos a 20 ciclos;
- (d) la amplificación en la etapa (iv) comprende realizar una reacción en cadena de la polimerasa (PCR) utilizando el segundo cebador adaptador y el segundo cebador específico de diana, opcionalmente en donde la PCR comprende de 5 ciclos a 20 ciclos;

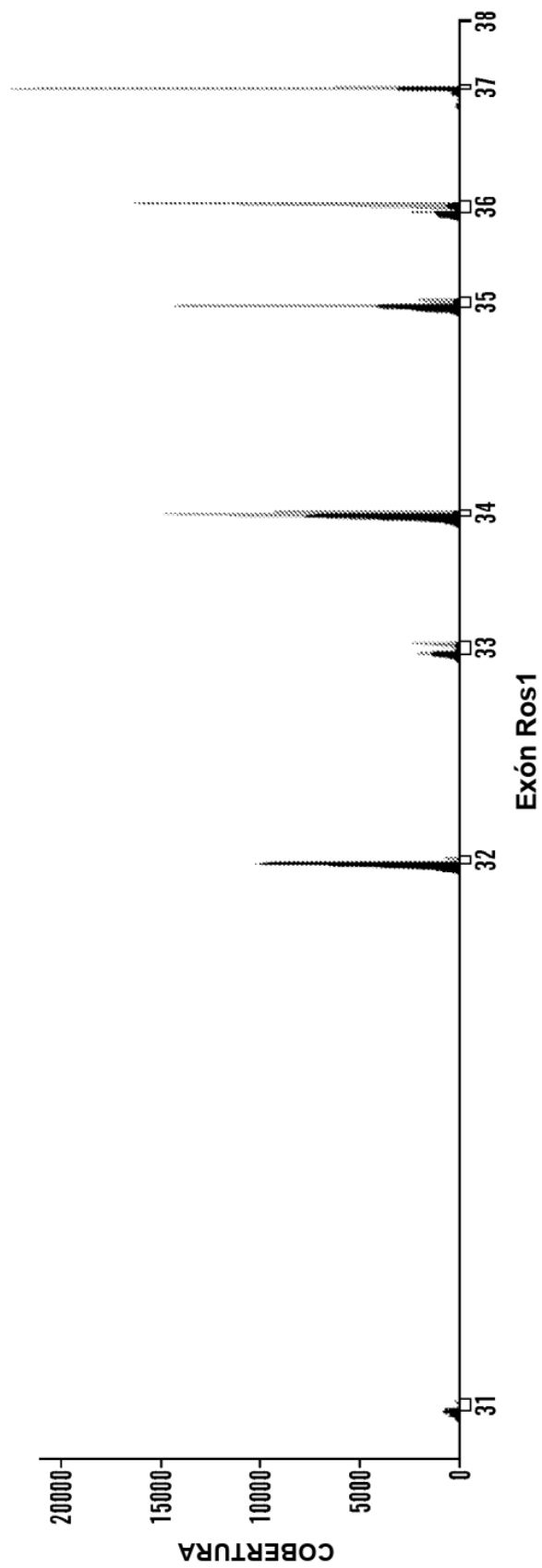
- (e) la amplificación en la etapa (iii) y/o la etapa (iv) comprende las etapas de separación de hebra, hibridación de cebador y extensión por polimerasa del primer cebador adaptador hibridado y el primer cebador específico de diana hibridado;
- 5 (f) la amplificación en la etapa (iii) comprende la amplificación exponencial del producto del ligamiento;
- (g) la amplificación en la etapa (iv) comprende amplificación exponencial del producto de amplificación de la etapa (iii);
- (h) el segundo cebador adaptador comprende un ácido nucleico que tiene una secuencia idéntica a una secuencia de una porción del primer cebador de secuenciación;
- 10 (i) el producto de amplificación de la etapa (iii) no se somete a un procedimiento de limpieza antes de ser amplificado en la etapa (iv);
- (j) los primeros y segundos cebadores de secuenciación son compatibles con la secuenciación de última generación; y/o
- (k) la secuenciación en la etapa (v) comprende la realización de un ensayo de secuenciación de última generación.
- 15 10. El método de la reivindicación 1, en donde el segundo cebador específico de diana comprende una porción 3' que comprende un ácido nucleico que puede hibridarse de forma específica con una porción de la secuencia nucleotídica diana conocida comprendida por el producto de amplificación de la etapa (iii), y una porción 5' que comprende un ácido nucleico que tiene una secuencia que es idéntica a una secuencia de un segundo cebador de secuenciación.
- 20 11. El método de la reivindicación 1, en donde el segundo cebador adaptador es anidado con respecto al primer cebador adaptador por al menos 3 nucleótidos, y/o en donde el segundo cebador adaptador es anidado con respecto al primer cebador adaptador por hasta 10 nucleótidos.
- 25 12. El método de la reivindicación 1, en donde la preparación de ácido nucleico se prepara a partir de una muestra obtenida de un sujeto sospechoso de tener un reordenamiento de genes.
13. El método de la reivindicación 1, en donde el adaptador en forma Y comprende además una porción de código de barras, y/o
- 30 en donde el método comprende además agrupar múltiples muestras de productos de ligamiento antes de la etapa (iii), y amplificar cada una de las múltiples muestras en la etapa (iii) para producir el producto de amplificación de la etapa (iii), teniendo cada muestra un adaptador en forma Y ligado que contiene una porción de código de barras única.
14. El método de la reivindicación 1, en donde el adaptador en forma Y está truncado y en donde el segundo cebador adaptador y el segundo cebador específico de diana son anidados por al menos tres nucleótidos en relación con el primer cebador adaptador y el primer cebador específico de diana, respectivamente.
- 35 15. El método de la reivindicación 14, en donde:
- (a) el adaptador en forma Y truncado comprende una porción monocatenaria y una porción bicatenaria, en donde la porción bicatenaria comprende hasta 17 pares de bases (pb);
- 40 (b) el adaptador en forma Y truncado comprende una porción monocatenaria y una porción bicatenaria, en donde la porción bicatenaria comprende al menos 7 pb;
- (c) el segundo cebador adaptador es anidado con respecto al primer cebador adaptador por hasta 10 nucleótidos; y/o
- 45 (d) (i) la preparación de ácido nucleico se prepara a partir de una muestra obtenida de un sujeto sospechoso de tener un reordenamiento de genes, y/o (ii) la secuencia nucleotídica diana conocida comprende una secuencia asociada a un reordenamiento de genes, opcionalmente en donde el reordenamiento de genes da como resultado una proteína de fusión.
- 50 16. Un método para determinar la secuencia nucleotídica contigua a una secuencia nucleotídica diana conocida, comprendiendo el método:
- (i) romper una preparación de ácido nucleico que comprende ADN genómico o ARNm para generar un fragmento de ácido nucleico diana que comprende una secuencia nucleotídica diana conocida, en donde la secuencia nucleotídica diana conocida comprende una secuencia asociada a un reordenamiento de genes, opcionalmente en donde el reordenamiento de genes da como resultado un oncogén de fusión;
- 55 (ii) ligar el fragmento de ácido nucleico diana que comprende la secuencia nucleotídica diana conocida con un adaptador de cola de oligonucleótido universal para producir un producto de ligamiento, en donde el adaptador de cola de oligonucleótido universal comprende un primer extremo dúplex que se puede ligar y un segundo extremo deshibridado;
- 60 (iii) amplificar el producto de ligamiento en una primera ronda de PCR utilizando un primer cebador específico de diana y un primer cebador adaptador, en donde el primer cebador adaptador comprende una secuencia que es idéntica a una región saliente en el segundo extremo deshibridado del adaptador de cola de oligonucleótido universal;
- 65 (iv) amplificar un producto de amplificación de (iii) en una segunda ronda de PCR utilizando un segundo cebador específico de diana y un segundo cebador adaptador, en donde el segundo cebador específico de diana es anidado

- con respecto al primer cebador específico de diana, en donde el segundo cebador adaptador comprende una secuencia que es idéntica a una región saliente en el segundo extremo deshibridado del adaptador de cola de oligonucleótido universal, en donde la amplificación del producto de amplificación de (iii) se realiza utilizando una enzima polimerasa de ácido nucleico capaz de iniciar la síntesis en un extremo 3' de un cebador hibridado con un molde y procediendo hacia un extremo 5' del molde, y en donde el molde es el producto de amplificación de (iii), por lo que se produce de este modo un producto de amplificación adicional que tiene un extremo 3'; y
- 5 (v) secuenciar el producto de amplificación adicional que tiene el extremo 3' de (iv) utilizando un primer y un segundo cebador de secuenciación.
- 10 17. El método de la reivindicación 16, en donde el segundo cebador específico de diana comprende una región 3' que hibrida en dirección 3' del primer cebador específico de diana y una región 5' que comprende una etiqueta.
18. El método de la reivindicación 16, que comprende además, después de la etapa (i) y antes de la etapa (ii):
- 15 (a) reparación de extremo del fragmento de ácido nucleico diana para dejar romos los extremos del fragmento de ácido nucleico diana;  
 (b) fosforilación del fragmento de ácido nucleico diana; y/o  
 (c) adenilación del fragmento de ácido nucleico diana para producir un saliente de adenosina 3' en el fragmento de ácido nucleico diana.

SECUENCIACIÓN DIRIGIDA ADN Y ARN



**FIG. 1**



**FIG. 2**

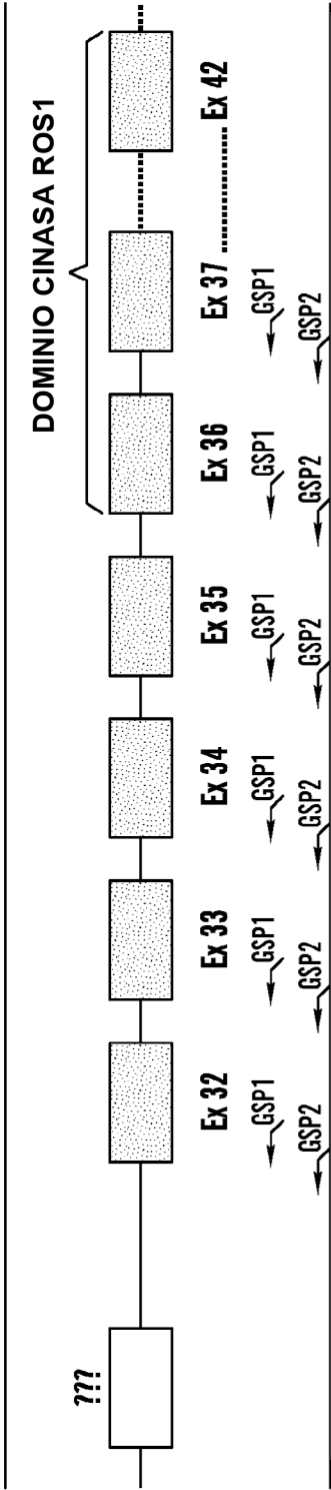


FIG. 3A

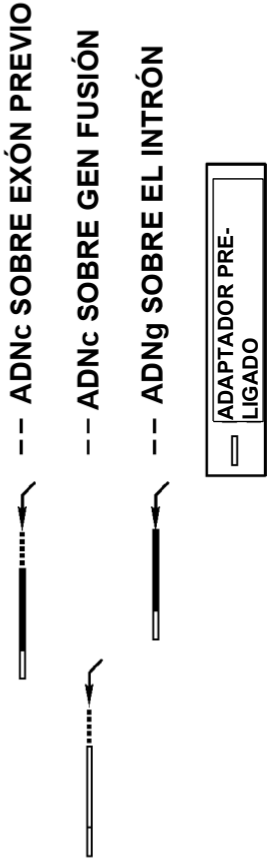
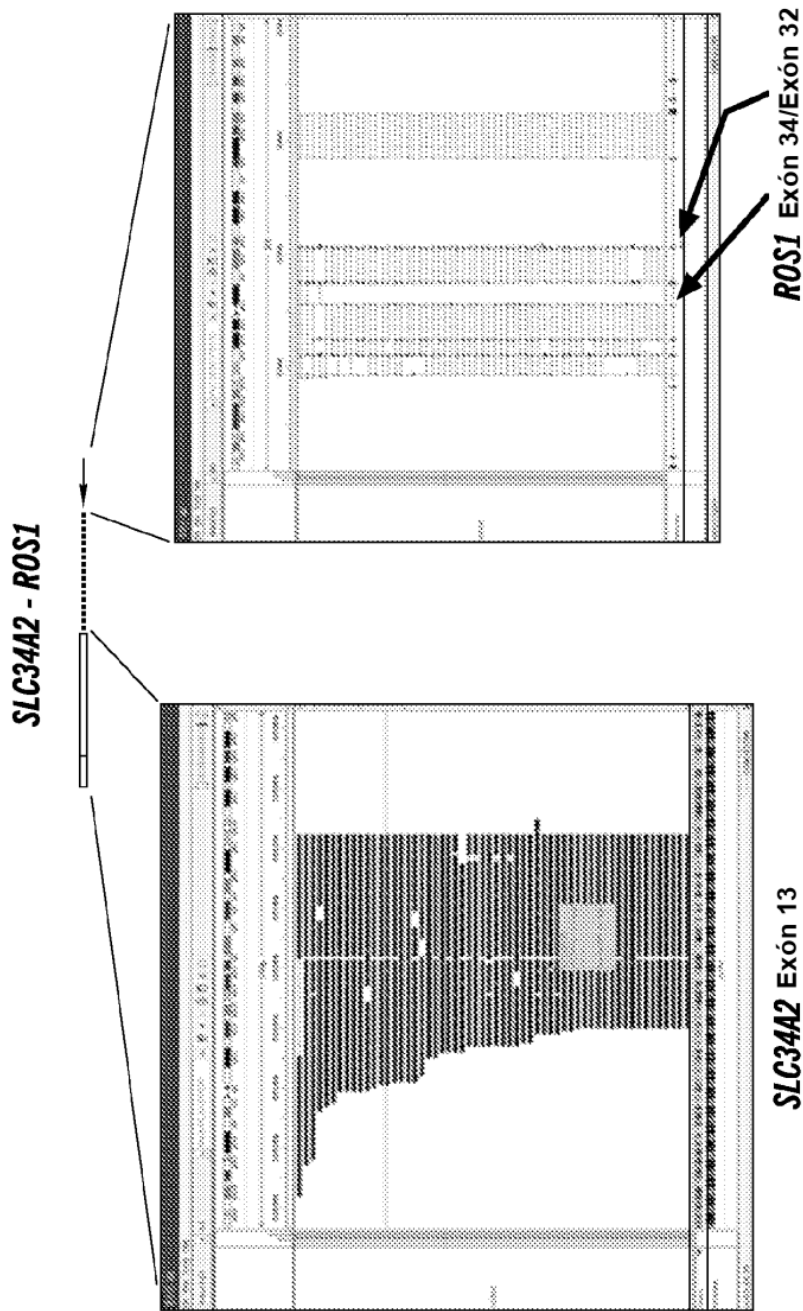


FIG. 3B



**FIG. 4A**

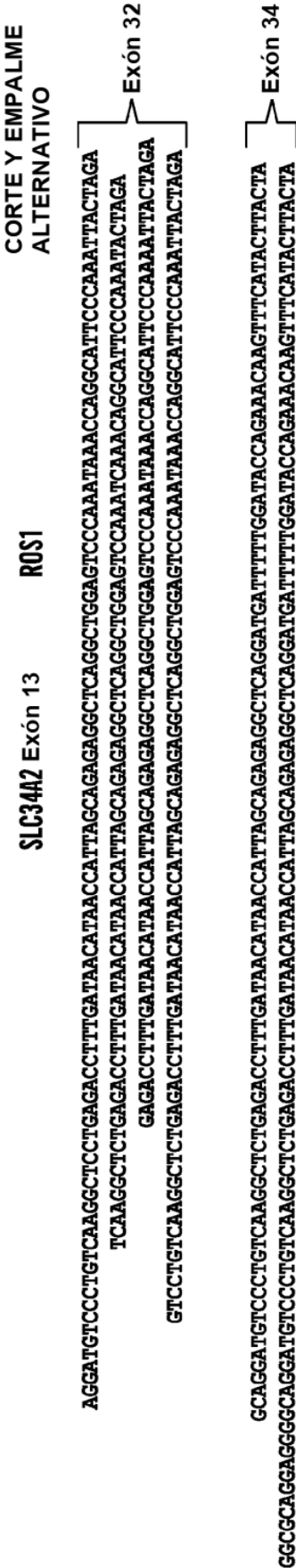


FIG. 4B

|              |                 |         |           |
|--------------|-----------------|---------|-----------|
| > MUESTRA_03 |                 | FUSIÓN  |           |
| FREQ         |                 |         |           |
| 164          | SLC34A2+Ex13.f2 | -- ROS1 | - Ex32.f1 |
| 103          | SLC34A2+Ex13.f2 | -- ROS1 | - Ex34.f1 |
| > MUESTRA_15 |                 | FUSIÓN  |           |
| FREQ         |                 |         |           |
| 41           | EZR+Ex10.f2     | --ROS1  | -Ex34.f1  |
| 8            | EZR+Ex10.f2     | --ROS1  | - Ex35.f1 |

FIG. 4C

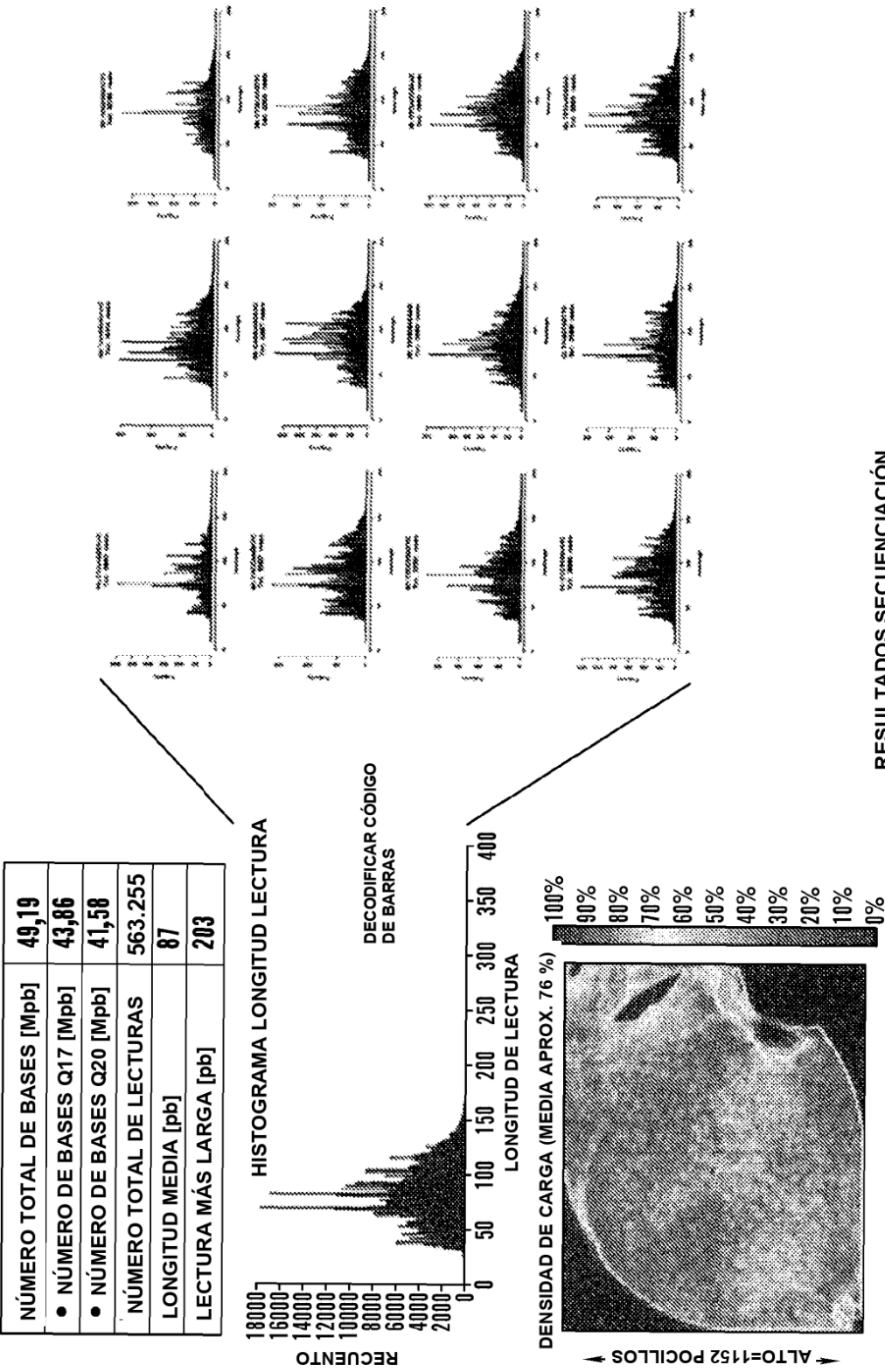
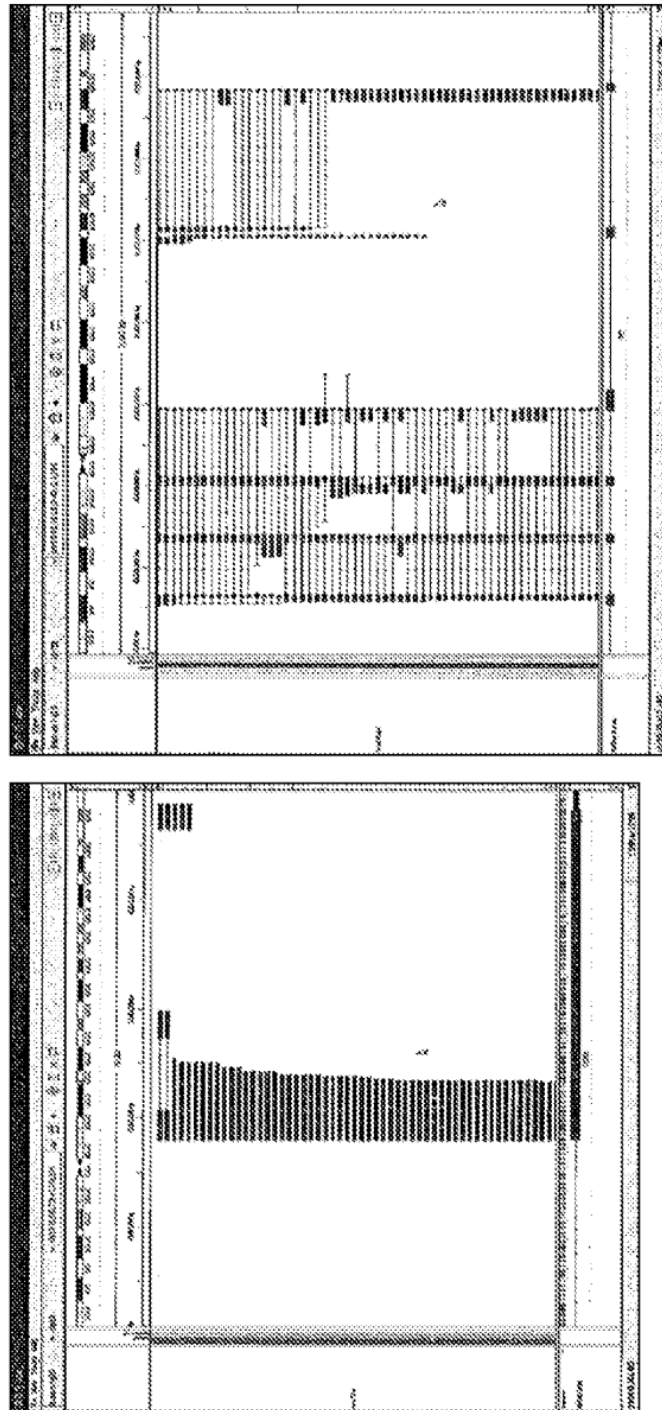


FIG. 5

CCDC6 -- RET



**FIG. 6**

EML4 -- ALK

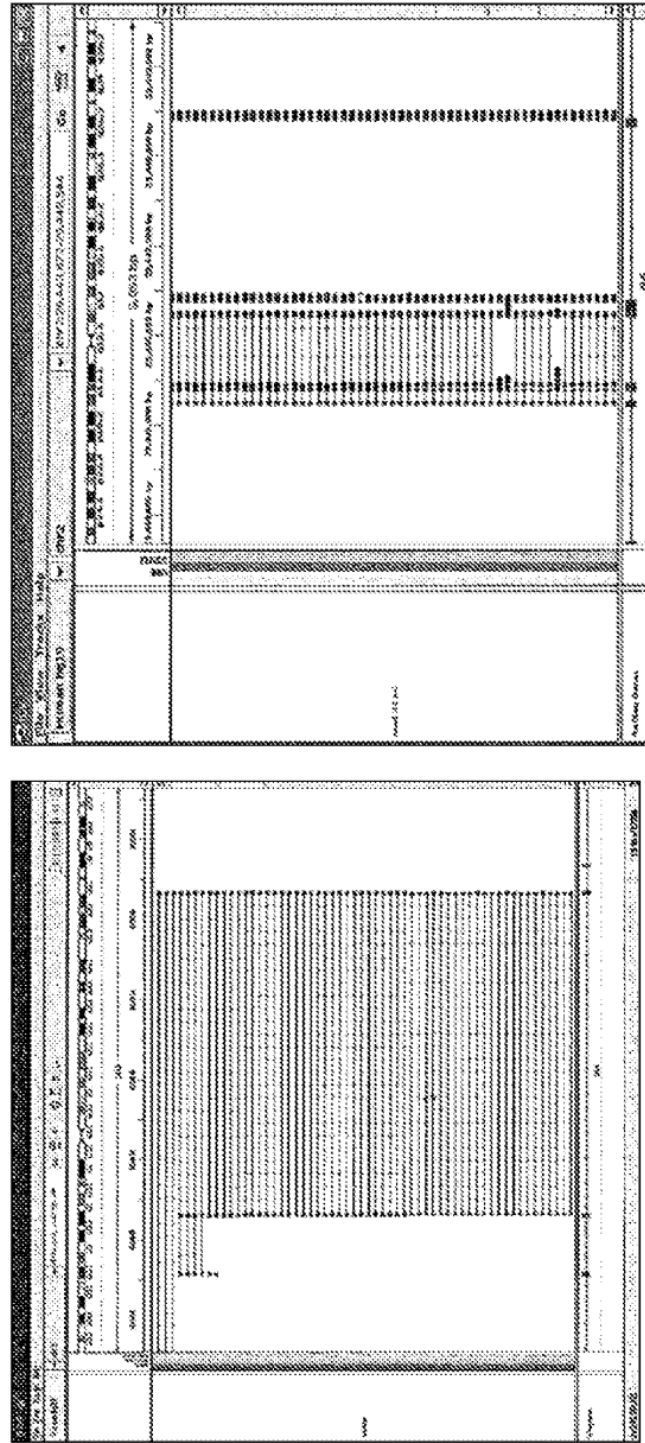
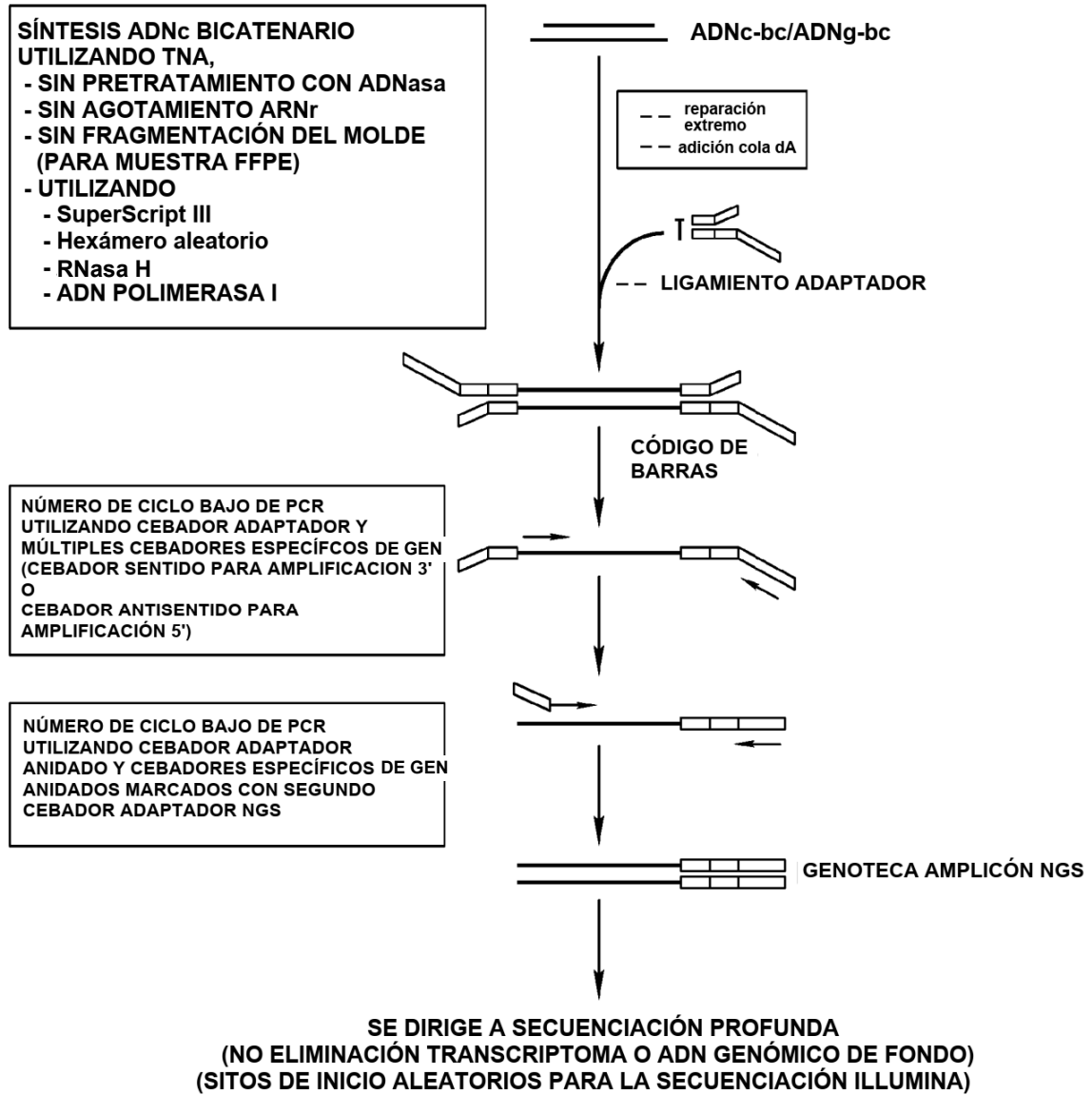
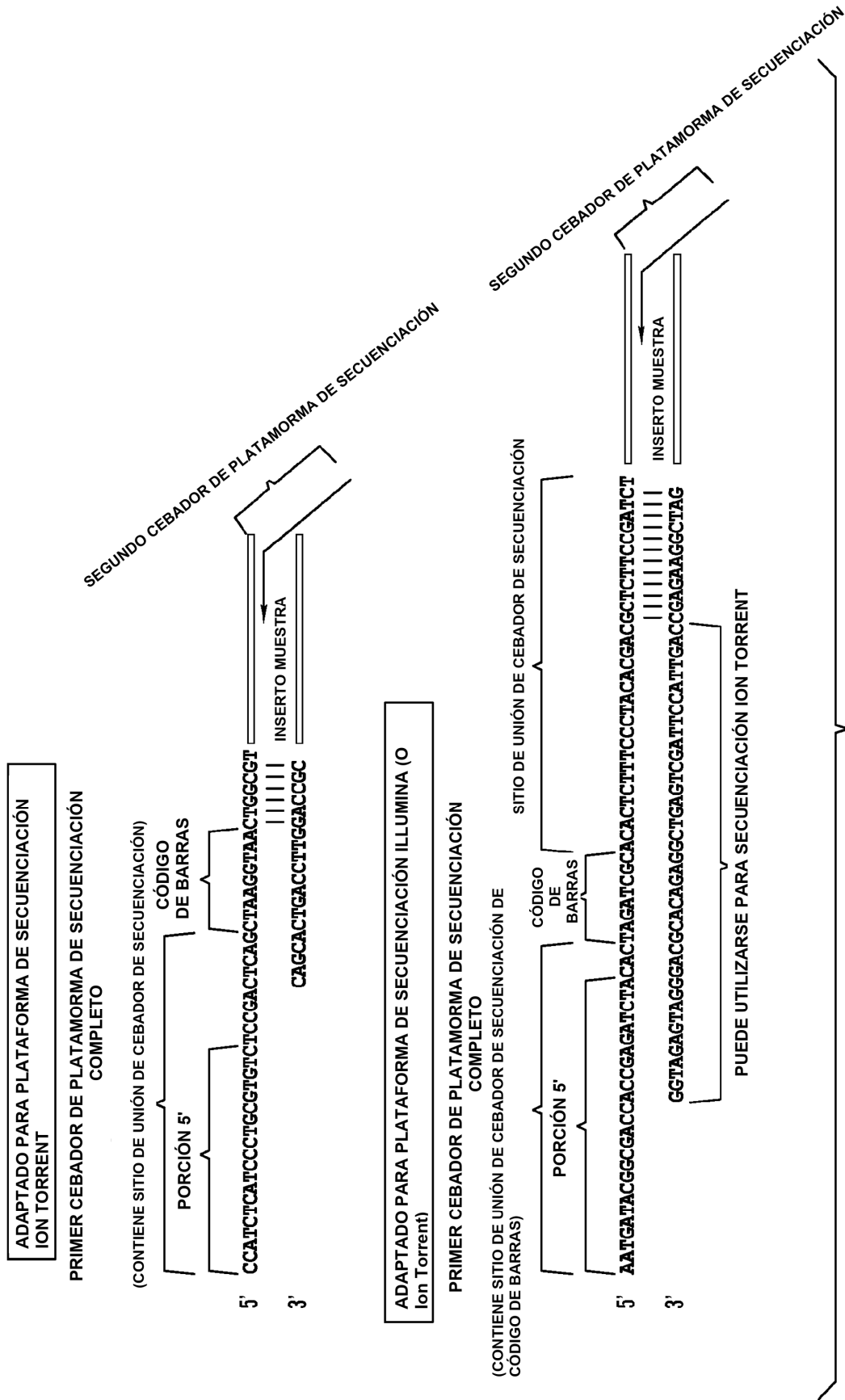


FIG. 7

SEC. DIRIGIDA ATN



**FIG. 8**



**FIG. 9**