

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 962 223**

51 Int. Cl.:

**C12Q 1/68** (2008.01)

**C12Q 1/6855** (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **13.04.2018** **E 21157081 (7)**

97 Fecha y número de publicación de la concesión europea: **23.08.2023** **EP 3885445**

54 Título: **Métodos para unir adaptadores a ácidos nucleicos de muestra**

30 Prioridad:

**14.04.2017 US 201762485769 P**  
**14.04.2017 WO PCT/US2017/027809**  
**18.04.2017 US 201762486663 P**  
**08.06.2017 US 201762517145 P**

45 Fecha de publicación y mención en BOPI de la  
traducción de la patente:  
**18.03.2024**

73 Titular/es:

**GUARDANT HEALTH, INC. (100.0%)**  
**3100 Hanover Street**  
**Palo Alto, CA 94304, US**

72 Inventor/es:

**KENNEDY, ANDREW y**  
**MORTIMER, STEFANIE ANN WARD**

74 Agente/Representante:

**IZQUIERDO BLANCO, María Alicia**

ES 2 962 223 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 unir adaptadores a ácidos nucleicos de muestra

### ANTECEDENTES

[0001] El cáncer es una de las principales causas de enfermedad a nivel mundial. Cada año, decenas de millones de personas son diagnosticadas de cáncer en todo el mundo, y más de la mitad acaban muriendo a causa del mismo. En muchos países, el cáncer es la segunda causa de muerte más frecuente tras las enfermedades cardiovasculares. La detección temprana de la enfermedad está asociada con unos resultados más positivos en muchos tipos de cáncer.

[0002] A menudo, el cáncer se detecta mediante biopsias de tumores, seguidas de un análisis de las células, los marcadores o el ADN extraído de las células. Sin embargo, más recientemente se ha propuesto que el cáncer también puede detectarse a partir de ácidos nucleicos libres de células que están presentes en los fluidos corporales, como la sangre o la orina (ver, por ejemplo, Siravegna et al., Nature Reviews, 2017). Estas pruebas tienen la ventaja de no ser invasivas y pueden realizarse sin tener que identificar mediante biopsia las células sospechosas de ser cancerosas. No obstante, la cantidad de ácido nucleico en los fluidos corporales es muy baja. Por consiguiente, estos análisis requieren métodos eficientes para convertir el ADN nativo libre de células -presente en los fluidos corporales- en formas susceptibles de analizarse.

[0003] Preparar las moléculas de ADN a partir de muestras de pacientes para su posterior análisis suele requerir que se reparen primero los salientes monocatenarios para permitir la ligación a los adaptadores para su amplificación y secuenciación. La reparación puede llevarse a cabo digiriendo la cadena saliente o extendiendo la cadena opuesta para producir un extremo romo, seguido de la fosforilación de los extremos 5' y la ligación del extremo romo a los adaptadores. De manera alternativa, tras la creación de extremos romos, los extremos romos pueden poliadenilarse con una polimerasa Taq. Los fragmentos poliadenilados se aparean y se ligan con adaptadores que incluyen una cola T de un solo nucleótido en el extremo 3'. Esta configuración favorece la ligación deseada de adaptador-molécula de ADN, pero la eficiencia de conversión global de las moléculas de ADN en una muestra con respecto a las moléculas que pueden secuenciarse puede seguir siendo inaceptablemente baja en el caso de las muestras en las que solo están disponibles pequeñas cantidades de ácido nucleico.

[0004] El documento WO 2015/044262 describe composiciones enzimáticas para la reparación, adenilación y fosforilación de extremos de ADN. El documento WO 2016/135300 desvela diversos métodos y kits para mejorar la eficiencia en la construcción de bibliotecas de secuenciación de próxima generación. Wu et al. (Nat. Methods, 9, i-ii, 2012) describen métodos para una creación más rápida de bibliotecas de secuenciación de próxima generación.

### RESUMEN

[0005] La invención proporciona un kit que comprende adaptadores que son al menos parcialmente bicatenarios con un saliente de un solo nucleótido T en el extremo 3' (adaptadores con cola T) y adaptadores que son al menos parcialmente bicatenarios con un saliente de un solo nucleótido C en el extremo 3' (adaptadores con cola C).

[0006] En algunas realizaciones, la relación molar de los adaptadores al menos parcialmente bicatenarios con un saliente de un solo nucleótido T con respecto al saliente de un solo nucleótido C es de entre 4:1 y 2:1, por ejemplo 3:1. En algunas realizaciones, el porcentaje de adaptadores de cola C con respecto a los adaptadores de cola T es de aproximadamente un 5-40% en moles, por ejemplo un 10-35%, un 15-25%, un 20-35%, un 25-35% o aproximadamente un 30%.

[0007] En algunas realizaciones, el kit también comprende adaptadores al menos parcialmente bicatenarios con extremos romos. En algunas realizaciones, la relación molar de los adaptadores de extremo romo respecto a los adaptadores con cola es de entre 1:5 y 1:500, por ejemplo de entre 1:10 y 1:100. En algunas realizaciones, los adaptadores de extremo romo están presentes con una relación molar de un 0,2-20%, o un 0,5-15% o un 1-10% respecto a los adaptadores con cola T y los adaptadores con cola C.

[0008] En algunas realizaciones, el kit también comprende una polimerasa T4 o un fragmento de Klenow grande; y/o una polimerasa Taq. En algunas realizaciones, el kit también comprende cuatro tipos de nucleótidos estándar.

[0009] En algunas realizaciones, el kit carece de adaptadores al menos parcialmente bicatenarios con colas 3' de un solo nucleótido A, y de adaptadores al menos parcialmente bicatenarios con colas 3' de un solo nucleótido G.

[0010] En algunas realizaciones, los adaptadores son adaptadores en forma de Y. En algunas realizaciones, los adaptadores comprenden oligonucleótidos de SEQ ID n.º 1 y 2, y 3 y 2.

**[0011]** En algunas realizaciones, los adaptadores comprenden un primer sitio de unión y/o un código de barras. En algunas realizaciones, los códigos de barras proporcionan entre 64 y 90 000, entre 400 y 22 500, entre 400 y 14 400 o entre 900 y 14 400 combinaciones de códigos de barras diferentes.

5 **[0012]** Un kit de la invención puede ser útil en un método para preparar ácidos nucleicos para su análisis, que incluye:  
 (a) crear ácidos nucleicos bicatenarios de extremos romos con salientes monocatenarios en una muestra mediante la acción de una o más enzimas que proporcionan actividad polimerasa 5'-3' y actividad de corrección de errores 3'-5', y cuatro tipos de nucleótidos estándar, de manera que los salientes monocatenarios con extremos 5' sirven como  
 10 plantillas para la extensión de una cadena complementaria mediante la actividad polimerasa, y los salientes monocatenarios con extremos 3' se digieren mediante la actividad de corrección de errores produciendo ácidos nucleicos de extremos romos; (b) sin separar los ácidos nucleicos de extremos romos del resto de componentes de la muestra, añadir colas en los extremos de los ácidos nucleicos de extremos romos mediante la acción de una polimerasa sin función de corrección de errores 3'-5' que lleva a cabo una adición dirigida sin plantilla de un nucleótido a los extremos 3' de los ácidos nucleicos de extremos romos, de manera que A se añade preferentemente a G o  
 15 preferentemente a C o T; (c) aparear los ácidos nucleicos del paso (c) con adaptadores al menos parcialmente bicatenarios con salientes de un solo nucleótido T o C en el extremo 3'; y (d) ligar los ácidos nucleicos con los adaptadores. Opcionalmente, el método también incluye desnaturalizar las una o más enzimas tras el paso (a). Opcionalmente, el método también incluye poner en contacto la muestra con las una o más enzimas, los cuatro tipos de nucleótidos estándar y la polimerasa sin una función de corrección de errores 3'-5'. Opcionalmente, la muestra se pone  
 20 en contacto a la vez con las una o más enzimas, los cuatro tipos de nucleótidos estándar y la polimerasa sin una función de corrección de errores 3'-5'. Opcionalmente, el paso (b) se lleva a cabo a una mayor temperatura que el paso (a). Opcionalmente, el paso (a) se lleva a cabo a temperatura ambiente y el paso (b) se lleva a cabo a una temperatura de más de 60°C. Opcionalmente, las una o más enzimas son una polimerasa con actividad polimerasa 5'-3' y actividad de corrección de errores 3'-5'. Opcionalmente, la polimerasa sin una función de corrección de errores 3'-5' es una polimerasa termotable y el método también incluye aumentar la temperatura de la muestra tras el paso (a) para inactivar la polimerasa con actividad polimerasa 5'-3' y actividad de corrección de errores 3'-5'. Opcionalmente, el método también incluye  
 25 (e) amplificar los ácidos nucleicos ligados a los adaptadores; y (f) analizar los ácidos nucleicos.

**[0013]** Opcionalmente, el método también incluye poner en contacto la muestra con adaptadores al menos  
 30 parcialmente bicatenarios de extremos romos que se ligan con ácidos nucleicos bicatenarios de extremos romos que no han recibido la adición dirigida sin plantilla de un nucleótido a los extremos 3' durante el paso de ligación. Opcionalmente, la primera polimerasa es una polimerasa T4 o un fragmento de Klenow grande. Opcionalmente, la segunda polimerasa es una polimerasa Taq. Opcionalmente, al menos los pasos (a)-(e) se llevan a cabo en un solo tubo. Opcionalmente, los pasos (a)-(f) o (a) a (g) se llevan a cabo en un solo tubo. Opcionalmente, la relación molar de los adaptadores al menos parcialmente bicatenarios con un solo nucleótido T con respecto a los de un solo nucleótido C es de entre 4:1 y 2:1, preferiblemente 3:1. Opcionalmente, la relación molar de los adaptadores de extremo romo respecto a los adaptadores con cola es de entre 1:5 y 1:500, preferiblemente de entre 1:10 y 1:100. Opcionalmente, al menos un 70% de los ácidos nucleicos bicatenarios de la muestra se unen a los adaptadores. Opcionalmente, se analizan al menos un 70% de los ácidos nucleicos bicatenarios disponibles en la muestra. Opcionalmente, el paso (f) comprende secuenciar los ácidos nucleicos ligados a los adaptadores. Opcionalmente, en la secuenciación se  
 40 secuencia un nucleótido que ha formado un saliente en el paso (c) o (d).

**[0014]** Un kit de la presente invención también puede ser útil en un método para convertir ADN bicatenario en ADN marcado con adaptadores, que incluye: (a) poner en contacto una población de moléculas de ADN bicatenario con una  
 45 población de adaptadores al menos parcialmente bicatenarios, de manera que: (i) la población de moléculas de ADN bicatenario comprende moléculas de ADN que comprenden un saliente con un solo nucleótido A y moléculas de ADN que comprenden un saliente con un solo nucleótido G, de manera que en la población los salientes con un solo nucleótido A son más abundantes (por ejemplo, 10 veces, 100 veces, 1000 veces) que los salientes con un solo nucleótido G, y (ii) la población de adaptadores al menos parcialmente bicatenarios comprende adaptadores que comprenden un saliente con un solo nucleótido T y adaptadores que comprenden un saliente con un solo nucleótido C; y (b) ligar los adaptadores con las moléculas de ADN, de manera que la ligación produce ADN marcado con adaptadores.

**[0015]** Opcionalmente, (i) la población de moléculas de ADN bicatenario también comprende al menos uno de los siguientes: moléculas de ADN que comprenden un saliente con un solo nucleótido C, moléculas de ADN que comprenden un saliente con un solo nucleótido T, o un extremo romo, y (ii) la población de adaptadores al menos parcialmente bicatenarios también comprende al menos uno de los siguientes: adaptadores que comprenden un saliente con un solo nucleótido G, adaptadores que comprenden un saliente con un solo nucleótido A, o un extremo romo. Opcionalmente, los adaptadores al menos parcialmente bicatenarios comprenden un sitio de unión del cebador de NGS ('secuenciación de próxima generación') y un código de barras de ADN. Opcionalmente, la población de adaptadores al menos parcialmente bicatenarios comprende múltiples códigos de barras de ADN diferentes. Opcionalmente, el número de combinaciones de códigos de barras que pueden unirse a ambos extremos de una molécula de ADN bicatenario es menor que el número de moléculas de ADN bicatenario presentes en la población, por ejemplo entre 5 y 10 000 combinaciones diferentes. Opcionalmente, el método también incluye: (c) amplificar el ADN marcado con adaptadores utilizando cebadores de amplificación que comprenden un código de barras de muestra y una secuencia de nucleótidos adaptada para hibridarse con un oligonucleótido inmovilizado en un soporte de celdas de flujo.

Opcionalmente, los adaptadores son adaptadores en forma de Y. Opcionalmente, la muestra es una muestra de fluido corporal, como sangre entera, suero o plasma. Opcionalmente, la población de ácido nucleico es una población de ácido nucleico libre de células. Opcionalmente, la muestra procede de un sujeto sospechoso de padecer cáncer. Opcionalmente, el paso de análisis detecta una variante somática o de línea germinal, una variación en el número de copias, una variación de un solo nucleótido (SNV), un indel o una fusión de genes.

**[0016]** Puede producirse una población de ácidos nucleicos adaptados utilizando un kit de la presente invención, de manera que la población comprende una pluralidad de moléculas de ácido nucleico y cada una de ellas comprende un fragmento de ácido nucleico flanqueado a ambos lados por un adaptador que incluye un código de barras con un par de bases A/T o G/C entre el fragmento de ácido nucleico y el adaptador. La pluralidad de moléculas de ácido nucleico puede incluir al menos 100 000 moléculas. La proporción de pares de bases A/T respecto a pares de bases G/C puede ser de entre 2:1 y 4:1. Al menos un 99% de las moléculas de ácido nucleico de la población pueden tener un fragmento de ácido nucleico flanqueado por adaptadores con diferentes códigos de barras.

## BREVE DESCRIPCIÓN DE LA FIGURA

**[0017]** La Figura 1 (FIG. 1) muestra adaptadores -de ADN de muestra- en forma de Y, de extremos romos, marcados y de unión a colas de T y colas de C.

## DEFINICIONES

**[0018]** El término 'sujeto' hace referencia a un animal, como una especie de mamífero (preferiblemente un ser humano) o una especie aviar (es decir, un pájaro), u otro organismo, por ejemplo una planta. Más específicamente, un sujeto puede ser un vertebrado, por ejemplo un mamífero, como un ratón, un primate, un simio o un ser humano. Los animales incluyen los animales de granja, los animales relacionados con el deporte y las mascotas. El sujeto puede ser un individuo sano, un individuo que padece una enfermedad o es sospechoso de padecer una enfermedad o tiene una predisposición a las enfermedades, o bien un individuo que necesita terapia o se sospecha que puede necesitar terapia.

**[0019]** Una variante genética es una alteración, una variante o un polimorfismo en una muestra de ácido nucleico o en el genoma de un sujeto. Dicha alteración, variante o polimorfismo puede producirse con respecto a un genoma de referencia, que puede ser un genoma de referencia del sujeto o de otro individuo. Las variaciones incluyen una o más variaciones de un solo nucleótido (SNVs), inserciones, deleciones, repeticiones, pequeñas inserciones, pequeñas deleciones, pequeñas repeticiones, uniones de variantes estructurales, repeticiones en tándem de longitud variable, y/o secuencias de flanqueo, variaciones en el número de copias (CNVs), transversiones y otros reordenamientos que también constituyen formas de variación genética. Una variación puede ser un cambio de base, una inserción, una deleción, una repetición, una variación del número de copias, una transversión o una combinación de las mismas.

**[0020]** Un marcador de cáncer es una variante genética asociada con la presencia o el riesgo de desarrollar cáncer. Un marcador de cáncer puede proporcionar un indicio de que un sujeto padece cáncer o tiene un mayor riesgo de desarrollar cáncer que un sujeto de la misma especie y de la misma edad y sexo. Un marcador de cáncer puede causar cáncer o no.

**[0021]** Una etiqueta de ácido nucleico es un ácido nucleico corto (por ejemplo, con una longitud de menos de 100, 50 o 10 nucleótidos), normalmente de una secuencia artificial y normalmente de ADN, que se usa para etiquetar los ácidos nucleicos de muestra a fin de diferenciar los ácidos nucleicos de diferentes muestras (por ejemplo, que constituyen un índice de muestras), de diferentes tipos o que han sido sometidos a diferentes procesos. Las etiquetas pueden ser de cadena simple o de cadena doble. Las etiquetas nucleicas pueden descodificarse para revelar información como la muestra de origen, la forma o el procesamiento de un ácido nucleico. Las etiquetas pueden usarse para permitir el agrupamiento y el procesamiento paralelo de múltiples ácidos nucleicos que portan diferentes etiquetas, de manera que los ácidos nucleicos se desgranaran posteriormente leyendo las etiquetas. Las etiquetas también pueden denominarse 'códigos de barras' o 'identificadores moleculares'.

**[0022]** Los adaptadores son ácidos nucleicos cortos (por ejemplo, con una longitud de menos de 500, 100 o 50 nucleótidos y normalmente de ADN) que normalmente son al menos parcialmente bicatenarios para formar enlaces con uno o ambos extremos de una molécula de ácido nucleico de muestra. Los adaptadores pueden incluir sitios de unión de cebadores para permitir la amplificación de una molécula de ácido nucleico de muestra flanqueada por adaptadores en ambos extremos, y/o un sitio de unión del cebador de secuenciación, incluidos los sitios de unión de cebadores para secuencias de próxima generación. Los adaptadores también pueden incluir sitios de unión para sondas de captura, como un oligonucleótido unido a un soporte de celdas de flujo. Los adaptadores también pueden incluir una etiqueta tal y como se ha descrito anteriormente. Preferiblemente, las etiquetas se posicionan en relación con el cebador y los sitios de unión del cebador de secuenciación, de tal manera que se incluye una etiqueta en los amplicones y las lecturas de secuenciación de un ácido nucleico de muestra. A los respectivos extremos de una molécula de muestra pueden unirse los mismos adaptadores o adaptadores diferentes. A veces, se une el mismo adaptador a los respectivos extremos, pero la etiqueta es diferente. Uno de los adaptadores preferidos es un adaptador en forma de Y en el que un extremo es un extremo romo o con cola, tal y como se describe en el presente documento, para unirse a un ácido nucleico de

muestra, que también tiene un extremo romo -o una cola- y un nucleótido complementario. Otro adaptador preferido es un adaptador en forma de campana que, de manera similar, tiene un extremo romo o con cola para unirse al ácido nucleico que ha de analizarse.

- 5 **[0023]** Los cuatro tipos de nucleótidos estándar son A, C, G, T para los desoxirribonucleótidos, y A, C, T y U para los ribonucleótidos.

## DESCRIPCIÓN DETALLADA

### 10 1. General

**[0024]** La preparación de muestras para las plataformas de secuenciación de nueva generación sigue a menudo un protocolo similar. Normalmente, las muestras contienen fragmentos de ácido nucleico bicatenario con salientes monocatenarios. Estos fragmentos pueden tener extremos romos y ligarse a los adaptadores directamente. Pero dichas ligaciones también generan subproductos en los que los adaptadores o fragmentos forman concatémicos. La formación de estos subproductos puede reducirse mediante un procedimiento alternativo en el que los fragmentos de extremos romos se poliadenilan y se ligan a adaptadores con cola T. Los kits comerciales que llevan a cabo la reparación de extremos y la adición de una cola en un solo tubo son rápidos y fáciles de usar y pueden usarse con adaptadores que están disponibles comercialmente. (Por ejemplo, NEBNext Ultra II; New England Biolabs, Ipswich, Massachusetts, Estados Unidos). Sin embargo, el uso de kits que no están optimizados para la poliadenilación puede dar como resultado la adición de colas de otros nucleótidos, como G, T y C. El resultado de una adición de colas ineficiente es una ligación ineficiente de los adaptadores y unas bibliotecas de baja complejidad.

**[0025]** Un kit de la invención puede ser útil en métodos mejorados para preparar ácidos nucleicos bicatenarios (preferiblemente ADN) con salientes monocatenarios para su amplificación y posterior análisis, particularmente una secuenciación. Se ha descubierto que poner en contacto ácidos nucleicos bicatenarios y de extremos romos con Taq en presencia de los cuatro tipos de nucleótidos estándar da como resultado una adición dirigida sin plantilla de un solo nucleótido a los extremos 3' del ácido nucleico, de tal manera que más frecuentemente se añade A, seguida de G y seguida de C y T. Si bien la inclusión de moléculas de ácido nucleico adicionales aumenta el potencial de reacciones secundarias fuera del objetivo, se ha descubierto que la proporción de adición de colas de G no es lo suficientemente alta con respecto a la adición de colas de una sola A y que la eficiencia de ligación de las moléculas de ácido nucleico de una muestra a los adaptadores puede aumentarse de manera considerable incluyendo una mezcla personalizada de adaptadores con adición de colas no solo de T (como en los métodos anteriores), sino también de C, de manera que dichos adaptadores se aparean respectivamente con los extremos 3' de las moléculas de ADN con colas de A y G. La eficiencia de ligación puede aumentarse aún más incluyendo también adaptadores de extremos romos (es decir, a los que no se les ha añadido ninguna cola con nucleótidos) para ligar las moléculas de ácido nucleico de extremos romos de la muestra que no han conseguido recibir una de adición de cola de ningún nucleótido.

### 40 2. Muestras

**[0026]** Una muestra puede ser cualquier muestra biológica aislada de un sujeto. Las muestras pueden incluir tejidos corporales, como tumores sólidos conocidos o sospechados, sangre entera, plaquetas, suero, plasma, heces, glóbulos rojos, glóbulos blancos o leucocitos, células endoteliales, biopsias tisulares, fluido cerebroespinal, fluido sinovial, fluido linfático, líquido ascítico, fluido intersticial o extracelular, el fluido de los espacios entre células, incluido el líquido crevicular gingival, médula ósea, efusiones pleurales, fluido cerebroespinal, saliva, mocos, esputos, semen, sudor u orina. Preferiblemente, las muestras son fluidos corporales, especialmente sangre y fracciones de esta, y orina. Una muestra puede estar aislada originalmente de un sujeto o puede haberse sometido a un procesamiento adicional para eliminar o añadir componentes, por ejemplo células, o para enriquecer un componente con respecto a otro. Por consiguiente, un fluido corporal preferido para su análisis es el plasma o suero que contiene ácidos nucleicos libres de células.

**[0027]** El volumen de plasma puede depender de la profundidad de lectura deseada para las regiones secuenciadas. Algunos volúmenes ejemplares son 0,4-40 mL, 5-20 mL, 10-20 mL. Por ejemplo, el volumen puede ser de 0,5 mL, 1 mL, 5 mL, 10 mL, 20 mL, 30 mL o 40 mL. El volumen de plasma muestreado puede ser, por ejemplo, de 5 a 20 mL.

**[0028]** Una muestra puede comprender diversas cantidades de ácido nucleico que contiene equivalentes genómicos. Por ejemplo, una muestra de aproximadamente 30 ng de ADN puede contener aproximadamente 10 000 equivalentes del genoma humano haploide y, en el caso del ADN libre de células, aproximadamente 200 000 millones de moléculas de ácido nucleico individuales. De manera similar, una muestra de aproximadamente 100 ng de ADN puede contener aproximadamente 30 000 equivalentes del genoma humano haploide y, en el caso del ADN libre de células, aproximadamente 600 000 millones de moléculas individuales. Algunas muestras contienen 1-500, 2-100, 5-150 ng de ADN libre de células, por ejemplo 5-30 ng o 10-150 ng de ADN libre de células.

**[0029]** Una muestra puede comprender ácidos nucleicos de diferentes fuentes. Por ejemplo, una muestra puede comprender ADN de línea germinal o ADN somático. Una muestra puede comprender ácidos nucleicos que son portadores de mutaciones. Por ejemplo, una muestra puede comprender ADN que porta mutaciones de la línea germinal

y/o mutaciones somáticas. Una muestra también puede comprender ADN que porta mutaciones asociadas al cáncer (por ejemplo, mutaciones somáticas asociadas al cáncer).

**[0030]** Las cantidades ejemplares de ácidos nucleicos libres de células presentes en una muestra antes de la amplificación son de entre aproximadamente 1 fg y aproximadamente 1 µg, por ejemplo de 1 pg a 200 ng, de 1 ng a 100 ng, o de 10 ng a 1000 ng. Por ejemplo, la cantidad puede ser de hasta aproximadamente 600 ng, de hasta aproximadamente 500 ng, de hasta aproximadamente 400 ng, de hasta aproximadamente 300 ng, de hasta aproximadamente 200 ng, de hasta aproximadamente 100 ng, de hasta aproximadamente 50 ng o de hasta aproximadamente 20 ng de moléculas de ácido nucleico libre de células. La cantidad puede ser de al menos 1 fg, al menos 10 fg, al menos 100 fg, al menos 1 pg, al menos 10 pg, al menos 100 pg, al menos 1 ng, al menos 10 ng, al menos 100 ng, al menos 150 ng o al menos 200 ng de moléculas de ácido nucleico libre de células. La cantidad puede ser de hasta 1 fentogramo (fg), 10 fg, 100 fg, 1 picogramo (pg), 10 pg, 100 pg, 1 ng, 10 ng, 100 ng, 150 ng o 200 ng de moléculas de ácido nucleico libre de células. El método puede comprender obtener de 1 fentogramo (fg) a 200 ng.

**[0031]** Una muestra ejemplar es de 5-10 ml de sangre entera, plasma o suero que incluye aproximadamente 30 ng de ADN o aproximadamente 10 000 equivalentes del genoma haploide.

**[0032]** Los ácidos nucleicos libres de células son ácidos nucleicos que no están contenidos en una célula o unidos de cualquier otra manera a una célula o, en otras palabras, son los ácidos nucleicos que quedan en una muestra de eliminación de células intactas. Los ácidos nucleicos libres de células incluyen el ADN, el ARN e híbridos de los mismos, lo que incluye el ADN genómico, el ADN mitocondrial, el siRNA, el mi-ARN, el ARN circulante (ARNc), el ARNt, el ARNr, el ARN pequeño nucleolar (snoRNA), el ARN asociado a Piwi (piRNA), el ARN no codificante largo (ncRNA largo), o fragmentos de cualquiera de estos. Los ácidos nucleicos libres de células pueden ser bicatenarios, monocatenarios o un híbrido de estos. Las moléculas de ADN bicatenario -de manera que al menos algunas de ellas tienen salientes monocatenarios- son una forma preferida de ADN libre de células para cualquier método desvelado en el presente documento. Un ácido nucleico libre de células puede liberarse en los fluidos corporales mediante secreción o mediante procesos de muerte celular, por ejemplo la necrosis y la apoptosis. Algunos ácidos nucleicos libres de células se liberan en los fluidos corporales desde células cancerosas, por ejemplo ADN circulante tumoral (ADNct). Otros se liberan desde células sanas.

**[0033]** Un ácido nucleico libre de células puede presentar una o más modificaciones epigenéticas; por ejemplo, un ácido nucleico libre de células puede estar acetilado, metilado, ubiquitinado, fosforilado, sumoilado, sumoilado, ribosilado y/o citrulinado.

**[0034]** Los ácidos nucleicos libres de células tienen una distribución de tamaño de unos 100-500 nucleótidos, particularmente de entre 110 y aproximadamente 230 nucleótidos, con un modo de unos 168 nucleótidos y un segundo pico menor de entre 240 y 440 nucleótidos.

**[0035]** Los ácidos nucleicos libres de células pueden aislarse de los fluidos corporales mediante un paso de partición en el que los ácidos nucleicos libres de células, tal y como se encuentran en la solución, se separan de las células intactas y otros componentes no solubles del fluido corporal. La partición puede incluir técnicas como la centrifugación o la filtración. De manera alternativa, las células de los fluidos corporales pueden lisarse, y los ácidos nucleicos celulares y los ácidos nucleicos libres de células pueden procesarse conjuntamente. Generalmente, tras añadir tampones y pasos de lavado, los ácidos nucleicos pueden precipitarse con un alcohol. También pueden usarse otros pasos de limpieza adicionales, como columnas a base de sílice, para eliminar las sales o los contaminantes. Por ejemplo, durante la reacción pueden añadirse ácidos nucleicos portadores de volumen no específicos para optimizar algunos aspectos del procedimiento, como el rendimiento.

**[0036]** Después del procesamiento, las muestras pueden incluir varias formas de ácidos nucleicos, incluidos el ADN bicatenario, el ADN monocatenario y el ARN monocatenario. De manera opcional, el ADN y el ARN monocatenarios pueden convertirse en formas bicatenarias para que puedan incluirse en posteriores pasos de análisis y procesamiento.

### 3. Unión de las moléculas de ácido nucleico de muestra a los adaptadores

**[0037]** El ácido nucleico presente en una muestra con o sin un procesamiento previo, tal y como se ha descrito anteriormente, suele contener una parte sustancial de moléculas en forma de moléculas parcialmente bicatenarias con salientes monocatenarios. Estas moléculas pueden convertirse en moléculas bicatenarias de extremos romos tratándose con una o más enzimas para proporcionar una polimerasa 5'-3' y una exonucleasa 3'-5' (o función correctora) en presencia de los cuatro tipos de nucleótidos estándar, tal y como se muestra en la parte superior de la Figura 1. Esta combinación de actividades puede extender cadenas con un extremo 3' encastrado, de manera que se alinean por el extremo con el extremo 5' de la cadena opuesta (en otras palabras, se crea un extremo romo), o bien puede digerir cadenas con salientes de 3', de manera que, igualmente, se alinean con el extremo 5' de la cadena opuesta. Opcionalmente, ambas actividades pueden obtenerse mediante una sola polimerasa. Preferiblemente, la polimerasa es sensible al calor, de manera que su actividad puede interrumpirse cuando se eleva la temperatura. Los ejemplos de polimerasas adecuadas incluyen los fragmentos de Klenow grandes y la polimerasa T4.

**[0038]** Preferiblemente, las una o más enzimas que proporcionan actividad polimerasa 5'-3' y actividad exonucleasa 3'-5' se desnaturalizan subiendo la temperatura o de cualquier otra manera. Por ejemplo, la desnaturalización puede llevarse a cabo subiendo la temperatura a 75°-80° C. Después, se actúa sobre las muestras con una polimerasa que carece de una función correctora (parte central de la Figura 1). Preferiblemente, esta polimerasa es termoestable, de manera que permanece activa a elevadas temperaturas. Las polimerasas Taq, Tth y los fragmentos grandes de Bst son ejemplos de una polimerasa de este tipo. La segunda polimerasa lleva a cabo una adición sin plantilla de un solo nucleótido a los extremos 3' de ácidos nucleicos de extremos romos. Si bien la mezcla de reacción normalmente contiene cantidades molares iguales de cada uno de los cuatro tipos de nucleótidos estándar del paso anterior, los cuatro tipos de nucleótidos no se añaden a los extremos 3' en igual proporción. En lugar de ello, A se añade más frecuentemente, seguida de G y seguida de C y T.

**[0039]** Después de añadir colas a las moléculas de muestra, y con o sin una posterior purificación de las moléculas de muestra con cola, las moléculas de muestra con cola se ponen en contacto con adaptadores que tienen colas con nucleótidos T y C complementarios en un extremo de los adaptadores (parte inferior de la Figura 1). Normalmente, los adaptadores se forman mediante la síntesis separada y el apareamiento de sus respectivas cadenas. Por lo tanto, las colas de T y C adicionales pueden añadirse como un nucleótido extra en la síntesis de una de las cadenas. Normalmente no se incluyen los adaptadores con colas de G y A, puesto que, si bien estos adaptadores pueden aparearse con moléculas de muestra con colas de C y T, respectivamente, también se aparearían con otros adaptadores. Las moléculas adaptadoras y las moléculas de muestra que contienen nucleótidos complementarios (por ejemplo, T-A y C-G) en sus extremos 3' se aparean y pueden ligarse unas con otras. El porcentaje de adaptadores con colas de C con respecto a los adaptadores con colas de T es de aproximadamente un 5-40% en moles, por ejemplo un 10-35%, un 15-25%, un 20-35%, un 25-35% o aproximadamente un 30%. Puesto que la adición dirigida sin plantilla de un solo nucleótido a los extremos 3' de las moléculas de muestra no llega a completarse del todo, la muestra también contiene algunas moléculas de muestra con extremos romos que no han recibido colas. Estas moléculas pueden recuperarse proporcionando también a la muestra adaptadores que tengan un -y, preferiblemente, sólo un- extremo romo. Normalmente, los adaptadores con extremos romos se proporcionan con un ratio molar de un 0,2-20% o un 0,5-15% o un 1-10% de los adaptadores con adaptadores con colas de T y C. Pueden proporcionarse adaptadores de extremos romos al mismo tiempo, antes o después de los adaptadores con colas de T y C. Los adaptadores de extremos romos que se ligan de nuevo con moléculas de muestra de extremos romos producen moléculas de muestra flanqueadas a ambos lados por adaptadores. Estas moléculas carecen de los pares de nucleótidos A-T o C-G entre las muestras y los adaptadores que están presentes cuando las moléculas de muestra con colas se ligan con los adaptadores con colas.

**[0040]** Preferiblemente, los adaptadores usados en estas reacciones tienen un -y sólo un- extremo con cola de T o C o un -y sólo un- extremo romo, de manera que pueden ligarse con las moléculas de muestra únicamente en una orientación. Los adaptadores pueden ser, por ejemplo, adaptadores con forma de Y en los que un extremo tiene cola o es romo y el otro extremo tiene dos cadenas simples. Los adaptadores con forma de Y ejemplares tienen las siguientes secuencias, de manera que '(6 bases)' indica una etiqueta. El oligonucleótido superior incluye una cola con una sola base T.

**[0041]** Adaptador universal:

5' AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT (SEQ ID NO. 1).

**[0042]**

Adaptador, Índice 1-12: 5' GATCGGAAGAGCACACGTCTGAACTCCAGTCAC (6 bases) ATCTCGTATGCCGTCTTCTGCTTG (SEQ ID NO. 2)

**[0043]** Otro adaptador en forma de Y con una cola de C tiene las siguientes secuencias:

5' AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCC (SEQ ID NO. 3) y el Adaptador, Índice 1-12: 5' GATCGGAAGAGCACACGTCTGAACTCCAGTCAC (6 bases) ATCTCGTATGCCGTCTTCTGCTTG (SEQ ID NO. 2)

**[0044]** Pueden sintetizarse combinaciones personalizadas de estos oligonucleótidos, incluidos los oligonucleótidos con ambas colas (de T y C), para usarse en los métodos de la presente invención.

**[0045]** Una versión truncada de estas secuencias adaptadoras se describe en Rohland et al., *Genome Res.*, mayo de 2012; 22(5): 939-946.

**[0046]** Los adaptadores también pueden tener forma de campana con un solo extremo, que tiene cola o es romo. Los adaptadores pueden incluir un sitio de unión de cebadores para amplificación, un sitio de unión para un cebador de secuenciación y/o una etiqueta de ácido nucleico con fines de identificación. En una sola reacción pueden usarse los mismos adaptadores o adaptadores diferentes.

**[0047]** Cuando los adaptadores incluyen una etiqueta de identificación y los ácidos nucleicos de una muestra se unen a los adaptadores en cada extremo, el número de potenciales combinaciones de identificadores aumenta exponencialmente con el número de etiquetas únicas proporcionadas (es decir,  $n^n$  combinaciones, donde  $n$  es el número de etiquetas de identificación únicas). En algunos métodos, el número de combinaciones de etiquetas únicas es suficiente para que sea estadísticamente probable que todas o prácticamente todas (por ejemplo, al menos un 90%) las diferentes moléculas de ADN bicatenario presentes en la muestra reciban una combinación de etiquetas diferente. En algunos métodos, el número de combinaciones únicas de etiquetas identificadoras es menor que el número de moléculas únicas de ADN bicatenario presentes en la muestra (por ejemplo, 5-10000 combinaciones de etiquetas diferentes).

**[0048]** Uno de los kits que proporcionan enzimas adecuadas para llevar a cabo los métodos anteriores es el kit 'NEBNext® Ultra™ II DNA Library Prep Kit' de Illumina®. El kit proporciona los siguientes reactivos:

**[0049]** La mezcla enzimática 'NEBNext Ultra II End Prep Enzyme Mix', el tampón de reacción 'NEBNext Ultra II End Prep Reaction Buffer', el potenciador de ligación 'NEBNext Ligation Enhancer', la mezcla 'NEBNext Ultra II Ligation Master Mix -20' y la mezcla 'NEBNext® Ultra II Q5® Master Mix'.

**[0050]** La formación de extremos romos y la adición de colas a ácidos nucleicos de muestra pueden llevarse a cabo en un solo tubo. No es necesario separar los ácidos nucleicos de extremos romos de la(s) enzima(s) que llevan a cabo la formación de extremos romos antes de que se produzca la reacción de adición de colas. Opcionalmente, todas las enzimas, nucleótidos y otros reactivos se suministran a la vez antes de que se produzca la reacción de formación de extremos romos. 'Suministrar a la vez' quiere decir que todos se introducen en la muestra lo suficientemente próximos en el tiempo como para que todos estén presentes cuando se produzca la incubación de la muestra a fin de que tenga lugar la formación de extremos romos. Opcionalmente, no se elimina nada de las muestras después de suministrar las enzimas, los nucleótidos y otros reactivos al menos hasta que se hayan completado las incubaciones para la formación de extremos romos y para la adición de colas. A menudo, la reacción de adición de colas se lleva a cabo a una temperatura mayor que la reacción de formación de extremos romos. Por ejemplo, la reacción de formación de extremos romos puede realizarse a temperatura ambiente, en la que la polimerasa 5'-3' y la exonucleasa 3'-5' están activas y la polimerasa termoestable está inactiva o mínimamente activa, y la reacción de adición de colas puede realizarse a una temperatura elevada, por ejemplo de más de 60° C, cuando la polimerasa 5'-3' y la exonucleasa 3'-5' están inactivas y la polimerasa termoestable está activa.

**[0051]** La unión de los adaptadores con colas de T y C tal y como se ha descrito da como resultado una población de ácidos nucleicos adaptados, de manera que la población comprende una pluralidad de moléculas de ácido nucleico y cada una de ellas comprende un fragmento de ácido nucleico flanqueado a ambos lados por un adaptador que incluye un código de barras con un par de bases A/T o G/C entre el fragmento de ácido nucleico y el adaptador. La pluralidad de moléculas de ácido nucleico puede incluir al menos 10 000, 100 000 o 1 000 000 de moléculas. La proporción de pares de bases A/T con respecto a pares de bases G/C en las regiones de unión entre los fragmentos y los adaptadores de flanco depende de la proporción entre los adaptadores con colas de T y los adaptadores con colas de C y puede ser, por ejemplo, de entre 2:1 y 4:1. La mayoría de los ácidos nucleicos de la población están flanqueados por adaptadores con diferentes códigos de barras (por ejemplo, al menos un 99%). Si también se incluyen adaptadores de extremos romos, la población incluye moléculas de ácido nucleico en un fragmento de ácido nucleico que se une directamente a un adaptador por un extremo o ambos (es decir, no interviene ningún par A/T o G/C).

#### 4. Amplificación

**[0052]** Los ácidos nucleicos de muestra flanqueados por adaptadores pueden amplificarse mediante PCR y otros métodos de amplificación normalmente cebados por cebadores que se unen a sitios de unión de cebadores en los adaptadores que flanquean un ácido nucleico que se va a amplificar. Los métodos de amplificación pueden incluir ciclos de extensión, desnaturalización y apareamiento derivados de un termociclado o pueden ser isotérmicos, como en la amplificación mediada por transcripción. Otros métodos de amplificación incluyen la reacción en cadena de la ligasa, la amplificación por desplazamiento de cadena, la amplificación basada en secuencias de ácidos nucleicos y la replicación autosostenida basada en secuencias.

**[0053]** Preferiblemente, los métodos en los que pueden usarse los kits de la invención provocan que al menos un 75, 80, 85, 90 o 95% de los ácidos nucleicos bicatenarios presentes en la muestra se unan a los adaptadores. Preferiblemente, el uso de colas de T y C aumenta el porcentaje de ácidos nucleicos bicatenarios presentes en la muestra que se unen a adaptadores, en comparación con los métodos de control realizados sólo con adaptadores con



colas de T, en al menos un 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10% (de manera que un aumento de rendimiento desde un 75% hasta un 80% se considera un aumento de un 5%). Preferiblemente, la adición de colas de T y C en combinación con adaptadores de extremos romos aumenta el porcentaje de ácidos nucleicos bicatenarios que se unen a adaptadores en al menos un 5, 10, 15, 20 o 25%. El porcentaje de ácidos nucleicos que se unen a adaptadores puede determinarse mediante una electroforesis comparativa en gel de la muestra original y la muestra procesada después de que se haya completado la unión a los adaptadores.

**[0054]** Preferiblemente, los métodos en los que pueden usarse los kits de la invención provocan que al menos un 75, 80, 85, 90 o 95% de las moléculas bicatenarias disponibles presentes en la muestra se secuencien. Preferiblemente, el uso de colas de T y C aumenta el porcentaje de ácidos nucleicos bicatenarios presentes en la muestra que se secuencian, en comparación con los métodos de control realizados sólo con adaptadores con colas de T, en al menos un 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10%. Preferiblemente, la adición de colas de T y C en combinación con adaptadores de extremos romos aumenta el porcentaje de ácidos nucleicos bicatenarios presentes en la muestra que se secuencian, en comparación con los métodos de control realizados sólo con adaptadores con colas de T, en al menos un 5, 10, 15, 20 o 25%. El porcentaje de ácidos nucleicos que se secuencian puede determinarse comparando el número de moléculas que se secuencian realmente en base al número que podría haberse secuenciado en base a los ácidos nucleicos de entrada y las regiones del genoma que son objeto de la secuenciación.

## 5. Etiquetas

**[0055]** Las etiquetas que proporcionan códigos de barras o identificadores moleculares pueden incorporarse o unirse a los adaptadores mediante ligación o PCR de extensión solapada, entre otros métodos. Generalmente, la asignación de identificadores únicos o no únicos -o códigos de barras moleculares- en las reacciones sigue los métodos y sistemas que se describen en las solicitudes de patente de EE. UU. 20010053519, 20030152490 y 20110160078 y en las patentes de EE. UU. nºs 6,582,908 y 7,537,898.

**[0056]** Las etiquetas pueden unirse a los ácidos nucleicos de muestra de manera aleatoria o no aleatoria. En algunos casos, se introducen en micropocillos en un ratio esperado de identificadores únicos. Por ejemplo, los identificadores únicos pueden cargarse de manera que se cargan más de aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, 100, 500, 1000, 5000, 10 000, 50 000, 100 000, 500 000, 1 000 000, 10 000 000, 50 000 000 o 1 000 000 000 de identificadores únicos por cada muestra de genoma. En algunos casos, los identificadores únicos pueden cargarse de manera que se cargan menos de aproximadamente 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, 100, 500, 1000, 5000, 10 000, 50000, 100 000, 500 000, 1 000 000, 10 000 000, 50 000 000 o 1 000 000 000 de identificadores únicos por cada muestra de genoma. En algunos casos, el número promedio de identificadores únicos que se cargan por cada muestra de genoma es menor que, o mayor que, aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, 100, 500, 1000, 5000, 10000, 50000, 100 000, 500 000, 1 000 000, 10 000 000, 50 000 000 o 1 000 000 000 de identificadores únicos por cada muestra de genoma.

**[0057]** En algunos casos, los identificadores únicos pueden ser oligonucleótidos de una secuencia predeterminada, aleatoria o semialeatoria. En otros casos, puede usarse una pluralidad de códigos de barras, de tal manera que los códigos de barras no son necesariamente únicos unos respecto a otros dentro de la pluralidad. En este ejemplo, los códigos de barras pueden estar ligados a moléculas individuales, de tal manera que la combinación del código de barras y la secuencia a la que puede estar ligado crea una secuencia única que puede rastrearse individualmente. Tal y como se describe en el presente documento, la detección de códigos de barras no únicos en combinación con los datos de secuencia de las porciones iniciales (comienzo) y finales (parada) de las lecturas de secuencia puede permitir asignar una identidad única a una molécula particular. La longitud, o el número de pares de bases, de una lectura de secuencia individual también puede usarse para asignar una identidad única a dicha molécula. Tal y como se describe en el presente documento, los fragmentos de una cadena simple de ácido nucleico al que se le ha asignado una identidad única pueden usarse, por tanto, para identificar posteriormente fragmentos de la cadena principal.

**[0001]** Los polinucleótidos de una muestra pueden etiquetarse con el número suficiente de etiquetas diferentes para que haya una alta probabilidad (por ejemplo, de al menos un 90%, al menos un 95%, al menos un 98%, al menos un 99%, al menos un 99,9% o al menos un 99,99%) de que todos los polinucleótidos que corresponden a una región genómica particular porten una etiqueta identificadora diferente (las moléculas de la región están básicamente etiquetadas de forma única). La región genómica a la que corresponden los polinucleótidos puede ser, por ejemplo, (1) todo el panel de los genes que se están secuenciando, (2) alguna parte de dicho panel, como la asignación a un solo gen, exón o intrón, (3) una única coordenada de nucleótidos (por ejemplo, al menos un nucleótido del polinucleótido corresponde a la coordenada, por ejemplo la posición de inicio, la posición de parada, un punto intermedio o cualquier punto entre ambas), o (4) un par particular de coordenadas de nucleótidos de inicio/parada (comienzo/final). El número de identificadores diferentes (recuento de etiquetas) que es necesario para etiquetar prácticamente de forma única los polinucleótidos es una función de cuántas moléculas de polinucleótidos originales de la muestra corresponden a la región. Esto, a su vez, es una función de varios factores. Uno de los factores es el número total de equivalentes del genoma haploide que se han incluido en el ensayo. Otro factor es el tamaño promedio de las moléculas de polinucleótidos. Otro factor es la distribución de las moléculas a lo largo de la región. Esto, a su vez, puede ser una función del patrón de escisión (cabe esperar que la escisión se produzca principalmente entre nucleosomas, de manera que haya más polinucleótidos asignados a lo largo de la ubicación de un nucleosoma que entre nucleosomas). Otro

factor es la distribución de los códigos de barras en el 'pool' y la eficiencia de ligación de los códigos de barras individuales, lo que potencialmente provoca diferencias en la concentración efectiva de un código de barras con respecto a otro. Otro factor es el tamaño de la región en la que están confinadas las moléculas que han de etiquetarse de forma única (por ejemplo, el mismo inicio/parada o el mismo exón).

**[0002]** El identificador puede ser un solo código de barras unido a un extremo de una molécula, o dos códigos de barras, cada uno de ellos unido a diferentes extremos de la molécula. Unir códigos de barras independientemente a ambos extremos de una molécula aumenta al cuadrado el número de posibles identificadores. En este caso, el número de códigos de barras diferentes se selecciona de tal manera que la combinación de los códigos de barras en cada extremo de un polinucleótido particular tiene una alta probabilidad de ser única con respecto a otros polinucleótidos asignados a la misma región genómica seleccionada.

**[0003]** En algunas realizaciones, el número de combinaciones de códigos de barras o identificadores diferentes (recuento de etiquetas) utilizado puede ser de al menos 64, 100, 400, 900, 1400, 2500, 5625, 10 000, 14 400, 22 500 o 40 000 y de no más de 90 000, 40 000, 22 500, 14 400 o 10 000. Por ejemplo, el número de combinaciones de códigos de barras o identificadores puede ser de entre 64 y 90 000, de entre 400 y 22 500, de entre 400 y 14 400 o de entre 900 y 14 400.

**[0004]** En una muestra que comprende ADN genómico fragmentado, por ejemplo ADN libre de células (ADNcf), de una pluralidad de genomas, hay bastantes probabilidades de que más de un polinucleótido de diferentes genomas tenga las mismas posiciones de inicio y parada ('duplicados' o 'análogos'). El número probable de que los duplicados comiencen en cualquier posición es una función del número de equivalentes del genoma haploide presentes en una muestra y de la distribución de los tamaños de los fragmentos. Por ejemplo, el ADNcf tiene un máximo de fragmentos en aproximadamente 160 nucleótidos y la mayoría de los fragmentos en este máximo oscilan entre aproximadamente 140 nucleótidos y 180 nucleótidos. Por consiguiente, el ADNcf de un genoma de aproximadamente 3000 millones de bases (por ejemplo, el genoma humano) puede estar compuesto de casi 20 millones ( $2 \times 10^7$ ) de fragmentos de polinucleótidos. Una muestra de aproximadamente 30 ng de ADN puede contener aproximadamente 10 000 equivalentes del genoma humano haploide. (De manera similar, una muestra de aproximadamente 100 ng de ADN puede contener aproximadamente 30 000 equivalentes del genoma humano haploide). Una muestra que contenga aproximadamente 10000 ( $10^4$ ) equivalentes del genoma haploide de este ADN puede contener aproximadamente 200 000 millones ( $2 \times 10^{11}$ ) de moléculas de polinucleótidos individuales. Se ha determinado empíricamente que en una muestra de aproximadamente 10 000 equivalentes del genoma haploide de ADN humano hay aproximadamente 3 polinucleótidos duplicados que comienzan en cualquier posición dada. Por consiguiente, dicha colección puede contener una variedad de aproximadamente  $6 \times 10^{10}$ - $8 \times 10^{10}$  (aproximadamente entre 60 000 millones y 80 000 millones; por ejemplo, aproximadamente 70 000 millones [ $7 \times 10^{10}$ ]) de moléculas de polinucleótidos secuenciadas diferentemente.

**[0005]** La probabilidad de identificar correctamente las moléculas depende del número inicial de equivalentes genómicos, de la distribución de longitud de las moléculas secuenciadas, de la uniformidad de secuencias y del número de etiquetas. El número puede calcularse usando una distribución de Poisson. Cuando el recuento de etiquetas es igual a uno, equivale a no tener ninguna etiqueta única o ningún etiquetado. La Tabla 1 de más abajo muestra las probabilidades de identificar correctamente una molécula como única suponiendo una distribución de tamaño libre de células típica como la mencionada anteriormente.

Tabla 1

Recuento de etiquetas	Porcentaje de etiquetas identificadas correctamente de forma única
1000 equivalentes del genoma haploide humano	
1	96,9643
4	99,2290
9	99,6539
16	99,8064
25	99,8741
100	99,9685
3000 equivalentes del genoma haploide humano	
1	91,7233
4	97,8178
9	99,0198
16	99,4424
25	99,6412
100	99,9107

**[0006]** En este caso, al secuenciar el ADN genómico, puede que no sea posible determinar qué lecturas de secuencia se derivan de qué moléculas originales. Este problema puede atenuarse etiquetando las moléculas originales con un número suficiente de identificadores únicos (es decir, el recuento de etiquetas), de tal manera que haya probabilidades de que dos moléculas duplicadas, es decir, moléculas que tienen las mismas posiciones de inicio y parada, porten identificadores únicos diferentes a fin de que las lecturas de secuencia puedan rastrearse hasta las moléculas originales particulares. Un enfoque para este problema consiste en etiquetar de forma única todas -o casi todas- las moléculas originales diferentes presentes en la muestra. Sin embargo, dependiendo del número de equivalentes génicos haploides y la distribución de los tamaños de los fragmentos en la muestra, esto podría requerir miles de millones de identificadores únicos diferentes.

**[0007]** Este método puede resultar caro y laborioso. Una población de polinucleótidos en una muestra de ADN genómico fragmentado puede etiquetarse con  $n$  identificadores únicos diferentes, de manera que  $n$  es al menos 2 y no más de  $100\,000 \times z$ , de manera que  $z$  es un indicativo de la tendencia central (es decir, la media, la mediana, la moda) del número esperado de moléculas duplicadas que tienen las mismas posiciones de inicio y parada.  $n$  puede ser al menos cualquiera de entre  $2 \times z$ ,  $3 \times z$ ,  $4 \times z$ ,  $5 \times z$ ,  $6 \times z$ ,  $7 \times z$ ,  $8 \times z$ ,  $9 \times z$ ,  $10 \times z$ ,  $11 \times z$ ,  $12 \times z$ ,  $13 \times z$ ,  $14 \times z$ ,  $15 \times z$ ,  $16 \times z$ ,  $17 \times z$ ,  $18 \times z$ ,  $19 \times z$ ,  $20 \times z$  o  $100 \times z$  (es decir, el límite inferior).  $n$  no puede ser mayor que  $100\,000 \times z$ ,  $10\,000 \times z$ ,  $2000 \times z$ ,  $1000 \times z$ ,  $500 \times z$  o  $100 \times z$  (es decir, el límite superior). Por consiguiente,  $n$  puede variar entre cualquier combinación de estos límites superior e inferior.  $n$  puede ser de entre  $100 \times z$  y  $1000 \times z$ , de entre  $5 \times z$  y  $15 \times z$ , de entre  $8 \times z$  y  $12 \times z$ , o de aproximadamente  $10 \times z$ . Por ejemplo, un equivalente del genoma humano haploide tiene aproximadamente 3 picogramos de ADN. Una muestra de aproximadamente 1 microgramo de ADN contiene aproximadamente 300 000 equivalentes del genoma humano haploide. El número  $n$  puede ser de entre 15 y 45, de entre 24 y 36, de entre 64 y 2500, de entre 625 y 31 000, o aproximadamente de entre 900 y 4000. Pueden obtenerse mejoras en la secuenciación siempre y cuando al menos algunos de los polinucleótidos duplicados o análogos porten identificadores únicos, es decir, porten etiquetas diferentes. No obstante, puede seleccionarse el número de etiquetas utilizadas, de tal manera que haya al menos un 95% de posibilidades de que todas las moléculas duplicadas que comiencen en una posición dada porten identificadores únicos. Por ejemplo, una muestra que comprende aproximadamente 10 000 equivalentes del genoma humano haploide de ADNcf puede etiquetarse con aproximadamente 36 identificadores únicos. Los identificadores únicos pueden comprender seis códigos de barras de ADN únicos. Unidos a ambos extremos de un polinucleótido, se producen 36 posibles identificadores únicos. Las muestras etiquetadas de este modo pueden ser aquellas con un rango de entre aproximadamente 10 ng y aproximadamente 100 ng, aproximadamente 1  $\mu$ g o aproximadamente 10  $\mu$ g de polinucleótidos fragmentados, por ejemplo ADN genómico, por ejemplo ADNcf.

**[0008]** Por consiguiente, usando los kits de la invención pueden obtenerse composiciones de polinucleótidos etiquetados. Los polinucleótidos pueden comprender ADN fragmentado, por ejemplo ADNcf. Un conjunto de polinucleótidos de la composición que se asignan a una posición de base asignable en un genoma pueden etiquetarse de forma no única, es decir, el número de identificadores diferentes puede ser de al menos 2 e inferior al número de polinucleótidos que se asignan a la posición de base asignable. Una composición de entre aproximadamente 10 ng y aproximadamente 10  $\mu$ g (por ejemplo, de aproximadamente 10 ng-1  $\mu$ g, de aproximadamente 10 ng-100 ng, de aproximadamente 100 ng-10  $\mu$ g, de aproximadamente 100 ng-1  $\mu$ g, o de aproximadamente 1  $\mu$ g-10  $\mu$ g) puede portar entre 2, 5, 10, 50 o 100 y 100, 1000, 10 000 o 100 000 identificadores diferentes. Por ejemplo, pueden usarse entre 5 y 100 o entre 100 y 4000 identificadores diferentes para etiquetar los polinucleótidos de dicha composición.

**[0009]** Los eventos en los que moléculas diferentes que se asignan a la misma coordenada (en este caso, que tienen las mismas posiciones de inicio/parada), y portan las mismas -y no diferentes- etiquetas, se denominan 'colisiones moleculares'. En algunos casos, el número real de colisiones moleculares puede ser mayor que el número de colisiones teóricas, calculado -por ejemplo- como se ha explicado anteriormente. Esto puede ser una función de la distribución desigual de moléculas a lo largo de las coordenadas, de las diferencias en la eficiencia de ligación entre códigos de barras y de otros factores. En este caso, pueden usarse métodos empíricos para determinar el número de códigos de barras que son necesarios para acercarse al número teórico de colisiones. Por ejemplo, un método para determinar el número de códigos de barras que es necesario para reducir las colisiones de códigos de barras para un determinado equivalente del genoma haploide puede estar basado en la distribución de longitud de las moléculas secuenciadas y la uniformidad de secuencias. El método incluye crear una pluralidad de conjuntos de moléculas de ácido nucleico; etiquetar cada conjunto con números de códigos de barras que aumentan progresivamente; y determinar el número óptimo de códigos de barras que reduce el número de colisiones de códigos de barras a un nivel teórico; por ejemplo, podría deberse a las diferencias en las concentraciones efectivas de los códigos de barras a causa de las diferencias en el agrupamiento y la eficiencia de ligación.

**[0010]** El número de identificadores que son necesarios para etiquetar prácticamente de forma única los polinucleótidos asignados a una región puede determinarse empíricamente. Por ejemplo, un número seleccionado de identificadores diferentes puede unirse a las moléculas de una muestra, y puede calcularse el número de identificadores diferentes para las moléculas asignadas a la región. Si se usa un número insuficiente de identificadores, algunos polinucleótidos asignados a la región portarán el mismo identificador. En ese caso, el número de identificadores contabilizado será inferior al número de moléculas originales de la muestra. El número de identificadores diferentes utilizados puede aumentarse repetidamente para un tipo de muestra hasta que no se detecte ningún identificador adicional que corresponda a nuevas moléculas originales. Por ejemplo, en una primera iteración pueden contabilizarse cinco identificadores diferentes que corresponden a al menos cinco moléculas originales diferentes. En una segunda

iteración, y usando más códigos de barras, se contabilizan siete identificadores diferentes que corresponden a al menos siete moléculas originales diferentes. En una tercera iteración, y usando más códigos de barras, se contabilizan diez identificadores diferentes que corresponden a al menos diez moléculas originales diferentes. En una cuarta iteración, y usando más códigos de barras, se contabilizan de nuevo 10 identificadores diferentes. En este punto, añadir más

## 6. Secuenciación

**[0060]** Los ácidos nucleicos de muestra flanqueados por adaptadores con o sin una amplificación previa pueden someterse a una secuenciación. Los métodos de secuenciación incluyen, por ejemplo, la secuenciación de Sanger, la secuenciación de alto rendimiento, la pirosecuenciación, la secuenciación mediante síntesis, la secuenciación de una sola molécula, la secuenciación de nanoporos, la secuenciación de semiconductores, la secuenciación por ligación, la secuenciación por hibridación, el método 'RNA-Seq' (Illumina), el método 'Digital Gene Expression' (Helicos), la secuenciación de próxima generación, el método 'Single Molecule Sequencing by Synthesis' (SMSS) (Helicos), la secuenciación masiva paralela, el método 'Clonal Single Molecule Array' (Solexa), la secuenciación de escopeta, la secuenciación por Ion Torrent, la secuenciación de nanoporos de Oxford Nanopore, la secuenciación de Roche Genia, la secuenciación Maxam-Gilbert, el paseo cromosómico y la secuenciación utilizando las plataformas de PacBio, SOLiD, Ion Torrent o Nanopore. Las reacciones de secuenciación pueden llevarse a cabo en una gran variedad de unidades de procesamiento de muestras, que pueden incluir múltiples carriles, múltiples canales, múltiples pocillos u otros medios para procesar múltiples conjuntos de muestras prácticamente de manera simultánea. Las unidades de procesamiento de muestras también pueden incluir múltiples cámaras de muestras que permiten procesar múltiples tandas simultáneamente.

**[0061]** Las reacciones de secuenciación pueden realizarse sobre uno o más tipos de fragmentos que se sabe que contienen marcadores de cáncer u otras enfermedades. Las reacciones de secuenciación también pueden realizarse sobre cualquier fragmento de ácido nucleico que esté presente en la muestra. Las reacciones de secuenciación pueden proporcionar una cobertura de la secuencia del genoma de al menos un 5%, un 10%, un 15%, un 20%, un 25%, un 30%, un 40%, un 50%, un 60%, un 70%, un 80%, un 90%, un 95%, un 99%, un 99,9% o un 100%. En otros casos, la cobertura de la secuencia del genoma puede ser inferior a un 5%, un 10%, un 15%, un 20%, un 25%, un 30%, un 40%, un 50%, un 60%, un 70%, un 80%, un 90%, un 95%, un 99%, un 99,9% o un 100%.

**[0062]** Pueden llevarse a cabo reacciones de secuenciación simultáneas usando la secuenciación multiplexada. En algunos casos, pueden secuenciarse ácidos nucleicos libres de células con al menos 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10 000, 50 000 o 100 000 reacciones de secuenciación. En otros casos, pueden secuenciarse polinucleótidos libres de células con menos de 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10 000, 50 000 o 100 000 reacciones de secuenciación. Las reacciones de secuenciación pueden realizarse secuencialmente o simultáneamente. Puede realizarse un análisis de datos posterior sobre todas las reacciones de secuenciación o sobre parte de las mismas. En algunos casos, el análisis de datos puede realizarse sobre al menos 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10 000, 50 000 o 100 000 reacciones de secuenciación. En otros casos, el análisis de datos puede realizarse sobre menos de 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10 000, 50 000 o 100 000 reacciones de secuenciación.

**[0063]** El método de secuenciación puede ser la secuenciación masiva paralela, es decir, la secuenciación simultánea (o en rápida sucesión) de al menos 100, 1000, 10 000, 100 000, 1 millón, 10 millones, 100 millones o 1000 millones de moléculas de ácido nucleico.

## 7. Análisis

**[0064]** Los kits de la invención pueden resultar útiles en métodos que se usan para diagnosticar la presencia de enfermedades -particularmente cáncer- en un sujeto, para caracterizar las enfermedades (por ejemplo, realizar la estadificación del cáncer o determinar la heterogeneidad del cáncer), para monitorizar la respuesta al tratamiento de una enfermedad y para realizar un pronóstico sobre el riesgo de desarrollar una enfermedad o sobre la evolución posterior de una enfermedad.

**[0065]** Usando estos métodos pueden detectarse diversos tipos de cáncer. Las células cancerosas, como la mayoría de las células, pueden caracterizarse por su tasa de renovación celular, por la que las células viejas mueren y son sustituidas por células nuevas. Generalmente, las células muertas, al entrar en contacto con el sistema vascular de un sujeto, pueden liberar ADN o fragmentos de ADN en el torrente sanguíneo. Y sucede lo mismo con las células cancerosas durante varias fases de la enfermedad. Dependiendo de la fase de la enfermedad, las células cancerosas también pueden caracterizarse por diversas aberraciones genéticas, como la variación en el número de copias, y por la aparición de raras mutaciones. Este fenómeno puede usarse para detectar la presencia o ausencia de cánceres individuales mediante estos métodos.

**[0066]** Los tipos y el número de cánceres que pueden detectarse incluyen los cánceres de la sangre, los cánceres cerebrales, los cánceres pulmonares, los cánceres de la piel, los cánceres nasales, los cánceres de garganta, los cánceres hepáticos, los cánceres óseos, los linfomas, los cánceres del páncreas, los cánceres de la piel, los cánceres

intestinales, los cánceres rectales, los cánceres de la tiroides, los cánceres de vejiga, los cánceres renales, los cánceres orales, los cánceres estomacales, los tumores sólidos, los tumores heterogéneos, los tumores homogéneos y similares.

**[0067]** Los cánceres pueden detectarse a partir de variaciones genéticas, lo que incluye las mutaciones, las mutaciones raras, los indeles, las variaciones en el número de copias, las transversiones, las translocaciones, las inversiones, las deleciones, la aneuploidia, la aneuploidia parcial, la poliploidia, la inestabilidad cromosómica, las alteraciones cromosómicas estructurales, las fusiones de genes, las fusiones cromosómicas, el truncamiento de genes, la amplificación génica, las duplicaciones de genes, las lesiones cromosómicas, las lesiones del ADN, los cambios anormales en las modificaciones químicas del ácido nucleico, los cambios anormales en los patrones epigenéticos, los cambios anormales en la metilación del ácido nucleico, las infecciones y el cáncer.

**[0068]** La información genética también puede usarse para caracterizar una forma específica de cáncer. Los cánceres son a menudo heterogéneos en cuanto a su composición y su estadificación. Los datos del perfil genético pueden permitir la caracterización de los subtipos de cáncer específicos, que puede ser importante para el diagnóstico o el tratamiento del subtipo específico en cuestión. Esta información también puede proporcionar pistas al sujeto o al médico en lo referente al pronóstico de un tipo de cáncer específico y puede permitir que el sujeto o el médico se adapten a las opciones de tratamiento según la evolución de la enfermedad. Algunos cánceres evolucionan y se vuelven más agresivos y genéticamente inestables. Otros cánceres pueden seguir siendo benignos o permanecer inactivos o latentes. Los kits de la invención pueden resultar útiles en métodos que pueden ser útiles para determinar la evolución de una enfermedad.

**[0069]** El presente análisis también es útil para determinar la eficacia de una opción de tratamiento particular. Las opciones de tratamiento eficaces pueden aumentar la cantidad de variaciones del número de copias o mutaciones raras que se detectan en la sangre del sujeto si el tratamiento tiene éxito, ya que más tipos de cáncer pueden desaparecer y liberar ADN. En otros ejemplos, esto puede no suceder. En otro ejemplo, quizás algunas opciones de tratamiento pueden correlacionarse con los perfiles genéticos de los cánceres con el paso del tiempo. Esta correlación puede ser útil para escoger una terapia. Asimismo, si se observa que un cáncer está en remisión tras el tratamiento, los kits de la invención pueden ser útiles en los métodos que se usan para monitorizar las enfermedades residuales o la reaparición de una enfermedad.

**[0070]** Los kits de la invención pueden ser útiles en los métodos que se usan para detectar variaciones genéticas en enfermedades diferentes al cáncer. Las células inmunitarias, como las células B, pueden experimentar una rápida expansión clonal ante la presencia de determinadas enfermedades. Las expansiones clonales pueden monitorizarse usando la detección de la variación del número de copias y también pueden monitorizarse determinados estados inmunitarios. En este ejemplo, a lo largo del tiempo puede realizarse un análisis de la variación del número de copias para obtener un perfil sobre cómo puede estar evolucionando una enfermedad particular. La detección de la variación del número de copias o incluso de mutaciones raras puede usarse para determinar cómo está cambiando una población de patógenos durante el transcurso de una infección. Esto puede ser particularmente importante durante las infecciones crónicas, como el SIDA/VIH o las infecciones por hepatitis, en las que los virus pueden cambiar los estados de su ciclo vital y/o mutar en formas más virulentas durante el curso de la infección. Los kits de la invención pueden ser útiles en los métodos que se usan para determinar o trazar un perfil de las actividades de rechazo del cuerpo huésped, a medida que las células inmunitarias tratan de destruir el tejido trasplantado, y para monitorizar el estado del tejido trasplantado y modificar el curso del tratamiento o la prevención del rechazo.

**[0071]** Asimismo, los kits de la invención pueden ser útiles en los métodos que se usan para caracterizar la heterogeneidad de una enfermedad anormal en un sujeto, de manera que los métodos incluyen generar un perfil genético de los polinucleótidos extracelulares del sujeto, y de manera que el perfil genético comprende múltiples datos derivados del análisis de la variación del número de copias y del análisis de las mutaciones raras. En algunos casos, lo que incluye -pero no se limita a- el cáncer, una enfermedad puede ser heterogénea. Las células patógenas pueden no ser idénticas. En el caso del cáncer, se sabe que algunos tumores comprenden diferentes tipos de células tumorales, de manera que algunas células están en diferentes estadios del cáncer. En otros ejemplos, la heterogeneidad puede comprender múltiples focos de enfermedad. De igual modo, en el caso del cáncer puede haber múltiples focos, de manera que uno o más focos tal vez sean el resultado de metástasis que se han propagado desde el sitio original.

**[0072]** Los kits de la invención pueden ser útiles en los métodos que se usan para generar o trazar un perfil de un conjunto de datos que es un compendio de la información genética obtenida a partir de diferentes células en una enfermedad heterogénea. Este conjunto de datos puede comprender el análisis de la variación del número de copias y el análisis de mutaciones raras, solos o de forma combinada.

**[0073]** Los kits de la invención pueden ser útiles en métodos que se usan para diagnosticar, pronosticar, monitorizar u observar cánceres u otras enfermedades de origen fetal. Es decir, estas metodologías pueden utilizarse en una mujer embarazada para diagnosticar, pronosticar, monitorizar u observar cánceres u otras enfermedades en un bebé nonato cuyo ADN y otros ácidos nucleicos pueden circular conjuntamente con las moléculas maternas.

**[0074]** 9. Kits

Los kits de la invención pueden usarse para poner en práctica cualquiera de los métodos anteriores. Los kits de la invención incluyen un par de adaptadores al menos parcialmente bicatenarios con colas 3' de nucleótidos simples T y C, respectivamente. Preferiblemente, los oligonucleótidos emparejados son idénticos salvo por las colas de T y C. Opcionalmente, el kit carece de adaptadores al menos parcialmente bicatenarios con colas 3' de nucleótidos simples A y G. Preferiblemente, los adaptadores tienen forma de Y, como los adaptadores que comprenden oligonucleótidos de las SEQ ID Nos. 1 y 2, y 3 y 2. Los kits también pueden incluir enzimas para poner en práctica los métodos, como una polimerasa T4 o un fragmento de Klenow grande, y/o una polimerasa Taq, y opcionalmente los cuatro tipos de nucleótidos estándar. Los kits también pueden incluir un embalaje, prospectos, CDs o similares que proporcionan instrucciones para poner en práctica los métodos reivindicados.

Ejemplos

**[0075]** El uso de adaptadores con colas de C y T contribuyó a aumentar la sensibilidad al capturar más moléculas en una muestra. Los adaptadores de C se evaluaron en proporciones de 0 a 1:2,75 (36%) con respecto a los adaptadores de T, tal y como se muestra en la Tabla 2.

Tabla 2

Muestra #	Contribución (ng)	Cola de T (40 $\mu$ M)	Cola de C (40 $\mu$ M)	% de LIG
1	20	3,25	0,5	80%
2	20	3,25	0,5	77%
3	20	3,25	1	79%
4	20	3,25	1	80%
5	20	2,75	0,5	79%
6	20	2,75	0,5	77%
7	20	2,75	1	80%
8	20	2,75	1	78%
9	20	3,25	-	75%
10	20	3,25	-	75%

**[0076]** Todas las muestras en las que estaban presentes los adaptadores con colas de C mostraron una mayor producción de ácidos nucleicos ligados a adaptadores (% de ligación) que las muestras en las que las colas de C estaban ausentes. La mejor producción fue para los cebadores de colas de C respecto a los de colas de T con un ratio de 1:3,25 (aproximadamente un 30%), pero se obtuvieron producciones mejoradas con ratios de 0,5:3,25 (aproximadamente un 15%) a 1:2,75 (un 36%).

**[0077]** Tras la secuenciación de ADN amplificado, se calculó la diversidad de cada preparación. La diversidad es el número de moléculas secuenciadas, calculado del siguiente modo: (tamaño promedio de la molécula de ADN, en bp) \* (# de moléculas únicas secuenciadas) / (tamaño de la región objetivo, en bp). De manera general, la diversidad fue mayor en las muestras en las que estaba presente el adaptador con colas de C. La secuenciación también indicó que la proporción de los adaptadores con colas de T -respecto a los adaptadores con colas de C- incorporados fue de aproximadamente un 10%.

# REIVINDICACIONES

1. Un kit, que comprende:  
5 adaptadores al menos parcialmente bicatenarios con un saliente de un solo nucleótido T en el extremo 3' (adaptadores con cola de T); y  
adaptadores al menos parcialmente bicatenarios con un saliente de un solo nucleótido C en el extremo 3' (adaptadores con cola de C).
- 10 2. El kit de la reivindicación 1, de manera que la proporción molar de los adaptadores al menos parcialmente bicatenarios con un saliente de un solo nucleótido T respecto a los adaptadores con un saliente de un solo nucleótido C es de 4:1 a 2:1, por ejemplo de 3:1.
- 15 3. El kit de la reivindicación 1, de manera que el porcentaje de adaptadores con cola de C respecto a los adaptadores con cola de T es de aproximadamente un 5-40% en moles, por ejemplo un 10-35%, un 15-25%, un 20-35%, un 25-35% o aproximadamente un 30%.
- 20 4. El kit de cualquiera de las reivindicaciones anteriores, que además comprende:  
adaptadores al menos parcialmente bicatenarios con extremos romos.
- 25 5. El kit de la reivindicación 4, de manera que la proporción molar de los adaptadores con extremos romos respecto a los adaptadores con colas es de 1:5 a 1:500, por ejemplo de 1:10 a 1:100.
6. El kit de la reivindicación 4, de manera que los adaptadores con extremos romos están en una proporción molar de un 0,2-20%, o un 0,5-15%, o un 1-10% de adaptadores con los adaptadores con cola de T y cola de C.
- 30 7. El kit de cualquiera de las reivindicaciones anteriores, que además comprende:  
una polimerasa T4 o un fragmento de Klenow grande; y/o  
una polimerasa Taq.
- 35 8. El kit de la reivindicación 7, que además comprende cuatro tipos de nucleótidos estándar.
9. El kit de cualquiera de las reivindicaciones anteriores, de manera que el kit carece de adaptadores al menos parcialmente bicatenarios con colas 3' de un solo nucleótido A, y de adaptadores al menos parcialmente bicatenarios con colas 3' de un solo nucleótido G.
- 40 10. El kit de cualquiera de las reivindicaciones anteriores, de manera que los adaptadores son adaptadores con forma de Y.
11. El kit de la reivindicación 10, de manera que los adaptadores comprenden oligonucleótidos de las SEQ ID Nos 1 y 2, y 3 y 2.
12. El kit de cualquiera de las reivindicaciones anteriores, de manera que los adaptadores comprenden un sitio de unión de cebadores y/o un código de barras.
- 45 13. El kit de la reivindicación 12, de manera que los códigos de barras proporcionan entre 64 y 90 000, entre 400 y 22500, entre 400 y 14 400, o entre 900 y 14 400 combinaciones de códigos de barras diferentes.

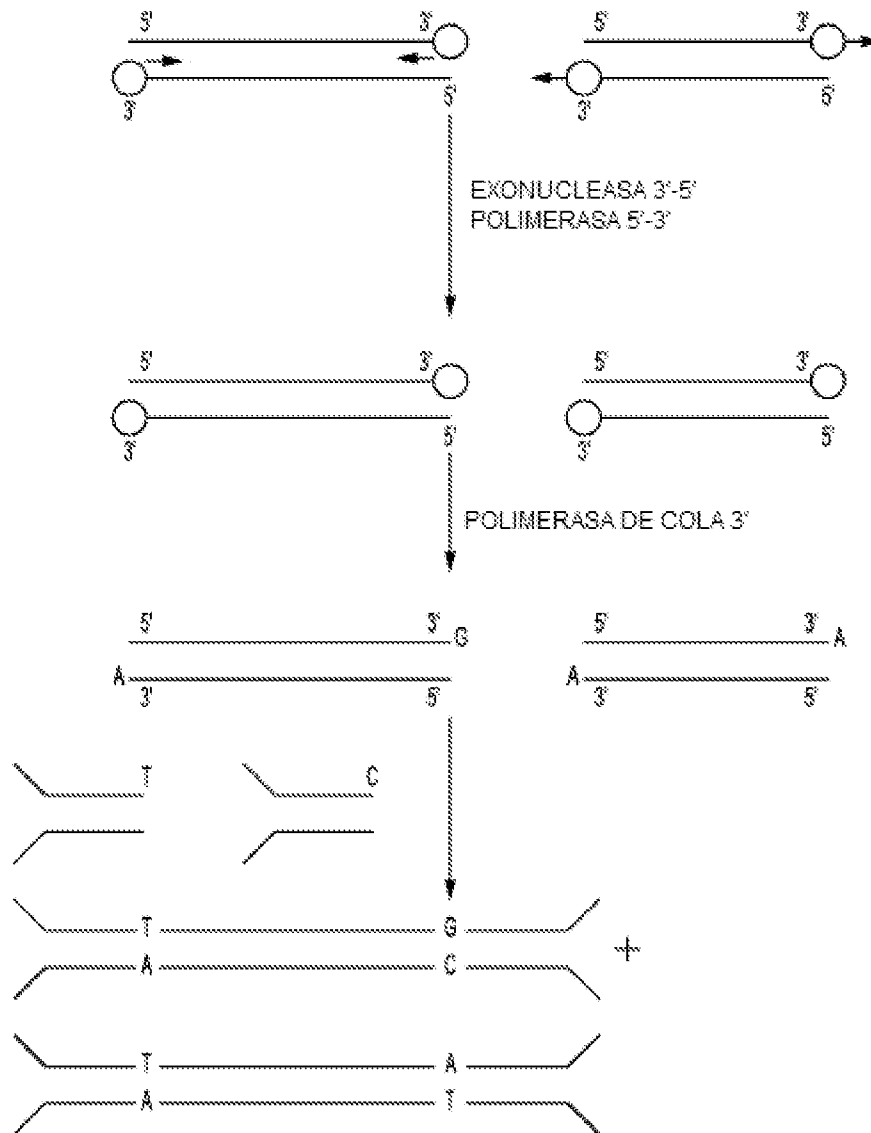


FIG. 1