

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 707 744**

51 Int. Cl.:

C12Q 1/6827 (2008.01)

C12Q 1/6874 (2008.01)

C12Q 1/6869 (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **11.12.2014 PCT/US2014/069848**

87 Fecha y número de publicación internacional: **18.06.2015 WO15089333**

96 Fecha de presentación y número de la solicitud europea: **11.12.2014 E 14868836 (9)**

97 Fecha y número de publicación de la concesión europea: **31.10.2018 EP 3080298**

54 Título: **Métodos para detectar variantes de secuencia raras**

30 Prioridad:

11.12.2013 US 201361914907 P
01.05.2014 US 201461987414 P
11.06.2014 US 201462010975 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
04.04.2019

73 Titular/es:

ACCURAGEN HOLDINGS LIMITED (100.0%)
P.O. Box 472, Harbour Place, 2nd Floor, 103
South Church Street, George Town
KY1-1106 Grand Cayman, KY

72 Inventor/es:

LIN, SHENGRONG;
SUN, ZHAOHUI;
ZHAO, GRACE QIZHI y
TANG, PAUL LING-FUNG

74 Agente/Representante:

VALLEJO LÓPEZ, Juan Pedro

ES 2 707 744 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 detectar variantes de secuencia raras

5 REFERENCIAS CRUZADAS

Esta solicitud reivindica el beneficio de la solicitud provisional de Estados Unidos N.º 61/914.907, presentada el 11 de diciembre de 2013; la solicitud provisional de Estados Unidos n.º 61/987.414, presentada el 1 de mayo de 2014; y la solicitud provisional de Estados Unidos n.º 62/010.975, presentada el 11 de junio de 2014.

10

Antecedentes de la invención

La identificación de la variación de secuencia dentro de poblaciones complejas es un campo en crecimiento activo, particularmente con la llegada de la secuenciación de ácidos nucleicos en paralelo a gran escala. Sin embargo, la secuenciación en paralelo a gran escala tiene limitaciones significativas en cuanto a que la frecuencia de error inherente en las técnicas de uso común es mayor que la frecuencia de muchas de las variaciones de secuencia reales en la población. Por ejemplo, se han indicado tasas de error de 0,1 a 1% en la secuenciación convencional de alto rendimiento. La detección de variantes de secuencia raras tiene altas tasas de falsos positivos cuando la frecuencia de las variantes es baja, como en o por debajo de la tasa de error.

20

Hay muchas razones para detectar variantes de secuencia raras. Por ejemplo, la detección de secuencias características raras se puede utilizar para identificar y distinguir la presencia de un contaminante dañino para el medio ambiente, tales como los taxones bacterianos. Una forma común de caracterizar los taxones bacterianos es identificar diferencias en una secuencia altamente conservada, tales como las secuencias de ARNr. Sin embargo, los enfoques típicos basados en la secuenciación para esto se enfrentan a desafíos relacionados con la gran cantidad de genomas diferentes en una muestra dada y el grado de homología entre los miembros, lo que presenta un problema complejo para procedimientos ya laboriosos. Los procedimientos mejorados tendrían el potencial de potenciar la detección de contaminación en varios entornos. Por ejemplo, los cuartos limpios utilizados para ensamblar componentes de satélites y otras naves espaciales se pueden estudiar con los sistemas y métodos actuales para comprender qué comunidades microbianas están presentes y desarrollar mejores técnicas de descontaminación y limpieza para evitar la introducción de microbios terrestres en otros planetas o muestras de los mismos o para desarrollar metodologías para distinguir los datos generados por los supuestos microorganismos extraterrestres de los generados por los microorganismos contaminantes terrestres. Las aplicaciones de monitorización de alimentos incluyen las pruebas periódicas de las líneas de producción en las plantas de procesamiento de alimentos, el estudio de mataderos, la inspección de las cocinas y las áreas de almacenamiento de alimentos de restaurantes, hospitales, escuelas, centros penitenciarios y otras instituciones para detectar patógenos transmitidos por los alimentos. Las reservas de agua y las plantas de procesamiento también se pueden monitorizar de manera similar.

40

La detección de variantes raras también puede ser importante para la detección temprana de mutaciones patológicas. Por ejemplo, la detección de mutaciones puntuales asociadas al cáncer en muestras clínicas puede mejorar la identificación de la enfermedad residual mínima durante la quimioterapia y detectar la aparición de células tumorales en pacientes con recidiva. La detección de mutaciones puntuales raras también es importante para la evaluación de la exposición a mutágenos ambientales, para monitorizar la reparación endógena del ADN y para estudiar la acumulación de mutaciones somáticas en personas de edad avanzada. Adicionalmente, los métodos más sensibles para detectar variantes raras pueden potenciar el diagnóstico prenatal, permitiendo la caracterización de las células fetales presentes en la sangre materna. Lou et al desvelan errores de secuenciación de ADN de alto rendimiento que se reducen mediante el uso de la secuenciación circular (PNAS; 110, 49, 15 de noviembre de 2013). El documento US2011/319229A1 describe métodos y composiciones para el análisis genético de regiones de interés a partir de muestras de polinucleótidos. El documento US2013/0217023A1 describe sistemas y métodos para generar productos compactos amplificados por clonación. El documento WO2007/140417A2 describe métodos y composiciones para la extracción y amplificación de ácidos nucleicos obtenidos de una muestra hospedadora.

50

Sumario de la invención

55

En vista de lo anterior, existe la necesidad de métodos mejorados para detectar variantes de secuencia raras. Las composiciones y métodos de la presente divulgación abordan esta necesidad y proporcionan también ventajas adicionales. En particular, los diversos aspectos de la divulgación proporcionan una detección altamente sensible de variantes de secuencia de ácido nucleico raras o de baja frecuencia (a veces denominadas mutaciones). Esto incluye la identificación y elucidación de variaciones de ácidos nucleicos de baja frecuencia (incluidas las sustituciones, inserciones y deleciones) en muestras que pueden contener cantidades bajas de secuencias variantes en un fondo de secuencias normales, así como la identificación de variaciones de baja frecuencia en un fondo de errores de secuenciación.

60

La invención proporciona un método para identificar una variante de secuencia en una muestra de ácido nucleico que comprende una pluralidad de polinucleótidos libres de células, en donde cada polinucleótido libre de células de la pluralidad tiene un extremo 5' y un extremo 3' y el método comprende: (a) circularizar polinucleótidos libres de células

65

individuales de dicha pluralidad para formar una pluralidad de polinucleótidos circulares, cada uno de los cuales tiene una unión entre el extremo 5' y el extremo 3'; (b) amplificar los polinucleótidos circulares de (a); (c) secuenciar los polinucleótidos amplificados para producir una pluralidad de lecturas de secuenciación; (d) identificar diferencias de secuencia entre lecturas de secuenciación y una secuencia de referencia; y (e) identificar una diferencia de secuencia como la variante de la secuencia solamente cuando la diferencia de secuencia se produce en al menos dos polinucleótidos circulares que tienen uniones diferentes. Varias realizaciones de la invención se exponen en las reivindicaciones adjuntas. En el presente documento también se desvela un método que comprende identificar diferencias de secuencia entre lecturas de secuenciación y una secuencia de referencia e identificar una diferencia de secuencia que se produce en al menos dos polinucleótidos circulares que tienen diferentes uniones como la variante de secuencia, en donde: (a) las lecturas de secuenciación corresponden a productos de amplificación de al menos los dos polinucleótidos circulares; y (b) cada uno de al menos los dos polinucleótidos circulares comprende una unión diferente formada al unir un extremo 5' y un extremo 3' de los polinucleótidos respectivos.

La pluralidad de polinucleótidos puede ser de cadena sencilla o de cadena doble. En algunas realizaciones, los polinucleótidos son de cadena sencilla. En algunas realizaciones, la circularización se efectúa sometiendo la pluralidad de polinucleótidos a una reacción de unión. En algunas realizaciones, un polinucleótido circular individual tiene una unión que es única entre los polinucleótidos circularizados. En algunas realizaciones, la variante de secuencia es un polimorfismo de un solo nucleótido (SNP). En algunas realizaciones, la secuencia de referencia es una secuencia consenso formada al alinear las lecturas de secuencia entre sí. En algunas realizaciones, la secuencia de referencia es una secuencia de referencia conocida, tal como un genoma de referencia o una parte del mismo. En algunas realizaciones, la circularización comprende la etapa de unir un polinucleótido adaptador al extremo 5', al extremo 3', o tanto al extremo 5' como al extremo 3' de un polinucleótido en la pluralidad de polinucleótidos. En algunas realizaciones, la amplificación se efectúa utilizando una polimerasa que tiene actividad de desplazamiento de cadena, tal como en la amplificación de círculo rodante (RCA). En algunas realizaciones, la amplificación comprende someter los polinucleótidos circulares a una mezcla de reacción de amplificación que comprende cebadores aleatorios. En algunas realizaciones, la amplificación comprende someter los polinucleótidos circulares a una mezcla de reacción de amplificación que comprende uno o más cebadores, cada uno de los cuales hibrida específicamente con una secuencia diana diferente mediante complementariedad de secuencia. En algunas realizaciones, un contaminante microbiano se identifica basándose en la etapa de identificación.

Los polinucleótidos amplificados pueden someterse a secuenciación con o sin enriquecimiento, tal como el enriquecimiento de uno o más polinucleótidos diana entre los polinucleótidos amplificados realizando una etapa de enriquecimiento antes de la secuenciación. En algunas realizaciones, la etapa de enriquecimiento comprende hibridar polinucleótidos amplificados a una pluralidad de sondas unidas a un sustrato. En algunas realizaciones, la etapa de enriquecimiento comprende amplificar una secuencia diana que comprende la secuencia A y la secuencia B orientada en una dirección de 5' a 3' en una mezcla de reacción de amplificación que comprende: (a) los polinucleótidos amplificados; (b) un primer cebador que comprende la secuencia A', en donde el primer cebador hibrida específicamente con la secuencia A de la secuencia diana mediante complementariedad de secuencia entre la secuencia A y la secuencia A'; (c) un segundo cebador que comprende la secuencia B, en donde el segundo cebador hibrida específicamente con la secuencia B presente en un polinucleótido complementario que comprende un complemento de la secuencia diana mediante complementariedad de secuencia entre B y B'; y (d) una polimerasa que extiende el primer cebador y el segundo cebador para producir polinucleótidos amplificados; en donde la distancia entre el extremo 5' de la secuencia A y el extremo 3' de la secuencia B de la secuencia diana es de 75 nt o menos.

En algunas realizaciones, la divulgación proporciona un método para amplificar en una mezcla de reacción una pluralidad de diferentes concatémicos que comprenden dos o más copias de una secuencia diana, en donde la secuencia diana comprende la secuencia A y la secuencia B orientadas en una dirección de 5' a 3'. En algunas realizaciones, el método comprende someter la mezcla de reacción a una reacción de amplificación de ácido nucleico, en donde la mezcla de reacción comprende: (a) la pluralidad de concatémicos, en donde los concatémicos individuales en la pluralidad comprenden diferentes uniones formadas mediante la circularización de polinucleótidos individuales que tienen un extremo 5' y un extremo 3'; (b) un primer cebador que comprende la secuencia A', en donde el primer cebador hibrida específicamente con la secuencia A de la secuencia diana mediante complementariedad de secuencia entre la secuencia A y la secuencia A'; (c) un segundo cebador que comprende la secuencia B, en donde el segundo cebador hibrida específicamente con la secuencia B presente en un polinucleótido complementario que comprende un complemento de la secuencia diana mediante complementariedad de secuencia entre la secuencia B y B'; y (d) una polimerasa que extiende el primer cebador y el segundo cebador para producir polinucleótidos amplificados; en donde la distancia entre el extremo 5' de la secuencia A y el extremo 3' de la secuencia B de la secuencia diana es de 75 nt o menos. En algunas realizaciones, el primer cebador comprende la secuencia C 5' con respecto a la secuencia A', el segundo cebador comprende la secuencia D 5' con respecto a la secuencia B, y ni la secuencia C ni la secuencia D hibridan con la pluralidad de concatémicos durante una primera fase de amplificación a una primera temperatura de hibridación. En algunas realizaciones, la amplificación comprende una primera fase y una segunda fase; la primera fase comprende una etapa de hibridación a una primera temperatura, durante la cual los cebadores primero y segundo hibridan con los concatémicos antes de la extensión de los cebadores; y la segunda fase comprende una etapa de hibridación a una segunda temperatura que es más alta que la primera temperatura, durante la cual los cebadores primero y segundo hibridan con productos de amplificación que comprenden el primer o segundo cebadores extendidos o complementos de los mismos. En algunas realizaciones, después de 5 ciclos de

hibridación a la segunda temperatura y la extensión de los cebadores, al menos el 5% de los polinucleótidos amplificados en la mezcla de reacción comprenden dos o más copias de la secuencia diana.

También se describe un método para amplificar en una mezcla de reacción una pluralidad de polinucleótidos circulares diferentes que comprenden una secuencia diana, en donde la secuencia diana comprende la secuencia A y la secuencia B orientadas en una dirección de 5' a 3'. Este método puede comprender someter la mezcla de reacción a una reacción de amplificación de ácido nucleico, en donde la mezcla de reacción comprende: (a) la pluralidad de polinucleótidos circulares, en donde los polinucleótidos circulares individuales en la pluralidad comprenden diferentes uniones formadas mediante la circularización de polinucleótidos individuales que tienen un extremo 5' y un extremo 3'; (b) un primer cebador que comprende la secuencia A', en donde el primer cebador hibrida específicamente con la secuencia A de la secuencia diana mediante complementariedad de secuencia entre la secuencia A y la secuencia A'; (c) un segundo cebador que comprende la secuencia B, en donde el segundo cebador hibrida específicamente con la secuencia B' presente en un polinucleótido complementario que comprende un complemento de la secuencia diana mediante complementariedad de secuencia entre la secuencia B y B'; y (d) una polimerasa que extiende el primer cebador y el segundo cebador para producir polinucleótidos amplificados; en donde la secuencia A y la secuencia B son secuencias endógenas y la distancia entre el extremo 5' de la secuencia A y el extremo 3' de la secuencia B de la secuencia diana es de 75 nt o menos. En algunas realizaciones, el primer cebador comprende la secuencia C 5' con respecto a la secuencia A', el segundo cebador comprende la secuencia D 5' con respecto a la secuencia B, y ni la secuencia C ni la secuencia D hibridan con la pluralidad de polinucleótidos circulares durante una primera fase de amplificación a una primera temperatura de hibridación. En algunas realizaciones, la amplificación comprende una primera fase y una segunda fase; la primera fase comprende una etapa de hibridación a una primera temperatura, durante la cual los cebadores primero y segundo hibridan con los polinucleótidos circulares o con los productos de la amplificación de los mismos antes de la extensión de los cebadores; y la segunda fase comprende una etapa de hibridación a una segunda temperatura que es más alta que la primera temperatura, durante la cual los cebadores primero y segundo hibridan con productos de amplificación que comprenden el primer o segundo cebadores extendidos o complementos de los mismos.

Esta divulgación proporciona una mezcla de reacción para realizar un método de acuerdo con los métodos de la divulgación. La mezcla de reacción puede comprender uno o más de los diversos componentes como se describe en el presente documento con respecto a cualquiera de los diversos métodos. En algunas realizaciones, la mezcla de reacción es una mezcla para amplificar una pluralidad de diferentes concatémoros que comprenden dos o más copias de una secuencia diana, en donde la secuencia diana comprende la secuencia A y la secuencia B orientadas en una dirección de 5' a 3', la mezcla de reacción: (a) la pluralidad de concatémoros, en donde los concatémoros individuales en la pluralidad comprenden diferentes uniones formadas mediante la circularización de polinucleótidos individuales que tienen un extremo 5' y un extremo 3'; (b) un primer cebador que comprende la secuencia A', en donde el primer cebador hibrida específicamente con la secuencia A de la secuencia diana mediante complementariedad de secuencia entre la secuencia A y la secuencia A'; (c) un segundo cebador que comprende la secuencia B, en donde el segundo cebador hibrida específicamente con la secuencia B presente en un polinucleótido complementario que comprende un complemento de la secuencia diana mediante complementariedad de secuencia entre B y B'; y; (d) una polimerasa que extiende el primer cebador y el segundo cebador para producir polinucleótidos amplificados; en donde la distancia entre el extremo 5' de la secuencia A y el extremo 3' de la secuencia B de la secuencia diana es de 75 nt o menos. En algunas realizaciones, el primer cebador comprende la secuencia C 5' con respecto a la secuencia A', el segundo cebador comprende la secuencia D 5' con respecto a la secuencia B, y ni la secuencia C ni la secuencia D hibridan con los dos o más concatémoros durante una primera etapa de amplificación en una reacción de amplificación.

Esta divulgación proporciona composiciones útiles en o producidas por métodos descritos en el presente documento, tales como en cualquiera de los otros diversos aspectos de la divulgación. En algunas realizaciones, la composición comprende una pluralidad de polinucleótidos circularizados que son de cadena sencilla y sustancialmente libres de ligasa. En algunas realizaciones, la composición comprende una pluralidad de concatémoros, en donde la pluralidad de concatémoros corresponde a un grupo de 10000 o menos polinucleótidos diana, y, en donde, adicionalmente, los concatémoros individuales en la pluralidad se caracterizan porque: (a) comprenden dos o más copias de una repetición de secuencia, en la que la totalidad de dichas copias corresponden al mismo polinucleótido diana; y (b) la unión entre las dos o más copias de la repetición de secuencia de un concatémoro individual es diferente de la de otro concatémoro individual en dicha composición.

Esta divulgación proporciona un sistema para detectar una variante de secuencia. En algunas realizaciones, el sistema comprende (a) un ordenador configurado para recibir una solicitud del usuario para realizar una reacción de detección en una muestra; (b) un sistema de amplificación que realiza una reacción de amplificación de ácido nucleico en la muestra o en una parte de la misma en respuesta a la solicitud del usuario, en donde la reacción de amplificación comprende las etapas de (i) circularizar los polinucleótidos individuales para formar una pluralidad de polinucleótidos circulares, cada uno de los cuales tiene una unión entre el extremo 5' y el extremo 3'; y (ii) amplificar los polinucleótidos circulares; (c) un sistema de secuenciación que genera lecturas de secuenciación para polinucleótidos amplificados por el sistema de amplificación, identifica las diferencias de secuencia entre las lecturas de secuenciación y una secuencia de referencia e identifica una diferencia de secuencia que se produce en al menos dos polinucleótidos circulares que tienen diferentes uniones como la variante de secuencia; y (d) un generador de informes que envía un informe a un destinatario, en el que el informe contiene resultados para la detección de la variante de secuencia. En

algunas realizaciones, el destinatario es el usuario.

La divulgación proporciona un medio legible por ordenador que comprende códigos que, al ser ejecutados por uno o más procesadores, implementan un método para detectar una variante de secuencia. En algunas realizaciones, el método implementado comprende: (a) recibir una solicitud del cliente para realizar una reacción de detección en una muestra; (b) realizar una reacción de amplificación de ácido nucleico en la muestra o en una parte de la misma en respuesta a la solicitud del cliente, en donde la reacción de amplificación comprende las etapas de (i) circularizar los polinucleótidos individuales para formar una pluralidad de polinucleótidos circulares, cada uno de los cuales tiene una unión entre el extremo 5' y el extremo 3'; y (ii) amplificar los polinucleótidos circulares; (c) realizar un análisis de secuenciación que comprende las etapas de (i) generar lecturas de secuenciación para polinucleótidos amplificados en la reacción de amplificación; (ii) identificar diferencias de secuencia entre lecturas de secuenciación y una secuencia de referencia; y (iii) identificar una diferencia de secuencia que se produce en al menos dos polinucleótidos circulares que tienen diferentes uniones como la variante de secuencia; y (d) generar un informe que contenga resultados para la detección de la variante de secuencia.

En algunas realizaciones de cualquiera de los diversos aspectos descritos en el presente documento, los métodos, composiciones y sistemas tienen aplicaciones terapéuticas, tales como la caracterización de una muestra de un paciente y, opcionalmente, el diagnóstico de una afección de un sujeto. En algunas realizaciones, una muestra es una muestra de un sujeto, tal como la orina, deposición, sangre, saliva, tejido, o fluido corporal. En algunas realizaciones, la muestra comprende células tumorales, tales como en una muestra de tejido tumoral de un sujeto. En algunas realizaciones, la muestra es una muestra fijada en formalina, embebida en parafina (FFGPE). En algunas realizaciones, un método comprende además la etapa de diagnosticar un sujeto basándose en una etapa de identificación. En algunas realizaciones, la variante de secuencia es una variante genética causal. En algunas realizaciones, la variante de secuencia está asociada con un tipo o etapa de cáncer. La pluralidad de polinucleótidos comprende polinucleótidos libres de células, tales como ADN libre de células o ADN tumoral circulante.

Breve descripción de los dibujos

Las nuevas características de la invención se exponen con particularidad en las reivindicaciones adjuntas. Se obtendrá una mejor comprensión de las características y ventajas de la presente invención haciendo referencia a la siguiente descripción detallada que expone realizaciones ilustrativas, en la que se utilizan los principios de la invención y los dibujos adjuntos de los cuales:

La FIG. 1 representa el esquema de una realización de métodos de acuerdo con la presente divulgación. Las cadenas de ADN se circularizan y se agregan los cebadores específicos correspondientes a los genes a investigar, junto con polimerasa, dNTP, tampones, etc., de manera que se produce la amplificación de círculo rodante (RCA) para formar concatémoros (por ejemplo, "multímeros") de la plantilla de ADN (por ejemplo, un "monómero"). Los concatémoros se tratan para sintetizar la cadena complementaria correspondiente y, a continuación, se agregan adaptadores para crear bibliotecas de secuenciación. Esta biblioteca resultante, que se secuenciará, a continuación, con tecnologías convencionales, generalmente contendrá tres especies: ADNn (ADN "normal") que no contiene una variante de secuencia rara (por ejemplo, una mutación); ADNn que contiene errores de secuenciación enzimática y ADN que contiene multímeros de variantes de secuencia "reales" o verdaderas que eran preexistentes en los polinucleótidos de muestra antes de la amplificación. La presencia de múltiples copias de la mutación eficazmente rara permite la detección e identificación de la variante de secuencia.

La FIG. 2 representa una estrategia similar a la de la Figura 1, pero con la adición de adaptadores para facilitar la circularización de polinucleótidos. La FIG. 2 también muestra el uso de cebadores específicos de diana.

La FIG. 3 es similar a la fig. 2, excepto en que se utilizan cebadores de adaptadores en la amplificación.

La FIG. 4 representa tres realizaciones asociadas con la formación de ADNcc circularizado. En la parte superior, el ADN de cadena sencilla (ADNcs) se circulariza en ausencia de adaptadores, mientras que el esquema del medio describe el uso de adaptadores, y el esquema inferior utiliza dos oligos de adaptadores (que producen secuencias diferentes en cada extremo) y puede incluir además un oligo splint (férula) que se hibrida a ambos adaptadores para acercar los dos extremos.

La FIG. 5 representa una realización para circularizar dianas específicas mediante el uso de una "abrazadera molecular" para llevar los dos extremos del ADN de cadena sencilla a una proximidad espacial para la unión.

Las FIG. 6A y 6B representan dos esquemas para la adición de adaptadores que usan extremos bloqueados de los ácidos nucleicos.

Las FIG. 7A, 7B y 7C representan tres formas diferentes de cebar una reacción de amplificación de círculo rodante (RCA). La FIG. 7A muestra el uso de cebadores específicos de diana, p.ej. los genes diana particulares o secuencias diana de interés. Esto generalmente da como resultado que solamente se amplifiquen las secuencias diana. La FIG. 7B representa el uso de cebadores aleatorios para realizar la amplificación del genoma completo (WGA), que generalmente amplificará todas las secuencias de muestra, que a continuación, se clasifican de manera bioinformática durante el procesamiento. La FIG. 7C representa el uso de cebadores de adaptadores cuando se usan adaptadores, lo que también da como resultado una amplificación general no específica de diana. La FIG. 8 representa un ejemplo de circularización y amplificación de ADN de doble cadena, de manera que ambas cadenas se amplifican, de acuerdo con una realización.

Las FIG. 9A, 9B, 9C y 9D representan varios esquemas para lograr la síntesis de cadenas complementarias para

la secuenciación posterior. La FIG. 9A representa el uso de cebado aleatorio de la cadena diana, seguido de unión. La FIG. 9B representa el uso de cebado con adaptadores de la cadena diana, seguido, de manera similar, de unión. La FIG. 9C representa el uso de un adaptador en "bucle", en el que el adaptador tiene dos secciones de secuencias que son complementarias, de modo que hibridan entre sí para crear un bucle (por ejemplo, estructuras de tallo-lazo). Tras la unión al final del concatémoro, el extremo libre del bucle sirve como cebador para la cadena complementaria. La FIG. 9D muestra el uso de cebadores aleatorios hiper-ramificados para lograr la síntesis de la segunda cadena.

La FIG. 10 muestra un método de PCR de acuerdo con una realización que promueve la secuenciación de polinucleótidos circulares o cadenas que contienen al menos dos copias de una secuencia de ácido nucleico diana, utilizando un par de cebadores que están orientados entre sí cuando están alineados dentro de un monómero de la secuencia diana (también conocida como "back to back", por ejemplo, orientados en dos direcciones pero no en los extremos del dominio a amplificar). En algunas realizaciones, estos conjuntos de cebadores se utilizan después de formarse los concatémeros para promover que los amplicones sean multímeros superiores, p. ej. dímeros, trímeros, etc., de la secuencia diana. Opcionalmente, el método puede incluir además una selección por tamaño para eliminar los amplicones que son más pequeños que los dímeros.

La FIG. 11 representa una realización en la que se utilizan cebadores back to back (B2B) con un paso de PCR de "retoque", de manera que se favorece menos la amplificación de productos cortos (como los monómeros). En este caso, los cebadores tienen dos dominios; un primer dominio que hibrida con la secuencia diana (flecha gris o negra) y un segundo dominio que es un dominio de enlace de "cebador universal" (rectángulos doblados; a veces también denominado adaptador) que no se hibrida a la secuencia diana original. En algunas realizaciones, las primeras rondas de PCR se realizan con una etapa de emparejamiento a baja temperatura, de modo que se unen las secuencias específicas de genes. El procesamiento a baja temperatura da como resultado productos de PCR de diversas longitudes, incluyendo productos cortos. Después de un número bajo de rondas, se aumenta la temperatura de emparejamiento, de modo que se favorece la hibridación del cebador completo, ambos dominios; como se muestra, estos se encuentran en los extremos de las plantillas, mientras que la unión interna es menos estable. Los productos más cortos, por lo tanto, se favorecen menos a la temperatura más alta con ambos dominios que a la temperatura más baja o con solo un dominio único.

Las FIG. 12A y 12B representan dos métodos diferentes de construcción de bibliotecas de secuenciación. La FIG. 12A ilustra un ejemplo del sistema de preparación de muestras Illumina® Nextera, mediante el cual el ADN se puede fragmentar y etiquetar simultáneamente con adaptadores de secuenciación en una sola etapa. En la Fig. 12B, los concatémeros se fragmentan por sonicación, seguida de la adición de adaptadores en ambos extremos (por ejemplo, mediante el uso de kits de KAPA Biosystems) y amplificación por PCR. Otros métodos están disponibles.

Las FIG. 13A-C proporcionan una ilustración de las ventajas de ejemplo del diseño de cebadores back-to-back (B2B) en comparación con el diseño de cebadores de PCR tradicional. El diseño tradicional de cebadores de PCR (izquierda) coloca los cebadores (flechas, A y B) en la región que flanquea una secuencia diana, que puede ser un punto caliente para las mutaciones (estrellas negras), y generalmente están separadas por al menos 60 pares de bases (pb), que da como resultado una huella típica de alrededor de 100 pb. En esta ilustración, el diseño de los cebadores B2B (derecha) coloca los cebadores en un lado de la secuencia diana. Los dos cebadores B2B están orientados en dirección opuesta, cualquiera puede superponerse (por ejemplo, aproximadamente o aproximadamente menos de 12 pb, 10 pb, 5 pb o menos). Dependiendo de la longitud de los cebadores B2B, la huella total en esta ilustración puede estar entre 28-50 pb. Debido a la mayor huella, es más probable que los eventos de fragmentación interrumpen la unión del cebador en el diseño tradicional, lo que lleva a la pérdida de información de secuencia, ya sea para fragmentos lineales (13A), ADN circularizado (13B) o productos de amplificación (13C). Por otra parte, tal como se ilustra en la figura 13C, el diseño de los cebadores B2B captura secuencias de unión (también conocidas como un "código de barras natural") que se pueden usar para distinguir diferentes polinucleótidos.

La FIG. 14 ilustra un método para generar plantillas para detectar variantes de secuencia, de acuerdo con una realización (por ejemplo, una implementación de ejemplo de un proceso que usa polinucleótidos circularizados, también denominado, en el presente documento, "Nebula"). El ADN de entrada se desnaturaliza en ADNcs, se circulariza por unión y el ADN no circularizado se degrada por digestión con exonucleasa. La eficacia de la unión se cuantifica mediante PCR cuantitativa (qPCR), que compara las cantidades de ADN de entrada y de ADN circularizado, lo que generalmente produce una eficacia de unión de al menos aproximadamente el 80%. El ADN circularizado se purifica para intercambiar el tampón, seguido de la amplificación del genoma completo (WGA) con cebadores aleatorios y polimerasa Phi29. Los productos de la WGA se purifican y los productos se fragmentan (por ejemplo, por sonicación) en fragmentos cortos de aproximadamente o aproximadamente menos de 400 pb. La tasa específica de ADN amplificado se cuantifica mediante qPCR comparando la misma cantidad del ADN del genoma de referencia con el ADN amplificado, que muestra normalmente una tasa específica promedio de aproximadamente o aproximadamente de más del 95%.

La FIG. 15 ilustra una implementación adicional de la amplificación con cebadores B2B de cola y la implementación de una segunda fase de "retoque" de la PCR a una temperatura más alta. Los cebadores B2B contienen una región específica de secuencia (línea negra gruesa) y una secuencia de adaptador (caja abierta). Con una temperatura de emparejamiento de fase uno más baja, la secuencia específica de diana se empareja a la plantilla para producir un monómero inicial y los productos de PCR contienen repeticiones en tándem (15A). En una segunda fase de amplificación a una temperatura más alta, tanto la hibridación específica de diana como la de la secuencia de adaptador se favorecen sobre la hibridación de secuencia específica de diana sola, disminuyendo el grado en que

los productos cortos se producen preferentemente (15B). Sin favorecer el cebador completo, el emparejamiento interno con las secuencias específicas de diana aumenta rápidamente la fracción de monómeros (15C, izquierda).

La FIG. 16 ilustra una comparación entre el ruido de fondo (frecuencia de variantes) detectado por los métodos de secuenciación de dianas usando un filtro Q30 con (línea inferior) y sin (línea superior) que requiera que se produzca una diferencia de secuencia en dos polinucleótidos diferentes (por ejemplo, identificados por uniones diferentes) para contarse como una variante. La aplicación de este filtro de validación también se conoce aquí como "Luciérnaga". El ADN genómico humano (12878, Instituto Coriell) se fragmentó en 100-200 pb, e incluyó un aumento del 2% de ADN genómico (19240, Instituto Coriell) que contiene un SNP conocido (CYP2C19). La señal de variante verdadera (pico marcado) no estaba significativamente por encima del fondo (gráfico superior, gris claro). El ruido de fondo se redujo a aproximadamente 0,1 aplicando el filtro de validación (gráfico inferior, negro).

La FIG. 17 ilustra la detección de variantes de secuencia añadidas a varias frecuencias bajas en la población de polinucleótidos (2%, 0,2 % y 0,02 %), que sin embargo son significativamente superiores al fondo, cuando se aplica un método de la divulgación.

La FIG. 18 ilustra los resultados de un análisis de la eficacia de unión y la tasa específica de una realización de la divulgación.

La FIG. 19 ilustra la conservación de las frecuencias alélicas y la ausencia sustancial de sesgo, en un método de acuerdo con una realización de la divulgación.

La FIG. 20 ilustra los resultados para la detección de variantes de secuencia en una pequeña muestra de entrada, de acuerdo con una realización.

La FIG. 21 ilustra un ejemplo de fondo alto en los resultados para la detección de variantes de secuencia obtenidas sin requerir que se produzca una diferencia de secuencia en dos polinucleótidos diferentes, de acuerdo con los métodos de secuenciación convencional.

La FIG. 22 proporciona gráficos que ilustran comparaciones entre las distribuciones de contenido de GC del genoma y las distribuciones de contenido de GC de los resultados de secuenciación producidos de acuerdo con un método de acuerdo con una realización de la divulgación ("Nebula-Luciérnaga"; izquierda), resultados de secuenciación utilizando un kit de construcción de una biblioteca de secuenciación alternativa (Rubicon, Rubicon Genomics; en el medio) y ADN libre de células (ADNcf) generalmente como se indica en la literatura para 32 ng (derecha).

La FIG. 23 proporciona un gráfico que ilustra la distribución de tamaño del ADN de entrada obtenido de las lecturas de secuenciación de un método de acuerdo con una realización.

La FIG. 24 proporciona un gráfico que ilustra la amplificación uniforme a través de múltiples dianas mediante un método de cebado aleatorio de acuerdo con una realización.

La FIG. 25 ilustra realizaciones para la formación de multímeros de polinucleótidos que tienen uniones identificables, en ausencia de circularización. Los polinucleótidos (tales como los fragmentos de polinucleótidos, o el ADN libre de células) se unen para formar multímeros que tienen uniones no naturales útiles para distinguir polinucleótidos independientes de acuerdo con las realizaciones de la divulgación (también denominados en el presente documento como "etiqueta automática"). En la FIG. 25A, los polinucleótidos se unen directamente entre sí mediante la unión de los extremos romos. En la Figura 25B, los polinucleótidos se unen a través de uno o más oligonucleótidos adaptadores intermedios, que pueden comprender además una secuencia de código de barras. A continuación, los multímeros se someten a amplificación mediante cualquiera de varios métodos, tal como por cebadores aleatorios (amplificación del genoma completo), por cebadores de adaptadores o por uno o más cebadores específicos de diana o pares de cebadores. Los procesos para la formación de multímeros que tienen uniones identificables a partir de una pluralidad de polinucleótidos separados también se denominan en el presente documento "Eclipse".

La FIG. 26 ilustra una variación de ejemplo en el proceso de la FIG. 25. Los polinucleótidos (por ejemplo, ADNcf u otros fragmentos de polinucleótidos) se reparan en el extremo, se añade una cola de A y se unen con un adaptador (por ejemplo, utilizando un kit convencional, como los kits de KAPA Biosystems). El ADN portador marcado con uracilo interno (U) se puede complementar para elevar la entrada de ADN total a los niveles deseados (por ejemplo, a aproximadamente o a aproximadamente más de 20 ng). Una variante de secuencia a detectar se indica con una "estrella". Cuando se completa la unión, el ADN portador puede degradarse mediante la adición de la enzima reactiva de escisión específica de uracilo (USER), que es una mezcla de ADN glicosilasa de uracilo (UDG) y la ADN glicosilasa-liaza Endonucleasa VIII. Los productos se purifican para eliminar fragmentos de ADN portador. Los productos purificados se amplifican (por ejemplo, mediante PCR, utilizando cebadores dirigidos a secuencias de adaptador). No es probable que cualquier ADN portador residual se amplifique debido a la degradación y la separación de un adaptador en al menos un extremo. Los productos amplificados se pueden purificar para eliminar fragmentos cortos de ADN.

La FIG. 27 ilustra una variación de ejemplo en el proceso de la FIG. 25. Los cebadores de amplificación específicos de diana comprenden una "cola" común en 5' que funciona como un adaptador (flecha gris). La amplificación inicial (por ejemplo, mediante PCR) se realiza durante unos pocos ciclos (por ejemplo, al menos aproximadamente 5, 10 o más ciclos). Los productos de PCR también pueden servir como cebadores, emparejados a otros productos de PCR (por ejemplo, cuando la temperatura de emparejamiento se reduce en una segunda fase) para producir concatémeros que tienen uniones identificables. La segunda fase puede comprender varios ciclos (por ejemplo, 5, 10, 15, 20 o más ciclos) y puede incluir una selección o variación de las condiciones que favorecen la formación y amplificación de concatémeros. Los métodos de acuerdo con este esquema también se conocen como "Relay Amp Seq", que puede encontrar un uso particular en una configuración compartimentada (por ejemplo, en una gota).

Las FIG. 28A-E ilustran ejemplos no limitantes de métodos para circularizar polinucleótidos. En la FIG. 28A, los

polinucleótidos de doble cadena (por ejemplo, ADNcd) se desnaturalizan en cadenas simples, seguido de circularización directa (por ejemplo, la unión de autounión por CircLigase). En la Fig. 28B, los polinucleótidos (por ejemplo, fragmentos de ADN) se reparan en el extremo y tienen una cola de A (añadiendo la extensión de adenosina de base única a los extremos 3') para mejorar la eficacia de la unión, seguido de la desnaturalización a cadenas simples y la circularización. En la FIG. 28C., los polinucleótidos se reparan en el extremo y se añade una cola de A (si son de doble cadena), se unen a adaptadores que tienen una extensión de timidina (T), se desnaturalizan en cadenas simples y se circularizan. En la FIG. 28D., los polinucleótidos se reparan en el extremo y se añade una cola de A (si son de doble cadena), ambos extremos están unidos a un adaptador que tiene tres elementos (extensión de T para la unión, complementariedad entre adaptadores y una cola 3'), las cadenas están desnaturalizadas y los polinucleótidos de cadena sencilla se circularizan (facilitado por la complementariedad entre las secuencias de adaptadores). En la FIG. 28E., los polinucleótidos de doble cadena se desnaturalizan a la forma de una sola cadena y se circularizan en presencia de una abrazadera molecular que acerca los extremos del polinucleótido para facilitar la unión.

La FIG. 29 ilustra un ejemplo de diseño de flujo de trabajo de un sistema de amplificación para identificar variantes de secuencia de acuerdo con los métodos de la divulgación, particularmente con respecto a polinucleótidos circularizados.

La FIG. 30 ilustra un ejemplo de diseño de flujo de trabajo de un sistema de amplificación para identificar variantes de secuencia de acuerdo con los métodos de la divulgación, particularmente con respecto a entradas de polinucleótidos lineales sin una etapa de circularización.

La FIG. 31 proporciona una ilustración de resumen de flujos de ejemplo para identificar variantes de secuencia de acuerdo con los métodos de la divulgación. A lo largo de la rama "Eclipse" (análisis de polinucleótidos lineales), el análisis puede incluir PCR digital (por ejemplo, PCR digital en gotas, ddPCR), PCR en tiempo real, enriquecimiento por captura de sonda (sec de captura) con análisis de secuencias de unión (etiqueta automática), secuenciación basada en secuencias de adaptador insertadas (inserción de código de barras), o secuenciación de Relay Amp. A lo largo del "Nebula" (análisis de polinucleótido circularizado), el análisis puede incluir PCR digital (por ejemplo, PCR digital en gotas, ddPCR), PCR en tiempo real, enriquecimiento por captura de sonda (sec de captura) con análisis de secuencias de unión (código de barras natural), enriquecimiento por captura de sonda o amplificación dirigida (por ejemplo, amplificación B2B), y análisis de secuencia con una etapa de validación para identificar una variante de secuencia como una diferencia que se produce en dos polinucleótidos diferentes (por ejemplo, polinucleótidos que tienen diferentes uniones).

La FIG. 32 muestra una ilustración de un sistema de acuerdo con una realización.

La FIG. 33 ilustra la eficacia de la captura y la cobertura a lo largo de las regiones diana según un ejemplo. > 90% de las bases diana están cubiertas más de 20x y > 50% de las bases diana tienen > 50x de cobertura.

Descripción detallada de la invención

La práctica de algunas realizaciones descritas en el presente documento emplea, a menos que se indique otra cosa, técnicas convencionales de inmunología, bioquímica, química, biología molecular, microbiología, biología celular, genómica y ADN recombinante, que están dentro de las habilidades en la técnica. Véase, por ejemplo, Sambrook y Green, *Molecular Cloning: A Laboratory Manual*, 4ª edición (2012); the series *Current Protocols in Molecular Biology* (F. M. Ausubel, et al. eds.); the series *Methods In Enzymology* (Academic Press, Inc.), PCR 2: A Practical Approach (M.J. MacPherson, B.D. Hames y G.R. Taylor eds. (1995)), Harlow y Lane, ed. (1988) *Antibodies, A Laboratory Manual*, and *Culture of Animal Cells: A Manual of Basic Technique and Specialized Applications*, 6ª Edition (R.I. Freshney, ed. (2010)).

El término "aproximadamente" o "de manera aproximada" significa dentro de un intervalo de error aceptable para el valor particular determinado por un experto en la materia, que dependerá en parte de cómo se mide o determina el valor, es decir, de las limitaciones del sistema de medida. Por ejemplo, "aproximadamente" puede significar dentro de 1 o más de 1 desviación estándar, por la practica en la técnica. Como alternativa, "aproximadamente" puede significar un intervalo un rango de hasta el 20%, hasta el 10 %, hasta el 5%, o hasta el 1% de un valor dado. Como alternativa, particularmente con respecto a los sistemas o procesos biológicos, el término puede significar dentro de un orden de magnitud, preferentemente entre 5 veces, y más preferentemente entre 2 veces, un valor. Cuando se describan valores particulares en la solicitud y reivindicaciones, a menos que se indique lo contrario, se debe asumir el término "aproximadamente" dentro de un intervalo de error aceptable para el valor particular.

Los términos "polinucleótido", "nucleótido", "secuencia de nucleótidos", "ácido nucleico" y "oligonucleótido" se usan indistintamente. Se refieren a una forma polimérica de nucleótidos de cualquier longitud, ya sean desoxirribonucleótidos o ribonucleótidos o análogos de los mismos. Los polinucleótidos pueden tener cualquier estructura tridimensional y pueden realizar cualquier función, conocida o desconocida. Los siguientes son ejemplos no limitantes de polinucleótidos: regiones codificantes o no codificantes de un gen o fragmento de gen, loci (locus) definidos a partir del análisis de enlace, exones, intrones, ARN mensajero (ARNm), ARN de transferencia (ARNt), ARN ribosómico (ARNr), ARN de interferencia pequeño (ARNip), ARN de horquilla corta (ARNhc), microARN (miARN), ribozimas, ADNc, polinucleótidos recombinantes, polinucleótidos ramificados, plásmidos, vectores, ADN aislado de cualquier secuencia, ARN aislado de cualquier secuencia, sondas de ácido nucleico y cebadores. Un polinucleótido puede comprender uno o más nucleótidos modificados, tales como nucleótidos metilados y análogos de nucleótidos. Si está presente, pueden impartirse modificaciones a la estructura de los nucleótidos antes o después del ensamblaje.

del polímero. La secuencia de nucleótidos puede estar interrumpida por componentes no nucleotídicos. Un polinucleótido puede modificarse adicionalmente después de la polimerización, tal como mediante conjugación con un componente de marcado.

- 5 En general, la expresión "polinucleótido diana" se refiere a una molécula de ácido nucleico o polinucleótido en una población inicial de moléculas de ácido nucleico que tiene una secuencia diana cuya presencia, cantidad y/o secuencia de nucleótidos, o cambios en uno o más de estos, se desea determinar. En general, la expresión "secuencia diana" se refiere a una secuencia de ácido nucleico en una cadena sencilla de ácido nucleico. La secuencia diana pueden ser una parte de un gen, una secuencia reguladora, ADN genómico, ADNc, ARN incluidos ARNm, miARN, ARNr y
10 otros. La secuencia diana puede ser una secuencia diana de una muestra o una diana secundaria tal como un producto de una reacción de amplificación.

- En general, una "sonda de nucleótidos", "sonda", u "oligonucleótido marcador" se refiere a un polinucleótido utilizado para detectar o identificar su polinucleótido diana correspondiente en una reacción de hibridación mediante hibridación con una secuencia diana correspondiente. Por lo tanto, una sonda de nucleótidos es hibridable a uno o más polinucleótidos diana. Los oligonucleótidos de etiqueta pueden ser perfectamente complementarios a uno o más polinucleótidos diana en una muestra, o contener uno o más nucleótidos que no están complementados por un nucleótido correspondiente en uno o más polinucleótidos diana en una muestra.

- 20 "Hibridación" se refiere a una reacción en la que uno o más polinucleótidos reaccionan para formar un complejo que se estabiliza a través de enlaces de hidrógeno entre las bases de los restos de nucleótidos. El enlace de hidrógeno puede ocurrir por el emparejamiento de bases de Watson Crick, la unión de Hoogsteen, o de cualquier otra forma específica de secuencia de acuerdo con la complementariedad de bases. El complejo puede comprender dos cadenas que forman una estructura dúplex, tres o más cadenas que forman un complejo de múltiples cadenas, una única
25 cadena de auto-hibridación, o cualquier combinación de éstas. Una reacción de hibridación puede constituir un paso en un proceso más extenso, como el inicio de la PCR, o la escisión de un polinucleótido por una endonucleasa. Una segunda secuencia que es complementaria de una primera secuencia se conoce como el "complemento" de la primera secuencia. El término "hibridable" aplicado a un polinucleótido se refiere a la capacidad del polinucleótido para formar un complejo que se estabiliza a través de enlaces de hidrógeno entre las bases de los restos de nucleótidos en una
30 reacción de hibridación.

- "Complementariedad" se refiere a la capacidad de un ácido nucleico para formar enlace(s) de hidrógeno con otra secuencia de ácido nucleico ya sea por Watson-Crick tradicional u otros tipos no tradicionales. Un porcentaje de complementariedad indica el porcentaje de restos en una molécula de ácido nucleico que puede formar enlaces de hidrógeno (por ejemplo, el emparejamiento de bases de Watson-Crick) con una segunda secuencia de ácido nucleico (por ejemplo, 5, 6, 7, 8, 9, 10 de cada 10 son el 50%, 60 %, 70 %, 80 %, el 90 %, y el 100 % complementarios, respectivamente). "Perfectamente complementario" significa que todos los restos contiguos de una secuencia de ácido nucleico se unirán por enlace de hidrógeno con el mismo número de restos contiguos en una segunda secuencia de ácido nucleico. "Substancialmente complementario" como se usa en el presente documento se refiere a un grado de complementariedad que es al menos el 60%, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 97 %, 98 %, el 99%, o el 100% en una región de 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50 o más nucleótidos, o se refiere a dos ácidos nucleicos que hibridan en condiciones rigurosas. La identidad de secuencia, tal como con el propósito de evaluar el porcentaje de complementariedad, puede medirse con cualquier algoritmo de alineamiento adecuado, incluido, pero sin limitación, el algoritmo de Needleman-Wunsch (véase, por ejemplo, el
45 alineador de agujas EMBOSS disponible en www.ebi.ac.uk/Tools/psa/emboss_needle/nucleotide.html, opcionalmente con la configuración predeterminada), el algoritmo BLAST (véase, por ejemplo, la herramienta de alineamiento BLAST disponible en blast.ncbi.nlm.nih.gov/Blast.cgi, opcionalmente con la configuración predeterminada), o el algoritmo Smith-Waterman (véase, por ejemplo, el alineador de agua EMBOSS disponible en www.ebi.ac.uk/Tools/psa/emboss_water/nucleotide.html, opcionalmente con la configuración predeterminada). El alineamiento óptimo se puede evaluar utilizando cualquier parámetro adecuado de un algoritmo elegido, incluidos los parámetros predeterminados.

- En general, Las "condiciones rigurosas" para la hibridación se refieren a las condiciones en las que un ácido nucleico que tiene complementariedad con una secuencia diana hibrida predominantemente con una secuencia diana, y
55 sustancialmente no hibrida con secuencias no diana. Las condiciones rigurosas son generalmente dependientes de la secuencia y varían dependiendo de una serie de factores. En general, cuanto más larga sea la secuencia, mayor será la temperatura a la que la secuencia se hibrida específicamente con su secuencia diana. Ejemplos no limitantes de condiciones rigurosas se describen en detalle en Tijssen(1993), Laboratory Techniques In Biochemistry And Molecular Biology-Hybridization With Nucleic Acid Probes Part I, Second Chapter "Overview of principles of hybridization and the strategy of nucleic acid probe assay", Elsevier, N. Y.

- En un aspecto, la divulgación proporciona un método de identificación de variante de secuencia, tal como una muestra de ácido nucleico. En algunas realizaciones, cada polinucleótido de la pluralidad tiene un extremo 5' y un extremo 3' y el método comprende: (a) circularizar polinucleótidos individuales de dicha pluralidad para formar una pluralidad de polinucleótidos circulares, cada uno de los cuales tiene una unión entre el extremo 5' y el extremo 3'; (b) amplificar los polinucleótidos circulares de (a); (c) secuenciar los polinucleótidos amplificados para producir una pluralidad de

lecturas de secuenciación; (d) identificar diferencias de secuencia entre lecturas de secuenciación y una secuencia de referencia; y (e) identificar una diferencia de secuencia que se produce en al menos dos polinucleótidos circulares que tienen diferentes uniones como la variante de secuencia. En algunas realizaciones, el método comprende identificar las diferencias de secuencia entre las lecturas de secuenciación y una secuencia de referencia e identifica una

5 diferencia de secuencia que se produce en al menos dos polinucleótidos circulares que tienen diferentes uniones como la variante de secuencia, en donde: (a) las lecturas de secuenciación corresponden a productos de amplificación de al menos los dos polinucleótidos circulares; y (b) cada uno de al menos los dos polinucleótidos circulares comprende una unión diferente formada al unir un extremo 5' y un extremo 3' de los polinucleótidos respectivos.

10 En general, la expresión "variante de secuencia" se refiere a cualquier variación en la secuencia relativa a una o más secuencias de referencia. Normalmente, la variante de secuencia se produce con una frecuencia menor que la secuencia de referencia para una población dada de individuos para quienes se conoce la secuencia de referencia. Por ejemplo, un género bacteriano particular puede tener una secuencia de referencia consenso para el gen 16S del ARNr, pero las especies individuales dentro de ese género pueden tener una o más variantes de secuencia dentro del

15 gen (o una porción del mismo) que son útiles para identificar esas especies en una población de bacterias. Como ejemplo adicional, las secuencias para múltiples individuos de la misma especie (o múltiples lecturas de secuenciación para el mismo individuo) pueden producir una secuencia consenso cuando están alineadas de manera óptima, y se pueden usar las variantes de secuencia con respecto a esa consenso para identificar mutantes en la población indicativos de contaminación peligrosa. En general, una "secuencia consenso" se refiere a una secuencia de

20 nucleótidos que refleja la elección más común de base en cada posición en la secuencia en la que la serie de ácidos nucleicos relacionados se ha sometido a un análisis matemático y/o de secuencia intensivo, tal como el alineamiento óptimo de secuencias según cualquiera de varios algoritmos de alineamiento de secuencias. Hay varios algoritmos de alineación disponibles, algunos de los cuales se describen en el presente documento. En algunas realizaciones, la secuencia de referencia es una secuencia de referencia única conocida, tal como la secuencia genómica de un solo

25 individuo. En algunas realizaciones, la secuencia de referencia es una secuencia consenso formada al alinear múltiples secuencias conocidas, tal como la secuencia genómica de múltiples individuos que sirven como población de referencia, o múltiples lecturas de secuenciación de polinucleótidos del mismo individuo. En algunas realizaciones, la secuencia de referencia es una secuencia consenso formada al alinear de manera óptima las secuencias de una muestra en análisis, de modo que una variante de secuencia representa una variación relativa a las secuencias

30 correspondientes en la misma muestra. En algunas realizaciones, la variante de secuencia se produce con una frecuencia baja en la población (también denominada variante de secuencia "rara"). Por ejemplo, la variante de secuencia se puede producir con una frecuencia de aproximadamente el o aproximadamente menos del 5%, 4 %, 3 %, 2 %, 1,5 %, 1 %, 0,75 %, 0,5 %, 0,25 %, 0,1 %, 0,075 %, 0,05 %, 0,04 %, 0,03 %, 0,02 %, 0,01 %, 0,005 %, 0,001 % o menos. En algunas realizaciones, la variante de secuencia se produce con una frecuencia de

35 aproximadamente el o aproximadamente menos del 0,1%.

Una variante de secuencia puede ser cualquier variación con respecto a una secuencia de referencia. Una variación de secuencia puede consistir en un cambio en, inserción de o eliminación de un solo nucleótido, o de una pluralidad de nucleótidos (por ejemplo, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más nucleótidos). Cuando una variante de secuencia comprende

40 dos o más diferencias de nucleótidos, los nucleótidos que son diferentes pueden ser contiguos entre sí, o discontinuos. Ejemplos no limitantes de tipos de variantes de secuencia incluyen polimorfismos de un solo nucleótido (SNP), polimorfismos de delección/inserción (DIP), variantes de número de copias (CNV), repeticiones cortas en tándem (STR), repeticiones de secuencia simple (SSR), número variable de repeticiones en tándem (VNTR), polimorfismos de longitud de fragmentos amplificados (AFLP), polimorfismos de inserción basada en retrotransposón, polimorfismo

45 amplificado específico de secuencia y diferencias en las marcas epigenéticas que pueden detectarse como variantes de secuencia (por ejemplo, diferencias en la metilación).

Las muestras de ácido nucleico que pueden someterse a los métodos descritos en el presente documento pueden derivar de cualquier fuente adecuada. En algunas realizaciones, las muestras utilizadas son muestras ambientales.

50 La muestra ambiental puede ser de cualquier fuente ambiental, por ejemplo, atmósfera de origen natural o artificial, sistemas de agua, suelo o cualquier otra muestra de interés. En algunas realizaciones, las muestras ambientales pueden obtenerse de, por ejemplo, sistemas de recolección de patógenos atmosféricos, sedimentos subterráneos, aguas subterráneas, aguas remotas en el interior del suelo, la superficie de contacto entre la raíz de la planta y el suelo de las praderas, las aguas costeras y las plantas de tratamiento de aguas residuales.

55 Los polinucleótidos de una muestra pueden ser cualquiera de varios polinucleótidos, incluyendo, pero sin limitación, ADN, ARN, ARN ribosómico (ARNr), ARN de transferencia (ARNt), microARN (miARN), ARN mensajero (ARNm), fragmentos de cualquiera de estos o combinaciones de cualquiera de dos o más de estos. En algunas realizaciones, las muestras comprenden ADN. En algunas realizaciones, las muestras comprenden ADN genómico. En algunas

60 realizaciones, las muestras comprenden ADN mitocondrial, ADN de cloroplasto, ADN plasmídico, cromosomas bacterianos artificiales, cromosomas de levadura artificiales, etiquetas de oligonucleótidos, o combinaciones de los mismos. En algunas realizaciones, las muestras comprenden ADN generado por amplificación, tal como las reacciones de extensión de cebadores que usan cualquier combinación adecuada de cebadores y una ADN polimerasa, que incluyen, pero sin limitación, reacción en cadena de la polimerasa (PCR), transcripción inversa y combinaciones de las mismas. Cuando la plantilla para la reacción de extensión del cebador es el ARN, el producto de la transcripción

65 inversa se denomina ADN complementario (ADNc). Los cebadores útiles en las reacciones de extensión de cebadores

pueden comprender secuencias específicas de una o más dianas, secuencias aleatorias, secuencias parcialmente aleatorias y combinaciones de las mismas. En general, los polinucleótidos de muestra comprenden cualquier polinucleótido presente en una muestra, que puede incluir o no polinucleótidos diana. Los polinucleótidos pueden ser de cadena sencilla, de doble cadena o una combinación de ellos. En algunas realizaciones, los polinucleótidos sometidos a un método de la divulgación son polinucleótidos de cadena sencilla, que pueden estar o no en presencia de polinucleótidos de doble cadena. En algunas realizaciones, los polinucleótidos son ADN de cadena sencilla. El ADN de cadena sencilla (ADNcs) puede ser un ADNcs que se aísla en una forma de cadena sencilla, o el ADN que se aísla en forma de doble cadena y posteriormente se hace de cadena sencilla para el propósito de uno o más pasos en un método de la divulgación.

En algunas realizaciones, los polinucleótidos se someten a etapas posteriores (por ejemplo, circularización y amplificación) sin una etapa de extracción y/o sin una etapa de purificación. Por ejemplo, una muestra de fluido se puede tratar para eliminar las células sin una etapa de extracción para producir una muestra de líquido purificado y una muestra de células, seguido del aislamiento del ADN de la muestra de fluido purificado. Se dispone de varios procedimientos para el aislamiento de polinucleótidos, tales como por precipitación o unión no específica a un sustrato seguido de lavado del sustrato para liberar polinucleótidos unidos. Cuando los polinucleótidos se aíslan de una muestra sin una etapa de extracción celular, los polinucleótidos serán en gran medida polinucleótidos extracelulares o "libres de células", que pueden corresponder a células muertas o dañadas. La identidad de dichas células se puede usar para caracterizar las células o la población de células de las que derivan, como células tumorales (por ejemplo, en la detección de cáncer), células fetales (por ejemplo, en el diagnóstico prenatal), células de tejido trasplantado (por ejemplo en la detección temprana de fracaso del trasplante), o miembros de una comunidad microbiana.

Si se trata una muestra para extraer polinucleótidos, tal como a partir de células en una muestra, hay varios métodos de extracción disponibles. Por ejemplo, los ácidos nucleicos pueden purificarse por extracción orgánica con fenol, fenol/cloroformo/alcohol isoamílico, o formulaciones similares, incluyendo TRIzol y TriReagent. Otros ejemplos no limitantes de técnicas de extracción incluyen: (1) extracción orgánica seguida de precipitación con etanol, por ejemplo, usando un reactivo orgánico de fenol/cloroformo (Ausubel et al., 1993), con o sin el uso de un extractor automático de ácido nucleico, por ejemplo, el Extractor de ADN Modelo 341 disponible en Applied Biosystems (Foster City, California); (2) métodos de adsorción de fase estacionaria (Patente de Estados Unidos N° 5.234.809; Walsh et al., 1991); y (3) métodos de precipitación de ácidos nucleicos inducidos por sales (Miller et al., (1988), denominándose a estos métodos de precipitación como métodos de "salado"). Otro ejemplo de aislamiento y/o purificación de ácidos nucleicos incluye el uso de partículas magnéticas a las que los ácidos nucleicos pueden unirse de manera específica o no, seguido del aislamiento de las perlas con un imán, y el lavado y elución de los ácidos nucleicos de las perlas (véase por ejemplo, la Patente de Estados Unidos N° 5.705.628). En algunas realizaciones, los métodos de aislamiento anteriores pueden estar precedidos por una etapa de digestión enzimática para ayudar a eliminar la proteína no deseada de la muestra, por ejemplo, la digestión con proteinasa K u otras proteasas similares. Véase, por ejemplo, la Patente de EE.UU n.º 7.001.724. Si se desea, pueden añadirse inhibidores de la ARNasa al tampón de lisis. Para ciertos tipos de células o muestras, puede ser deseable añadir al protocolo una etapa de desnaturalización/digestión de proteínas. Los métodos de purificación pueden dirigirse para aislar ADN, ARN o ambos. Cuando tanto el ADN como el ARN se aíslan juntos durante o después de un procedimiento de extracción, se pueden emplear etapas adicionales para purificar uno o ambos por separado. También se pueden generar subfracciones de ácidos nucleicos extraídos, por ejemplo, purificación por tamaño, secuencia u otra característica física o química. Además de una etapa inicial de aislamiento de ácidos nucleicos, la purificación de los ácidos nucleicos se puede realizar después de cualquier etapa en los métodos descritos, como la eliminación de reactivos en exceso o no deseados, reactivos o productos. Se dispone de varios métodos para determinar la cantidad y/o la pureza de los ácidos nucleicos en una muestra, como la absorbancia (por ejemplo, la absorbancia de la luz a 260 nm, 280 nm y una proporción de estos) y la detección de un marcador (por ejemplo, colorantes fluorescentes y agentes intercalantes, tales como SYBR verde, SYBR azul, DAPI, yoduro de propidio, tinción Hoechst, SYBR oro, bromuro de etidio).

En los casos en los que se desee, los polinucleótidos de una muestra pueden fragmentarse antes de continuar con el procesamiento. La fragmentación se puede lograr por cualquiera de varios métodos, incluyendo fragmentación química, enzimática y mecánica. En algunas realizaciones, los fragmentos tienen una longitud media o mediana de aproximadamente 10 a aproximadamente 1.000 nucleótidos de longitud, tal como entre 10-800, 10-500, 50-500, 90-200 o 50-150 nucleótidos. En algunas realizaciones, los fragmentos tienen una longitud media o mediana de aproximadamente o aproximadamente menos de 100, 200, 300, 500, 600, 800, 1000 o 1500 nucleótidos. En algunas realizaciones, los fragmentos varían desde aproximadamente 90-200 nucleótidos y/o tienen una longitud promedio de aproximadamente 150 nucleótidos. En algunas realizaciones, la fragmentación se realiza mecánicamente y comprende someter los polinucleótidos de la muestra a sonicación acústica. En algunas realizaciones, la fragmentación comprende tratar los polinucleótidos de muestra con una o más enzimas en condiciones adecuadas para que una o más enzimas generen roturas de ácido nucleico de doble cadena. Los ejemplos de enzimas útiles en la generación de fragmentos de polinucleótidos incluyen nucleasas específicas de secuencia y no específicas de secuencia. Los ejemplos no limitantes de nucleasas incluyen DNasa I, Fragmentasa, endonucleasas de restricción, variantes de las mismas y combinaciones de las mismas. Por ejemplo, la digestión con DNasa I puede inducir roturas aleatorias de doble cadena en el ADN en ausencia de Mg ++ y en presencia de Mn++. En algunas realizaciones, la fragmentación comprende tratar los polinucleótidos de muestra con una o más endonucleasas de restricción. La fragmentación puede producir fragmentos que tienen salientes 5', salientes 3', extremos romos o una combinación de

los mismos. En algunas realizaciones, tal como cuando la fragmentación comprende el uso de una o más endonucleasas de restricción, la escisión de polinucleótidos de muestra deja salientes que tienen una secuencia predecible. Los polinucleótidos fragmentados pueden someterse a una etapa de selección por tamaño de los fragmentos mediante métodos convencionales tales como la purificación en columna o el aislamiento de un gel de agarosa.

De acuerdo con algunas realizaciones, los polinucleótidos entre la pluralidad de polinucleótidos de una muestra están circularizados. La circularización puede incluir unir el extremo 5' de un polinucleótido al extremo 3' del mismo polinucleótido, al extremo 3' de otro polinucleótido en la muestra, o al extremo 3' de un polinucleótido de una fuente diferente (por ejemplo, un polinucleótido, tal como un adaptador de oligonucleótido). En algunas realizaciones, el extremo 5' de un polinucleótido se une al extremo 3' del mismo polinucleótido (también denominado "autounión"). En algunas realizaciones, las condiciones de la reacción de circularización se seleccionan para favorecer la autounión de polinucleótidos dentro de un intervalo concreto de longitudes, para producir una población de polinucleótidos circularizados de una longitud promedio concreta. Por ejemplo, las condiciones de reacción de circularización pueden seleccionarse para favorecer la autounión de polinucleótidos más cortos que aproximadamente 5000, 2500, 1000, 750, 500, 400, 300, 200, 150, 100, 50 o menos nucleótidos de longitud. En algunas realizaciones, se favorecen los fragmentos que tienen longitudes entre 50-5000 nucleótidos, 100-2500 nucleótidos, o 150-500 nucleótidos, de tal manera que la longitud promedio de los polinucleótidos circularizados entra dentro del respectivo intervalo. En algunas realizaciones, el 80% o más de los fragmentos circularizados tienen entre 50-500 nucleótidos de longitud, tal como entre 50-200 nucleótidos de longitud. Las condiciones de reacción que pueden optimizarse incluyen el tiempo asignado para una reacción de unión, la concentración de varios reactivos y la concentración de polinucleótidos a unir. En algunas realizaciones, una reacción de circularización conserva la distribución de las longitudes de los fragmentos presentes en una muestra antes de la circularización. Por ejemplo, uno o más de la media, la mediana, la moda y la desviación estándar de las longitudes de los fragmentos en una muestra antes de la circularización y de los polinucleótidos circularizados están dentro del 75%, 80 %, 85 %, 90 %, 95 % o más entre sí.

En lugar de formar preferentemente productos de circularización de autounión, se usan uno o más oligonucleótidos adaptadores, de manera que el extremo 5' y el extremo 3' de un polinucleótido en la muestra se unen mediante uno o más oligonucleótidos adaptadores intermedios para formar un polinucleótido circular. Por ejemplo, el extremo 5' de un polinucleótido puede unirse al extremo 3' de un adaptador, y el extremo 5' del mismo adaptador puede unirse al extremo 3' del mismo polinucleótido. Un oligonucleótido adaptador incluye cualquier oligonucleótido que tenga una secuencia, que al menos se conozca una parte de la misma, que pueda unirse a un polinucleótido de muestra. Los oligonucleótidos adaptadores pueden comprender ADN, ARN, análogos de nucleótidos, nucleótidos no canónicos, nucleótidos marcados, nucleótidos modificados o combinaciones de los mismos. Los oligonucleótidos adaptadores pueden ser de cadena sencilla, de doble cadena o dúplex parcial. En general, un adaptador dúplex parcial comprende una o más regiones de cadena sencillas y una o más regiones de doble cadena. Los adaptadores de doble cadena pueden comprender dos oligonucleótidos separados hibridados entre sí (también llamados "dúplex de oligonucleótidos") y la hibridación puede dejar uno o más extremos romos, uno o más salientes 3', uno o más salientes 5', uno o más protuberancias resultantes de nucleótidos desapareados y/o no emparejados o cualquier combinación de estos. Cuando dos regiones hibridadas de un adaptador están separadas una de otra por una región no hibridada, se produce una estructura de "burbuja". Se pueden utilizar adaptadores de diferentes tipos en combinación, tales como adaptadores de diferentes secuencias. Se pueden unir diferentes adaptadores a los polinucleótidos de muestra en reacciones secuenciales o simultáneamente. En algunas realizaciones, se añaden adaptadores idénticos a ambos extremos de un polinucleótido diana. Por ejemplo, los adaptadores primero y segundo se pueden añadir a la misma reacción. Los adaptadores pueden manipularse antes de combinarlos con polinucleótidos de muestra. Por ejemplo, se pueden añadir o eliminar los fosfatos terminales.

Cuando se usan oligonucleótidos adaptadores, los oligonucleótidos adaptadores pueden contener uno o más de varios elementos de secuencia, incluyendo, pero sin limitación, una o más secuencias de emparejamiento de cebadores de amplificación o complementos de las mismas, una o más secuencias de emparejamiento de cebadores de secuenciación o de las mismas, uno o más secuencias de códigos de barras, una o más secuencias comunes compartidas entre múltiples adaptadores diferentes o subconjuntos de adaptadores diferentes, uno o más sitios de reconocimiento de enzimas de restricción, uno o más salientes complementarios a uno o más salientes de polinucleótidos diana, uno o más sitios de unión de la sonda (por ejemplo, para unirse a una plataforma de secuenciación, tal como una celda de flujo para una secuenciación masiva paralela, tal como las celdas de flujo desarrolladas por Illumina, Inc.), una o más secuencias aleatorias o casi aleatorias (por ejemplo, uno o más nucleótidos seleccionados al azar de un conjunto de dos o más nucleótidos diferentes en una o más posiciones, con cada uno de los diferentes nucleótidos seleccionados en una o más posiciones representadas en un conjunto de adaptadores que comprenden la secuencia aleatoria), y combinaciones de las mismas. En algunos casos, los adaptadores se pueden usar para purificar los círculos que contienen los adaptadores, por ejemplo, mediante el uso de perlas (particularmente perlas magnéticas para facilitar su manejo) que están recubiertas con oligonucleótidos que comprenden una secuencia complementaria al adaptador, que pueden "capturar" los círculos cerrados con los adaptadores correctos mediante la hibridación a los mismos, lavar los círculos que no contienen los adaptadores y los componentes no ligados y liberar, a continuación, los círculos capturados de las perlas. Además, en algunos casos, el complejo de la sonda de captura hibridada y el círculo diana se pueden usar directamente para generar concatémicos, tal como por amplificación directa por círculo rodante (RCA). En algunas realizaciones, los adaptadores en los círculos también pueden usarse como

cebador de secuenciación. Dos o más elementos de secuencia pueden ser no adyacentes entre sí (por ejemplo, estar separados por uno o más nucleótidos), adyacentes entre sí, parcialmente superpuestos o completamente superpuestos. Por ejemplo, una secuencia de emparejamiento de cebadores de amplificación también puede servir como una secuencia de emparejamiento de cebadores de secuenciación. Los elementos de secuencia se pueden ubicar en o cerca del extremo 3', en o cerca del extremo 5', o en el interior del oligonucleótido adaptador. Un elemento de secuencia puede tener cualquier longitud adecuada, tal como aproximadamente o aproximadamente menos de 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50 o más nucleótidos de longitud. Los oligonucleótidos adaptadores pueden tener cualquier longitud adecuada, al menos suficiente para alojar uno o más elementos de secuencia de los que están comprendidos. En algunas realizaciones, los adaptadores son aproximadamente o aproximadamente menos de 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 90, 100, 200, o más nucleótidos de longitud. En algunas realizaciones, un oligonucleótido adaptador está en el intervalo de aproximadamente 12 a 40 nucleótidos de longitud, tal como de aproximadamente 15 a 35 nucleótidos de longitud.

En algunas realizaciones, los oligonucleótidos adaptadores unidos a los polinucleótidos fragmentados de una muestra comprenden una o más secuencias comunes a todos los oligonucleótidos adaptadores y un código de barras que es exclusivo de los adaptadores unidos a los polinucleótidos de esa muestra en particular, de manera que la secuencia del código de barras se puede usar para distinguir polinucleótidos originados de una muestra o de reacción de unión de adaptador, de polinucleótidos que se originan de otra muestra o de reacción de unión de adaptador. En algunas realizaciones, un oligonucleótido adaptador comprende un saliente 5', un saliente 3', o ambos que es complementario a uno o más salientes de polinucleótidos diana. Los salientes complementarios pueden tener uno o más nucleótidos de longitud, incluidos, pero sin limitación, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 o más nucleótidos de longitud. Los salientes complementarios pueden comprender una secuencia fija. Los salientes complementarios de un oligonucleótido adaptador pueden comprender una secuencia aleatoria de uno o más nucleótidos, de modo que uno o más nucleótidos se seleccionan al azar de un conjunto de dos o más nucleótidos diferentes en una o más posiciones, con cada uno de los diferentes nucleótidos seleccionados en una o más posiciones representadas en un conjunto de adaptadores con salientes complementarios que comprenden la secuencia aleatoria. En algunas realizaciones, un saliente de adaptador es complementario a un saliente de un polinucleótido diana producido por la digestión con endonucleasas de restricción. En algunas realizaciones, un saliente de adaptador consiste en una adenina o una timina.

Se dispone de varios métodos para la circularización de polinucleótidos. En algunas realizaciones, la circularización comprende una reacción enzimática, tal como el uso de una ligasa (por ejemplo, una ARN o ADN ligasa). Hay varias ligasas disponibles, incluyendo, pero sin limitación, CircLigase™ (Epicentre; Madison, WI), ARN ligasa, ARN ligasa 1 de T4 (ARNss ligasa, que funciona tanto en el ADN como en el ARN). Además, la ADN ligasa de T4 también puede ligar ADNcs si no hay plantillas de ADNcd, aunque esto generalmente es una reacción lenta. Otros ejemplos no limitantes de ligasas incluyen ligasas dependientes de NAD que incluyen Tag ADN ligasa, ADN ligasa de *Thermus filiformis*, ADN ligasa de *Escherichia coli*, Tth ADN ligasa, ADN ligasa (I y II) de *Thermus scotoductus*, ligasa termoestable, ADN ligasa ampligasa termoestable, ligasa tipo VanC, 9° N ADN ligasa, Tsp ADN ligasa y nuevas ligasas descubiertas mediante bioprospección; ligasas dependientes de ATP incluyendo la ARN ligasa de T4, ADN ligasa de T4, ADN ligasa de T3, ADN ligasa de T7, Pfu ADN ligasa, ADN ligasa 1, ADN ligasa II, ADN ligasa IV y nuevas ligasas descubiertas mediante bioprospección; e isoformas mutantes de tipo silvestre y variantes de ingeniería genética de las mismas. Cuando se desea una auto-unión, la concentración de polinucleótidos y enzimas puede ajustarse para facilitar la formación de círculos intramoleculares en lugar de estructuras intermoleculares. También se pueden ajustar las temperaturas y tiempos de reacción. En algunas realizaciones, se utiliza 60 °C para facilitar los círculos intramoleculares. En algunas realizaciones, los tiempos de reacción son entre 12-16 horas. Las condiciones de reacción pueden ser aquellas especificadas por el fabricante de la enzima seleccionada. En algunas realizaciones, se puede incluir una etapa de exonucleasa para digerir cualquier ácido nucleico no unido después de la reacción de circularización. Es decir, los círculos cerrados no contienen un extremo libre 5' o 3', y por lo tanto, la introducción de una exonucleasa 5' o 3' no digerirá los círculos cerrados, sino que digerirá los componentes no unidos. Esto puede encontrar un uso particular en sistemas multiplexados.

En general, la unión de los extremos de un polinucleótido entre sí para formar un polinucleótido circular (ya sea directamente o con uno o más oligonucleótidos adaptadores intermedios) produce una unión que tiene una secuencia de unión. Cuando el extremo 5' y el extremo 3' de un polinucleótido se unen a través de un polinucleótido adaptador, el término "unión" puede referirse a una unión entre el polinucleótido y el adaptador (por ejemplo, uno de la unión del extremo 5' o de la unión del extremo 3'), o a la unión entre el extremo 5' y el extremo 3' del polinucleótido formado por e incluyendo el polinucleótido adaptador. Cuando el extremo 5' y el extremo 3' de un polinucleótido se unen sin un adaptador intermedio (por ejemplo, el extremo 5' y el extremo 3' de un ADN de cadena sencilla), el término "unión" se refiere al punto en el que estos dos extremos están unidos. Una unión puede identificarse por la secuencia de nucleótidos que comprende la unión (también denominada "secuencia de unión"). En algunas realizaciones, las muestras comprenden polinucleótidos que tienen una mezcla de extremos formados por procesos de degradación natural (como la lisis celular, la muerte celular y otros procesos mediante los cuales el ADN se libera de una célula a su entorno circundante en el que puede degradarse aún más, tal como en los polinucleótidos libres de células), la fragmentación que es un subproducto del procesamiento de la muestra (como los procedimientos de fijación, tinción y/o almacenamiento), y la fragmentación por métodos que escinden el ADN sin restricción a secuencias diana específicas (p. ej., fragmentación mecánica, tal como por sonicación; tratamiento con nucleasa no específica de

secuencia, tal como la DNasa I, fragmentasa). Cuando las muestras comprenden polinucleótidos que tienen una mezcla de extremos, la probabilidad de que dos polinucleótidos tengan el mismo extremo 5' o el mismo extremo 3' es baja, y la probabilidad de que dos polinucleótidos tengan independientemente tanto el mismo extremo 5' como el mismo extremo 3' es extremadamente baja. En consecuencia, en algunas realizaciones, las uniones pueden usarse para distinguir diferentes polinucleótidos, incluso cuando los dos polinucleótidos comprenden una porción que tiene la misma secuencia diana. Cuando los extremos de los polinucleótidos se unen sin un adaptador intermedio, se puede identificar una secuencia de unión mediante el alineamiento con una secuencia de referencia. Por ejemplo, donde el orden de las dos secuencias de componentes parece invertirse con respecto a la secuencia de referencia, el punto en el que parece ocurrir la reversión puede ser una indicación de una unión en ese punto. Cuando los extremos de los polinucleótidos se unen a través de una o más secuencias de adaptadores, se puede identificar una unión por la proximidad a la secuencia de adaptador conocida, o por el alineamiento como anteriormente, si una lectura de secuenciación tiene la longitud suficiente para obtener la secuencia de los extremos tanto 5' como 3' del polinucleótido circularizado. En algunas realizaciones, la formación de una unión particular es un evento lo suficientemente raro como para que sea único entre los polinucleótidos circularizados de una muestra.

La FIG. 4 ilustra tres ejemplos no limitantes de métodos para circularizar polinucleótidos. En la parte superior, los polinucleótidos se circularizan en ausencia de adaptadores, mientras que el esquema del medio representa el uso de adaptadores, y el esquema inferior utiliza dos adaptadores. Cuando se usan dos adaptadores, uno puede unirse al extremo 5' del polinucleótido, mientras que el segundo adaptador puede unirse al extremo 3' del mismo polinucleótido. En algunas realizaciones, la unión del adaptador puede comprender el uso de dos adaptadores diferentes junto con un ácido nucleico "splint" (férula) que es complementario a los dos adaptadores para facilitar la unión. También se pueden usar adaptadores bifurcados o en "Y". Cuando se usan dos adaptadores, los polinucleótidos que tienen el mismo adaptador en ambos extremos se pueden eliminar en pasos posteriores debido al autoemparejamiento.

La FIG. 6 ilustra ejemplos no limitantes adicionales de métodos para circularizar polinucleótidos, tal como ADN de cadena sencilla. El adaptador se puede añadir asimétricamente al extremo 5' o 3' de un polinucleótido. Como se muestra en la FIG. 6A, el ADN de cadena sencilla (ADNcs) tiene un grupo hidroxilo libre en el extremo 3' y el adaptador tiene un extremo 3' bloqueado, de manera que en presencia de una ligasa, una reacción preferida une el extremo 3' del ADNcs al 5' extremo del adaptador. En esta realización, puede ser útil usar agentes tales como polietilenglicoles (PEG) para dirigir la unión intermolecular de un único fragmento de ADNcs y un único adaptador, antes que la unión intramolecular para formar un círculo. También se puede hacer el orden inverso de los extremos (3' bloqueado, 5' libre, etc.). Una vez que se logra la unión lineal, las piezas unidas pueden tratarse con una enzima para eliminar el resto de bloqueo, tal como mediante el uso de una quinasa u otras enzimas o sustancias químicas adecuadas. Una vez que se elimina el resto de bloqueo, la adición de una enzima de circularización, tal como CircLigase, permite una reacción intramolecular para formar el polinucleótido circularizado. Como se muestra en la figura 6B, al usar un adaptador de doble cadena con una cadena que tiene un extremo 5' o 3' bloqueado, se puede formar una estructura de doble cadena, que al unirse produce un fragmento de doble cadena con muescas. Las dos cadenas se pueden separar a continuación, se elimina el resto de bloqueo y se circulariza el fragmento de una sola cadena para formar un polinucleótido circularizado.

En algunas realizaciones, las abrazaderas moleculares se utilizan para unir dos extremos de un polinucleótido (por ejemplo, un ADN de cadena sencilla) para aumentar la tasa de circularización intramolecular. Una ilustración de ejemplo de uno de tales procesos se ilustra en la FIG. 5. Esto se puede hacer con o sin adaptadores. El uso de abrazaderas moleculares puede ser particularmente útil en casos donde el fragmento de polinucleótido promedio es mayor de aproximadamente 100 nucleótidos de longitud. En algunas realizaciones, la sonda de abrazadera molecular comprende tres dominios: un primer dominio, un dominio intermedio y un segundo dominio. El primer y segundo dominio hibridarán primero con las secuencias correspondientes en un polinucleótido diana mediante la complementariedad de secuencia. El dominio intermedio de la sonda de abrazadera molecular no hibrida significativamente con la secuencia diana. La hibridación de la abrazadera con el polinucleótido diana trae de este modo los dos extremos de la secuencia diana a una mayor proximidad, lo que facilita la circularización intramolecular de la secuencia diana en presencia de una enzima de circularización. En algunas realizaciones, esto es adicionalmente útil ya que la abrazadera molecular puede servir también como un cebador de amplificación.

Después de la circularización, los productos de reacción se pueden purificar antes de la amplificación o secuenciación para aumentar la concentración o pureza relativa de los polinucleótidos circularizados disponibles para participar en las etapas posteriores (por ejemplo, mediante el aislamiento de polinucleótidos circulares o la eliminación de una o más moléculas en la reacción). Por ejemplo, una reacción de circularización o componentes de la misma pueden tratarse para eliminar polinucleótidos de cadena sencilla (no circularizados), tal como por tratamiento con una exonucleasa. Como ejemplo adicional, una reacción de circularización o parte de ella puede someterse a cromatografía de exclusión por tamaño, por lo que los reactivos pequeños se retienen y descartan (por ejemplo, adaptadores sin reaccionar), o los productos de circularización se retienen y liberan en un volumen separado. Hay disponibles varios kits para limpiar las reacciones de unión, tales como los kits proporcionados por los kits de purificación de oligo Zymo fabricados por Zymo Research. En algunas realizaciones, la purificación comprende un tratamiento para eliminar o degradar la ligasa usada en la reacción de circularización y/o para purificar polinucleótidos circularizados lejos de dicha ligasa. En algunas realizaciones, el tratamiento para degradar la ligasa comprende el tratamiento con una proteasa, tal como la proteinasa K. El tratamiento con proteinasa K puede seguir los protocolos

del fabricante o los protocolos convencionales (por ejemplo, según se proporciona en Sambrook y Green, *Molecular Cloning: A Laboratory Manual*, 4ª edición (2012)). El tratamiento con proteasa también puede seguirse por extracción y precipitación. En un ejemplo, los polinucleótidos circularizados se purifican mediante tratamiento con proteinasa K (Qiagen) en presencia de SDS al 0,1% y EDTA 20 mM, se extraen con fenol/cloroformo 1:1 y cloroformo y se precipitan con etanol o isopropanol. En algunas realizaciones, la precipitación es en etanol.

La circularización se puede seguir directamente secuenciando los polinucleótidos circularizados. Como alternativa, la secuenciación puede estar precedida por una o más reacciones de amplificación. En general, "amplificación" se refiere a un proceso mediante el cual se hacen una o más copias de un polinucleótido diana o una porción del mismo. Hay disponibles varios métodos para amplificar polinucleótidos (por ejemplo, ADN y/o ARN). La amplificación puede ser lineal, exponencial o involucrar fases tanto lineales como exponenciales en un proceso de amplificación multifase. Los métodos de amplificación pueden implicar cambios en la temperatura, tal como una etapa de desnaturalización por calor, o pueden ser procesos isotérmicos que no requieren desnaturalización por calor. La reacción en cadena de la polimerasa (PCR) utiliza múltiples ciclos de desnaturalización, emparejamiento de pares de cebadores en cadenas opuestas y la extensión del cebador para aumentar exponencialmente los números de copias de la secuencia diana. La desnaturalización de las cadenas de ácido nucleico emparejadas se puede lograr mediante la aplicación de calor, aumentando las concentraciones locales de iones metálicos (por ejemplo, la Patente de los Estados Unidos n.º 6.277.605), la radiación de ultrasonido (por ejemplo, el documento WO/2000/049176), la aplicación de voltaje (por ejemplo, la Patente de Estados Unidos n.º 5.527.670, la Patente de EE.UU. n.º 6.033.850, la Patente de EE.UU. n.º 5.939.291 y la Patente de EE.UU. n.º 6.333.157), y la aplicación de un campo electromagnético en combinación con cebadores unidos a un material magnéticamente sensible (por ejemplo, la Patente de EE.UU. n.º 5.545.540). En una variación denominada RT-PCR, se usa transcriptasa inversa (RT) para preparar un ADN complementario (ADNc) a partir de ARN y el ADNc se amplifica a continuación, mediante PCR para producir múltiples copias de ADN (por ejemplo, la Patente de Estados Unidos N.º 5.322.770 y la Patente de Estados Unidos N.º 5.310.652). Un ejemplo de un método de amplificación isotérmica es la amplificación por desplazamiento de cadena, comúnmente conocida como SDA, que utiliza ciclos de emparejamiento de secuencias de cebadores para cadenas opuestas de una secuencia diana, la extensión del cebador en presencia de un dNTP para producir un producto de extensión de cebador hemifosforotioato dúplex, el corte mediado por endonucleasa de un sitio de reconocimiento de endonucleasa de restricción hemimodificado y la extensión de cebador mediada por polimerasa desde el extremo 3' del corte para desplazar una cadena existente y producir una cadena para la siguiente ronda de emparejamiento de cebador, corte y desplazamiento de la cadena, dando como resultado la amplificación geométrica del producto (por ejemplo, la Patente de Estados Unidos N.º 5.270.184 y la Patente de Estados Unidos N.º 5.455.166). La SDA termofílica (tSDA) usa endonucleasas y polimerasas termofílicas a temperaturas más altas en esencialmente el mismo método (Patente Europea N.º 0 684 315). Otros métodos de amplificación incluyen la amplificación de círculo rodante (RCA) (por ejemplo, Lizardi, "Rolling Circle Replication Reporter Systems," Patente de EE.UU. n.º 5.854.033); amplificación dependiente de helicasa (HDA) (por ejemplo, Kong et al., "Helicase Dependent Amplification Nucleic Acids," solicitud de patente de Estados Unidos Pub. N.º US 2004-0058378 A1); y amplificación isotérmica mediada por bucle (LAMP) (por ejemplo, Notomi et al., "Process for Synthesizing Nucleic Acid," Patente de EE.UU. n.º 6.410.278). En algunos casos, la amplificación isotérmica utiliza la transcripción por una ARN polimerasa de una secuencia promotora, tal como puede incorporarse en un cebador de oligonucleótido. Los métodos de amplificación basados en la transcripción incluyen la amplificación basada en la secuencia de ácido nucleico, también conocida como NASBA (por ejemplo, la Patente de EE. UU. N.º 5,130,238); métodos que se basan en el uso de una ARN replicasa para amplificar la propia molécula de sonda, comúnmente conocida como replicasa Q β (p.ej, Lizardi, P. et al. (1988) *BioTechnol.* 6, 1197-1202); replicación de secuencia auto sostenida (por ejemplo, Guatelli, J. et al. (1990) *Proc. Natl. Acad. Sci. USA* 87, 1874-1878; Landgren (1993) *Trends in Genetics* 9, 199-202; y HELEN H. LEE et al., NUCLEIC ACID AMPLIFICATION TECHNOLOGIES (1997)); y métodos para generar plantillas de transcripción adicionales (por ejemplo, la Patente de EE.UU. N.º 5.480.784 y la patente de EE.UU. N.º 5.399.491). Otros métodos de amplificación de ácido nucleico isotérmica incluyen el uso de cebadores que contienen nucleótidos no canónicos (por ejemplo, uracilo o nucleótidos de ARN) en combinación con una enzima que escinde ácidos nucleicos en los nucleótidos no canónicos (por ejemplo, ADN glicosilasa o RNasaH) para exponer los sitios de unión para cebadores adicionales (por ejemplo, la Patente de Estados Unidos N.º 6.251.639, patente de Estados Unidos n.º 6.946.251 y patente de Estados Unidos n.º 7.824.890). Los procesos de amplificación isotérmica pueden ser lineales o exponenciales.

En algunas realizaciones, la amplificación comprende la amplificación por círculo rodante (RCA). Una mezcla de reacción de RCA típica comprende uno o más cebadores, una polimerasa y dNTP y produce concatémeros. Normalmente, la polimerasa en una reacción de RCA es una polimerasa que tiene actividad de desplazamiento de cadena. Hay disponibles varias tales polimerasas, ejemplos no limitantes de estas incluyen exonucleasa menos Fragmento grande de ADN Polimerasa I (Klenow), ADN polimerasa Phi29, Taq ADN polimerasa y similares. En general, un concatémero es un producto de amplificación de polinucleótido que comprende dos o más copias de una secuencia diana de un polinucleótido plantilla (por ejemplo, aproximadamente de o aproximadamente más de 2, 3, 4, 5, 6, 7, 8, 9, 10 o más copias de la secuencia diana; en algunas realizaciones, aproximadamente o aproximadamente más de 2 copias). Los cebadores de amplificación pueden ser de cualquier longitud adecuada, tal como aproximadamente o al menos aproximadamente 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 90, 100, o más nucleótidos, cualquiera de sus porciones o todas ellas pueden ser complementarias a la secuencia diana correspondiente con la que el cebador hibrida (por ejemplo, aproximadamente, o al menos aproximadamente 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, o más nucleótidos). La FIG. 7 ilustra tres ejemplos no limitantes de cebadores adecuados.

La FIG. 7A muestra el uso de no adaptadores y un cebador específico de diana, que se puede utilizar para la detección de la presencia o ausencia de una variante de secuencia dentro de secuencias diana específicas. En algunas realizaciones, en la misma reacción se utilizan múltiples cebadores específicos de diana para una pluralidad de dianas. Por ejemplo, se pueden usar cebadores específicos de diana para aproximadamente 10, 50, 100, 150, 200, 250, 300, 400, 500, 1000, 2500, 5000, 10000, 15000 o más secuencias diana diferentes en una sola reacción de amplificación para amplificar un número correspondiente de secuencias diana (si están presentes) en paralelo. Las secuencias diana múltiples pueden corresponder a diferentes partes del mismo gen, a genes diferentes o a secuencias no génicas. Cuando múltiples cebadores se dirigen a múltiples secuencias diana en un solo gen, los cebadores pueden estar espaciados a lo largo de la secuencia del gen (por ejemplo, separados por aproximadamente o aproximadamente al menos 50 nucleótidos, cada 50-150 nucleótidos, o cada 50-100 nucleótidos) para cubrir todo o una porción específica de un gen diana. En la Figura 7C se ilustra el uso de un cebador que hibrida con una secuencia adaptadora (que en algunos casos puede ser un oligonucleótido adaptador en sí mismo).

La FIG. 7B ilustra un ejemplo de amplificación por cebadores aleatorios. En general, un cebador aleatorio comprende una o más secuencias aleatorias o casi aleatorias (por ejemplo, uno o más nucleótidos seleccionados al azar de un conjunto de dos o más nucleótidos diferentes en una o más posiciones, con cada uno de los diferentes nucleótidos seleccionados en una o más posiciones representadas en un conjunto de adaptadores que comprenden la secuencia aleatoria). De esta manera, los polinucleótidos (por ejemplo, todos o sustancialmente todos los polinucleótidos circularizados) pueden amplificarse en una secuencia de forma no específica. Dichos procedimientos pueden denominarse "amplificación del genoma completo" (WGA); sin embargo, los protocolos de WGA típicos (que no implican una etapa de circularización) no amplifican de manera eficaz los polinucleótidos cortos, tales como los fragmentos de polinucleótidos contemplados en la presente divulgación. Para un mayor análisis ilustrativo de los procedimientos de WGA, véase, por ejemplo, Li et al (2006) *J Mol. Diagn.* 8(1):22-30.

Cuando los polinucleótidos circularizados se amplifican antes de la secuenciación, los productos amplificados se pueden someter a secuenciación directamente sin enriquecimiento, o después de una o más etapas de enriquecimiento. El enriquecimiento puede comprender la purificación de uno o más componentes de reacción, tal como la retención de productos de amplificación o la eliminación de uno o más reactivos. Por ejemplo, los productos de amplificación pueden purificarse por hibridación a una pluralidad de sondas unidas a un sustrato, seguida de la liberación de polinucleótidos capturados, tal como por ejemplo una etapa de lavado. Como alternativa, los productos de amplificación pueden marcarse con un miembro de un par de unión seguido de la unión al otro miembro del par de unión unido a un sustrato y lavar para liberar el producto de amplificación. Los posibles sustratos incluyen, pero sin limitación, vidrio y vidrio modificado o funcionalizado, plásticos (incluidos acrílicos, poliestireno y copolímeros de estireno y otros materiales, polipropileno, polietileno, polibutileno, poliuretanos, Teflon™, etc.), polisacáridos, nilón y nitrocelulosa, cerámicas, resinas, sílice o materiales a base de sílice, incluidos el silicio y el silicio modificado, carbono, metales, vidrios inorgánicos, plásticos, paquetes de fibra óptica, y varios otros polímeros. En algunas realizaciones, el sustrato está en forma de una perla u otra partícula pequeña, discreta, que puede ser una perla magnética o paramagnética para facilitar el aislamiento mediante la aplicación de un campo magnético. En general, "par de unión" se refiere a uno de un primer y un segundo resto, en el que el primer y el segundo resto tienen una afinidad de unión específica entre sí. Los pares de unión adecuados incluyen, pero sin limitación, antígenos/anticuerpos (por ejemplo, digoxigenina/anti-digoxigenina, dinitrofenil (DNP)/anti-DNP, dansil-X-anti-dansil, fluoresceína/anti-fluoresceína, Lucifer amarillo/anti-lucifer amarillo y rodamina anti-rodamina); biotina/avidina (o biotina/estreptavidina); proteína de unión a la calmodulina (CBP)/calmodulina; hormona/receptor de hormona; lectina/carbohidrato; péptido/receptor de membrana celular; proteína A anticuerpo; hapteno/antihapteno; enzima/cofactor; y enzima/sustrato.

En algunas realizaciones, el enriquecimiento después de la amplificación de polinucleótidos circularizados comprende una o más reacciones de amplificación adicionales. En algunas realizaciones, el enriquecimiento comprende amplificar una secuencia diana que comprende la secuencia A y la secuencia B (orientada en una dirección de 5' a 3') en una mezcla de reacción de amplificación que comprende (a) el polinucleótido amplificado; (b) un primer cebador que comprende la secuencia A', en donde el primer cebador hibrida específicamente con la secuencia A de la secuencia diana mediante complementariedad de secuencia entre la secuencia A y la secuencia A'; (c) un segundo cebador que comprende la secuencia B, en donde el segundo cebador hibrida específicamente con la secuencia B presente en un polinucleótido complementario que comprende un complemento de la secuencia diana mediante complementariedad de secuencia entre B y B'; y (d) una polimerasa que extiende el primer cebador y el segundo cebador para producir polinucleótidos amplificados; en donde la distancia entre el extremo 5' de la secuencia A y el extremo 3' de la secuencia B de la secuencia diana es de 75 nt o menos. La FIG. 10 ilustra una disposición de ejemplo los cebadores primero y segundo con respecto a una secuencia diana en el contexto de una sola repetición (que normalmente no se amplificará a menos que sea circular) y los concatémicos que comprenden múltiples copias de la secuencia diana. Dada la orientación de los cebadores con respecto a un monómero de la secuencia diana, esta disposición puede denominarse cebadores "back to back" (B2B) o "invertidos". La amplificación con cebadores B2B facilita el enriquecimiento de productos de amplificación circular y/o concatémica. Por otra parte, esta orientación combinada con una huella relativamente más pequeña (distancia total abarcada por un par de cebadores) permite la amplificación de una variedad más amplia de eventos de fragmentación alrededor de una secuencia diana, ya que es menos probable que se produzca una unión entre los cebadores que en la disposición de los cebadores encontrados en una reacción de amplificación típica (enfrentadas, que abarca una secuencia diana). En algunas realizaciones, la distancia entre el extremo 5' de la secuencia A y el extremo 3' de la secuencia B es aproximadamente de o aproximadamente menos

de 200, 150, 100, 75, 50, 40, 30, 25, 20, 15 o menos nucleótidos. En algunas realizaciones, la secuencia A es el complemento de la secuencia B. En algunas realizaciones, se usan múltiples pares de cebadores B2B dirigidos a una pluralidad de secuencias diana diferentes en la misma reacción para amplificar una pluralidad de secuencias diana diferentes en paralelo (por ejemplo, aproximadamente o aproximadamente al menos 10, 50, 100, 150, 200, 250, 300, 400, 500, 1000, 2500, 5000, 10000, 15000, o más secuencias diana diferentes). Los cebadores pueden ser de cualquier longitud adecuada, tal como se describe en otra parte en el presente documento. La amplificación puede comprender cualquier reacción de amplificación adecuada en condiciones apropiadas, tal como una reacción de amplificación descrita en el presente documento. En algunas realizaciones, la amplificación es una reacción en cadena de la polimerasa.

En algunas realizaciones, Los cebadores B2B comprenden al menos dos elementos de secuencia, un primer elemento que hibrida con una secuencia diana mediante complementariedad de secuencia y una "cola" 5' que no hibrida con la secuencia diana durante una primera fase de amplificación a una primera temperatura de hibridación durante la cual el primer elemento hibrida (por ejemplo, debido a la falta de complementariedad de secuencia entre la cola y la porción de la secuencia diana inmediatamente 3' con respecto a donde se une el primer elemento). Por ejemplo, el primer cebador comprende la secuencia C 5' con respecto a la secuencia A', el segundo cebador comprende la secuencia D 5' con respecto a la secuencia B, y ni la secuencia C ni la secuencia D hibridan con la pluralidad de concatémicos durante una primera fase de amplificación a una primera temperatura de hibridación. En algunas realizaciones en las que se usan tales cebadores con cola, la amplificación puede comprender una primera fase y una segunda fase; la primera fase comprende una etapa de hibridación a una primera temperatura, durante la cual los cebadores primero y segundo hibridan con los concatémicos (o con los polinucleótidos circularizados) y la extensión de los cebadores; y la segunda fase comprende una etapa de hibridación a una segunda temperatura que es más alta que la primera temperatura, durante la cual los cebadores primero y segundo hibridan con productos de amplificación que comprenden el primer o segundo cebador extendido o complementos de los mismos y la extensión de los cebadores. La temperatura más alta favorece la hibridación a lo largo del primer elemento y el elemento de cola del cebador en productos de extensión del cebador sobre fragmentos más cortos formados por hibridación entre solo el primer elemento en un cebador y una secuencia diana interna dentro de un concatémico. En consecuencia, la amplificación de dos fases puede usarse para reducir el grado en que los productos de amplificación cortos podrían ser favorecidos, manteniendo así una proporción relativamente mayor de productos de amplificación que tienen dos o más copias de una secuencia diana. Por ejemplo, después de 5 ciclos (por ejemplo, al menos 5, 6, 7, 8, 9, 10, 15, 20 o más ciclos) de hibridación a la segunda temperatura y de extensión del cebador, al menos el 5 % (por ejemplo, al menos el 5 %, 6 %, 7 %, 8 %, 9 %, 10 %, 15 %, 20 %, 25 %, al menos el 30 % o más) de los polinucleótidos amplificados en la mezcla de reacción comprenden dos o más copias de la secuencia diana. Una ilustración de una realización de acuerdo con este proceso de amplificación de dos fases de cebador B2B con cola se ilustra en la FIG. 11.

En algunas realizaciones, los enriquecimientos comprenden la amplificación en condiciones que están sesgadas para aumentar la longitud de los amplicones de los concatémicos. Por ejemplo, la concentración del cebador puede reducirse, de modo que no todos los sitios de cebado hibridarán un cebador, lo que hace que los productos de PCR sean más largos. De manera similar, la disminución del tiempo de hibridación del cebador durante los ciclos permitirá, de manera similar, que se hibriden menos cebadores, por lo que también aumenta el tamaño promedio del amplicón de la PCR. Además, aumentar la temperatura y/o el tiempo de extensión de los ciclos aumentará de manera similar la longitud promedio de los amplicones de PCR. Se puede usar cualquier combinación de estas técnicas.

En algunas realizaciones, particularmente cuando se ha realizado una amplificación con cebadores B2B, los productos de amplificación se tratan para filtrar los amplicones resultantes en función del tamaño para reducir y/o eliminar el número de monómeros de una mezcla que comprende concatémicos. Esto se puede realizar utilizando varias técnicas, incluyendo, pero sin limitación, la escisión del fragmento de los geles y la filtración del gel (por ejemplo, para enriquecer los fragmentos mayores de aproximadamente 300, 400, 500 o más nucleótidos de longitud); así como las perlas SPRI (Agencourt AMPure XP) para la selección del tamaño mediante ajuste fino de la concentración del tampón de unión. Por ejemplo, el uso de un tampón de unión de 0,6x durante la mezcla con fragmentos de ADN puede usarse para unir preferentemente fragmentos de ADN mayores de aproximadamente 500 pares de bases (pb).

En algunas realizaciones, donde la amplificación da como resultado concatémicos de cadena sencilla, las hebras sencillas se convierten en construcciones de doble cadena antes o como parte de la formación de bibliotecas de secuenciación que se generan para las reacciones de secuenciación. Se dispone de varios métodos adecuados para generar una construcción de doble cadena a partir de un ácido nucleico de cadena sencilla. En la Figura 9 se representan varios métodos posibles, aunque también se pueden usar otros métodos. Como se muestra en la FIG. 9A, por ejemplo, el uso de cebadores aleatorios, polimerasa, dNTP y una ligasa dará como resultado dobles cadenas. La FIG. 9B representa la síntesis de la segunda cadena cuando el concatémico contiene secuencias adaptadoras, que se pueden usar como los cebadores en la reacción. La FIG. 9C representa el uso de un "bucle," en donde un extremo del adaptador en bucle se agrega al extremo de los concatémicos, en donde el adaptador de bucle tiene una pequeña sección de ácidos nucleicos de auto-hibridación. En este caso, la unión del adaptador en bucle da como resultado el bucle que se auto hibrida y sirve como plantilla de cebador de polimerasa. La FIG. 9D muestra el uso de cebadores hiper-ramificados, en general, es de mayor utilidad en los casos en que se conoce la secuencia diana, donde se forman múltiples cadenas, particularmente cuando se usa una polimerasa con una función de desplazamiento de cadena fuerte.

- De acuerdo con algunas realizaciones, los polinucleótidos circularizados (o sus productos de amplificación, que pueden haber sido enriquecidos opcionalmente) se someten a una reacción de secuenciación para generar lecturas de secuenciación. Las lecturas de secuenciación producidas por tales métodos pueden usarse de acuerdo con otros métodos desvelados en este documento. Hay disponibles varias metodologías de secuenciación, particularmente metodologías de secuenciación de alto rendimiento. Los ejemplos incluyen, sin limitación, sistemas de secuenciación fabricados por Illumina (sistemas de secuenciación tales como HiSeq® y MiSeq®), Life Technologies (Ion Torrent®, SOLiD®, etc.), sistemas 454 Life Sciences de Roche, Sistemas Pacific Biosciences, etc. En algunas realizaciones, la secuenciación comprende el uso de los sistemas HiSeq® y MiSeq® para producir lecturas de aproximadamente o aproximadamente más de 50, 75, 100, 125, 150, 175, 200, 250, 300 o más nucleótidos de longitud. En algunas realizaciones, la secuenciación comprende una secuenciación mediante un proceso de síntesis, donde los nucleótidos individuales se identifican de manera iterativa, a medida que se agregan al producto de extensión de cebador en crecimiento. La pirosecuenciación es un ejemplo de una secuencia mediante un proceso de síntesis que identifica la incorporación de un nucleótido mediante el ensayo de la mezcla de síntesis resultante para detectar la presencia de subproductos de la reacción de secuenciación, es decir, pirofosfato. En particular, un complejo de cebador/plantilla/polimerasa se pone en contacto con un solo tipo de nucleótido. Si se incorpora ese nucleótido, la reacción de polimerización escinde el nucleósido trifosfato entre los fosfatos α y β de la cadena trifosfato, liberando pirofosfato. La presencia de pirofosfato liberado se identifica a continuación, utilizando un sistema indicador de enzimas quimioluminiscentes que convierte el pirofosfato, con AMP, en ATP, a continuación, mide el ATP utilizando una enzima luciferasa para producir señales luminosas medibles. Donde se detecta luz, se incorpora la base, donde no se detecta luz, la base no se incorpora. Después de los pasos de lavado apropiados, las diversas bases se ponen en contacto cíclicamente con el complejo para identificar secuencialmente las bases posteriores en la secuencia plantilla. Véase, por ejemplo, la Patente de EE.UU. n.º 6.210.891.
- En procesos de secuenciación relacionados, el complejo cebador/plantilla/polimerasa se inmoviliza sobre un sustrato y el complejo se pone en contacto con nucleótidos marcados. La inmovilización del complejo puede ser a través de la secuencia del cebador, la secuencia plantilla y/o la enzima polimerasa, y puede ser covalente o no covalente. Por ejemplo, la inmovilización del complejo puede ser a través de un enlace entre la polimerasa o el cebador y la superficie del sustrato. En configuraciones alternativas, los nucleótidos están provistos con y sin grupos terminadores removibles. Tras la incorporación, el marcador se acopla con el complejo y, por lo tanto, es detectable. En el caso de los nucleótidos que portan terminador, los cuatro nucleótidos diferentes, que llevan marcadores identificables individualmente, se ponen en contacto con el complejo. La incorporación de nucleótidos marcados detiene la extensión, en virtud de la presencia del terminador y agrega el marcador al complejo, permitiendo la identificación del nucleótido incorporado. El marcador y el terminador se retiran del nucleótido incorporado a continuación, y siguiendo los pasos de lavado apropiados, se repite el proceso. En el caso de los nucleótidos no terminadores, se agrega un solo tipo de nucleótido marcado al complejo para determinar si se incorporará, como con la pirosecuenciación. Tras la eliminación del grupo marcador en el nucleótido y las etapas de lavado apropiadas, los diversos nucleótidos diferentes se ciclan a través de la mezcla de reacción en el mismo proceso. Véase, por ejemplo, la Patente de EE.UU. n.º 6.833.246. Por ejemplo, el sistema analizador de genoma de Illumina se basa en la tecnología descrita en el documento WO 98/44151, en donde las moléculas de ADN se unen a una plataforma de secuenciación (celda de flujo) a través de un sitio de unión de sonda de anclaje (también denominado sitio de unión de celda de flujo) y se amplifican in situ en un portaobjetos de cristal. Una superficie sólida en la que se amplifican las moléculas de ADN normalmente comprende una pluralidad de oligonucleótidos unidos primero y segundo, el primero complementario a una secuencia cerca o en un extremo de un polinucleótido diana y el segundo complementario a una secuencia cerca o en el otro extremo de un polinucleótido diana. Esta disposición permite la amplificación de puente, tal como se describe en el documento US20140121116. Las moléculas de ADN se emparejan a continuación, en un cebador de secuenciación y se secuencian en paralelo base por base utilizando un enfoque de terminación reversible. La hibridación de un cebador de secuenciación puede estar precedida por la escisión de una hebra de un polinucleótido puente de doble cadena en un sitio de escisión en uno de los oligonucleótidos unidos que anclan el puente, dejando así una hebra única no unida al sustrato sólido que puede eliminarse por desnaturalización y la otra hebra unida y disponible para hibridación a un cebador de secuenciación. Normalmente, El sistema analizador de genoma de Illumina utiliza celdas de flujo con 8 canales, generando lecturas de secuenciación de 18 a 36 bases de longitud, generando > 1.3 Gpb de datos de alta calidad por ejecución (véase www.illumina.com).
- En aún otra secuencia adicional por proceso de síntesis, la incorporación de nucleótidos marcados de manera diferente se observa en tiempo real a medida que se lleva a cabo la síntesis dependiente de plantilla. En particular, se observa un complejo de cebador/plantilla/polimerasa inmovilizado individual cuando se incorporan nucleótidos marcados con fluorescencia, lo que permite la identificación en tiempo real de cada base agregada a medida que se agrega. En este proceso, los grupos marcadores están unidos a una porción del nucleótido que se escinde durante la incorporación. Por ejemplo, al unir el grupo marcador a una porción de la cadena de fosfato eliminada durante la incorporación, es decir, es decir, un β , γ u otro grupo fosfato terminal en un nucleósido polifosfato, el marcador no se incorpora a la cadena naciente, y en su lugar, se produce ADN natural. La observación de moléculas individuales implica normalmente el confinamiento óptico del complejo dentro de un volumen de iluminación muy pequeño. Al confinar ópticamente el complejo, se crea una región monitorizada en la que los nucleótidos de difusión aleatoria están presentes durante un período de tiempo muy corto, mientras que los nucleótidos incorporados se retienen dentro del volumen de observación durante más tiempo a medida que se incorporan. Esto da como resultado una señal

característica asociada con el evento de incorporación, que también se caracteriza por un perfil de señal que es característico de la base que se agrega. En aspectos relacionados, se proporcionan componentes marcadores interactivos, como los pares de tinte de transferencia de energía resonante fluorescente (FRET), sobre la polimerasa u otra porción del complejo y el nucleótido incorporador, de modo que el evento de incorporación pone los componentes de marcado en proximidad interactiva y da como resultado una señal característica, que es de nuevo, también característica de la base que se incorpora (Véase, por ejemplo, las patentes de los EE.UU. n.º 6.917.726, 7.033.764, 7.052.847, 7.056.676, 7.170.050, 7.361.466 y 7.416.844; y el documento US 20070134128).

En algunas realizaciones, los ácidos nucleicos en la muestra pueden ser secuenciados por unión. Este método generalmente utiliza una enzima ADN ligasa para identificar la secuencia diana, por ejemplo, como se usa en el método de polonía y en la tecnología SOLiD (Applied Biosystems, ahora Invitrogen). En general, se proporciona un conjunto de todos los oligonucleótidos posibles de una longitud fija, marcados de acuerdo con la posición secuenciada. Los oligonucleótidos son emparejados y unidos; la unión preferencial por ADN ligasa para secuencias coincidentes da como resultado una señal correspondiente a la secuencia complementaria en esa posición.

De acuerdo con algunas realizaciones, una diferencia de secuencia entre las lecturas de secuenciación y una secuencia de referencia se llama variante de secuencia genuina (por ejemplo, que existe en la muestra antes de la amplificación o secuenciación, y no es el resultado de ninguno de estos procesos) si sucede en al menos dos polinucleótidos diferentes (por ejemplo, dos polinucleótidos circulares diferentes, que pueden distinguirse como resultado de tener uniones diferentes). Debido a que es poco probable que las variantes de secuencia que son el resultado de errores de amplificación o secuenciación se dupliquen de manera exacta (por ejemplo, posición y tipo) en dos polinucleótidos diferentes que comprenden la misma secuencia diana, la adición de este parámetro de validación reduce en gran medida el fondo de variantes de secuencia erróneas, con un aumento concurrente de la sensibilidad y la precisión de la detección de la variación de secuencia real en una muestra. En algunas realizaciones, una variante de secuencia tiene una frecuencia de aproximadamente el o aproximadamente menos del 5 %, 4 %, 3 %, 2 %, 1,5 %, 1 %, 0,75 %, 0,5 %, 0,25 %, 0,1 %, 0,075 %, 0,05 %, 0,04 %, 0,03 %, 0,02 %, 0,01 %, 0,005 %, 0,001 %, o menos, está suficientemente por encima del fondo para permitir una identificación precisa. En algunas realizaciones, la variante de secuencia se produce con una frecuencia de aproximadamente el o aproximadamente menos del 0,1 %. En algunas realizaciones, la frecuencia de una variante de secuencia está suficientemente por encima del fondo cuando dicha frecuencia está estadísticamente muy por encima de la tasa de error de fondo (por ejemplo, con un valor de p de aproximadamente o aproximadamente menos de 0,05, 0,01, 0,001, 0,0001 o inferior). En algunas realizaciones, la frecuencia de una variante de secuencia está suficientemente por encima del fondo cuando dicha frecuencia está aproximadamente o al menos aproximadamente 2 veces, 3 veces, 4 veces, 5 veces, 6 veces, 7 veces, 8 veces, 9 veces, 10 veces, 25 veces, 50 veces, 100 veces, o más por encima de la tasa de error de fondo (por ejemplo, al menos 5 veces más). En algunas realizaciones, la tasa de error de fondo para determinar con precisión la secuencia en una posición dada es aproximadamente o aproximadamente inferior al 1 %, 0,5 %, 0,1 %, 0,05 %, 0,01 %, 0,005 %, 0,001 %, 0,0005 % o menos. En algunas realizaciones, la tasa de error es inferior al 0,001 %.

En algunas realizaciones, la identificación de una variante de secuencia genuina (también denominada "identificación" o "hacer una identificación") comprende alinear de manera óptima una o más lecturas de secuencia con una secuencia de referencia para identificar diferencias entre las dos, así como para identificar uniones. En general, la alineación implica colocar una secuencia a lo largo de otra secuencia, introducir espacios de manera iterativa a lo largo de cada secuencia, puntuar cómo de bien coinciden las dos secuencias y, preferentemente, repetir para varias posiciones a lo largo de la referencia. La coincidencia de mejor puntuación se considera la alineación y representa una inferencia sobre el grado de relación entre las secuencias. En algunas realizaciones, una secuencia de referencia con la que se comparan las lecturas de secuenciación es un genoma de referencia, como el genoma de un miembro de la misma especie que el sujeto. El genoma de referencia puede ser completo o incompleto. En algunas realizaciones, un genoma de referencia consiste solo en regiones que contienen polinucleótidos diana, tal como por un genoma de referencia o de un consenso generado a partir de lecturas de secuenciación en análisis. En algunas realizaciones, una secuencia de referencia comprende o consiste en secuencias de polinucleótidos de uno o más organismos, tales como secuencias de una o más bacterias, arqueas, virus, protistas, hongos u otro organismo. En algunas realizaciones, la secuencia de referencia consiste en solo una parte de un genoma de referencia, tal como las regiones correspondientes a una o más secuencias diana en análisis (por ejemplo, uno o más genes, o porciones de los mismos). Por ejemplo, para la detección de un patógeno (tal como en el caso de detección de contaminación), el genoma de referencia es el genoma completo del patógeno (por ejemplo, VIH, VPH o una cepa bacteriana dañina, por ejemplo, *E. coli*), o una porción del mismo útil en la identificación, tal como de una cepa o serotipo particular. En algunas realizaciones, Las lecturas de secuenciación se alinean con múltiples secuencias de referencia diferentes, tal como para detectar múltiples organismos o cepas diferentes.

En una alineación típica, una base en una lectura de secuenciación junto con una base no coincidente en la referencia indica que se ha producido una mutación por sustitución en ese punto. De manera similar, cuando una secuencia incluye un espacio junto a una base en la otra secuencia, se infiere que se produce una mutación por inserción o delección (una "indel"). Cuando se desea especificar que una secuencia se está alineando entre sí, la alineación a veces se denomina alineación por pares. La alineación de secuencias múltiples generalmente se refiere a la alineación de dos o más secuencias, incluyendo, por ejemplo, por una serie de alineaciones por pares. En algunas realizaciones, puntuar una alineación implica establecer valores para las probabilidades de sustituciones e indeles. Cuando las bases

individuales están alineadas, una coincidencia o falta de coincidencia contribuye a la puntuación de alineación mediante una probabilidad de sustitución, que podría ser, por ejemplo, 1 para una coincidencia y 0,33 para una falta de coincidencia. Una indel deduce de una puntuación de alineación por una penalización de hueco, que podría ser, por ejemplo, -1. Las penalizaciones por hueco y las probabilidades de sustitución pueden basarse en el conocimiento empírico o en suposiciones a priori sobre cómo mutan las secuencias. Sus valores afectan la alineación resultante. Ejemplos de algoritmos para realizar alineaciones incluyen, sin limitación, el algoritmo Smith-Waterman (SW), el algoritmo Needleman-Wunsch (NW), algoritmos basados en la transformación de Burrows-Wheeler (BWT) y alineadores de funciones hash tal como Novoalign (Novocraft Technologies; disponible en www.novocraft.com), ELAND (Illumina, San Diego, Calif.), SOAP (disponible en soap.genomics.org.cn), y Maq (disponible en maq.sourceforge.net). Un programa de alineación a modo de ejemplo, que implementa un enfoque BWT, es Burrows-Wheeler Aligner (BWA) disponible en el sitio web de SourceForge mantenido por Geeknet (Fairfax, Virginia). BWT ocupa normalmente 2 bits de memoria por nucleótido, lo que hace posible indexar secuencias de nucleótidos tan larga como 4G pares de bases con un ordenador de sobremesa o portátil común. El procesamiento previo incluye la construcción de BWT (es decir, la indexación de la referencia) y las estructuras de datos auxiliares de soporte. BWA incluye dos algoritmos diferentes, ambos basados en BWT. La alineación por BWA puede continuar usando el algoritmo bwa-short, diseñado para consultas cortas de hasta aproximadamente 200 con una tasa de error baja (<3%) (Li H. y Durbin R. *Bioinformatics*, 25: 1754-60 (2009)). El segundo algoritmo, BWA-SW, está diseñado para lecturas largas con más errores (Li H. y Durbin R. (2010). La alineación de lectura larga rápida y precisa con transformación de Burrows-Wheeler. *Bioinformatics*, Epub.). El alineador bwa-sw a veces se denomina "bwa-long", "Algoritmo bwa long", o similar. Un programa de alineación que implementa una versión del algoritmo Smith-Waterman es MUMmer, disponible en el sitio web de SourceForge mantenido por Geeknet (Fairfax, Virginia). MUMmer es un sistema para alinear rápidamente genomas completos, ya sea de forma completa o en borrador (Kurtz, S., et al., *Genome Biology*, 5:R12 (2004); Delcher, A. L., et al., *Nucl. Acids Res.*, 27:11 (1999)). Por ejemplo, MUMmer 3.0 puede encontrar todas las coincidencias exactas de 20 pares de bases o más largas entre un par de genomas de 5 megabases en 13,7 segundos, utilizando 78 MB de memoria, en un ordenador de sobremesa Linux de 2,4 GHz. MUMmer también puede alinear genomas incompletos; puede manejar fácilmente los cientos o miles de contigs de un proyecto de secuenciación de escopeta y los alinearán con otro conjunto de contigs o un genoma usando el programa NUCmer incluido con el sistema. Otros ejemplos no limitantes de programas de alineación incluyen: BLAT de Kent Informatics (Santa Cruz, Calif.) (Kent, W. J., *Genome Research* 4: 656-664 (2002)); SOAP2, de Beijing Genomics Institute (Beijing, Conn.) o BGI Americas Corporation (Cambridge, Mass.); Bowtie (Langmead, et al., *Genome Biology*, 10:R25 (2009)); La alineación eficiente a gran escala de bases de datos de nucleótidos (ELAND) o el componente ELANDv2 del software de evaluación de consenso de secuencia y variación (CASAVA) (Illumina, San Diego, Calif.); Investigador RTG de Real Time Genomics, Inc. (San Diego, Calif.); Novoalign de Novocraft (Selangor, Malaysia); Exonerate, European Bioinformatics Institute (Hinxton, UK) (Slater, G., y Birney, E., *BMC Bioinformatics* 6:31 (2005)), Clustal Omega, del University College Dublin (Dublin, Irlanda) (Sievers F., et al., *Mol Syst Biol* 7, artículo 539 (2011)); ClustalW o ClustalX del University College Dublin (Dublin, Irlanda) (Larkin M. A., et al., *Bioinformatics*, 23, 2947-2948 (2007)); y FASTA, Instituto Europeo de Bioinformática (Hinxton, Reino Unido) (Pearson W. R., et al., *PNAS* 85(8):2444-8 (1988); Lipman, D. J., *Science* 227(4693): 1435-41 (1985)).

Normalmente, los datos de secuenciación se adquieren a gran escala, reacciones de secuenciación paralelas. Muchos de los sistemas de secuenciación de alto rendimiento de próxima generación exportan datos como archivos FASTQ, aunque se pueden usar otros formatos. En algunas realizaciones, las secuencias se analizan para identificar la longitud de la unidad de repetición (por ejemplo, la longitud del monómero), la unión formada por circularización y cualquier variación verdadera con respecto a una secuencia de referencia, normalmente a través de la alineación de la secuencia. La identificación de la longitud de la unidad de repetición puede incluir el cálculo de las regiones de las unidades repetidas, la búsqueda de los loci de referencia de las secuencias (por ejemplo, cuando una o más secuencias se dirigen particularmente a la amplificación, enriquecimiento y/o secuenciación), los límites de cada región repetida y/o el número de repeticiones dentro de cada ejecución de secuenciación. El análisis de secuencia puede incluir el análisis de datos de secuencia para ambas cadenas de un dúplex. Como se ha señalado anteriormente, en algunas realizaciones, una variante idéntica que aparece en las secuencias de lecturas de diferentes polinucleótidos de la muestra (por ejemplo, polinucleótidos circularizados que tienen diferentes uniones) se considera una variante confirmada. En algunas realizaciones, una variante de secuencia también se puede considerar una variante confirmada o genuina si se produce en más de una unidad repetida del mismo polinucleótido, ya que es poco probable que la misma variación de secuencia se produzca en la misma posición en una secuencia diana repetida dentro del mismo concatémoro. La puntuación de calidad de una secuencia se puede considerar al identificar variantes y variantes confirmadas, por ejemplo, se pueden filtrar la secuencia y las bases con puntuaciones de calidad inferiores a un umbral. Se pueden usar otros métodos de bioinformática para aumentar aún más la sensibilidad y especificidad de las identificaciones de variantes.

En algunas realizaciones, los análisis estadísticos pueden aplicarse a la determinación de variantes (mutaciones) y cuantificar la proporción de la variante en muestras de ADN total. La medición total de una base particular se puede calcular utilizando los datos de secuenciación. Por ejemplo, a partir de los resultados de alineación calculados en las etapas anteriores, se puede calcular el número de "lecturas efectivas", es decir, el número de lecturas confirmadas para cada locus. La frecuencia alélica de una variante se puede normalizar mediante el recuento de lecturas efectivas para el locus. Se puede calcular el nivel de ruido general, que es la tasa promedio de las variantes observadas en todos los loci. La frecuencia de una variante y el nivel de ruido general, combinados con otros factores, pueden

utilizarse para determinar el intervalo de confianza de la identificación de la variante. Los modelos estadísticos, como las distribuciones de Poisson, se pueden utilizar para evaluar el intervalo de confianza de las identificaciones de variantes. La frecuencia alélica de las variantes también se puede utilizar como un indicador de la cantidad relativa de la variante en la muestra total.

En algunas realizaciones, un contaminante microbiano se identifica basándose en la etapa de identificación. Por ejemplo, una variante de secuencia particular puede indicar contaminación por un microbio potencialmente infeccioso. Las variantes de secuencia pueden identificarse dentro de un polinucleótido altamente conservado con el propósito de identificar un microbio. Ejemplos de polinucleótidos altamente conservados útiles en la caracterización filogenética y la identificación de microbios comprenden secuencias de nucleótidos encontradas en el gen de ARNr 16S, gen de ARNr 23S, gen de ARNr 5S, gen de ARNr 5.8S, gen de ARNr 12S, gen de ARNr 18S, gen de ARNr 28S, gen *gyrB*, gen *rpoB*, gen *fusA*, gen *recA*, gen *coxI* y gen *nifD*. Con los eucariotas, el gen de ARNr puede ser nuclear, mitocondrial o ambos. En algunas realizaciones, las variantes de secuencia en el espaciador transcrito interno (ITS) del gen ARNr 16S-23S se pueden usar para la diferenciación e identificación de taxones estrechamente relacionados con o sin el uso de otros genes ARNr. Debido a las restricciones estructurales del ARNr 16S, las regiones específicas de todo el gen tienen una secuencia polinucleotídica altamente conservada, aunque los segmentos no estructurales pueden tener un alto grado de variabilidad. La identificación de variantes de secuencia se puede usar para identificar unidades taxonómicas operativas (OTU) que representan un subgénero, un género, una subfamilia, una familia, un suborden, un orden, una subclase, una clase, un subfilo, un filo, un subreino, o un reino, y opcionalmente determinan su frecuencia en una población. La detección de variantes de secuencia particulares se puede usar en la detección de presencia y opcionalmente, de cantidad (relativa o absoluta) de un microbio indicativo de contaminación. Las aplicaciones de ejemplo incluyen pruebas de calidad del agua para determinar la contaminación fecal u otra, pruebas para patógenos animales o humanos, localización de fuentes de contaminación del agua, pruebas de agua recuperada o reciclada, pruebas de corrientes de descarga de aguas residuales que incluyen plumas de descarga oceánica, monitorización de instalaciones acuícolas para detectar patógenos, monitorización de playas, áreas de baño u otras instalaciones recreativas relacionadas con el agua y predecir la proliferación de algas tóxicas. Las aplicaciones de monitorización de alimentos incluyen las pruebas periódicas de las líneas de producción en las plantas de procesamiento de alimentos, el estudio de mataderos, la inspección de las cocinas y las áreas de almacenamiento de alimentos de restaurantes, hospitales, escuelas, instalaciones correccionales y otras instituciones para patógenos transmitidos por los alimentos, como las cepas O157:H7 o O111:B4 de *E. coli*, *Listeria monocytogenes* o *Salmonella enterica* subsp. *enterica* serovar *Enteritidis*. Los mariscos y las aguas productoras de mariscos pueden examinarse en busca de algas responsables de la intoxicación paralítica por mariscos, la intoxicación neurotóxica por mariscos, la intoxicación diarreica por mariscos y la intoxicación amnésica por mariscos. Adicionalmente, los alimentos importados se pueden analizar en la aduana antes de liberarlos para garantizar la seguridad alimentaria. Las aplicaciones de monitorización de patógenos de plantas incluyen horticultura y monitorización de viveros, por ejemplo, la monitorización de *Phytophthora ramorum*, el microorganismo responsable de la muerte súbita del roble, la vigilancia de patógenos de cultivos y el manejo de enfermedades y la vigilancia de patógenos forestales y el manejo de enfermedades. Los entornos de fabricación de productos farmacéuticos, dispositivos médicos y otros consumibles o componentes críticos en los que la contaminación microbiana es un problema importante de seguridad pueden estudiarse para detectar la presencia de patógenos específicos como *Pseudomonas aeruginosa* o *Staphylococcus aureus*, la presencia de microorganismos más comunes asociados con humanos, microorganismos asociados con la presencia de agua u otros que representan la carga biológica que se identificó previamente en ese entorno en particular o en otros similares. De manera similar, las áreas de construcción y ensamblaje para equipos sensibles, incluidas las naves espaciales, se pueden monitorizar para detectar microorganismos previamente identificados que habitan o que se introducen más comúnmente en dichos entornos.

En un aspecto, la invención proporciona un método para identificar una variante de secuencia en una muestra de ácido nucleico que comprende menos de 50 ng de polinucleótidos, teniendo cada polinucleótido un extremo 5' y un extremo 3'. En algunas realizaciones, el método comprende: (a) circularizar con una ligasa polinucleótidos individuales en dicha muestra para formar una pluralidad de polinucleótidos circulares; (b) después de separar dicha ligasa de dichos polinucleótidos circulares, amplificar los polinucleótidos circulares para formar concatémeros; (c) secuenciar los concatémeros para producir una pluralidad de lecturas de secuenciación; (d) identificar las diferencias de secuencia entre la pluralidad de lecturas de secuenciación y una secuencia de referencia; y (e) identificar una diferencia de secuencia que se produce con una frecuencia de 0,05% o más en dicha pluralidad de lecturas de dicha muestra de ácido nucleico de menos de 50 ng de polinucleótidos como la variante de secuencia.

La cantidad inicial de polinucleótidos en una muestra puede ser pequeña. En algunas realizaciones, la cantidad de polinucleótidos de partida es menor que 50 ng, tal como menos de 45 ng, 40 ng, 35 ng, 30 ng, 25 ng, 20 ng, 15 ng, 10 ng, 5 ng, 4 ng, 3 ng, 2 ng, 1 ng, 0,5 ng, 0,1 ng, o menos. En algunas realizaciones, la cantidad eficaz de polinucleótidos de partida se encuentra en el intervalo de aproximadamente 0,1-100 ng, tal como entre 1-75 ng, 5-50 ng, o 10 - 20 ng. En general, un material de partida inferior aumenta la importancia de aumentar la recuperación de varias etapas de procesamiento. Los procesos que reducen la cantidad de polinucleótidos en una muestra para participar en una reacción posterior disminuyen la sensibilidad con la que se pueden detectar mutaciones raras. Por ejemplo, tal como se describe por Lou et al. (PNAS, 2013, 110 (49)) se espera recuperar solo el 10-20% del material de partida. Para grandes cantidades de material de partida (por ejemplo, como bacterias cultivadas en el laboratorio purificadas), esto puede no ser un obstáculo importante. Sin embargo, para muestras donde el material de partida es sustancialmente

más bajo, la recuperación en este intervalo bajo puede ser un obstáculo sustancial para la detección de variantes lo suficientemente raras. En consecuencia, en algunas realizaciones, la recuperación de la muestra de una etapa a otra en un método de la divulgación (por ejemplo, la fracción de masa de la entrada en una etapa de circularización disponible para la entrada en una etapa de amplificación o etapa de secuenciación posterior) es aproximadamente o más de aproximadamente el 50%, 60 %, 75 %, 80 %, 85 %, 90 %, el 95 % o más. La recuperación de una etapa particular puede estar cerca del 100%. La recuperación puede ser con respecto a una forma particular, como la recuperación de polinucleótidos circulares de una entrada de polinucleótidos no circulares.

Los polinucleótidos pueden ser de cualquier muestra adecuada, tal como una muestra descrita en el presente documento con respecto a los diversos aspectos de la divulgación. Los polinucleótidos de una muestra pueden ser cualquiera de varios polinucleótidos, incluyendo, pero sin limitación, ADN, ARN, ARN ribosómico (ARNr), ARN de transferencia (ARNt), microARN (miARN), ARN mensajero (ARNm), fragmentos de cualquiera de estos o combinaciones de cualquiera de dos o más de estos. En algunas realizaciones, las muestras comprenden ADN. En algunas realizaciones, los polinucleótidos son de cadena sencilla, ya sea como se obtienen o a través de tratamiento (por ejemplo, desnaturalización). Se describen en el presente documento ejemplos adicionales de polinucleótidos adecuados, tal como con respecto a cualquiera de los diversos aspectos de la divulgación. En algunas realizaciones, los polinucleótidos se someten a etapas posteriores (por ejemplo, circularización y amplificación) sin una etapa de extracción y/o sin una etapa de purificación. Por ejemplo, una muestra de fluido se puede tratar para eliminar las células sin una etapa de extracción para producir una muestra de líquido purificado y una muestra de células, seguido del aislamiento del ADN de la muestra de fluido purificado. Se dispone de varios procedimientos para el aislamiento de polinucleótidos, tales como por precipitación o unión no específica a un sustrato seguido de lavado del sustrato para liberar polinucleótidos unidos. Cuando los polinucleótidos se aíslan de una muestra sin una etapa de extracción celular, los polinucleótidos serán en gran medida polinucleótidos extracelulares o "libres de células", que pueden corresponder a células muertas o dañadas. La identidad de dichas células se puede usar para caracterizar las células o la población de células de las que derivan, tal como en una comunidad microbiana. Si se trata una muestra para extraer polinucleótidos, tal como a partir de células en una muestra, hay varios métodos de extracción disponibles, ejemplos de los cuales se proporcionan en este documento (por ejemplo, con respecto a cualquiera de los diversos aspectos de la divulgación).

La variante de secuencia en la muestra de ácido nucleico puede ser cualquiera de varias variantes de secuencia. Los ejemplos no limitantes de variantes de secuencia se describen a continuación, tal como con respecto a cualquiera de los diversos aspectos de la divulgación. En algunas realizaciones la variante de secuencia es un polimorfismo de un solo nucleótido (SNP). En algunas realizaciones, la variante de secuencia se produce con una frecuencia baja en la población (también denominada variante de secuencia "rara"). Por ejemplo, la variante de secuencia se puede producir con una frecuencia de aproximadamente el o aproximadamente menos del 5%, 4 %, 3 %, 2 %, 1,5 %, 1 %, 0,75 %, 0,5 %, 0,25 %, 0,1 %, 0,075 %, 0,05 %, 0,04 %, 0,03 %, 0,02 %, 0,01 %, 0,005 %, 0,001 % o menos. En algunas realizaciones, la variante de secuencia se produce con una frecuencia de aproximadamente el o aproximadamente menos del 0,1%.

De acuerdo con algunas realizaciones, los polinucleótidos de una muestra se circularizan, tal como mediante el uso de una ligasa. La circularización puede incluir unir el extremo 5' de un polinucleótido al extremo 3' del mismo polinucleótido, al extremo 3' de otro polinucleótido en la muestra, o al extremo 3' de un polinucleótido de una fuente diferente (por ejemplo, un polinucleótido, tal como un adaptador de oligonucleótido). En algunas realizaciones, el extremo 5' de un polinucleótido se une al extremo 3' del mismo polinucleótido (también denominado "autounión"). Ejemplos no limitantes de procesos de circularización (por ejemplo, con y sin oligonucleótidos adaptadores), reactivos (por ejemplo, tipos de adaptadores, uso de ligasas), condiciones de reacción (por ejemplo, favoreciendo la autounión) y procesamiento adicional opcional (por ejemplo, purificación posterior a la reacción) se proporcionan en el presente documento, tal como con respecto a cualquiera de los diversos aspectos de la divulgación.

En general, la unión de los extremos de un polinucleótido entre sí para formar un polinucleótido circular (ya sea directamente o con uno o más oligonucleótidos adaptadores intermedios) produce una unión que tiene una secuencia de unión. Cuando el extremo 5' y el extremo 3' de un polinucleótido se unen a través de un polinucleótido adaptador, el término "unión" puede referirse a una unión entre el polinucleótido y el adaptador (por ejemplo, uno de la unión del extremo 5' o de la unión del extremo 3'), o a la unión entre el extremo 5' y el extremo 3' del polinucleótido formado por e incluyendo el polinucleótido adaptador. Cuando el extremo 5' y el extremo 3' de un polinucleótido se unen sin un adaptador intermedio (por ejemplo, el extremo 5' y el extremo 3' de un ADN de cadena sencilla), el término "unión" se refiere al punto en el que estos dos extremos están unidos. Una unión puede identificarse por la secuencia de nucleótidos que comprende la unión (también denominada "secuencia de unión"). En algunas realizaciones, las muestras comprenden polinucleótidos que tienen una mezcla de extremos formados por fragmentación de procesos de degradación natural (como la lisis celular, la muerte celular y otros procesos mediante los cuales el ADN se libera de una célula a su entorno circundante en el que puede degradarse aún más, tal como en los polinucleótidos libres de células) que es un subproducto del procesamiento de la muestra (como los procedimientos de fijación, tinción y/o almacenamiento), y la fragmentación por métodos que escinden el ADN sin restricción a secuencias diana específicas (p. ej., fragmentación mecánica, como por sonicación; tratamiento con nucleasa no específica de secuencia, tal como la DNasa I, fragmentasa). Cuando las muestras comprenden polinucleótidos que tienen una mezcla de extremos, la probabilidad de que dos polinucleótidos tengan el mismo extremo 5' o el mismo extremo 3' es baja, y la probabilidad

de que dos polinucleótidos tengan independientemente tanto el mismo extremo 5 'como el mismo extremo 3' es extremadamente baja. En consecuencia, en algunas realizaciones, las uniones pueden usarse para distinguir diferentes polinucleótidos, incluso cuando los dos polinucleótidos comprenden una porción que tiene la misma secuencia diana. Cuando los extremos de los polinucleótidos se unen sin un adaptador intermedio, se puede identificar una secuencia de unión mediante el alineamiento con una secuencia de referencia. Por ejemplo, donde el orden de las dos secuencias de componentes parece invertirse con respecto a la secuencia de referencia, el punto en el que parece ocurrir la reversión puede ser una indicación de una unión en ese punto. Cuando los extremos de los polinucleótidos se unen a través de una o más secuencias de adaptadores, se puede identificar una unión por la proximidad a la secuencia de adaptador conocida, o por el alineamiento como anteriormente, si una lectura de secuenciación tiene la longitud suficiente para obtener la secuencia de los extremos tanto 5 'como 3' del polinucleótido circularizado. En algunas realizaciones, la formación de una unión particular es un evento lo suficientemente raro como para que sea único entre los polinucleótidos circularizados de una muestra.

Después de la circularización, los productos de reacción se pueden purificar antes de la amplificación o secuenciación para aumentar la concentración o pureza relativa de los polinucleótidos circularizados disponibles para participar en las etapas posteriores (por ejemplo, mediante el aislamiento de polinucleótidos circulares o la eliminación de una o más moléculas en la reacción). Por ejemplo, una reacción de circularización o componentes de la misma pueden tratarse para eliminar polinucleótidos de cadena sencilla (no circularizados), tal como por tratamiento con una exonucleasa. Como ejemplo adicional, una reacción de circularización o parte de ella puede someterse a cromatografía de exclusión por tamaño, por lo que los reactivos pequeños se retienen y descartan (por ejemplo, adaptadores sin reaccionar), o los productos de circularización se retienen y liberan en un volumen separado. Hay disponibles varios kits para limpiar las reacciones de unión, tales como los kits proporcionados por los kits de purificación de oligo Zymo fabricados por Zymo Research. En algunas realizaciones, la purificación comprende un tratamiento para eliminar o degradar la ligasa usada en la reacción de circularización y/o para purificar polinucleótidos circularizados lejos de dicha ligasa. En algunas realizaciones, el tratamiento para degradar la ligasa comprende el tratamiento con una proteasa. Las proteasas adecuadas están disponibles en procariotas, virus y eucariotas. Los ejemplos de proteasas incluyen proteinasa K (de *Tritirachium album*), pronasa E (de *Streptomyces griseus*), proteasa de *Bacillus polymyxa*, thermolysin (de bacterias termofílicas), tripsina, subtilisina, furina y similares. En algunas realizaciones, la proteasa es proteinasa K. El tratamiento con proteasa puede seguir los protocolos del fabricante o someterse a los protocolos convencionales (por ejemplo, según se proporciona en Sambrook y Green, *Molecular Cloning: A Laboratory Manual*, 4ª edición (2012)). El tratamiento con proteasa también puede seguirse por extracción y precipitación. En un ejemplo, los polinucleótidos circularizados se purifican mediante tratamiento con proteinasa K (Qiagen) en presencia de SDS al 0,1% y EDTA 20 mM, se extraen con fenol/cloroformo 1:1 y cloroformo y se precipitan con etanol o isopropanol. En algunas realizaciones, la precipitación es en etanol.

Como se describe con respecto a otros aspectos de la divulgación, la circularización puede seguirse directamente secuenciando los polinucleótidos circularizados. Como alternativa, la secuenciación puede estar precedida por una o más reacciones de amplificación. Hay disponibles varios métodos para amplificar polinucleótidos (por ejemplo, ADN y/o ARN). La amplificación puede ser lineal, exponencial o involucrar fases tanto lineales como exponenciales en un proceso de amplificación multifase. Los métodos de amplificación pueden implicar cambios en la temperatura, tal como una etapa de desnaturalización por calor, o pueden ser procesos isotérmicos que no requieren desnaturalización por calor. Los ejemplos no limitantes de procesos de amplificación adecuados se describen, tal como con respecto a cualquiera de los diversos aspectos de la divulgación. En algunas realizaciones, la amplificación comprende la amplificación por círculo rodante (RCA). Tal como se describe en otra parte en el presente documento, una mezcla de reacción de RCA típica comprende uno o más cebadores, una polimerasa y dNTP y produce concatémeros. Normalmente, la polimerasa en una reacción de RCA es una polimerasa que tiene actividad de desplazamiento de cadena. Hay disponibles varias tales polimerasas, ejemplos no limitantes de estas incluyen exonucleasa menos Fragmento grande de ADN Polimerasa I (Klenow), ADN polimerasa Phi29, Taq ADN polimerasa y similares. En general, un concatémero es un producto de amplificación de polinucleótido que comprende dos o más copias de una secuencia diana de un polinucleótido plantilla (por ejemplo, aproximadamente de o aproximadamente más de 2, 3, 4, 5, 6, 7, 8, 9, 10 o más copias de la secuencia diana; en algunas realizaciones, aproximadamente o aproximadamente más de 2 copias). Los cebadores de amplificación pueden ser de cualquier longitud adecuada, tal como aproximadamente o al menos aproximadamente 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 90, 100, o más nucleótidos, cualquiera de sus porciones o todas ellas pueden ser complementarias a la secuencia diana correspondiente con la que el cebador hibrida (por ejemplo, aproximadamente, o al menos aproximadamente 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, o más nucleótidos). En el presente documento se describen ejemplos de diversos procesos de RCA, como el uso de cebadores aleatorios, cebadores específicos de diana y cebadores dirigidos por adaptador, algunos de los cuales se ilustran en la FIG. 7.

Cuando los polinucleótidos circularizados se amplifican antes de la secuenciación (p.ej. para producir concatémeros), los productos amplificados se pueden someter a secuenciación directamente sin enriquecimiento, o después de una o más etapas de enriquecimiento. Los ejemplos no limitantes de procesos de enriquecimiento adecuados se describen en el presente documento, tal como con respecto a cualquiera de los diversos aspectos de la divulgación (por ejemplo, el uso de cebadores B2B para una segunda etapa de amplificación). De acuerdo con algunas realizaciones, los polinucleótidos circularizados (o sus productos de amplificación, que pueden haber sido enriquecidos opcionalmente) se someten a una reacción de secuenciación para generar lecturas de secuenciación. Las lecturas de secuenciación

producidas por tales métodos pueden usarse de acuerdo con otros métodos desvelados en este documento. Hay disponibles varias metodologías de secuenciación, particularmente metodologías de secuenciación de alto rendimiento. Los ejemplos incluyen, sin limitación, sistemas de secuenciación fabricados por Illumina (sistemas de secuenciación tales como HiSeq® y MiSeq®), Life Technologies (Ion Torrent®, SOLiD®, etc.), sistemas 454 Life Sciences de Roche, Sistemas Pacific Biosciences, etc. En algunas realizaciones, la secuenciación comprende el uso de los sistemas HiSeq® y MiSeq® para producir lecturas de aproximadamente o aproximadamente más de 50, 75, 100, 125, 150, 175, 200, 250, 300 o más nucleótidos de longitud. Se describen en el presente documento ejemplos adicionales no limitantes de plataformas y metodologías de amplificación, tal como con respecto a cualquiera de los diversos aspectos de la divulgación.

De acuerdo con algunas realizaciones, una diferencia de secuencia entre las lecturas de secuenciación y una secuencia de referencia se llama variante de secuencia genuina (por ejemplo, que existe en la muestra antes de la amplificación o secuenciación, y no es el resultado de ninguno de estos procesos) si sucede en al menos dos polinucleótidos diferentes (por ejemplo, dos polinucleótidos circulares diferentes, que pueden distinguirse como resultado de tener uniones diferentes). Debido a que es poco probable que las variantes de secuencia que son el resultado de errores de amplificación o secuenciación se dupliquen de manera exacta (por ejemplo, posición y tipo) en dos polinucleótidos diferentes que comprenden la misma secuencia diana, la adición de este parámetro de validación reduce en gran medida el fondo de variantes de secuencia erróneas, con un aumento concurrente de la sensibilidad y la precisión de la detección de la variación de secuencia real en una muestra. En algunas realizaciones, una variante de secuencia tiene una frecuencia de aproximadamente el o aproximadamente menos del 5 %, 4 %, 3 %, 2 %, 1,5 %, 1 %, 0,75 %, 0,5 %, 0,25 %, 0,1 %, 0,075 %, 0,05 %, 0,04 %, 0,03 %, 0,02 %, 0,01 %, 0,005 %, 0,001%, o menos, está suficientemente por encima del fondo para permitir una identificación precisa. En algunas realizaciones, la variante de secuencia se produce con una frecuencia de aproximadamente el o aproximadamente menos del 0,1%. En algunas realizaciones, el método comprende identificar como una variante de secuencia genuina, las diferencias de secuencia que tienen una frecuencia en el intervalo de aproximadamente 0,0005% a aproximadamente 3%, tal como entre 0,001% -2%, o 0,01% -1%. En algunas realizaciones, la frecuencia de una variante de secuencia está suficientemente por encima del fondo cuando dicha frecuencia está estadísticamente muy por encima de la tasa de error de fondo (por ejemplo, con un valor de p de aproximadamente o aproximadamente menos de 0,05, 0,01, 0,001, 0,0001 o inferior). En algunas realizaciones, la frecuencia de una variante de secuencia está suficientemente por encima del fondo cuando dicha frecuencia está aproximadamente o al menos aproximadamente 2 veces, 3 veces, 4 veces, 5 veces, 6 veces, 7 veces, 8 veces, 9 veces, 10 veces, 25 veces, 50 veces, 100 veces, o más por encima de la tasa de error de fondo (por ejemplo, al menos 5 veces más). En algunas realizaciones, la tasa de error de fondo para determinar con precisión la secuencia en una posición dada es aproximadamente o aproximadamente inferior al 1%, 0,5 %, 0,1 %, 0,05 %, 0,01 %, 0,005 %, 0,001 %, 0,0005 % o menos. En algunas realizaciones, la tasa de error es inferior al 0,001%. Los métodos para determinar la frecuencia y la tasa de error se describen en el presente documento, tal como con respecto a cualquiera de los diversos aspectos de la divulgación.

En algunas realizaciones, la identificación de una variante de secuencia genuina (también denominada "identificación" o "hacer una identificación") comprende alinear de manera óptima una o más lecturas de secuencia con una secuencia de referencia para identificar diferencias entre las dos, así como para identificar uniones. En general, la alineación implica colocar una secuencia a lo largo de otra secuencia, introducir espacios de manera iterativa a lo largo de cada secuencia, puntuar cómo de bien coinciden las dos secuencias y, preferentemente, repetir para varias posiciones a lo largo de la referencia. La coincidencia de mejor puntuación se considera la alineación y representa una inferencia sobre el grado de relación entre las secuencias. Está disponible varios algoritmos de alineación y alineadores que los implementan están disponibles, se describen en el presente documento ejemplos no limitantes de los mismos, tal como con respecto a cualquiera de los diversos aspectos de la divulgación. En algunas realizaciones, una secuencia de referencia con la que se comparan es una secuencia de referencia conocida, tal como un genoma de referencia (p.ej. el genoma de un miembro de la misma especie que el sujeto). El genoma de referencia puede ser completo o incompleto. En algunas realizaciones, un genoma de referencia consiste solo en regiones que contienen polinucleótidos diana, tal como por un genoma de referencia o de un consenso generado a partir de lecturas de secuenciación en análisis. En algunas realizaciones, una secuencia de referencia comprende o consiste en secuencias de polinucleótidos de uno o más organismos, tales como secuencias de una o más bacterias, arqueas, virus, protistas, hongos u otro organismo. En algunas realizaciones, la secuencia de referencia consiste en solo una parte de un genoma de referencia, tal como las regiones correspondientes a una o más secuencias diana en análisis (por ejemplo, uno o más genes, o porciones de los mismos). Por ejemplo, para la detección de un patógeno (tal como en el caso de detección de contaminación), el genoma de referencia es el genoma completo del patógeno (por ejemplo, VIH, VPH o una cepa bacteriana dañina, por ejemplo, *E. coli*), o una porción del mismo útil en la identificación, tal como de una cepa o serotipo particular. En algunas realizaciones, Las lecturas de secuenciación se alinean con múltiples secuencias de referencia diferentes, tal como para detectar múltiples organismos o cepas diferentes. En el presente documento se describen ejemplos adicionales no limitantes de secuencias de referencia con respecto a las cuales se pueden identificar las diferencias de secuencia (y se denominan variantes de secuencia), tal como con respecto a cualquiera de los diversos aspectos de la divulgación.

En un aspecto, la divulgación proporciona un método para amplificar en una mezcla de reacción una pluralidad de diferentes concatémicos que comprenden dos o más copias de una secuencia diana, en donde la secuencia diana comprende la secuencia A y la secuencia B orientadas en una dirección de 5' a 3'. En algunas realizaciones, el método comprende someter la mezcla de reacción a una reacción de amplificación de ácido nucleico, en donde la mezcla de

reacción comprende: (a) la pluralidad de concatémoros, en donde los concatémoros individuales en la pluralidad comprenden diferentes uniones formadas mediante la circularización de polinucleótidos individuales que tienen un extremo 5' y un extremo 3'; (b) un primer cebador que comprende la secuencia A', en donde el primer cebador hibrida específicamente con la secuencia A de la secuencia diana mediante complementariedad de secuencia entre la secuencia A y la secuencia A'; (c) un segundo cebador que comprende la secuencia B, en donde el segundo cebador hibrida específicamente con la secuencia B' presente en un polinucleótido complementario que comprende un complemento de la secuencia diana mediante complementariedad de secuencia entre la secuencia B y B'; y (d) una polimerasa que extiende el primer cebador y el segundo cebador para producir polinucleótidos amplificados; en donde la distancia entre el extremo 5' de la secuencia A y el extremo 3' de la secuencia B de la secuencia diana es de 75 nt o menos.

En un aspecto relacionado, la divulgación proporciona un método para amplificar en una mezcla de reacción una pluralidad de polinucleótidos circulares diferentes que comprenden una secuencia diana, en donde la secuencia diana comprende la secuencia A y la secuencia B orientadas en una dirección de 5' a 3'. En algunas realizaciones, el método comprende someter la mezcla de reacción a una reacción de amplificación de ácido nucleico, en donde la mezcla de reacción comprende: (a) la pluralidad de polinucleótidos circulares, en donde los polinucleótidos circulares individuales en la pluralidad comprenden diferentes uniones formadas mediante la circularización de polinucleótidos individuales que tienen un extremo 5' y un extremo 3'; (b) un primer cebador que comprende la secuencia A', en donde el primer cebador hibrida específicamente con la secuencia A de la secuencia diana mediante complementariedad de secuencia entre la secuencia A y la secuencia A'; (c) un segundo cebador que comprende la secuencia B, en donde el segundo cebador hibrida específicamente con la secuencia B' presente en un polinucleótido complementario que comprende un complemento de la secuencia diana mediante complementariedad de secuencia entre la secuencia B y B'; y (d) una polimerasa que extiende el primer cebador y el segundo cebador para producir polinucleótidos amplificados; en donde la secuencia A y la secuencia B son secuencias endógenas y la distancia entre el extremo 5' de la secuencia A y el extremo 3' de la secuencia B de la secuencia diana es de 75 nt o menos.

Ya sea que amplifiquen polinucleótidos circulares o concatémoros, dichos polinucleótidos pueden provenir de cualquier fuente de muestra adecuada (ya sea directa o indirectamente, tal como por amplificación). Varias fuentes de muestra adecuadas, procesos de extracción opcionales, tipos de polinucleótidos y tipos de variantes de secuencia se describen en el presente documento, tal como con respecto a cualquiera de los diversos aspectos de la divulgación. Los polinucleótidos circulares pueden derivar de polinucleótidos no circulares circularizantes. Ejemplos no limitantes de procesos de circularización (por ejemplo, con y sin oligonucleótidos adaptadores), reactivos (por ejemplo, tipos de adaptadores, uso de ligasas), condiciones de reacción (por ejemplo, favoreciendo la autounión), procesamiento adicional opcional (por ejemplo, la purificación posterior a la reacción), y las uniones formadas de este modo se proporcionan en el presente documento, tal como con respecto a cualquiera de los diversos aspectos de la divulgación. Los concatémoros pueden derivar de la amplificación de polinucleótidos circulares. Hay disponibles varios métodos para amplificar polinucleótidos (por ejemplo, ADN y/o ARN, también se describen en el presente documento ejemplos no limitantes de los mismos. En algunas realizaciones, los concatémoros son generados por la amplificación por círculo rodante de polinucleótidos circulares.

La FIG. 10 ilustra una disposición de ejemplo los cebadores primero y segundo con respecto a una secuencia diana en el contexto de una sola repetición (que normalmente no se amplificará a menos que sea circular) y los concatémoros que comprenden múltiples copias de la secuencia diana. Como se señaló con respecto a otros aspectos descritos en el presente documento, esta disposición de cebadores puede denominarse cebadores "back to back" (B2B) o "invertidos". La amplificación con cebadores B2B facilita el enriquecimiento de plantillas circulares y/o concatémérica. Por otra parte, esta orientación combinada con una huella relativamente más pequeña (distancia total abarcada por un par de cebadores) permite la amplificación de una variedad más amplia de eventos de fragmentación alrededor de una secuencia diana, ya que es menos probable que se produzca una unión entre los cebadores que en la disposición de los cebadores encontrados en una reacción de amplificación típica (enfrentadas, que abarca una secuencia diana). En algunas realizaciones, la distancia entre el extremo 5' de la secuencia A y el extremo 3' de la secuencia B es aproximadamente de o aproximadamente menos de 200, 150, 100, 75, 50, 40, 30, 25, 20, 15 o menos nucleótidos. En algunas realizaciones, la secuencia A es el complemento de la secuencia B. En algunas realizaciones, se usan múltiples pares de cebadores B2B dirigidos a una pluralidad de secuencias diana diferentes en la misma reacción para amplificar una pluralidad de secuencias diana diferentes en paralelo (por ejemplo, aproximadamente o aproximadamente al menos 10, 50, 100, 150, 200, 250, 300, 400, 500, 1000, 2500, 5000, 10000, 15000, o más secuencias diana diferentes). Los cebadores pueden ser de cualquier longitud adecuada, tal como se describe en otra parte en el presente documento. La amplificación puede comprender cualquier reacción de amplificación adecuada en condiciones apropiadas, tal como una reacción de amplificación descrita en el presente documento. En algunas realizaciones, la amplificación es una reacción en cadena de la polimerasa.

En algunas realizaciones, Los cebadores B2B comprenden al menos dos elementos de secuencia, un primer elemento que hibrida con una secuencia diana mediante complementariedad de secuencia y una "cola" 5' que no hibrida con la secuencia diana durante una primera fase de amplificación a una primera temperatura de hibridación durante la cual el primer elemento hibrida (por ejemplo, debido a la falta de complementariedad de secuencia entre la cola y la porción de la secuencia diana inmediatamente 3' con respecto a donde se une el primer elemento). Por ejemplo, el primer cebador comprende la secuencia C 5' con respecto a la secuencia A', el segundo cebador comprende la secuencia D

5' con respecto a la secuencia B, y ni la secuencia C ni la secuencia D hibridan con la pluralidad de concatémeros (o polinucleótidos circulares) durante una primera fase de amplificación a una primera temperatura de hibridación. En algunas realizaciones en las que se usan tales cebadores con cola, la amplificación puede comprender una primera fase y una segunda fase; la primera fase comprende una etapa de hibridación a una primera temperatura, durante la cual los cebadores primero y segundo hibridan con los concatémeros (o con los polinucleótidos circulares) y la extensión de los cebadores; y la segunda fase comprende una etapa de hibridación a una segunda temperatura que es más alta que la primera temperatura, durante la cual los cebadores primero y segundo hibridan con productos de amplificación que comprenden el primer o segundo cebador extendido o complementos de los mismos y la extensión de los cebadores. El número de ciclos de amplificación en cada una de las dos temperaturas se puede ajustar en función de los productos deseados. Normalmente, la primera temperatura se utilizará para un número relativamente bajo de ciclos, tal como aproximadamente de o aproximadamente menos de 15, 10, 9, 8, 7, 6, 5 o menos ciclos. El número de ciclos a la temperatura más alta se puede seleccionar independientemente del número de ciclos a la primera temperatura, pero normalmente serán tantos o más ciclos, como aproximadamente o al menos aproximadamente 5, 6, 7, 8, 9, 10, 15, 20, 25, o más ciclos. La temperatura más alta favorece la hibridación a lo largo del primer elemento y el elemento de cola del cebador en productos de extensión del cebador sobre fragmentos más cortos formados por hibridación entre solo el primer elemento en un cebador y una secuencia diana interna dentro de un concatémero. En consecuencia, la amplificación de dos fases puede usarse para reducir el grado en que los productos de amplificación cortos podrían ser favorecidos, manteniendo así una proporción relativamente mayor de productos de amplificación que tienen dos o más copias de una secuencia diana. Por ejemplo, después de 5 ciclos (por ejemplo, al menos 5, 6, 7, 8, 9, 10, 15, 20 o más ciclos) de hibridación a la segunda temperatura y de extensión del cebador, al menos el 5 % (por ejemplo, al menos el 5 %, 6 %, 7 %, 8 %, 9 %, 10 %, 15 %, 20 %, 25 %, al menos el 30 % o más) de los polinucleótidos amplificados en la mezcla de reacción comprenden dos o más copias de la secuencia diana. Una ilustración de una realización de acuerdo con este proceso de amplificación de dos fases, el proceso de amplificación del cebador B2B con cola, se ilustra en la FIG. 11.

En algunas realizaciones, la amplificación se produce en condiciones que están sesgadas para aumentar la longitud de los amplicones de los concatémeros. Por ejemplo, la concentración del cebador puede reducirse, de modo que no todos los sitios de cebado hibridarán un cebador, lo que hace que los productos de PCR sean más largos. De manera similar, la disminución del tiempo de hibridación del cebador durante los ciclos permitirá, de manera similar, que se hibriden menos cebadores, por lo que también aumenta el tamaño promedio del amplicón de la PCR. Además, aumentar la temperatura y/o el tiempo de extensión de los ciclos aumentará de manera similar la longitud promedio de los amplicones de PCR. Se puede usar cualquier combinación de estas técnicas.

En algunas realizaciones, particularmente cuando se ha realizado una amplificación con cebadores B2B, los productos de amplificación se tratan para filtrar los amplicones resultantes en función del tamaño para reducir y/o eliminar el número de monómeros de una mezcla que comprende concatémeros. Esto se puede realizar utilizando varias técnicas, incluyendo, pero sin limitación, la escisión del fragmento de los geles y la filtración del gel (por ejemplo, para enriquecer los fragmentos mayores de aproximadamente 300, 400, 500 o más nucleótidos de longitud); así como las perlas SPRI (Agencourt AMPure XP) para la selección del tamaño mediante ajuste fino de la concentración del tampón de unión. Por ejemplo, el uso de un tampón de unión de 0,6x durante la mezcla con fragmentos de ADN puede usarse para unir preferentemente fragmentos de ADN mayores de aproximadamente 500 pares de bases (pb).

En algunas realizaciones, el primer cebador comprende la secuencia C 5' con respecto a la secuencia A', el segundo cebador comprende la secuencia D 5' con respecto a la secuencia B, y ni la secuencia C ni la secuencia D hibridan con la pluralidad de polinucleótidos circulares durante una primera fase de amplificación a una primera temperatura de hibridación. La amplificación puede comprender una primera fase y una segunda fase; en donde la primera fase comprende una etapa de hibridación a una primera temperatura, durante la cual los cebadores primero y segundo hibridan con los polinucleótidos circulares o con los productos de la amplificación de los mismos antes de la extensión de los cebadores; y la segunda fase comprende una etapa de hibridación a una segunda temperatura que es más alta que la primera temperatura, durante la cual los cebadores primero y segundo hibridan con productos de amplificación que comprenden el primer o segundo cebadores extendidos o complementos de los mismos. Por ejemplo, la primera temperatura puede seleccionarse como aproximadamente o más que aproximadamente la T_m de la secuencia A', de la secuencia B, o el promedio de estas, o una temperatura que sea mayor que 1 °C, 2 °C, 3 °C, 4 °C, 5 °C, 6 °C, 7 °C, 8 °C, 9 °C, 10 °C, o más alta que estas T_m . En este ejemplo, la segunda temperatura se puede seleccionar para que sea aproximadamente o aproximadamente más que la T_m de la secuencia combinada (A' + C), la secuencia de combinación (B + D), o el promedio de estas, o una temperatura que sea mayor que 1 °C, 2 °C, 3 °C, 4 °C, 5 °C, 6 °C, 7 °C, 8 °C, 9 °C, 10 °C, o más alta que estas T_m . El término " T_m " también se conoce como la "temperatura de fusión," y generalmente representa la temperatura a la que el 50% de un oligonucleótido consiste en una secuencia de referencia (que de hecho puede ser una subsecuencia dentro de un polinucleótido más grande) y su secuencia complementaria se hibridan (o separan). En general, La T_m aumenta al aumentar la longitud, y como tal, se espera que la T_m de la secuencia A' sea más baja que la T_m de la secuencia de combinación (A' + C).

En un aspecto, esta divulgación proporciona una mezcla de reacción para realizar un método de acuerdo con los métodos de la divulgación. La mezcla de reacción puede comprender uno o más de los diversos componentes como se describe en el presente documento con respecto a cualquiera de los diversos métodos, incluyendo mezclas de reacción descritas en los métodos descritos en el presente documento. En algunas realizaciones, la mezcla de

reacción es una mezcla para amplificar una pluralidad de diferentes concatémoros que comprenden dos o más copias de una secuencia diana, o un polinucleótido circular que comprende una o más copias de las secuencias diana (por ejemplo, un monómero circular), en donde la secuencia diana comprende la secuencia A y la secuencia B orientadas en una dirección de 5' a 3', la mezcla de reacción: (a) la pluralidad de concatémoros (o polinucleótidos circulares), en donde los concatémoros (o polinucleótidos circulares) individuales en la pluralidad comprenden diferentes uniones formadas mediante la circularización de polinucleótidos individuales que tienen un extremo 5' y un extremo 3'; (b) un primer cebador que comprende la secuencia A', en donde el primer cebador hibrida específicamente con la secuencia A de la secuencia diana mediante complementariedad de secuencia entre la secuencia A y la secuencia A'; (c) un segundo cebador que comprende la secuencia B, en donde el segundo cebador hibrida específicamente con la secuencia B presente en un polinucleótido complementario que comprende un complemento de la secuencia diana mediante complementariedad de secuencia entre B y B'; y; (d) una polimerasa que extiende el primer cebador y el segundo cebador para producir polinucleótidos amplificados; en donde la distancia entre el extremo 5' de la secuencia A y el extremo 3' de la secuencia B de la secuencia diana es de 75 nt o menos. Las muestras, polinucleótidos, imprimadores, polimerasas, otros reactivos y las condiciones de reacción pueden ser cualquiera de los descritos en el presente documento, tal como con respecto a cualquiera de los diversos aspectos, que puede incluirse en la mezcla de reacción en cualquier combinación adecuada. En algunas realizaciones, el primer cebador comprende la secuencia C 5' con respecto a la secuencia A', el segundo cebador comprende la secuencia D 5' con respecto a la secuencia B, y ni la secuencia C ni la secuencia D hibridan con los dos o más concatémoros durante una primera etapa de amplificación en una reacción de amplificación.

En un aspecto, esta divulgación proporciona composiciones útiles en o producidas por métodos descritos en el presente documento, tales como en cualquiera de los otros diversos aspectos de la divulgación. En algunas realizaciones, la composición comprende una pluralidad de polinucleótidos circularizados que son de cadena sencilla y sustancialmente libres de ligasa. En algunas realizaciones, la composición comprende una pluralidad de concatémoros, en donde la pluralidad de concatémoros corresponde a un grupo de 10000 o menos polinucleótidos diana, y, en donde, adicionalmente, los concatémoros individuales en la pluralidad se caracterizan porque: (a) comprenden dos o más copias de una repetición de secuencia, en la que la totalidad de dichas copias corresponden al mismo polinucleótido diana; y (b) la unión entre las dos o más copias de la repetición de secuencia de un concatémoro individual es diferente de la de otro concatémoro individual en dicha composición. Las muestras, polinucleótidos, imprimadores, polimerasas y otros reactivos pueden ser cualquiera de los descritos en el presente documento, tal como con respecto a cualquiera de los diversos aspectos, que puede incluirse en la composición en cualquier combinación adecuada. La composición puede comprender uno o más pares de cebadores, tales como cebadores B2B descritos en el presente documento, diseñados para amplificar una o más secuencias diana. Las composiciones pueden proporcionarse en forma de kits. Los reactivos y otros materiales en un kit pueden estar contenidos en cualquier envase adecuado, y pueden estar en una forma de uso inmediato o requerir una combinación con otros reactivos en el kit o reactivos suministrados por un usuario (por ejemplo, dilución de una composición concentrada o reconstitución de una composición liofilizada). Un kit puede proporcionar tampones, cuyos ejemplos no limitantes incluyen el tampón de carbonato de sodio, un tampón de bicarbonato de sodio, un tampón de borato, un tampón Tris, un tampón MOPS, un tampón HEPES y combinaciones de los mismos. Los kits pueden comprender además instrucciones para la realización de uno o más métodos descritos en el presente documento con respecto a cualquiera de los diversos aspectos. Las instrucciones se pueden proporcionar en uno o más idiomas (por ejemplo, 2, 3, 4, 5 o más idiomas).

En un aspecto, esta divulgación proporciona un sistema para detectar una variante de secuencia. En algunas realizaciones, el sistema comprende (a) un ordenador configurado para recibir una solicitud del usuario para realizar una reacción de detección en una muestra; (b) un sistema de amplificación que realiza una reacción de amplificación de ácido nucleico en la muestra o en una parte de la misma en respuesta a la solicitud del usuario, en donde la reacción de amplificación comprende las etapas de (i) circularizar los polinucleótidos individuales para formar una pluralidad de polinucleótidos circulares, cada uno de los cuales tiene una unión entre el extremo 5' y el extremo 3'; y (ii) amplificar los polinucleótidos circulares; (c) un sistema de secuenciación que genera lecturas de secuenciación para polinucleótidos amplificados por el sistema de amplificación, identifica las diferencias de secuencia entre las lecturas de secuenciación y una secuencia de referencia e identifica una diferencia de secuencia que se produce en al menos dos polinucleótidos circulares que tienen diferentes uniones como la variante de secuencia; y (d) un generador de informes que envía un informe a un destinatario, en el que el informe contiene resultados para la detección de la variante de secuencia. En algunas realizaciones, el destinatario es el usuario. La FIG. 32 ilustra un ejemplo no limitante de un sistema útil en los métodos de la presente divulgación.

Un ordenador para usar en el sistema puede comprender uno o más procesadores. Los procesadores se pueden asociar con uno o más controladores, unidades de cálculo y/u otras unidades de un sistema informático, o se pueden implantar en el firmware según se desee. Si se implementa en el software, las rutinas se pueden almacenar en cualquier memoria legible por ordenador, como en la memoria RAM, ROM, memoria flash, un disco magnético, un disco láser u otro medio de almacenamiento adecuado. Asimismo, este software se puede suministrar a un dispositivo informático a través de cualquier método de suministro conocido que incluye, por ejemplo, a través de un canal de comunicación tal como una línea telefónica, internet, una conexión inalámbrica, etc., o a través de un medio transportable, tal como un disco legible por ordenador, unidad flash, etc. Los distintos pasos pueden implementarse como varios bloques, operaciones, herramientas, módulos y técnicas que, a su vez, pueden implementarse en

hardware, firmware, software o cualquier combinación de hardware, firmware y/o software. Cuando se implementa en hardware, algunos o todos los bloques, operaciones, técnicas, etc. se pueden implementar en, por ejemplo, un circuito integrado personalizado (IC), un circuito integrado específico de la aplicación (ASIC), una matriz lógica programable de campo (FPGA), una matriz lógica programable (PLA), etc. Se puede usar una arquitectura de base de datos relacional cliente-servidor en realizaciones del sistema. una arquitectura cliente-servidor es una arquitectura de red en la que cada ordenador o proceso en la red es un cliente o un servidor. Los ordenadores de servidor suelen ser potentes ordenadores dedicados a administrar unidades de disco (servidores de archivos), impresoras (servidores de impresión) o tráfico de red (servidores de red). Los ordenadores cliente incluyen PC (ordenadores personales) o estaciones de trabajo en las que los usuarios ejecutan aplicaciones, así como ejemplos de dispositivos de salida como se describe en el presente documento. Los ordenadores cliente dependen de los ordenadores del servidor para obtener recursos, tales como archivos, dispositivos, e incluso poder de procesamiento. En algunas realizaciones, el ordenador del servidor maneja toda la funcionalidad de la base de datos. El ordenador cliente puede tener un software que maneja toda la administración de datos de interfaz y también puede recibir datos ingresados por los usuarios.

el sistema se puede configurar para recibir una solicitud del usuario para realizar una reacción de detección en una muestra. La solicitud del usuario puede ser directa o indirecta. Los ejemplos de solicitud directa incluyen aquellas transmitidas por medio de un dispositivo de entrada, tal como un teclado, un ratón o una pantalla táctil. Los ejemplos de solicitudes indirectas incluyen la transmisión a través de un medio de comunicación, tal como a través de Internet (ya sea por cable o inalámbrico).

El sistema puede comprender adicionalmente un sistema de amplificación que realiza una reacción de amplificación de ácido nucleico en la muestra o en una parte de la misma en respuesta a la solicitud del usuario. Hay disponibles varios métodos para amplificar polinucleótidos (por ejemplo, ADN y/o ARN). La amplificación puede ser lineal, exponencial o involucrar fases tanto lineales como exponenciales en un proceso de amplificación multifase. Los métodos de amplificación pueden implicar cambios en la temperatura, tal como una etapa de desnaturalización por calor, o pueden ser procesos isotérmicos que no requieren desnaturalización por calor. Los ejemplos no limitantes de procesos de amplificación adecuados se describen, tal como con respecto a cualquiera de los diversos aspectos de la divulgación. En algunas realizaciones, la amplificación comprende la amplificación por círculo rodante (RCA). Hay disponibles varios sistemas para amplificar polinucleótidos y pueden variar según el tipo de reacción de amplificación a realizar. Por ejemplo, para los métodos de amplificación que comprenden ciclos de cambios de temperatura, el sistema de amplificación puede comprender un termociclador. Un sistema de amplificación puede comprender un instrumento de amplificación y detección en tiempo real, como los sistemas fabricados por Applied Biosystems, Roche y Strategene. En algunas realizaciones, la reacción de amplificación comprende las etapas de (i) circularizar los polinucleótidos individuales para formar una pluralidad de polinucleótidos circulares, cada uno de los cuales tiene una unión entre el extremo 5' y el extremo 3'; y (ii) amplificar los polinucleótidos circulares. Las muestras, polinucleótidos, imprimadores, polimerasas y otros reactivos pueden ser cualquiera de los descritos en el presente documento, tal como con respecto a cualquiera de los diversos aspectos. Ejemplos no limitantes de procesos de circularización (por ejemplo, con y sin oligonucleótidos adaptadores), reactivos (por ejemplo, tipos de adaptadores, uso de ligasas), condiciones de reacción (por ejemplo, favoreciendo la autounión), procesamiento adicional opcional (por ejemplo, la purificación posterior a la reacción), y las uniones formadas de este modo se proporcionan en el presente documento, tal como con respecto a cualquiera de los diversos aspectos de la divulgación. Los sistemas se pueden seleccionar y/o diseñar para ejecutar cualquiera de estos métodos.

Los sistemas pueden comprender adicionalmente, un sistema de secuenciación que genera lecturas de secuenciación para polinucleótidos amplificados por el sistema de amplificación, identifica las diferencias de secuencia entre las lecturas de secuenciación y una secuencia de referencia e identifica una diferencia de secuencia que se produce en al menos dos polinucleótidos circulares que tienen diferentes uniones como la variante de secuencia. El sistema de secuenciación y el sistema de amplificación pueden ser iguales, o comprender equipos superpuestos. Por ejemplo, tanto el sistema de amplificación como el sistema de secuenciación pueden utilizar el mismo termociclador. Hay disponibles varias plataformas de secuenciación para uso en el sistema y pueden seleccionarse según el método de secuenciación seleccionado. Se describen en el presente documento ejemplos de métodos de secuenciación. La amplificación y la secuenciación pueden implicar el uso de manipuladores de líquidos. Se pueden utilizar varios sistemas de manejo de líquidos disponibles en el mercado para ejecutar la automatización de estos procesos (véase, por ejemplo, los manipuladores de líquidos de Perkin-Elmer, Beckman Coulter, Caliper Life Sciences, Tecan, Eppendorf, Apricot Design, Velocity 11 como ejemplos). Varias máquinas de secuenciación automatizadas están disponibles comercialmente, e incluyen secuenciadores fabricados por Life Technologies (plataforma SOLiD y detección basada en pH), Roche (plataforma 454), Illumina (por ejemplo, sistemas basados en celdas de flujo, tal como dispositivos analizadores de genoma). La transferencia entre 2, 3, 4, 5 o más dispositivos automatizados (por ejemplo, entre uno o más manipuladores de líquidos y un dispositivo de secuenciación) puede ser manual o automatizada.

Los métodos para identificar diferencias de secuencia e identificar variantes de secuencia con respecto a una secuencia de referencia se describen en el presente documento, tal como con respecto a cualquiera de los diversos aspectos de la divulgación. El sistema de secuenciación normalmente comprenderá un software para realizar estas etapas en respuesta a una entrada de datos de secuencia y la entrada de parámetros deseados (por ejemplo, la selección de un genoma de referencia). Los ejemplos de algoritmos de alineación y alineadores que implementan

estos algoritmos se describen en este documento y pueden formar parte del sistema de secuenciación.

El sistema puede comprender adicionalmente un generador de informes que envía un informe a un destinatario, en el que el informe contiene resultados para la detección de la variante de secuencia. Se puede generar un informe en tiempo real, como durante una lectura de secuencia o mientras se analizan los datos de secuenciación, con actualizaciones periódicas a medida que avanza el proceso. Además, o alternativamente, se puede generar un informe al final del análisis. Se puede generar un informe de manera automática, como cuando el sistema de secuenciación completa el paso de identificar todas las variantes de secuencia. En algunas realizaciones, el informe se genera en respuesta a las instrucciones de un usuario. Además de los resultados de la detección de la variante de secuencia, un informe también puede contener un análisis basado en una o más variantes de secuencia. Por ejemplo, cuando una o más variantes de secuencia están asociadas con un contaminante o fenotipo particular, el informe puede incluir información sobre esta asociación, como la probabilidad de que el contaminante o fenotipo esté presente, a qué nivel, y opcionalmente una sugerencia basada en esta información (por ejemplo, pruebas adicionales, monitoreo, o medidas correctivas). El informe puede adoptar varias formas. Se prevé que los datos relacionados con la presente divulgación puedan transmitirse a través de dichas redes o conexiones (o cualquier otro medio adecuado para transmitir información, que incluye, pero sin limitación, el envío de un informe físico, como una impresión) para su recepción y/o para su revisión por un receptor. El receptor puede ser, pero sin limitación, un sistema individual o electrónico (por ejemplo, uno o más ordenadores y/o uno o más servidores).

En un aspecto, la divulgación proporciona un medio legible por ordenador que comprende códigos que, al ser ejecutados por uno o más procesadores, implementan un método para detectar una variante de secuencia. En algunas realizaciones, el método implementado comprende: (a) recibir una solicitud del cliente para realizar una reacción de detección en una muestra; (b) realizar una reacción de amplificación de ácido nucleico en la muestra o en una parte de la misma en respuesta a la solicitud del cliente, en donde la reacción de amplificación comprende las etapas de (i) circularizar los polinucleótidos individuales para formar una pluralidad de polinucleótidos circulares, cada uno de los cuales tiene una unión entre el extremo 5' y el extremo 3'; y (ii) amplificar los polinucleótidos circulares; (c) realizar un análisis de secuenciación que comprende las etapas de (i) generar lecturas de secuenciación para polinucleótidos amplificados en la reacción de amplificación; (ii) identificar diferencias de secuencia entre lecturas de secuenciación y una secuencia de referencia; y (iii) identificar una diferencia de secuencia que se produce en al menos dos polinucleótidos circulares que tienen diferentes uniones como la variante de secuencia; y (d) generar un informe que contenga resultados para la detección de la variante de secuencia.

Un medio legible por máquina que comprende un código ejecutable por ordenador puede tomar muchas formas, incluyendo, pero sin limitación, un medio de almacenamiento tangible, un medio de onda portadora o un medio de transmisión físico. Los medios de almacenamiento no volátiles incluyen, por ejemplo, discos ópticos o magnéticos, tal como cualquiera de los dispositivos de almacenamiento en cualquier ordenador o similares, como los que se pueden usar para implementar las bases de datos, etc. Los medios de almacenamiento volátiles incluyen memoria dinámica, como la memoria principal de dicha plataforma de ordenador. Los medios de transmisión tangibles incluyen cables coaxiales; cable de cobre y fibras ópticas, incluidos los cables que forman un bus dentro de un sistema informático. Los medios de transmisión de onda portadora pueden tomar la forma de señales eléctricas o electromagnéticas, u ondas acústicas o de luz, como las generadas durante las comunicaciones de datos por radiofrecuencia (RF) e infrarrojos (IR). Las formas comunes de medios legibles por ordenador incluyen por ejemplo: un disquete, un disco flexible, disco duro, cinta magnética, cualquier otro medio magnético, un CD-ROM, DVD o DVD-ROM, cualquier otro medio óptico, tarjetas de papel perforadas, cualquier otro medio de almacenamiento físico con patrones de orificios, una RAM, una ROM, una PROM y una EPROM, una FLASH-EPROM, cualquier otro chip o cartucho de memoria, una onda portadora que transmita datos o instrucciones, cables o enlaces que transporten dicha onda portadora, o cualquier otro medio desde el cual un ordenador pueda leer código y/o datos de programación. Muchas de estas formas de medios legibles por ordenador pueden estar implicadas en llevar una o más secuencias de una o más instrucciones a un procesador para su ejecución.

El código ejecutable por ordenador sujeto se puede ejecutar en cualquier dispositivo adecuado que comprenda un procesador, incluido un servidor, un PC o un dispositivo móvil como un teléfono inteligente o una tableta. Cualquier controlador u ordenador incluye opcionalmente un monitor, que puede ser una pantalla de tubo de rayos catódicos ("CRT"), una pantalla plana (por ejemplo, dispositivo de cristal líquido de matriz activa, pantalla de cristal líquido, etc.) u otros. Los circuitos de ordenador a menudo se colocan en una caja, que incluye numerosos chips de circuitos integrados, como un microprocesador, memoria, circuitos de interfaz y otros. La caja también incluye opcionalmente una unidad de disco duro, una unidad de disquete, una unidad extraíble de alta capacidad, como un CD-ROM grabable, y otros elementos periféricos comunes. Los dispositivos de entrada, tales como el teclado, el ratón o la pantalla táctil, proporcionan opcionalmente la entrada de un usuario. El ordenador puede incluir el software apropiado para recibir instrucciones del usuario, ya sea en forma de entrada del usuario en un conjunto de campos de parámetros, por ejemplo, en una GUI, o en forma de instrucciones preprogramadas, por ejemplo, preprogramadas para varias diferentes operaciones específicas.

En algunas realizaciones de cualquiera de los diversos aspectos descritos en el presente documento, los métodos, composiciones y sistemas tienen aplicaciones terapéuticas, tales como la caracterización de una muestra de un paciente y, opcionalmente, el diagnóstico de una afección de un sujeto. Las aplicaciones terapéuticas también pueden

incluir informar la selección de terapias a las que un paciente puede ser más receptivo (también conocido como "teragnósticos"), y el tratamiento real de un sujeto que lo necesite, basándose en los resultados de un método descrito en el presente documento. En particular, los métodos y composiciones desvelados en el presente documento se pueden usar para diagnosticar la presencia de tumores, la progresión y/o la metástasis de los tumores, especialmente cuando los polinucleótidos analizados comprenden o consisten en ADNcf, ADNct o ADN de tumor fragmentado. En algunas realizaciones, se monitoriza un sujeto para la eficacia del tratamiento. Por ejemplo, al monitorizar el ADNct a lo largo del tiempo, se puede usar una disminución en el ADNct como una indicación de un tratamiento eficaz, mientras que los aumentos pueden facilitar la selección de diferentes tratamientos o diferentes dosificaciones. Otros usos incluyen la evaluación del rechazo de órganos en receptores de trasplantes (donde se utilizan como indicadores tempranos del rechazo del trasplante los aumentos en la cantidad de ADN circulante correspondiente al genoma del donante de trasplante) y el genotipado/isotipado de infecciones por patógenos, como infecciones víricas o bacterianas. La detección de variantes de secuencia en el ADN fetal circulante puede usarse para diagnosticar una afección en un feto.

Como se usa en el presente documento, "tratamiento" o "tratar", "paliar" o "mejorar" se usan indistintamente. Estos términos se refieren a un enfoque para obtener resultados beneficiosos o deseados que incluyen, pero sin limitación, un beneficio terapéutico y/o un beneficio profiláctico. Por beneficio terapéutico se entiende cualquier mejora o efecto terapéuticamente relevante en una o más enfermedades, afecciones o síntomas en tratamiento. Para el beneficio profiláctico, las composiciones pueden administrarse a un sujeto con riesgo de desarrollar una enfermedad, afección o síntoma particular, o a un sujeto que presente uno o más de los síntomas fisiológicos de una enfermedad, aunque la enfermedad, afección o síntoma puede que aún no se haya manifestado. Normalmente, el beneficio profiláctico incluye reducir la incidencia y/o empeoramiento de una o más enfermedades, afecciones o síntomas en tratamiento (por ejemplo, entre poblaciones tratadas y no tratadas, o entre estados de un sujeto tratados y no tratados). La mejora del resultado de un tratamiento puede incluir el diagnóstico de una afección de un sujeto para identificar al sujeto como aquel que se beneficiará o no del tratamiento con uno o más agentes terapéuticos u otra intervención terapéutica (como la cirugía). En tales aplicaciones de diagnóstico, la tasa global de tratamiento exitoso con uno o más agentes terapéuticos puede mejorarse, en relación con su eficacia entre pacientes agrupados sin diagnóstico de acuerdo con un método de la presente divulgación (por ejemplo, una mejora en una medida de eficacia terapéutica de al menos aproximadamente el 10%, 20 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 %, el 95 % o más).

Los términos "sujeto", "individuo", "paciente" se usan indistintamente en el presente documentos para referirse a un vertebrado, preferentemente un mamífero, más preferentemente, un ser humano. Los mamíferos incluyen, pero sin limitación, murinos, simios, seres humanos, animales de granja, animales de deporte y mascotas. También se incluyen tejidos, células y su progenie de una entidad biológica obtenidos in vivo o cultivados in vitro.

La expresión "agente terapéutico", "agente con capacidad terapéutica" o "agente de tratamiento" se usan indistintamente y se refieren a una molécula o compuesto que confiere algún efecto beneficioso al administrarse a un sujeto. El efecto beneficioso incluye la habilitación de determinaciones diagnósticas; la mejora de una enfermedad, síntoma, trastorno o afección patológica; reducir o prevenir, la aparición de una enfermedad, síntoma, trastorno o afección; y generalmente contrarrestando una enfermedad, síntoma, sintoma, trastorno o afección patológica.

En algunas realizaciones de los diversos métodos descritos en el presente documento, la muestra es de un sujeto. Un sujeto puede ser cualquier organismo, ejemplos no limitantes de estos incluyen plantas, animales, hongos, protistas, moneras, virus, mitocondrias y cloroplastos. Los polinucleótidos de muestra pueden aislarse de un sujeto, como una muestra celular, muestra de tejido, muestra de fluido corporal o una muestra de órgano (o cultivos celulares derivados de cualquiera de estos), incluyendo, por ejemplo, líneas celulares cultivadas, biopsia, muestra de sangre, frotis de mejilla o muestra de líquido que contiene una célula (por ejemplo, la saliva). En algunos casos, la muestra no comprende células intactas, se trata para eliminar células o se aíslan polinucleótidos sin una etapa de extracción celular (por ejemplo, para aislar polinucleótidos libres de células, como ADN libre de células). Otros ejemplos de fuentes de muestra incluyen aquellos de sangre, orina, heces, narinas, los pulmones, el intestino, otros fluidos corporales o excreciones, materiales derivados de ellos o combinaciones de los mismos. El sujeto puede ser un animal, incluyendo, pero sin limitación, una vaca, cardo, un ratón, una rata, un pollo, un gato, un perro, etc., y es habitualmente un mamífero, tal como se un ser humano. En algunas realizaciones, la muestra comprende células tumorales, tales como en una muestra de tejido tumoral de un sujeto. En algunas realizaciones, la muestra es una muestra de sangre o una porción de la misma (por ejemplo, plasma sanguíneo o suero). El suero y el plasma pueden ser de particular interés, debido al enriquecimiento relativo del ADN tumoral asociado con la tasa más alta de muerte de células malignas entre dichos tejidos. Una muestra puede ser una muestra nueva o una muestra sometida a uno o más procesos de almacenamiento (por ejemplo, muestras embebidas en parafina, particularmente una muestra fijada en formalina, embebida en parafina (FFPE). En algunas realizaciones, una muestra de un solo individuo se divide en múltiples muestras separadas (por ejemplo, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más muestras separadas) que se someten a métodos de la divulgación de manera independiente, Tal como el análisis por duplicado, triplicado, cuadruplicado o más. Cuando la muestra es de un sujeto, la secuencia de referencia también puede derivar del sujeto, como una secuencia consenso de la muestra en análisis o la secuencia de polinucleótidos de otra muestra o tejido del mismo sujeto. Por ejemplo, una muestra de sangre puede analizarse para detectar mutaciones de ADNct, mientras que el ADN celular de otra muestra (por ejemplo, muestra bucal o de piel) se analiza para determinar la secuencia de referencia.

Los polinucleótidos pueden extraerse de una muestra, con o sin extracción de células en una muestra, de acuerdo con cualquier método adecuado. Hay varios kits disponibles para la extracción de polinucleótidos, cuya selección puede depender del tipo de muestra o del tipo de ácido nucleico que se va a aislar. Se proporcionan en el presente documento ejemplos de métodos de extracción, tal como aquellos descritos con respecto a cualquiera de los diversos aspectos de la divulgación. En un ejemplo, la muestra puede ser una muestra de sangre, como una muestra recolectada en un tubo EDTA (por ejemplo, BD Vacutainer). El plasma se puede separar de las células de la sangre periférica mediante centrifugación (por ejemplo, 10 minutos a 1900xg a 4 °C). La separación de plasma realizada de esta manera en una muestra de sangre de 6 ml producirá normalmente de 2,5 a 3 ml de plasma. El ADN libre de células circulantes se puede extraer de una muestra de plasma, por ejemplo, utilizando un kit de ácido nucleico circulante QIAmp (Qiagen), de acuerdo con el protocolo del fabricante. El ADN se puede cuantificar (por ejemplo, en un bioanalizador Agilent 2100 con kit de ADN de alta sensibilidad (Agilent)). Como ejemplo, el rendimiento del ADN circulante de dicha muestra de plasma de una persona sana puede variar de 1 ng a 10 ng por ml de plasma, con mucho más en las muestras de pacientes con cáncer.

Los polinucleótidos también pueden derivar de muestras almacenadas, tales como muestras congeladas o archivadas. Un método común para almacenar muestras es fijar en formalina y embebida en parafina. Sin embargo, este proceso también está asociado con la degradación de los ácidos nucleicos. Los polinucleótidos procesados y analizados a partir de una muestra FFPE pueden incluir polinucleótidos cortos, tal como fragmentos en el intervalo de 50-200 pares de bases, o más cortos. Existen varias técnicas para la purificación de ácidos nucleicos a partir de muestras fijadas, embebidas en parafina, como las descritas en el documento WO2007133703, y los métodos descritos por Foss, et al Diagnostic Molecular Pathology, (1994) 3: 148-155 y Paska, C., et al Diagnostic Molecular Pathology, (2004) 13:234-240. Los kits disponibles comercialmente se pueden usar para purificar polinucleótidos de muestras de FFPE, tal como el kit de aislamiento de Ácido Nucleico Total de Recuperación de Ambion. Los métodos típicos comienzan con una etapa que elimina la parafina del tejido mediante extracción con xileno u otro disolvente orgánico, seguido de un tratamiento con calor y una proteasa como la proteinasa K, que escinde el tejido y las proteínas y ayuda a liberar el material genómico del tejido. Los ácidos nucleicos liberados pueden capturarse, a continuación, en una membrana o precipitarse de la solución, lavarse para eliminar las impurezas y, en el caso del aislamiento del ARNm, a veces se añade una etapa de tratamiento con DNasa para degradar ADN no deseado. Hay disponibles otros métodos para extraer el ADN de FFPE y se pueden usar en los métodos de la presente divulgación.

En algunas realizaciones, la pluralidad de polinucleótidos libres de células son ADN libre de células (ADNcf) o ADN tumoral circulante (ADNct). El ADN libre de células circula tanto en individuos sanos como enfermos. El ADNcf de los tumores (ADNct) no se limita a ningún tipo de cáncer específico, pero parece ser un hallazgo común en diferentes tumores malignos. De acuerdo con algunas mediciones, la concentración de ADN circulante libre en plasma es de aproximadamente 14-18 ng/ml en sujetos control y aproximadamente 180-318 ng/ml en pacientes con neoplasias. La muerte celular apoptótica y necrótica contribuye al ADN circulante libre de células en los fluidos corporales. Por ejemplo, se han observado niveles significativamente mayores de ADN circulante en el plasma de pacientes con cáncer de próstata y otras enfermedades de la próstata, como la hiperplasia benigna de próstata y las prostatitis. Además, el ADN tumoral circulante está presente en los fluidos que se originan en los órganos donde se produce el tumor primario. Por lo tanto, la detección de cáncer de mama se puede lograr en los lavados ductales; la detección del cáncer colorrectal en las heces; la detección del cáncer de pulmón en el esputo y la detección del cáncer de próstata en la orina o en el eyaculado. El ADN libre de células se puede obtener de varias fuentes. Una fuente común son muestras de sangre de un sujeto. Sin embargo, el ADNcf u otro ADN fragmentado puede derivar de varias otras fuentes. Por ejemplo, las muestras de orina y heces pueden ser una fuente de ADNcf, incluido el ADNct.

En algunas realizaciones, los polinucleótidos se someten a etapas posteriores (por ejemplo, circularización y amplificación) sin una etapa de extracción y/o sin una etapa de purificación. Por ejemplo, una muestra de fluido se puede tratar para eliminar las células sin una etapa de extracción para producir una muestra de líquido purificado y una muestra de células, seguido del aislamiento del ADN de la muestra de fluido purificado. Se dispone de varios procedimientos para el aislamiento de polinucleótidos, tales como por precipitación o unión no específica a un sustrato seguido de lavado del sustrato para liberar polinucleótidos unidos. Cuando los polinucleótidos se aíslan de una muestra sin una etapa de extracción celular, los polinucleótidos serán en gran medida polinucleótidos extracelulares o "libres de células". Por ejemplo, los polinucleótidos libres de células pueden incluir ADN libre de células (también denominado ADN "circulante"). En algunas realizaciones, el ADN circulante es ADN tumoral circulante (ADNct) de células tumorales, tal como de un fluido corporal o excreción (por ejemplo, muestra de sangre). Los tumores con frecuencia muestran apoptosis o necrosis, de modo que los ácidos nucleicos del tumor se liberan al cuerpo, incluido el torrente sanguíneo de un sujeto, a través de diversos de mecanismos, en diferentes formas y a diferentes niveles. Normalmente, el tamaño del ADNct puede variar entre concentraciones más altas de fragmentos más pequeños, generalmente de 70 a 200 nucleótidos de longitud, a concentraciones más bajas de fragmentos grandes de hasta miles de kilobases.

En algunas realizaciones de cualquiera de los diversos aspectos descritos en el presente documento, la detección de una variante de secuencia comprende la detección de mutaciones (por ejemplo, mutaciones somáticas raras) con respecto a una secuencia de referencia o en un fondo sin mutaciones, donde la variante de secuencia está correlacionada con la enfermedad. En general, las variantes de secuencia para las cuales hay evidencia estadística,

biológica y/o funcional de asociación con una enfermedad o rasgo se conocen como "variantes genéticas causales". Una única variante genética causal puede estar asociada con más de una enfermedad o rasgo. En algunas realizaciones, una variante genética causal puede asociarse con un rasgo mendeliano, un rasgo no mendeliano o ambos. Las variantes genéticas causales pueden manifestarse como variaciones en un polinucleótido, como 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50 o más diferencias de secuencia (como entre un polinucleótido que comprende variante genética causal y un polinucleótido que carece de la variante genética causal en la misma posición genómica relativa). Ejemplos no limitantes de tipos de variantes genéticas causales incluyen polimorfismos de un solo nucleótido (SNP), polimorfismos de delección/inserción (DIP), variantes de número de copias (CNV), repeticiones cortas en tándem (STR), polimorfismos de longitud de fragmentos de restricción (RFLP), repeticiones de secuencia simple (SSR), número variable de repeticiones en tándem (VNTR), ADN polimórfico amplificado al azar (RAPD), polimorfismos de longitud de fragmentos amplificados (AFLP), polimorfismos amplificados entre retrotransposones (IRAP), elementos intercalados largos y cortos (LÍNEA/SINE), repeticiones en tándem largas (LTR), elementos móviles, polimorfismos amplificados de micro satélite de retrotransposón, polimorfismos de inserción basada en retrotransposón, polimorfismos amplificados específicos de secuencia y modificación genética hereditaria (por ejemplo, metilación de ADN). Una variante genética causal también puede ser un conjunto de variantes genéticas causales estrechamente relacionadas. Algunas variantes genéticas causales pueden ejercer influencia como variaciones de secuencia en polinucleótidos de ARN. A este nivel, algunas variantes genéticas causales también están indicadas por la presencia o ausencia de una especie de polinucleótidos de ARN. También, algunas variantes genéticas causales dan como resultado variaciones de secuencia en polipéptidos de proteínas. Se han indicado varias variantes genéticas causales. Un ejemplo de una variante genética causal que es un SNP, es la variante de hemoglobina Hb S que causa anemia de células falciformes. Un ejemplo de una variante genética causal que es un DIP, es la mutación delta508 del gen CFTR que causa la fibrosis quística. Un ejemplo de una variante genética causal que es una CNV, es la trisomía 21, que causa el síndrome de Down. Un ejemplo de una variante genética causal que es una STR, es la repetición en tándem que causa la enfermedad de Huntington. En la Tabla 4 se proporcionan ejemplos no limitantes de variantes genéticas causales y enfermedades con las que están asociadas. Se describen ejemplos adicionales no limitantes de variantes genéticas causales en el documento WO2014015084. Otros ejemplos de genes en los que las mutaciones están asociadas con enfermedades y en las que se pueden detectar variantes de secuencia de acuerdo con un método de la divulgación, se proporcionan en la Tabla 5.

Tabla 4

Enfermedad	Gen	Nombre de la variante
Deficiencia de 21-hidroxilasa	CYP21A2	F306+t
Deficiencia de 21-hidroxilasa	CYP21A2	F306+t
Deficiencia de 21-hidroxilasa	CYP21A3	g.655A />G
Deficiencia de 21-hidroxilasa	CYP21A4	g.655A />G
Deficiencia de 21-hidroxilasa	CYP21A6	G110de18nt
Deficiencia de 21-hidroxilasa	CYP21A5	G110de18nt
Deficiencia de 21-hidroxilasa	CYP21A7	I172N, rs34607927
Deficiencia de 21-hidroxilasa	CYP21A2	I236N
Deficiencia de 21-hidroxilasa	CYP21A2	M239K, rs6476
Deficiencia de 21-hidroxilasa	CYP21A2	P30L
Deficiencia de 21-hidroxilasa	CYP21A2	P453S
Deficiencia de 21-hidroxilasa	CYP21A2	Q318X
Deficiencia de 21-hidroxilasa	CYP21A2	R356W
Deficiencia de 21-hidroxilasa	CYP21A2	V237E, rs12530380
Deficiencia de 21-hidroxilasa	CYP21A2	V281L, rs6471
Hiperinsulinismo relacionado con ABCC8	ABCC8	3992-9G>A
Hiperinsulinismo relacionado con ABCC8	ABCC8	delF1388

Hiperinsulinismo relacionado con ABCC8	ABCC8	delF1388
Hiperinsulinismo relacionado con ABCC8	ABCC8	V187D
Acondroplasia	FGFR3	G375C
Acondroplasia	FGFR3	G380R, rs28931614
Acromatopsia	CNGB3	c.1148delC
Acromatopsia	CNGB3	c.1148delC
Acromatopsia	CNGB3	c.819-826de18
Acromatopsia	CNGB3	c.819-826de18
Acromatopsia	CNGB3	c.886-896del11insT
Acromatopsia	CNGB3	c.886-896del11insT
Acromatopsia	CNGB3	c.991-3T>G
Acromatopsia	CNGB3	p.Arg403Gln
Acromatopsia	CNGB3	p.Glu336X
Adenosina Monofosfato desaminasa 1	AMPD1	P48L
Adenosina Monofosfato desaminasa 1	AMPD1	Q12X, rs17602729
Agenesia del Cuerpo Calloso con Neuronopatía	SLC12A6	c.2436delG
Agenesia del Cuerpo Calloso con Neuronopatía	SLC12A6	c.2436delG
Alcaptonuria	HGD	c.174delA
Alcaptonuria	HGD	c.174delA
Alcaptonuria	HGD	c.457_458insG
Alcaptonuria	HGD	c.457_458insG
Alcaptonuria	HGD	G161R
Alcaptonuria	HGD	G270R
Alcaptonuria	HGD	IVS1-1G>A
Alcaptonuria	HGD	IVS5+1G>A
Alcaptonuria	HGD	Met368Val
Alcaptonuria	HGD	P230S
Alcaptonuria	HGD	S47L
Alcaptonuria	HGD	V300G
Deficiencia de Alfa-1 Antitripsina	SERPINA1	Arg101His, rs709932
Deficiencia de Alfa-1 Antitripsina	SERPINA1	Glu264Val
Deficiencia de Alfa-1 Antitripsina	SERPINA1	Glu342Lys, rs28929474
Deficiencia de Alfa-1 Antitripsina	SERPINA1	Glu376Asp, rs1303
Alfa-manosidosis	MAN2B1	IVS14+1G>C
Alfa-manosidosis	MAN2B1	P.L809P
Alfa-manosidosis	MAN2B1	P.R750W
Alfa-sarcoglicanopatía	SGCA	R77C, rs28933693
Alfa Talasemia	HBA2	H19D
Alfa Talasemia	HBA1	HbQ
Alfa Talasemia	HBA1,HBA2	Delección alfa-2 de 3.7 kb (tipo I)
Alfa Talasemia	HBA1,HBA2	Delección alfa-2 de 3.7 kb (tipo I)
Alfa Talasemia	HBA1,HBA2	ID de base de datos HbVar nº 1086, --(SEA); delección de -20 kb incluyendo ambos genes alfa-Thal-1 de alfa-globina
Alfa Talasemia	HBA1,HBA2	ID de base de datos HbVar nº 1086, --(SEA); delección de -20 kb incluyendo ambos genes alfa-Thal-1 de alfa-globina
Alfa Talasemia	HBA1,HBA2	ID de base de datos HbVar nº 1087, --(MED-I); delección de -17,5 kb incluyendo ambos genes alfa-Thal-1 de alfa-globina
Alfa Talasemia	HBA1,HBA2	ID de base de datos HbVar nº 1087, --(MED-I); delección de ~ 17,5 kb incluyendo ambos genes alfa-Thal-1 de alfa-globina
Alfa Talasemia	HBA1,HBA2	ID de base de datos HbVar nº 1088, -(alfa)20,5; esta delección de 20,5 kb involucra alfa2 y el extremo 5' de alfa1; alfa-Thal-1
Alfa Talasemia	HBA1,HBA2	ID de base de datos HbVar nº 1088, -(alfa)20,5; esta delección de 20,5 kb involucra alfa2 y el extremo 5' de alfa1; alfa-Thal-1

Alfa Talasemia	HBA1,HBA2	ID de base de datos HbVar nº 1094, -(FIL); una delección de 30-34 kb que involucra a los genes alfa-Thal-1 de alfa1, alfa2 y zeta
Alfa Talasemia	HBA1,HBA2	ID de base de datos HbVar nº 1094, -(FIL); una delección de 30-34 kb que involucra a los genes alfa-Thal-1 de alfa1, alfa2 y zeta
Alfa Talasemia	HBA1,HBA2	ID de base de datos HbVar nº 1095, -(THAI); una delección de 34-38 kb que involucra a los genes alfa-Thal-1 de alfa1, alfa2 y zeta
Alfa Talasemia	HBA1,HBA2	ID de base de datos HbVar nº 1095, -(THAI); una delección de 34-38 kb que involucra a los genes alfa-Thal-1 de alfa1, alfa2 y zeta
Alfa Talasemia	HBA1,HBA2	ID de base de datos HbVar nº 1097, -(MED-II); una delección de 26,5 kb que involucra a los genes alfa-Thal-1 de alfa y zeta
Alfa Talasemia	HBA1,HBA2	ID de base de datos HbVar nº 1097, -(MED-II); una delección de 26,5 kb que involucra a los genes alfa-Thal-1 de alfa y zeta
Alfa Talasemia	HBA2	ID de base de datos HbVar nº 187
Alfa Talasemia	HBA2	ID de base de datos HbVar nº 2598, IVS I-5 (G>A)
Alfa Talasemia	HBA1,HBA2	ID de base de datos HbVar nº 703
Alfa Talasemia	HBA1,HBA2	ID de base de datos HbVar nº 704
Alfa Talasemia	HBA1,HBA2	ID de base de datos HbVar nº 705, Hb Koya Dora
Alfa Talasemia	HBA1,HBA2	ID de base de datos HbVar nº 707, rs41412046
Alfa Talasemia	HBA1	ID de base de datos HbVar nº 87
Alfa Talasemia	HBA1,HBA2	ID de base de datos HbVar nº 969, Poly A (A->G); AATAAA->AATGAA beta+
Alfa Talasemia	HBA1,HBA2	ID de base de datos HbVar nº 971, Poly A (A->G); AATAAA->AATAAG beta+
Alfa Talasemia	HBA2	M1T
Alfa Talasemia	HBA1	W14X
Receptor de Angiotensina II, Type 1	AGTR1	A1166C
Genotipado de Apolipoproteína E	APOE	P.C112R, rs429358
Genotipado de Apolipoproteína E	APOE	P.R158C, rs7412
Aciduria argininosuccínica	ASL	R385C
ARSACS	SACS	5254C>T
ARSACS	SACS	6594delT
ARSACS	SACS	6594delT
Aspartilglicosaminuria	AGA	c.199_200delGA
Aspartilglicosaminuria	AGA	c.199_200delGA
Aspartilglicosaminuria	AGA	C163S
Ataxia con Deficiencia de Vitamina E	TTPA	744delA
Ataxia con Deficiencia de Vitamina E	TTPA	744delA
Ataxia Telangiectasia	ATM	R35X
Síndrome de poliendocrinopatía autoinmunitaria de tipo 1	AIRE	c.1163_1164insA
Síndrome de poliendocrinopatía autoinmunitaria de tipo 1	AIRE	c.1163_1164insA
Síndrome de poliendocrinopatía autoinmunitaria de tipo 1	AIRE	c.769C>T
Síndrome de poliendocrinopatía autoinmunitaria de tipo 1	AIRE	c.967_979del
Síndrome de poliendocrinopatía autoinmunitaria de tipo 1	AIRE	c.967_979del
Síndrome de poliendocrinopatía autoinmunitaria de tipo 1	AIRE	Y85C
Síndrome de Bardet-Biedl	BBS1	M390R

Síndrome de Bardet-Biedl	BBS10	p.C91LfsX4
Síndrome de Bardet-Biedl	BBS10	p.C91LfsX4
Distrofia Macular Viteliforme de Best	BEST1	C.G383C
Beta-Sarcoglicanopatía	SGCB	S114F
Beta Talasemia	HBB	-28 (A->G) beta+
Beta Talasemia	HBB	-29 (A->G) beta+
Beta Talasemia	HBB	-29A>G
Beta Talasemia	HBB	-30 (T->A) beta+
Beta Talasemia	HBB	-87 (C->G) beta+
Beta Talasemia	HBB	-88C>T
Beta Talasemia	HBB	CAP+1 (A->C) beta+
Beta Talasemia	HBB	Codón 15 (G->A); TGG(Trp)->TAG(codón de parada) beta0, rs34716011
Beta Talasemia	HBB	Codón 15 (G->A); TGG(Trp)->TAG(codón de parada) beta0, rs34716011
Beta Talasemia	HBB	Codón 16 (-C); GGC(Gly)->GG-beta0
Beta Talasemia	HBB	Codón 16 (-C); GGC(Gly)->GG-beta0
Beta Talasemia	HBB	Codón 17 (A->T); AAG(Lys)->TAG(codón de parada) beta0
Beta Talasemia	HBB	Codón 24 (T->A); GGT(Gly)->GGA(Gly) beta+
Beta Talasemia	HBB	Codón 39 (C->T); CAG(Gln)->TAG(codón de parada) beta0
Beta Talasemia	HBB	Codón 5 (-CT); CCT(Pro)->C-beta0
Beta Talasemia	HBB	Codón 5 (-CT); CCT(Pro)->C--beta0
Beta Talasemia	HBB	Codón 6 (-A); GAG(Glu)->G-G beta0
Beta Talasemia	HBB	Codón 6 (-A); GAG(Glu)->G-G beta0
Beta Talasemia	HBB	Codón 8 (-AA); AAG(Lys)->-G beta0
Beta Talasemia	HBB	Codón 8 (-AA); AAG(Lys)->-G beta0
Beta Talasemia	HBB	Codones 41/42 (-TTCT); TTCTTT(Phe-Phe)->----TT beta0
Beta Talasemia	HBB	Codones 41/42 (-TTCT); TTCTTT(Phe-Phe)->----TT beta0
Beta Talasemia	HBB	Codones 71/72 (+A); TTT AGT(Phe Ser)->TTT A AGT; beta0
Beta Talasemia	HBB	Codones 71/72 (+A); TTT AGT(Phe Ser)->TTT A AGT; beta0
Beta Talasemia	HBB	Codones 8/9 (+G); AAG TCT(Lys;Ser)->AAG G TCT beta0
Beta Talasemia	HBB	Codones 8/9 (+G); AAG TCT(Lys;Ser)->AAG G TCT beta0
Beta Talasemia	HBB	ID de base de datos HbVar n° 889, IVS-II-654 (C->T); AAGGCAATA->AAG^GTAATA beta+(grave)
Beta Talasemia	HBB	ID de base de datos HbVar n° 890, IVS-II-705 (T->G); GATGTAAGA->GAG^GTAAGA beta+
Beta Talasemia	HBB	ID de base de datos HbVar n° 891, IVS-II-745 (C->G); CAGCTACCAT->CAG^GTACCAT beta+
Beta Talasemia	HBB	ID de base de datos HbVar n.º 979, delección beta0 de 619 pb
Beta Talasemia	HBB	delección beta0 de 619 pb
Beta Talasemia	HBB	IVS-I-1 (G->A); AG^GTTGGT->AGATTGGT beta0
Beta Talasemia	HBB	IVS-I-1 (G->T); AG^GTTGGT->AGTTTGGT beta0
Beta Talasemia	HBB	IVS-I-110 (G->A) beta+; la mutación es 21 nucleótidos 5' al sitio de corte y empalme del aceptor AG^GC
Beta Talasemia	HBB	IVS-I-5 (G->C) beta+(grave)
Beta Talasemia	HBB	IVS-II-1 (G->A); beta0
Beta Talasemia	HBB	IVS-II-844 (C->G); beta+
Beta Talasemia	HBB	IVS1+6T>C
Beta Talasemia	HBB	IVS11-849A>C
Beta Talasemia	HBB	IVS11-849A>G
Deficiencia de biotinidasa	BTBD	A171T, rs13073139
Deficiencia de biotinidasa	BTBD	D252G, rs28934601
Deficiencia de biotinidasa	BTBD	D444H, rs13078881
Deficiencia de biotinidasa	BTBD	F403V
Deficiencia de biotinidasa	BTBD	G98:d7i3
Deficiencia de biotinidasa	BTBD	G98:d7i3
Deficiencia de biotinidasa	BTBD	Q456H
Deficiencia de biotinidasa	BTBD	R157H
Deficiencia de biotinidasa	BTBD	R538C
Síndrome de Blau	NOD2	E383K
Síndrome de Blau	NOD2	L469F
Síndrome de Blau	NOD2	R334Q

Síndrome de Blau	NOD2	R334W
Síndrome de Bloom	BLM	2407insT
Síndrome de Bloom	BLM	2407insT
Síndrome de Bloom	BLM	736delATCTGAinsTAGATTC (2281del6/ins7)
Síndrome de Bloom	BLM	736delATCTGAinsTAGATTC (2281del6/ins7)
BRCA1 Cáncer de mama/ovario hereditario	BRCA1	185delAG
BRCA1 Cáncer de mama/ovario hereditario	BRCA1	185delAG
BRCA1 Cáncer de mama/ovario hereditario	BRCA1	5382insC
BRCA1 Cáncer de mama/ovario hereditario	BRCA1	5382insC
BRCA1 Cáncer de mama/ovario hereditario	BRCA1	Tyr978X
BRCA2 Cáncer de mama/ovario hereditario	BRCA2	6174delT
BRCA2 Cáncer de mama/ovario hereditario	BRCA2	6174delT
BRCA2 Cáncer de mama/ovario hereditario	BRCA2	8765delAG
BRCA2 Cáncer de mama/ovario hereditario	BRCA2	8765delAG
Enfermedad de Canavan	ASPA	A305E (914C>A), rs28940574
Enfermedad de Canavan	ASPA	E285A (854A>C), rs28940279
Enfermedad de Canavan	ASPA	IVS2-2A>G (433-2A>G)
Enfermedad de Canavan	ASPA	Y231X (693C>A)
Deficiencia de Carnitina palmitoiltransferasa IA	CPT1A	G710E
Deficiencia de Carnitina palmitoiltransferasa IA	CPT1A	P479L
Deficiencia de Carnitina palmitoiltransferasa II	CPT2	G549D
Deficiencia de Carnitina palmitoiltransferasa II	CPT2	L178F534 ins/25 pb del
Deficiencia de Carnitina palmitoiltransferasa II	CPT2	L178F534 ins/25 pb del
Deficiencia de Carnitina palmitoiltransferasa II	CPT2	P227L
Deficiencia de Carnitina palmitoiltransferasa II	CPT2	P50H
Deficiencia de Carnitina palmitoiltransferasa II	CPT2	P604S
Deficiencia de Carnitina palmitoiltransferasa II	CPT2	Q413fs
Deficiencia de Carnitina palmitoiltransferasa II	CPT2	Q413fs
Deficiencia de Carnitina palmitoiltransferasa II	CPT2	Q550R
Deficiencia de Carnitina palmitoiltransferasa II	CPT2	R124X
Deficiencia de Carnitina palmitoiltransferasa II	CPT2	R503C
Deficiencia de Carnitina palmitoiltransferasa II	CPT2	R631C
Deficiencia de Carnitina palmitoiltransferasa II	CPT2	S113L
Deficiencia de Carnitina palmitoiltransferasa II	CPT2	s38fs
Deficiencia de Carnitina palmitoiltransferasa II	CPT2	s38fs
Deficiencia de Carnitina palmitoiltransferasa II	CPT2	Y628S, rs28936673

Hipoplasia Cartilago-Cabello	RMRP	g.262G>T
Hipoplasia Cartilago-Cabello	RMPR	g.70A>G
Trastornos relacionados con CFTR	CFTR	1811+1.6kbA->G
Trastornos relacionados con CFTR	CFTR	2183AA>G
Trastornos relacionados con CFTR	CFTR	2183AA>G
Trastornos relacionados con CFTR	CFTR	3849+10kbC>T
Trastornos relacionados con CFTR	CFTR	A455E
Trastornos relacionados con CFTR	CFTR	A559T
Trastornos relacionados con CFTR	CFTR	C524X
Trastornos relacionados con CFTR	CFTR	574delA, 574delA
Trastornos relacionados con CFTR	CFTR	574delA, 574delA
Trastornos relacionados con CFTR	CFTR	2108delA, 2108delA
Trastornos relacionados con CFTR	CFTR	2108delA, 2108delA
Trastornos relacionados con CFTR	CFTR	3171delC, 3171delC
Trastornos relacionados con CFTR	CFTR	3171delC, 3171delC
Trastornos relacionados con CFTR	CFTR	621+1G->T
Trastornos relacionados con CFTR	CFTR	2105-2117del13insAGAAA
Trastornos relacionados con CFTR	CFTR	2105-2117del13insAGAAA
Trastornos relacionados con CFTR	CFTR	711+1G->T
Trastornos relacionados con CFTR	CFTR	711+5G->A
Trastornos relacionados con CFTR	CFTR	712-1G->T
Trastornos relacionados con CFTR	CFTR	1288insTA, 1288insTA
Trastornos relacionados con CFTR	CFTR	1288insTA, 1288insTA
Trastornos relacionados con CFTR	CFTR	936delTA
Trastornos relacionados con CFTR	CFTR	936delTA
Trastornos relacionados con CFTR	CFTR	[delta]F311
Trastornos relacionados con CFTR	CFTR	[delta]F311
Trastornos relacionados con CFTR	CFTR	1078delT, 1078delT
Trastornos relacionados con CFTR	CFTR	1078delT, 1078delT
Trastornos relacionados con CFTR	CFTR	1161delC, 1161delC
Trastornos relacionados con CFTR	CFTR	1161delC, 1161delC
Trastornos relacionados con CFTR	CFTR	1609delCA, 1609delCA

ES 2 707 744 T3

Trastornos relacionados con CFTR	CFTR	1609delCA, 1609delCA
Trastornos relacionados con CFTR	CFTR	[delta] 1507
Trastornos relacionados con CFTR	CFTR	[delta] 1507
Trastornos relacionados con CFTR	CFTR	rs332, [delta]F508
Trastornos relacionados con CFTR	CFTR	rs332, [delta]F508
Trastornos relacionados con CFTR	CFTR	1677delTA, 1677delTA
Trastornos relacionados con CFTR	CFTR	1677delTA, 1677delTA
Trastornos relacionados con CFTR	CFTR	1717-1G->A
Trastornos relacionados con CFTR	CFTR	1812-1G->A
Trastornos relacionados con CFTR	CFTR	1898+1G->A
Trastornos relacionados con CFTR	CFTR	1898+1G->T
Trastornos relacionados con CFTR	CFTR	1898+5G->T
Trastornos relacionados con CFTR	CFTR	1949del84, 1949del84
Trastornos relacionados con CFTR	CFTR	1949del84, 1949del84
Trastornos relacionados con CFTR	CFTR	2043delG, 2043delG
Trastornos relacionados con CFTR	CFTR	2043delG, 2043delG
Trastornos relacionados con CFTR	CFTR	2055del9->A
Trastornos relacionados con CFTR	CFTR	2055del9->A
Trastornos relacionados con CFTR	CFTR	2143delT, 2143delT
Trastornos relacionados con CFTR	CFTR	2143delT, 2143delT
Trastornos relacionados con CFTR	CFTR	2184delA, 2184delA
Trastornos relacionados con CFTR	CFTR	2184delA, 2184delA
Trastornos relacionados con CFTR	CFTR	2184insA, 2184insA
Trastornos relacionados con CFTR	CFTR	2184insA, 2184insA
Trastornos relacionados con CFTR	CFTR	2307insA, 2307insA
Trastornos relacionados con CFTR	CFTR	2307insA, 2307insA
Trastornos relacionados con CFTR	CFTR	296+12T->C
Trastornos relacionados con CFTR	CFTR	2789+5G->A
Trastornos relacionados con CFTR	CFTR	2869insG, 2869insG
Trastornos relacionados con CFTR	CFTR	2869insG, 2869insG
Trastornos relacionados con CFTR	CFTR	3120G->A
Trastornos relacionados con CFTR	CFTR	3120+1G->A

ES 2 707 744 T3

Trastornos relacionados con CFTR	CFTR	3272-26A->G
Trastornos relacionados con CFTR	CFTR	3659delC, 3659delC
Trastornos relacionados con CFTR	CFTR	3659delC, 3659delC
Trastornos relacionados con CFTR	CFTR	3667del4, 3667del4
Trastornos relacionados con CFTR	CFTR	3667del4, 3667del4
Trastornos relacionados con CFTR	CFTR	3791delC, 3791delC
Trastornos relacionados con CFTR	CFTR	3791delC, 3791delC
Trastornos relacionados con CFTR	CFTR	3821delT, 3821delT
Trastornos relacionados con CFTR	CFTR	3821delT, 3821delT
Trastornos relacionados con CFTR	CFTR	3905insT, 3905insT
Trastornos relacionados con CFTR	CFTR	3905insT, 3905insT
Trastornos relacionados con CFTR	CFTR	4016insT, 4016insT
Trastornos relacionados con CFTR	CFTR	4016insT, 4016insT
Trastornos relacionados con CFTR	CFTR	394delTT, 394delTT
Trastornos relacionados con CFTR	CFTR	394delTT, 394delTT
Trastornos relacionados con CFTR	CFTR	405+1G->A
Trastornos relacionados con CFTR	CFTR	405+3A->C
Trastornos relacionados con CFTR	CFTR	444delA
Trastornos relacionados con CFTR	CFTR	444delA
Trastornos relacionados con CFTR	CFTR	3876delA, 3876delA
Trastornos relacionados con CFTR	CFTR	3876delA, 3876delA
Trastornos relacionados con CFTR	CFTR	457TAT->G
Trastornos relacionados con CFTR	CFTR	457TAT->G
Trastornos relacionados con CFTR	CFTR	3199del6, 3199del6
Trastornos relacionados con CFTR	CFTR	3199del6, 3199del6
Trastornos relacionados con CFTR	CFTR	406-1G->A
Trastornos relacionados con CFTR	CFTR	663delT, 663delT
Trastornos relacionados con CFTR	CFTR	663delT, 663delT
Trastornos relacionados con CFTR	CFTR	935delA, 935delA
Trastornos relacionados con CFTR	CFTR	935delA, 935delA
Trastornos relacionados con CFTR	CFTR	CFTR dele2,3 (21kb)
Trastornos relacionados con CFTR	CFTR	CFTR dele2,3 (21kb)

Trastornos relacionados con CFTR	CFTR	D1152H
Trastornos relacionados con CFTR	CFTR	E60X
Trastornos relacionados con CFTR	CFTR	E92X
Trastornos relacionados con CFTR	CFTR	F508C, rs1800093
Trastornos relacionados con CFTR	CFTR	G178R
Trastornos relacionados con CFTR	CFTR	G330X
Trastornos relacionados con CFTR	CFTR	G480C
Trastornos relacionados con CFTR	CFTR	G542X
Trastornos relacionados con CFTR	CFTR	G551D
Trastornos relacionados con CFTR	CFTR	G622D
Trastornos relacionados con CFTR	CFTR	G85E
Trastornos relacionados con CFTR	CFTR	G91R
Trastornos relacionados con CFTR	CFTR	I148T, rs35516286
Trastornos relacionados con CFTR	CFTR	I506V
Trastornos relacionados con CFTR	CFTR	IVS8-5T
Trastornos relacionados con CFTR	CFTR	IVS8-7T
Trastornos relacionados con CFTR	CFTR	IVS8-9T
Trastornos relacionados con CFTR	CFTR	K710X
Trastornos relacionados con CFTR	CFTR	L206W
Trastornos relacionados con CFTR	CFTR	M1101K, rs36210737
Trastornos relacionados con CFTR	CFTR	N1303K
Trastornos relacionados con CFTR	CFTR	P574H
Trastornos relacionados con CFTR	CFTR	Q1238X
Trastornos relacionados con CFTR	CFTR	Q359K/T360K_wt
Trastornos relacionados con CFTR	CFTR	Q493X
Trastornos relacionados con CFTR	CFTR	Q552X
Trastornos relacionados con CFTR	CFTR	Q890X
Trastornos relacionados con CFTR	CFTR	R1066C
Trastornos relacionados con CFTR	CFTR	R1070Q
Trastornos relacionados con CFTR	CFTR	R1158X
Trastornos relacionados con CFTR	CFTR	R1162X
Trastornos relacionados con CFTR	CFTR	R117C

Trastornos relacionados con CFTR	CFTR	R117H
Trastornos relacionados con CFTR	CFTR	R1283M
Trastornos relacionados con CFTR	CFTR	R334W
Trastornos relacionados con CFTR	CFTR	R347H
Trastornos relacionados con CFTR	CFTR	R347P
Trastornos relacionados con CFTR	CFTR	R352Q
Trastornos relacionados con CFTR	CFTR	R553X
Trastornos relacionados con CFTR	CFTR	R560T
Trastornos relacionados con CFTR	CFTR	R709X
Trastornos relacionados con CFTR	CFTR	R75X
Trastornos relacionados con CFTR	CFTR	R764X
Trastornos relacionados con CFTR	CFTR	S1196X
Trastornos relacionados con CFTR	CFTR	S1235R, rs34911792
Trastornos relacionados con CFTR	CFTR	S1251N
Trastornos relacionados con CFTR	CFTR	S1255X
Trastornos relacionados con CFTR	CFTR	S364P
Trastornos relacionados con CFTR	CFTR	S549I
Trastornos relacionados con CFTR	CFTR	S549N
Trastornos relacionados con CFTR	CFTR	S549R
Trastornos relacionados con CFTR	CFTR	S549R
Trastornos relacionados con CFTR	CFTR	T338I
Trastornos relacionados con CFTR	CFTR	V520F
Trastornos relacionados con CFTR	CFTR	W1089X
Trastornos relacionados con CFTR	CFTR	W1204X
Trastornos relacionados con CFTR	CFTR	W1204X
Trastornos relacionados con CFTR	CFTR	W1282X
Trastornos relacionados con CFTR	CFTR	Y1092X
Trastornos relacionados con CFTR	CFTR	Y122X
Coroideremia	CHM	c.1609+2dupT
Coroideremia	CHM	c.1609+2dupT
Ceroide lipofuscinosis neuronal relacionada con CLN3	CLN3	c.461_677del
Ceroide lipofuscinosis neuronal relacionada con CLN3	CLN3	c.461_677del

Ceroide lipofuscinosis neuronal relacionada con CLN3	CLN3	c.791_1056del
Ceroide lipofuscinosis neuronal relacionada con CLN3	CLN3	c.791_1056del
Ceroide lipofuscinosis neuronal relacionada con CLN5	CLN5	c.1175_1176delAT
Ceroide lipofuscinosis neuronal relacionada con CLN5	CLN5	c.1175_1176delAT
Ceroide lipofuscinosis neuronal relacionada con CLN5	CLN5	c.225G>A
Ceroide lipofuscinosis neuronal relacionada con CLN8	CLN8	c.70C>G
Síndrome de Cohen	VPS13B	c.3348_3349delCT
Síndrome de Cohen	VPS13B	c.3348_3349delCT
Cataratas congénitas, Dimorfismo facial y Neuropatía	CTDP1	IVS6+389C>T
Trastornos congénitos de la glicosilación Ia	PMM2	p.F119L
Trastornos congénitos de la glicosilación Ia	PMM2	p.R141H
Trastornos congénitos de la glicosilación Ib	MPI	R295H, rs28928906
Nefrosis Congénita Finlandesa	NPHS1	c.121_122del
Nefrosis Congénita Finlandesa	NPHS1	c.121_122del
Nefrosis Congénita Finlandesa	NPHS1	c.3325C>T
Enfermedad de Crohn	NOD2	3020 ins C
Enfermedad de Crohn	NOD2	3020 ins C
Enfermedad de Crohn	NOD2	G908R, rs2066845
Enfermedad de Crohn	NOD2	R702W, rs2066844
Cistinosis	CTNS	1035insC
Cistinosis	CTNS	1035insC
Cistinosis	CTNS	537del21
Cistinosis	CTNS	537del21
Cistinosis	CTNS	deleción de 57kb
Cistinosis	CTNS	deleción de 57kb
Cistinosis	CTNS	D205N
Cistinosis	CTNS	L158P
Cistinosis	CTNS	W138X
DFNA 9 (COCH)	COCH	P51S
Diabetes y Pérdida de audición	ADNmt	3234A>G
Diabetes y Pérdida de audición	ADNmt	3271T>C
Diabetes y Pérdida de audición	ADNmt	G8363A
Diabetes y Pérdida de audición	ADNmt	T14709C
Distonía primaria de inicio temprano (DYT1)	TOR1A	904_906delGAG
Distonía primaria de inicio temprano (DYT1)	TOR1A	904_906delGAG
Epidermolisis Bullosa juntural, tipo Herlitz-Pearson	LAMB3	3024delT

Epidermolisis Bullosa juntural, tipo Herlitz-Pearson	LAMB3	3024delT
Epidermolisis Bullosa juntural, tipo Herlitz-Pearson	LAMB3	P.Q243X
Epidermolisis Bullosa juntural, tipo Herlitz-Pearson	LAMB3	R144X
Epidermolisis Bullosa juntural, tipo Herlitz-Pearson	LAMB3	R42X
Epidermolisis Bullosa juntural, tipo Herlitz-Pearson	LAMB3	R635X
Epidermolisis Bullosa juntural, tipo Herlitz-Pearson	LAMA3	R650X
Epidermolisis Bullosa juntural, tipo Herlitz-Pearson	LAMC2	R95X
Factor V Leiden Trombofilia	F5	H1299R
Factor V Leiden Trombofilia	F5	R506Q, rs6025
Mutación del Factor V R2 Trombofilia	F5	rs6027
Deficiencia de factor XI	F11	E117X (576G>T)
Deficiencia de factor XI	F11	F283L (1074T>C)
Deficiencia de factor XI	F11	IVS14 +1G>A
Deficiencia de factor XI	F11	IVS14del14
Deficiencia de factor XI	F11	IVS14del14
Deficiencia de factor XIII	F13A1	V34L, rs5985
Poliposis Adenomatosa Familiar	APC	I1307K, rs1801155
Disautonomia familiar	IKBKAP	2507+6T>C
Disautonomia familiar	IKBKAP	P914L
Disautonomia familiar	IKBKAP	R696P
Hipercolesterolemia Familiar Tipo B	APOB	R3500Q, rs5742904
Hipercolesterolemia Familiar Tipo B	APOB	R3500W
Hipercolesterolemia Familiar Tipo B	APOB	R3531C, rs12713559
Fiebre mediterránea familiar	MEFV	A744S (2230G>T)
Fiebre mediterránea familiar	MEFV	del1692 (del2076_2078)
Fiebre mediterránea familiar	MEFV	del1692 (del2076_2078)
Fiebre mediterránea familiar	MEFV	E148Q (442 G>C), rs3743930
Fiebre mediterránea familiar	MEFV	E167D (501 G>C)
Fiebre mediterránea familiar	MEFV	F479L (1437 C>G)
Fiebre mediterránea familiar	MEFV	K695R (2084A>G)
Fiebre mediterránea familiar	MEFV	M680I (2040G>C)
Fiebre mediterránea familiar	MEFV	M694I (2082G>A), rs28940578
Fiebre mediterránea familiar	MEFV	M694V (2080A>G)
Fiebre mediterránea familiar	MEFV	P369S (1105 C>T), rs11466023
Fiebre mediterránea familiar	MEFV	R408Q (1223G>A), rs11466024
Fiebre mediterránea familiar	MEFV	R653H (1958G>A)
Fiebre mediterránea familiar	MEFV	R761H (2282G>A)
Fiebre mediterránea familiar	MEFV	T267I (800 C>T)
Fiebre mediterránea familiar	MEFV	V726A (2177T>C), rs28940579
Anemia de Fanconi relacionada con FANCC	FANCC	322delG
Anemia de Fanconi relacionada con FANCC	FANCC	322delG
Anemia de Fanconi relacionada con FANCC	FANCC	IVS4+4A>T (711 +4A>T)
Anemia de Fanconi relacionada con FANCC	FANCC	Q13X (37C>T)
Anemia de Fanconi relacionada con FANCC	FANCC	R547X
Craneosinostosis relacionada con FGFR1	FGFR1	P252R

Craneosinostosis relacionada con FGFR2	FGFR2	P253R
Craneosinostosis relacionada con FGFR2	FGFR2	S252W
Craneosinostosis relacionada con FGFR3	FGFR3	A391E, rs28931615
Craneosinostosis relacionada con FGFR3	FGFR3	P250R, rs4647924
Trastornos de almacenamiento de ácido siálico libre	SLC17A5	c.1007_1008delTA
Trastornos de almacenamiento de ácido siálico libre	SLC17A5	c.1007_1008delTA
Trastornos de almacenamiento de ácido siálico libre	SLC17A5	c.115C>T
Demencia frontotemporal con parkinsonismo-17	MAPT	IVS10+16
Demencia frontotemporal con parkinsonismo-17	MAPT	P301L
Demencia frontotemporal con parkinsonismo-17	MAPT	P301S
Demencia frontotemporal con parkinsonismo-17	MAPT	R406W
Deficiencia de fumarasa	FH	c.1431_1433dupAAA
Deficiencia de fumarasa	FH	c.1431_1433dupAAA
Galactosemia	GALT	deleción de génicos de 5,0Kb
Galactosemia	GALT	deleción de génicos de 5,0Kb
Galactosemia	GALT	5'UTR-119del
Galactosemia	GALT	5'UTR-119del
Galactosemia	GALT	IVS2-2 A>G
Galactosemia	GALT	K285N
Galactosemia	GALT	L195P T>C
Galactosemia	GALT	L218L
Galactosemia	GALT	N314D, rs2070074
Galactosemia	GALT	Phel71Ser
Galactosemia	GALT	Q169K
Galactosemia	GALT	Q188R
Galactosemia	GALT	S135L
Galactosemia	GALT	T138M C>T
Galactosemia	GALT	X380R
Galactosemia	GALT	Y209C A>G
Enfermedad de Gaucher	GBA	1035insG
Enfermedad de Gaucher	GBA	1035insG
Enfermedad de Gaucher	GBA	84insG
Enfermedad de Gaucher	GBA	84insG
Enfermedad de Gaucher	GBA	D409H, rs1064651
Enfermedad de Gaucher	GBA	D409V
Enfermedad de Gaucher	GBA	IVS2(+1)G>A
Enfermedad de Gaucher	GBA	L444P (1448T>C), rs35095275
Enfermedad de Gaucher	GBA	N370S
Enfermedad de Gaucher	GBA	R463C
Enfermedad de Gaucher	GBA	R463H
Enfermedad de Gaucher	GBA	R496H (1604G>A)
Enfermedad de Gaucher	GBA	V394L
Pérdida de Audición No sindrómica y Sordera DFNA 3 Relacionada con GJB2	GJB2	167delT
Pérdida de Audición No sindrómica y Sordera DFNA 3 Relacionada con GJB2	GJB2	167delT

Pérdida de Audición No sindrómica y Sordera DFNA 3 Relacionada con GJB2	GJB2	235delC
Pérdida de Audición No sindrómica y Sordera DFNA 3 Relacionada con GJB2	GJB2	235delC
Pérdida de Audición No sindrómica y Sordera DFNA 3 Relacionada con GJB2	GJB2	35delG
Pérdida de Audición No sindrómica y Sordera DFNA 3 Relacionada con GJB2	GJB2	35delG
Pérdida de Audición No sindrómica y Sordera DFNA 3 Relacionada con GJB2	GJB2	IVS1+1G>A
Pérdida de Audición No sindrómica y Sordera DFNB 1 Relacionada con GJB2	GJB2	101delAG
Pérdida de Audición No sindrómica y Sordera DFNB 1 Relacionada con GJB2	GJB2	313del14
Pérdida de Audición No sindrómica y Sordera DFNB 1 Relacionada con GJB2	GJB2	313del14
Pérdida de Audición No sindrómica y Sordera DFNB 1 Relacionada con GJB2	GJB2	delE120
Pérdida de Audición No sindrómica y Sordera DFNB 1 Relacionada con GJB2	GJB2	delE120
Pérdida de Audición No sindrómica y Sordera DFNB 1 Relacionada con GJB2	GJB2	M34T, rs35887622
Pérdida de Audición No sindrómica y Sordera DFNB 1 Relacionada con GJB2	GJB2	Q124X
Pérdida de Audición No sindrómica y Sordera DFNB 1 Relacionada con GJB2	GJB2	R184P
Pérdida de Audición No sindrómica y Sordera DFNB 1 Relacionada con GJB2	GJB2	V37I
Pérdida de Audición No sindrómica y Sordera DFNB 1 Relacionada con GJB2	GJB2	W24X
Pérdida de Audición No sindrómica y Sordera DFNB 1 Relacionada con GJB2	GJB2	W77R
Pérdida de Audición No sindrómica y Sordera DFNB 1 Relacionada con GJB2	GJB2	W77X
Deficiencia de Glucosa-6- fosfato Deshidrogenasa	G6PD	A335V
Deficiencia de Glucosa-6- fosfato Deshidrogenasa	G6PD	R459L
Deficiencia de Glucosa-6- fosfato Deshidrogenasa	G6PD	R459P
Deficiencia de Glucosa-6- fosfato Deshidrogenasa	G6PD	rs1050828, rs1050828
Deficiencia de Glucosa-6- fosfato Deshidrogenasa	G6PD	rs1050829, rs1050829
Deficiencia de Glucosa-6- fosfato Deshidrogenasa	G6PD	rs5030868, rs5030868
Aciduria Glutárica Tipo 1	GCDH	A421V
Aciduria Glutárica Tipo 1	GCDH	R402W

Enfermedad del almacenamiento del glucógeno Tipo 1a	G6PC	459insTA
Enfermedad del almacenamiento del glucógeno Tipo 1a	G6PC	459insTA
Enfermedad del almacenamiento del glucógeno Tipo 1a	G6PC	727G/T
Enfermedad del almacenamiento del glucógeno Tipo 1a	G6PC	del F327
Enfermedad del almacenamiento del glucógeno Tipo 1a	G6PC	del F327
Enfermedad del almacenamiento del glucógeno Tipo 1a	G6PC	G188R
Enfermedad del almacenamiento del glucógeno Tipo 1a	G6PC	G270V
Enfermedad del almacenamiento del glucógeno Tipo 1a	G6PC	Q242X
Enfermedad del almacenamiento del glucógeno Tipo 1a	G6PC	Q27fsdelC
Enfermedad del almacenamiento del glucógeno Tipo 1a	G6PC	Q27fsdelC
Enfermedad del almacenamiento del glucógeno Tipo 1a	G6PC	Q347X
Enfermedad del almacenamiento del glucógeno Tipo 1a	G6PC	R83C
Enfermedad del almacenamiento del glucógeno Tipo 1a	G6PC	R83H
Enfermedad del almacenamiento del glucógeno Tipo 1b	G6PT1	1211delCT
Enfermedad del almacenamiento del glucógeno Tipo 1b	G6PT1	A367T
Enfermedad del almacenamiento del glucógeno Tipo 1b	G6PT1	G339C
Enfermedad del almacenamiento del glucógeno Tipo 1b	G6PT1	G339D
Enfermedad del almacenamiento del glucógeno Tipo 1b	G6PT1	W118R
Enfermedad del almacenamiento del glucógeno Tipo II	GAA	Arg854X
Enfermedad del almacenamiento del glucógeno Tipo II	GAA	Asp645Glu, rs28940868
Enfermedad del almacenamiento del glucógeno Tipo II	GAA	IVS1(-13t>g)

Enfermedad del almacenamiento del glucógeno Tipo III	AGL	1484delT
Enfermedad del almacenamiento del glucógeno Tipo III	AGL	1484delT
Enfermedad del almacenamiento del glucógeno Tipo III	AGL	17delAG
Enfermedad del almacenamiento del glucógeno Tipo III	AGL	17delAG
Enfermedad del almacenamiento del glucógeno Tipo III	AGL	Q6X
Enfermedad del almacenamiento del glucógeno Tipo V	PYGM	G204S
Enfermedad del almacenamiento del glucógeno Tipo V	PYGM	K542T
Enfermedad del almacenamiento del glucógeno Tipo V	PYGM	K542X
Enfermedad del almacenamiento del glucógeno Tipo V	PYGM	R49X
Miopatías relacionadas con GNE	GNE	M712T, rs28937594
Síndrome Gracile	BCS1L	c.232A>G, rs28937590
Hemoglobina S Beta-Talasemia	HBB	c.19G>A
Hemoglobina S Beta-Talasemia	HBB	c.20A>T
Hemoglobina S Beta-Talasemia	HBB	c.79G>A
Hemoglobina S Beta-Talasemia	HBB	Hb CS
Hemoglobina S Beta-Talasemia	HBB	Hb D
Hemoglobina S Beta-Talasemia	HBB	Hb O
Intolerancia a la Fructosa Hereditaria	ALDOB	A149P, rs1800546
Intolerancia a la Fructosa Hereditaria	ALDOB	A174D
Intolerancia a la Fructosa Hereditaria	ALDOB	Delta4E4
Intolerancia a la Fructosa Hereditaria	ALDOB	Delta4E4
Intolerancia a la Fructosa Hereditaria	ALDOB	N334K
Intolerancia a la Fructosa Hereditaria	ALDOB	Y203X
Pancreatitis hereditaria	PRSS1	A16V
Pancreatitis hereditaria	SPINK1	M1T
Pancreatitis hereditaria	PRSS1	N29I
Pancreatitis hereditaria	SPINK1	N34S, rs17107315
Pancreatitis hereditaria	PRSS1	R122C
Pancreatitis hereditaria	PRSS1	R122H
Timina-Uracilouria	DPYD	rs3918290
Deficiencia de Hexosaminidasa A	HEXA	1278insTATC

Deficiencia de Hexosaminidasa A	HEXA	1278insTATC
Deficiencia de Hexosaminidasa A	HEXA	G269S (805G>A)
Deficiencia de Hexosaminidasa A	HEXA	IVS12 +1G>C
Deficiencia de Hexosaminidasa A	HEXA	IVS7+1G>A
Deficiencia de Hexosaminidasa A	HEXA	IVS9 +1G>A
Deficiencia de Hexosaminidasa A	HEXA	R178C
Deficiencia de Hexosaminidasa A	HEXA	R178H
Deficiencia de Hexosaminidasa A	HEXA	R247W (739C>T)
Deficiencia de Hexosaminidasa A	HEXA	R249W (745C>T)
Hemocromatosis hereditaria asociada a HFE	HFE	E168Q
Hemocromatosis hereditaria asociada a HFE	HFE	E168X
Hemocromatosis hereditaria asociada a HFE	HFE	HM971246, H63H
Hemocromatosis hereditaria asociada a HFE	HFE	P160delC
Hemocromatosis hereditaria asociada a HFE	HFE	P160delC
Hemocromatosis hereditaria asociada a HFE	HFE	Q127H, rs28934595
Hemocromatosis hereditaria asociada a HFE	HFE	Q283P
Hemocromatosis hereditaria asociada a HFE	HFE	rs1799945, rs1799945
Hemocromatosis hereditaria asociada a HFE	HFE	rs1800562,rs1800562
Hemocromatosis hereditaria asociada a HFE	HFE	rs1800730, rs1800730
Hemocromatosis hereditaria asociada a HFE	HFE	V53M
Hemocromatosis hereditaria asociada a HFE	HFE	V59M
Hemocromatosis hereditaria asociada a HFE	HFE	W169X
Displasia Ectodérmica Hidróica 2	GJB6	A88V, rs28937872
Displasia Ectodérmica Hidróica 2	GJB6	G11R
Displasia Ectodérmica Hidróica 2	GJB6	V37E
Homocistinuria Causada por la Deficiencia de Cistationina Beta-Sintasa	CBS	G307S 919G->A
Homocistinuria Causada por la Deficiencia de Cistationina Beta-Sintasa	CBS	I278T 833T->C, rs5742905
Parálisis Periódica Hipercalémica Tipo 1	SCN4A	I693T
Parálisis Periódica Hipercalémica Tipo 1	SCN4A	L689I
Parálisis Periódica Hipercalémica Tipo 1	SCN4A	L689V
Parálisis Periódica Hipercalémica Tipo 1	SCN4A	M1360V

Parálisis Periódica Hipercalémica Tipo 1	SCN4A	M1592V
Parálisis Periódica Hipercalémica Tipo 1	SCN4A	P.A1156T
Parálisis Periódica Hipercalémica Tipo 1	SCN4A	p.M1370V
Parálisis Periódica Hipercalémica Tipo 1	SCN4A	p.R1448C
Parálisis Periódica Hipercalémica Tipo 1	SCN4A	p.T1313M
Parálisis Periódica Hipercalémica Tipo 1	SCN4A	R675G
Parálisis Periódica Hipercalémica Tipo 1	SCN4A	R675Q
Parálisis Periódica Hipercalémica Tipo 1	SCN4A	R675W
Parálisis Periódica Hipercalémica Tipo 1	SCN4A	T704M
Parálisis Periódica Hipercalémica Tipo 1	SCN4A	V781I
El Síndrome de hiperornitinemia, homocitrulinuria e hiperamonemia	SLC25A15	F188del
El Síndrome de hiperornitinemia, homocitrulinuria e hiperamonemia	SLC25A15	F188del
Hiperoxaluria, Primaria, Type 1	AGXT	33insC
Hiperoxaluria, Primaria, Type 1	AGXT	33insC
Hiperoxaluria, Primaria, Type 1	AGXT	F152I
Hiperoxaluria, Primaria, Type 1	AGXT	G170R
Hiperoxaluria, Primaria, Type 1	AGXT	I244T
Hiperoxaluria, Primaria, Type 2	GRHPR	103delG
Hiperoxaluria, Primaria, Type 2	GRHPR	103delG
Hipocondroplasia	FGFR3	Asn328Ile
Hipocondroplasia	FGFR3	I538V
Hipocondroplasia	FGFR3	K650M
Hipocondroplasia	FGFR3	K650N 1950G>T
Hipocondroplasia	FGFR3	K650Q
Hipocondroplasia	FGFR3	N540K 1620C>A
Hipocondroplasia	FGFR3	N540S
Hipocondroplasia	FGFR3	N540T
Parálisis Periódica Hipocalémica Tipo 1	CACNA1S	R528G
Parálisis Periódica Hipocalémica Tipo 1	CACNA1S	R528H
Parálisis Periódica Hipocalémica Tipo 1	CACNA1S	rs28930068, rs28930068
Parálisis Periódica Hipocalémica Tipo 1	CACNA1S	rs28930069, rs28930069
Parálisis Periódica Hipocalémica Tipo 2	SCN4A	R669H
Parálisis Periódica Hipocalémica Tipo 2	SCN4A	R672C
Parálisis Periódica Hipocalémica Tipo 2	SCN4A	R672G

Parálisis Periódica Hipocalémica Tipo 2	SCN4A	R672H
Parálisis Periódica Hipocalémica Tipo 2	SCN4A	R672S
Hipofosfatasa	ALPL	Asp361Val
Hipofosfatasa	ALPL	c.1559delT
Hipofosfatasa	ALPL	c.1559delT
Hipofosfatasa	ALPL	E174K
Hipofosfatasa	ALPL	G317D
Hipofosfatasa	ALPL	Phe310Leu
Acidemia Isovalérica	IVD	A282V
Acidemia Isovalérica	IVD	rs28940889
Enfermedad de Krabbe	GALC	EX11-17DEL
Enfermedad de Krabbe	GALC	EX11-17DEL
Enfermedad de Krabbe	GALC	G270D
Enfermedad de Krabbe	GALC	rs1805078, rs1805078
Enfermedad de Krabbe	GALC	rs398607
Neuropatía Óptica Hereditaria De Leber	ADNmt	14484T>C
Neuropatía Óptica Hereditaria De Leber	ADNmt	15257G>A
Neuropatía Óptica Hereditaria De Leber	ADNmt	G14459A
Neuropatía Óptica Hereditaria De Leber	ADNmt	G3460A
Neuropatía Óptica Hereditaria De Leber	ADNmt	m.11778G>A
Neuropatía Óptica Hereditaria De Leber	ADNmt	m.13708G>A
Neuropatía Óptica Hereditaria De Leber	ADNmt	m.15812G>A
Neuropatía Óptica Hereditaria De Leber	ADNmt	m.3394T>C
Neuropatía Óptica Hereditaria De Leber	ADNmt	m.4216T>C
Neuropatía Óptica Hereditaria De Leber	ADNmt	m.4917A>G
Síndrome de Leigh, Tipo Franco-canadiense	LRPPRC	A354V
LGMD2I	FKRP	L276I, rs28937900
Deficiencia de 3-hidroxiacil- CoA deshidrogenasa de cadena larga	HADHA	E474Q c.1528G>C
Deficiencia de 3-hidroxiacil- CoA deshidrogenasa de cadena larga	HADHA	Q342X 1132C>T
Enfermedad Urinaria del Sirope de Arce Tipo 1A	BCKDHA	Y438N
Enfermedad Urinaria del Sirope de Arce Tipo 1B	BCKDHB	E372X
Enfermedad Urinaria del Sirope de Arce Tipo 1B	BCKDHB	G278S
Enfermedad Urinaria del Sirope de Arce Tipo 1B	BCKDHB	R183P
Síndrome de McCune- Albright	GNAS	R201C
Síndrome de McCune- Albright	GNAS	R201G
Síndrome de McCune- Albright	GNAS	R201H
Deficiencia de Acil- coenzima A Deshidrogenasa de cadena media	ACADM	244insT

Deficiencia de Acil-coenzima A Deshidrogenasa de cadena media	ACADM	244insT
Deficiencia de Acil-coenzima A Deshidrogenasa de cadena media	ACADM	250C>T
Deficiencia de Acil-coenzima A Deshidrogenasa de cadena media	ACADM	583G>A
Deficiencia de Acil-coenzima A Deshidrogenasa de cadena media	ACADM	616C>T
Deficiencia de Acil-coenzima A Deshidrogenasa de cadena media	ACADM	617G>A
Deficiencia de Acil-coenzima A Deshidrogenasa de cadena media	ACADM	799G>A
Deficiencia de Acil-coenzima A Deshidrogenasa de cadena media	ACADM	K304E
Deficiencia de Acil-coenzima A Deshidrogenasa de cadena media	ACADM	Y42H
Leucoencefalopatía megalencefálica con quistes subcorticales	MLC1	135insC
Leucoencefalopatía megalencefálica con quistes subcorticales	MLC1	135insC
MELAS	ADNmt	3243A>G
MELAS	ADNmt	3250T>C
MELAS	ADNmt	3252A>G
MELAS	ADNmt	A12770G
MELAS	ADNmt	C3256T
MELAS	ADNmt	G13513A
MELAS	ADNmt	T3291C
MELAS	ADNmt	T8356C
MELAS	ADNmt	T9957C
MERRF	ADNmt	8361G>A
MERRF	ADNmt	A8296G
MERRF	ADNmt	m.8344A>G
Leucodistrofia metacromática	ARSA	c.459+1G>A
Leucodistrofia metacromática	ARSA	P.P426L, rs28940893
Leucodistrofia metacromática	ARSA	p.T274M
Leucodistrofia metacromática	ARSA	P377L
Cardiomiopatía Mitocondrial	ADNmt	A3260T
Cardiomiopatía Mitocondrial	ADNmt	A4300G
Cardiomiopatía Mitocondrial	ADNmt	C3303T
Cardiomiopatía Mitocondrial	ADNmt	T9997C

Síndrome de Leigh y NARP asociado al ADN mitocondrial	ADNmt	5537insT
Síndrome de Leigh y NARP asociado al ADN mitocondrial	ADNmt	5537insT
Síndrome de Leigh y NARP asociado al ADN mitocondrial	ADNmt	8993T>C
Síndrome de Leigh y NARP asociado al ADN mitocondrial	ADNmt	8993T>G
Síndrome de Leigh y NARP asociado al ADN mitocondrial	ADNmt	C11777A
Síndrome de Leigh y NARP asociado al ADN mitocondrial	ADNmt	T10158C
Síndrome de Leigh y NARP asociado al ADN mitocondrial	ADNmt	T10191C
Síndrome de Leigh y NARP asociado al ADN mitocondrial	ADNmt	T8851C
Síndrome de Leigh y NARP asociado al ADN mitocondrial	ADNmt	T9176C
Síndrome de Leigh y NARP asociado al ADN mitocondrial	ADNmt	T9176G
Deficiencia de MTHFR	MTHFR	1298A>C
Deficiencia de MTHFR	MTHFR	rs1801133, rs1801133
Pérdida de Audición y Sordera Relacionada con MTRNR1	ADNmt	1095T>C
Pérdida de Audición y Sordera Relacionada con MTRNR1	ADNmt	1494C>T
Pérdida de Audición y Sordera Relacionada con MTRNR1	ADNmt	1555A>G
Pérdida de Audición y Sordera Relacionada con MTRNR1	ADNmt	961T>G
Pérdida de Audición y Sordera Relacionada con MTRNR1	ADNmt	A7445G
Pérdida de Audición y Sordera Relacionada con MTTS1	ADNmt	7443A>G
Pérdida de Audición y Sordera Relacionada con MTTS1	ADNmt	7444G>A
Pérdida de Audición y Sordera Relacionada con MTTS1	ADNmt	7472insC
Pérdida de Audición y Sordera Relacionada con MTTS1	ADNmt	7472insC
Pérdida de Audición y Sordera Relacionada con MTTS1	ADNmt	7510T>C
Pérdida de Audición y Sordera Relacionada con MTTS1	ADNmt	7511T>C

Pérdida de Audición y Sordera Relacionada con MTTS1	ADNmt	7512T>C
Mucopolipidosis IV	MCOLN1	delEx1 3 Ex7 (511>6944del)
Mucopolipidosis IV	MCOLN1	delEx1 3 Ex7 (511>6944del)
Mucopolipidosis IV	MCOLN1	IVS-2A>G
Mucopolisacaridosis Tipo I	IDUA	c.46_57del
Mucopolisacaridosis Tipo I	IDUA	c.46_57del
Mucopolisacaridosis Tipo I	IDUA	p.A327P
Mucopolisacaridosis Tipo T	IDUA	P.P533R
Mucopolisacaridosis Tipo I	IDUA	Q70X
Mucopolisacaridosis Tipo I	IDUA	W402X
Mucopolisacaridosis Tipo IIIA	SGSH	p.R245H
Mucopolisacaridosis Tipo IIIA	SGSH	p.R74C
Mucopolisacaridosis Tipo IIIA	SGSH	p.S66W
Mucopolisacaridosis Tipo VII	GUSB	p.D152N
Neoplasia Endocrina Múltiple Tipo 2	RET	2047T>A
Neoplasia Endocrina Múltiple Tipo 2	RET	2047T>A
Neoplasia Endocrina Múltiple Tipo 2	RET	2047T>C
Neoplasia Endocrina Múltiple Tipo 2	RET	2047T>G
Neoplasia Endocrina Múltiple Tipo 2	RET	2048G>A
Neoplasia Endocrina Múltiple Tipo 2	RET	A883F 2647 G>T
Neoplasia Endocrina Múltiple Tipo 2	RET	Glu768Asp G>C
Neoplasia Endocrina Múltiple Tipo 2	RET	M918T
Enfermedad músculo-ojo-cerebro	POMGNT1	c.1539+1G>A
Poliposis Asociada con MYH	MUTYH	c.1376C>A
Poliposis Asociada con MYH	MUTYH	c.494A>G, rs34612342
Poliposis Asociada con MYH	GENE_SYMBOL_TBD	rs36053993, rs36053993
Enfermedad de Niemann-Pick por Deficiencia de Esfingomielinasa	SMPD1	c.990delC
Enfermedad de Niemann-Pick por Deficiencia de Esfingomielinasa	SMPD1	c.990delC
Enfermedad de Niemann-Pick por Deficiencia de Esfingomielinasa	SMPD1	fsP330
Enfermedad de Niemann-Pick por Deficiencia de Esfingomielinasa	SMPD1	fsP330
Enfermedad de Niemann-Pick por Deficiencia de Esfingomielinasa	SMPD1	L302P
Enfermedad de Niemann-Pick por Deficiencia de Esfingomielinasa	SMPD1	R496L

Enfermedad de Niemann-Pick por Deficiencia de Esfingomielinasa	SMPD1	R608del
Enfermedad de Niemann-Pick Tipo C1	NPC1	I1061T
Síndrome de rotura de Nijmegen	NBN	657del5
Síndrome de rotura de Nijmegen	NBN	657del5
Síndrome de Pallister-Hall	GLI3	2012delG
Síndrome de Pallister-Hall	GLI3	2012delG
Síndrome de Pallister-Hall	GLI3	2023delG
Síndrome de Pallister-Hall	GLI3	2023delG
Síndrome de Pendred	SLC26A4	1197delT
Síndrome de Pendred	SLC26A4	1197delT
Síndrome de Pendred	SLC26A4	E384G
Síndrome de Pendred	SLC26A4	IV58+1(G->A)
Síndrome de Pendred	SLC26A4	L236P
Síndrome de Pendred	SLC26A4	T416P
Deficiencia enzimática bifuncional peroxisomal	HSD17B4	c.302+1G>C
Deficiencia enzimática bifuncional peroxisomal	HSD17B4	c.303-1G>A
Trastornos profundos del desarrollo	NLGN4X	D396X
Trastornos profundos del desarrollo	NLGN4X	D396X
Trastornos profundos del desarrollo	NLGN4X	NLGN4X: 1253delAG
Trastornos profundos del desarrollo	NLGN4X	NLGN4X: 1253delAG
Trastornos profundos del desarrollo	NLGN3	R451C
Deficiencia de Fenilalanina Hidroxilasa	PAH	G272X
Deficiencia de Fenilalanina Hidroxilasa	PAH	I65T
Deficiencia de Fenilalanina Hidroxilasa	PAH	IVS12+1G>T
Deficiencia de Fenilalanina Hidroxilasa	PAH	L48S, rs5030841
Deficiencia de Fenilalanina Hidroxilasa	PAH	R158Q, rs5030843
Deficiencia de Fenilalanina Hidroxilasa	PAH	R252W, rs5030847
Deficiencia de Fenilalanina Hidroxilasa	PAH	R261Q, rs5030849
Deficiencia de Fenilalanina Hidroxilasa	PAH	R408Q, rs5030859
Deficiencia de Fenilalanina Hidroxilasa	PAH	R408W, rs5030858
Deficiencia de Fenilalanina Hidroxilasa	PAH	rs5030855
Deficiencia de Fenilalanina Hidroxilasa	PAH	rs5030861
Deficiencia de Fenilalanina Hidroxilasa	PAH	Y414C, rs5030860
Inhibidor Activador del Plasminógeno I	SERPINE1	-844 G>A
Inhibidor Activador del Plasminógeno I	SERPINE1	4G/5G
Enfermedad Renal Poliquística, Autosómica Recesiva	PKHD1	c.10412T>G

Enfermedad Renal Poliquística, Autosómica Recesiva	PKHD1	c.107C>T
Enfermedad Renal Poliquística, Autosómica Recesiva	PKHD1	c.1486C>T
Enfermedad Renal Poliquística, Autosómica Recesiva	PKHD1	c.5895dupA
Enfermedad Renal Poliquística, Autosómica Recesiva	PKHD1	c.5895dupA
Enfermedad Renal Poliquística, Autosómica Recesiva	PKHD1	c.9689delA
Enfermedad Renal Poliquística, Autosómica Recesiva	PKHD1	c.9689delA
Ceroide lipofuscinosis neuronal relacionada con PPT1	PPT1	c.364A>T
Ceroide lipofuscinosis neuronal relacionada con PPT1	PPT1	p.L10X
Ceroide lipofuscinosis neuronal relacionada con PPT1	PPT1	p.R151X
Ceroide lipofuscinosis neuronal relacionada con PPT1	PPT1	T75P
Deficiencia de la hormona pituitaria relacionada con PROP1	PROP1	301-302delAG
Deficiencia de la hormona pituitaria relacionada con PROP1	PROP1	301-302delAG
Protrombina Trombofilia	F2	rs1799963
Protrombina Trombofilia	F2	rs6025, rs6025
Raquitismo por Deficiencia de Pseudovitamina D	CYP27B1	duplicación de 7bp en el exón8
Raquitismo por Deficiencia de Pseudovitamina D	CYP27B1	duplicación de 7bp en el exón8
Raquitismo por Deficiencia de Pseudovitamina D	CYP27B1	958delG
Raquitismo por Deficiencia de Pseudovitamina D	CYP27B1	958delG
Síndrome de Rett	MECP2	806delG
Síndrome de Rett	MECP2	806delG
Síndrome de Rett	MECP2	A140V, rs28934908
Síndrome de Rett	MECP2	P152R
Síndrome de Rett	MECP2	P225R
Síndrome de Rett	MECP2	R106W, rs28934907
Síndrome de Rett	MECP2	R133C
Síndrome de Rett	MECP2	R168X
Síndrome de Rett	MECP2	R255X
Síndrome de Rett	MECP2	R270X
Síndrome de Rett	MECP2	R294X
Síndrome de Rett	MECP2	R306C, rs28935468
Síndrome de Rett	MECP2	S134C
Síndrome de Rett	MECP2	T158M, rs28934906
Condrodisplasia Rizomélica Puntual Tipo 1	PEX7	P.A218V
Condrodisplasia Rizomélica Puntual Tipo 1	PEX7	P.G217R

ES 2 707 744 T3

Condrodisplasia Rizomélica Puntual Tipo 1	PEX7	p.L292X, rs1805137
Deficiencia de acil-CoA deshidrogenasa de cadena corta	ACADS	c.511C>T, rs1800556
Deficiencia de acil-CoA deshidrogenasa de cadena corta	ACADS	c.625G>A
Deficiencia de acil-CoA deshidrogenasa de cadena corta	ACADS	R107C
Síndrome de Shwachman-Diamond	SBDS	183_184TA>CT
Síndrome de Shwachman-Diamond	SBDS	183_184TA>CT
Síndrome de Shwachman-Diamond	SBDS	258+2T>C
Síndrome de Sjogren-Larsson	ALDH3A2	c.943C>T
Síndrome de Smith-Lemli-Opitz	DHCR7	C380Y
Síndrome de Smith-Lemli-Opitz	DHCR7	IVS8-1G>C
Síndrome de Smith-Lemli-Opitz	DHCR7	L109P
Síndrome de Smith-Lemli-Opitz	DHCR7	L157P
Síndrome de Smith-Lemli-Opitz	DHCR7	R352Q
Síndrome de Smith-Lemli-Opitz	DHCR7	R352W
Síndrome de Smith-Lemli-Opitz	DHCR7	R404C
Síndrome de Smith-Lemli-Opitz	DHCR7	R446Q
Síndrome de Smith-Lemli-Opitz	DHCR7	T93M
Síndrome de Smith-Lemli-Opitz	DHCR7	V326L
Síndrome de Smith-Lemli-Opitz	DHCR7	W151X
Síndrome de Smith-Lemli-Opitz	DHCR7	W151X
Paraplejia espástica 13	HSPD1	V72I
Osteocondrodisplasia Relacionada con el Transportador de Sulfato	SLC26A2	340delV
Osteocondrodisplasia Relacionada con el Transportador de Sulfato	SLC26A2	340delV
Osteocondrodisplasia Relacionada con el Transportador de Sulfato	SLC26A2	c.837C>T
Osteocondrodisplasia Relacionada con el Transportador de Sulfato	SLC26A2	C653S
Osteocondrodisplasia Relacionada con el Transportador de Sulfato	SLC26A2	IVS1+2T>C
Osteocondrodisplasia Relacionada con el Transportador de Sulfato	SLC26A2	R178X
Hemocromatosis Hereditaria Relacionada con TFR2	TFR2	AVAQ594-597del

Hemocromatosis Hereditaria Relacionada con TFR2	TFR2	AVAQ594-597del
Hemocromatosis Hereditaria Relacionada con TFR2	TFR2	E60X
Hemocromatosis Hereditaria Relacionada con TFR2	TFR2	E60X
Hemocromatosis Hereditaria Relacionada con TFR2	TFR2	M172K
Hemocromatosis Hereditaria Relacionada con TFR2	TFR2	Y250X
Displasia tanatofórica	FGFR3	G370C
Displasia tanatofórica	FGFR3	K650E
Displasia tanatofórica	FGFR3	R248C
Displasia tanatofórica	FGFR3	S249C
Displasia tanatofórica	FGFR3	S371C
Displasia tanatofórica	FGFR3	X807C A>T
Displasia tanatofórica	FGFR3	X807G
Displasia tanatofórica	FGFR3	X807L
Displasia tanatofórica	FGFR3	X807R
Displasia tanatofórica	FGFR3	X807S
Displasia tanatofórica	FGFR3	X807W
Displasia tanatofórica	FGFR3	Y373C
Ceroide lipofuscinosis neuronal relacionada con TPP1	TPP1	c.509-1G>A
Ceroide lipofuscinosis neuronal relacionada con TPP1	TPP1	c.509-1G>C
Ceroide lipofuscinosis neuronal relacionada con TPP1	TPP1	G284V
Ceroide lipofuscinosis neuronal relacionada con TPP1	TPP1	p.R208X
Amiloidosis Transtiretina	TTR	c.148G>A
Tirosina Hidroxilasa con deficiencia de DRD	TH	L205P
Tirosina Hidroxilasa con deficiencia de DRD	TH	R202H
Tirosinemia Tipo I	FAH	E357X
Tirosinemia Tipo I	FAH	IVS12+5 G>A
Tirosinemia Tipo I	FAH	IVS7-6 T>G
Tirosinemia Tipo I	FAH	IVS8-1G>C
Tirosinemia Tipo I	FAH	P.W262X
Tirosinemia Tipo I	FAH	P261L
Tirosinemia Tipo I	FAH	Q64H
Enfermedad de Wilson	ATP7B	1340del4
Enfermedad de Wilson	ATP7B	3402delC
Enfermedad de Wilson	ATP7B	3402delC
Enfermedad de Wilson	ATP7B	H1069Q
Enfermedad de Wilson	ATP7B	R778G
Enfermedad de Wilson	ATP7B	W779X
Enfermedad de Wilson	ATP7B	W779X
Retinosquiasis Juvenil Asociada a X	RS1	E72K
Retinosquiasis Juvenil Asociada a X	RS1	G109R
Retinosquiasis Juvenil Asociada a X	RS1	G74V
Síndrome de Zellweger Spectrum	PEX1	c.2097_2098insT
Síndrome de Zellweger Spectrum	PEX1	c.2097_2098insT

Síndrome de Zellweger Spectrum	PEX1	c.2916delA
Síndrome de Zellweger Spectrum	PEX1	c.2916delA
Síndrome de Zellweger Spectrum	PEX1	p.G843D

Tabla 5

Enfermedad/Trastorno	Gen(es)
Neoplasia	PTEN; ATM; ATR; EGFR; FRBB2; ERBB3; ERBB4; Notch1; Notch2; Notch3; Notch4; AKT; AKT2; AKT3; HIF; HIF1a; HIF3a; Met; HRG; Bcl2; PPAR alfa; PPAR gamma; WT1 (Tumor de Wilms); Miembros de la familia de receptores de FGF (5 miembros: 1, 2, 3, 4, 5); CDKN2a; APC; RB (retinoblastoma); MEN1: VHL; BRCA1; BRCA2; AR (Receptor de Andr6genos); TSG101; IGF; Receptor de IGF; Igf1(4 variantes); Igf2(3 variantes); Receptor de Igf 1; Receptor de Igf 2; Bax; Bcl2; familia de caspasas (9 miembros: 1, 2, 3, 4, 6, 7, 8, 9, 12); Kras; Apc
Degeneraci3n macular relacionada con la edad	Aber; Ccl2; Cc2; cp (ceruloplasmina); Timp3; catepsinaD; Vldlr; Ccr2
Esquizofrenia	Neuregulina1 (1); Erb4 (receptor para Neuregulina); Coraplexina1 (Cplx1); Tph1 Tript6fano Hidroxilasa; Tph2 Tript6fano Hidroxilasa 2; Neurexina 1; GSK3; GSK3a; GSK3b
Trastornos de la repetic3n del trinucle3tido	HTT (Dx de Huntington); SBMA/SMAX1/AR (Dx de Kennedy); FXN/X25 (Atazi de Friedrich); ATX3 (Dx de Machado-Joseph); ATXN1 y ATXN2 (ataxias espinocerebelares); DMPK (distrofia miot3nica); Atropfina-1 y Atn1 (Dx de DRPLA); CBP (inestabilidad global de Creb-BP); VLDLR (Alzheimer); Atxn7; Atxn10
S3ndrome de X fr3gil	FMR2; FXR1; FXR2; mGLUR5
Trastornos relacionado con Secretasa	APH-1 (alfa y beta); Presenilina (Psen1); nicastrina (Ncstn); PEN-2
Trastornos relacionados con priones	Prp
ALS	SOD1; ALS2; STEX; FUS; TARDBP; VEGF (VEGF-a;VEGF-b; VEGF-c)
Drogadicci3n	Prkce (alcohol); Drd2; Drd4; ABAT (alcohol); GRIA2; Grm5; Grin1; Htr1b; Grin2a; Drd3; Pdyn; Grial (alcohol)
Autismo	Mecp2; BZRAP1; MDGA2; Sema5A; Neurexina 1; X Fr3gil (FMR2 (AFF2); FXR1; FXR2; Mglur5)
Enfermedad de Alzheimer	E1; CHIP; UCH; UBB; Tau; LRP; PICALM; Clusterina; PS 1: SORL1; CR1; Vldlr; Uhal; Uba3; CHIP28 (Aqp1, Aquaporina 1); Uchl1; Uchl3; APP
Inflamaci3n	IL-10; IL-1 (IL-1a; IL-1b); IL-13; IL-17 (IL-17a (CTLA8); IL-17b; IL-17c; IL-17d; IL-17f); IL-23; Cx3cr1; ptpn22; TNFa; NOD2/CARD 15 para IBD; IL-6; IL-12 (IL-12a; IL-12b); CTLA4; Cx3c11
Enfermedad de Parkinson	x-Sinucle3na; DJ-1; LRRK2; Parkin; PINK1

En algunas realizaciones, un m3todo comprende adem3s la etapa de diagnosticar un sujeto bas3ndose en una etapa de identificaci3n, tal como diagnosticar al sujeto con una enfermedad asociada con una variante gen3tica causal detectada, o indicar una probabilidad de que el paciente tenga o vaya a desarrollar dicha enfermedad. Se proporcionan en el presente documento ejemplos de enfermedades, genes asociados y variantes de secuencia asociadas. En algunas realizaciones, se informa un resultado a trav3s de un generador de informes, tal como se describe en el presente documento.

En algunas realizaciones, una o m3s variantes gen3ticas causales son variantes de secuencia asociadas con un tipo o etapa particular de c3ncer, o de c3ncer que tiene una caracter3stica particular (por ejemplo, potencial metast3tico, resistencia a f3rmacos, capacidad de respuesta a f3rmacos). En algunas realizaciones, la divulgaci3n proporciona m3todos para la determinaci3n del pron3stico, tal como cuando se sabe que ciertas mutaciones est3n asociadas con los resultados del paciente. Por ejemplo, ADNct ha demostrado ser un mejor biomarcador para el pron3stico del c3ncer de mama que el ant3geno de c3ncer tradicional 53 (CA-53) y para la enumeraci3n de c3lulas tumorales circulantes (v3ase, por ejemplo, Dawson, et al., N Engl J Med 368:1199 (2013)). Adicionalmente, pueden usarse m3todos de la presente divulgaci3n en decisiones terap3uticas, orientaci3n y seguimiento, as3 como desarrollo y ensayos cl3nicos de terapias contra el c3ncer. Por ejemplo, la eficacia del tratamiento se puede monitorizar comparando muestras de ADNct de pacientes de antes, durante y despu3s del tratamiento con terapias particulares como terapias dirigidas moleculares (f3rmacos monoclonales), f3rmacos quimioterap3uticos, protocolos de radiaci3n, etc. o combinaciones de estos. Por ejemplo, el ADNct se puede monitorizar para ver si ciertas mutaciones aumentan o disminuyen, aparecen nuevas mutaciones, etc., despu3s del tratamiento, lo que puede permitir a un m3dico alterar un tratamiento (continuar, detener o cambiar el tratamiento, por ejemplo) en un per3odo mucho m3s corto de el tiempo que ofrecen los m3todos de monitorizaci3n que rastrean los s3ntomas del paciente. En algunas realizaciones, un m3todo comprende adem3s la etapa de diagnosticar un sujeto bas3ndose en una etapa de identificaci3n, como diagnosticar al sujeto con una

etapa o tipo de cáncer en particular asociado con una variante de secuencia detectada, o informar de una probabilidad de que el paciente tenga o vaya a desarrollar dicho cáncer.

Por ejemplo, para terapias dirigidas específicamente a pacientes basadas en marcadores moleculares (p. ej., estado de Herceptin y her2/neu), se evalúa a los pacientes para determinar si ciertas mutaciones están presentes en su tumor, y estas mutaciones pueden usarse para predecir la respuesta o la resistencia a la terapia y oriente la decisión de utilizar la terapia. Por lo tanto, la detección y el seguimiento de ADNct durante el curso del tratamiento pueden ser muy útiles para orientar las selecciones de tratamientos. Algunas mutaciones de cáncer primario (antes del tratamiento) o secundario (después del tratamiento) son responsables de la resistencia de los cánceres a algunos tratamientos (Misale et al., *Nature* 486 (7404): 532 (2012)).

Se conocen varias variantes de secuencia que están asociadas con uno o más tipos de cáncer que pueden ser útiles en el diagnóstico, pronóstico o decisiones de tratamiento. Las secuencias diana adecuadas de importancia oncológica que encuentran uso en los métodos de la divulgación incluyen, pero sin limitación, alteraciones en el gen de TP53, el gen AFR, el gen KRAS, el gen de PIK3CA, el gen BRAF, el gen EGFR y el gen KIT. Una secuencia diana que puede amplificarse específicamente y/o analizarse específicamente para detectar variantes de secuencia puede ser la totalidad o parte de un gen asociado con el cáncer. En algunas realizaciones, se identifican una o más variantes de secuencia en el gen TP53. TP53 es uno de los genes mutados con más frecuencia en los cánceres humanos, por ejemplo, las mutaciones TP53 se encuentran en el 45% de los cánceres de ovario, el 43% de los cánceres de intestino grueso y el 42% de los cánceres de la vía aerodigestiva superior (véase, por ejemplo, M. Olivier, et al. TP53 Mutations in Human Cancers: Origins, Consequences, and Clinical Use. *Cold of TP53 Spring Harb Perspect Biol.* 2010 January; 2(1)). La caracterización del estado de mutación de TP53 puede ayudar en el diagnóstico clínico, proporcionar valor pronóstico e influir en el tratamiento para los pacientes con cáncer. Por ejemplo, las mutaciones de TP53 se pueden usar como un predictor de un mal pronóstico para pacientes en tumores del SNC derivados de células gliales y un predictor de progresión rápida de la enfermedad en pacientes con leucemia linfocítica crónica (ver, por ejemplo, McLendon RE, et al. *Cancer*. 15 oct 2005; 104(8): 1693-9; Dicker F, et al. *Leukemia*. 2009 Jan; 23(1): 117-24). La variación de secuencia puede ocurrir en cualquier parte dentro del gen. Por lo tanto, la totalidad o parte del gen TP53 se puede evaluar en el presente documento. Es decir, como se describe en otra parte en el presente documento, cuando se usan componentes específicos de la diana (por ejemplo, cebadores específicos de diana), se puede usar una pluralidad de secuencias específicas de TP53, por ejemplo, para amplificar y detectar fragmentos que abarcan el gen, en lugar de solo una o más subsecuencias seleccionadas (como los "puntos calientes" de mutación) como se puede usar para dianas seleccionadas. Como alternativa, se pueden diseñar cebadores específicos de diana que hibridan cadena arriba o cadena abajo de una o más subsecuencias seleccionadas (por ejemplo, un nucleótido o una región de nucleótido asociada con un aumento de la tasa de mutación entre una clase de sujetos, que también abarca la expresión "punto caliente"). Se pueden diseñar cebadores convencionales que abarcan una subsecuencia de este tipo, y/o se pueden diseñar cebadores B2B que hibridan cadena arriba o cadena abajo de dicha subsecuencia.

En algunas realizaciones, se identifican una o más variantes de secuencia en la totalidad o parte del gen ALK. Se han indicado fusiones de ALK en hasta un 7% de los tumores de pulmón, algunos de los cuales están asociados con la resistencia al inhibidor de la tirosina quinasa (TKI) EGFR (véase, por ejemplo, Shaw et al., *J Clin Oncol.* 10 de Septiembre de 2009; 27(26): 4247-4253). Hasta 2013, se han encontrado varias mutaciones puntuales diferentes en todo el dominio ALK tirosina quinasa en pacientes con resistencia secundaria al inhibidor de la tirosina quinasa ALK (TKI) (Katayama R 2012 *Sci Transl Med.* 8 de Febrero de 2012; 4(120)). Por lo tanto, la detección de mutaciones en el gen ALK se puede usar para ayudar a tomar decisiones sobre el tratamiento del cáncer.

En algunas realizaciones, se identifican una o más variantes de secuencia en la totalidad o parte del gen KRAS. Se ha indicado que aproximadamente el 15-25% de los pacientes con adenocarcinoma de pulmón y el 40% de los pacientes con cáncer colorrectal albergan mutaciones de KRAS asociadas a tumores (véase, por ejemplo, Neuman 2009, *Pathol Res Pract.* 2009; 205(12): 858-62). La mayoría de las mutaciones se encuentran en los codones 12, 13 y 61 del gen KRAS. Estas mutaciones activan las vías de señalización de KRAS, que desencadenan el crecimiento y la proliferación de células tumorales. Algunos estudios indican que es poco probable que los pacientes con tumores que albergan mutaciones en KRAS se beneficien de la terapia con anticuerpos anti-EGFR solo o en combinación con quimioterapia (véase, por ejemplo, Amado et al. 2008 *J Clin. col.*, 1 de abril de 2008; 26(10): 1626-34, Bokemeyer et al. 2009 *J Clin Oncol.* 10 de Febrero de 2009; 27(5): 663-71). Un "punto caliente" particular para la variación de secuencia que puede ser diana para identificar la variación de secuencia es la posición 35 del gen. La identificación de variantes de la secuencia KRAS se puede utilizar en la selección de tratamientos, tal como en la selección de tratamientos para un sujeto con cáncer colorrectal.

En algunas realizaciones, se identifican una o más variantes de secuencia en la totalidad o parte del gen PIK3CA. Las mutaciones somáticas en PIK3CA se han encontrado con frecuencia en varios tipos de cáncer, por ejemplo, en el 10-30% de los cánceres colorrectales (véase, por ejemplo, Samuels et al. 2004 *Science.* 23 de abril de 2004; 304(5670): 554). Estas mutaciones se ubican más comúnmente dentro de dos áreas de "punto caliente" dentro del exón 9 (el dominio helicoidal) y el exón 20 (el dominio quinasa), que pueden ser específicamente dirigidas a la amplificación y/o análisis de las variantes de secuencia de detección. La posición 3140 también puede ser específicamente dirigida.

En algunas realizaciones, se identifican una o más variantes de secuencia en la totalidad o parte del gen BRAF. Se ha indicado que cerca del 50% de todos los melanomas malignos albergan mutaciones somáticas en BRAF (véase, por ejemplo, Maldonado et al., J Natl Cancer Inst. 17 diciembre de 2003;95(24): 1878-90). Las mutaciones de BRAF se encuentran en todos los subtipos de melanoma, pero son más frecuentes en los melanomas derivados de la piel sin daño crónico inducido por el sol. Entre las mutaciones de BRAF más comunes en el melanoma están las mutaciones sin sentido V600E, que sustituyen la valina en la posición 600 con glutamina. Las mutaciones V600E de BRAF están asociadas con el beneficio clínico de la terapia con inhibidores de BRAF. La detección de la mutación de BRAF se puede utilizar en la selección del tratamiento del melanoma y en los estudios de resistencia a la terapia dirigida.

En algunas realizaciones, se identifican una o más variantes de secuencia en la totalidad o parte del gen EGFR. Las mutaciones de EGFR se asocian frecuentemente con el cáncer de pulmón no microcítico (aproximadamente el 10% en los EE. UU. y el 35% en Asia oriental; Véase, por ejemplo, Pao et al., Proc Natl Acad Sci US A. 7 de Septiembre de 2004;101(36):13306-11). Estas mutaciones ocurren normalmente dentro de los exones de EGFR 18-21, y generalmente son heterocigotas. Aproximadamente el 90% de estas mutaciones son delecciones del exón 19 o mutaciones puntuales del exón 21 L858R.

En algunas realizaciones, se identifican una o más variantes de secuencia en la totalidad o parte del gen KIT. Se ha indicado que cerca del 85% de los tumores del estroma gastrointestinal (GIST) albergan mutaciones en KIT (véase, por ejemplo, Heinrich et al. 2003 J Clin Oncol. 1 de diciembre de 2003;21(23):4342-9). La mayoría de las mutaciones de KIT se encuentran en el dominio yuxtamembrana (exón 11, 70%), motivo de dimerización extracelular (exón 9, 10-15%), dominio de tirosina quinasa I (TKI) (exón 13, 1-3%) y dominio de tirosina quinasa 2 (TK2) y bucle de activación (exón 17, 1-3%). Las mutaciones secundarias de KIT se identifican comúnmente después de la terapia de diana con imatinib y después de que los pacientes hayan desarrollado resistencia a la terapia.

Los ejemplos adicionales no limitantes de genes asociados con el cáncer, todos o una parte de los cuales pueden analizarse para detectar variantes de secuencia de acuerdo con un método descrito en el presente documento incluyen, pero sin limitación, PTEN; ATM; ATR; EGFR; ERBB2; ERBB3; ERBB4; Notch1; Notch2; Notch3; Notch4; AKT; AKT2; AKT3; HIF; HIF1a; HIF3a; Met; HRG; Bcl2; PPAR alfa; PPAR gamma; WT1 (Tumor de Wilms); Miembros de la familia de receptores de FGF (5 miembros: 1, 2, 3, 4, 5); CDKN2a; APC; RB (retinoblastoma); MEN1; VHL; BRCA1; BRCA2; AR; (Receptor de Andrógenos); TSG101; IGF; Receptor de IGF; Igf1(4 variantes); Igf2(3 variantes); Receptor de Igf 1; Receptor de Igf 2; Bax; Bcl2; familia de caspasas (9 miembros: 1, 2, 3, 4, 6, 7, 8, 9, 12); Kras; y Apc. Se proporcionan ejemplos adicionales en otra parte en el presente documento. Los ejemplos de cánceres que se pueden diagnosticar basándose en la identificación de una o más variantes de secuencia de acuerdo con un método descrito en el presente documento incluyen, sin limitación, Acanthoma, Carcinoma de células Acínicas, Neuroma acústico, Melanoma acral lentiginoso, Acrospiroma, Leucemia eosinofílica aguda, Leucemia linfoblástica aguda, Leucemia megacarioblástica aguda, Leucemia monocítica aguda, Leucemia mieloblástica aguda con maduración, Leucemia mieloide aguda de células dendríticas, Leucemia mieloide aguda, Leucemia promielocítica aguda, Adamantinoma, Adenocarcinoma, Carcinoma adenoideo quístico, Adenoma, Tumor odontogénico adenomatoide, Carcinoma adrenocortical, Leucemia de linfocitos T del adulto, Leucemia agresiva de células NK, Cánceres relacionados con SIDA, Linfoma relacionado con SIDA, Sarcoma alveolar de las partes blandas, Fibroma ameloblástico, Cáncer de ano, Linfoma anaplásico de células grandes, Cáncer de tiroides anaplásico, Linfoma angioinmunoblástico de linfocitos T, Angiomolipoma, Angiosarcoma, Cáncer de apéndice, Astrocitoma, Tumor teratoideo/rabdoide atípico, Carcinoma de células basales, Carcinoma de tipo basaloide, Leucemia de linfocitos B, Linfoma de linfocitos B, Carcinoma del conducto biliar, Cáncer de las vías biliares, Cáncer de vejiga, Blastoma, Cáncer óseo, Tumor óseo, Glioma del tronco encefálico, Tumor cerebral, Cáncer de mama, Tumor de Brenner, Tumores bronquial, Carcinoma bronquioalveolar, Tumor pardo, Linfoma de Burkitt, Cáncer de sitio primario desconocido, Tumor carcinoide, Carcinoma, Carcinoma in situ, Carcinoma del pene, Carcinoma de sitio primario desconocido, Carcinosarcoma, Enfermedad de Castleman, Tumor embrionario del sistema nervioso central, Astrocitoma cerebeloso, Astrocitoma cerebral, Cáncer de cuello uterino, Colangiocarcinoma, Condroma, Condrosarcoma, Cordoma, Coriocarcinoma, Papiloma del plexo coroideo, Leucemia linfocítica crónica, Leucemia monocítica crónica, Leucemia mielógena crónica, Trastorno mieloproliferativo crónico, Leucemia neutrofílica crónica, Tumor de células claras, Cáncer de colon, Cáncer colorrectal, Craneofaringioma, Linfoma cutáneo de linfocitos T, Enfermedad de Degos, Dermatofibrosarcoma protuberans, Quiste dermoide, Tumor desmoplásico de células redondas pequeñas, Linfoma difuso de linfocitos B grandes, Tumor neuroepitelial disembrionario, Carcinoma embrionario, Tumor del seno endodérmico, Cáncer endometrial, Cáncer endometrial uterino, Tumor endometriode, Linfoma de linfocitos T asociado a enteropatía, Ependimoblastoma, Ependimoma, Sarcoma epiteloide, Eritroleucemia, Cáncer de esófago, Estesioneuroblastoma, Familia de tumor de Ewing, Familia del sarcoma de Ewing, Sarcoma de Ewing, Tumor de células germinales extracraneales, Tumor de células germinales extragonadales, Cáncer del conducto biliar extrahepático, Enfermedad de Paget extramamaria, Cáncer de las trompas de Falopio, Fetus in fetu, Fibroma, Fibrosarcoma, Linfoma folicular, Cáncer de tiroides folicular, Cáncer de vesícula biliar, Cáncer de vesícula biliar, Ganglioglioma, Ganglioneuroma, Cáncer gástrico, Linfoma gástrico, Cáncer gastrointestinal, Tumor carcinoide gastrointestinal, Tumor estromal gastrointestinal, Tumor de células germinales, Germinoma, Coriocarcinoma gestacional, Tumor trofoblástico gestacional, Tumor de células gigantes, Glioblastoma multiforme, Glioma, Gliomatosis cerebri, Tumor del glomus, Glucagonoma, Gonadoblastoma, Tumor de células de la granulosa, Leucemia de células pilosas, Leucemia de células pilosas, Cáncer de cabeza y cuello, Cáncer de cabeza y cuello, Cáncer cardíaco,

Hemangioblastoma, Hemangiopericitoma, Hemangiosarcoma, Neoplasia hematológica, Carcinoma hepatocelular, Linfoma hepatosplénico de linfocitos T, Síndrome hereditario de cáncer de mama y ovario, Linfoma de Hodgkin, Linfoma de Hodgkin, Cáncer hipofaríngeo, Glioma hipotalámico, Cáncer de mama inflamatorio, Melanoma intraocular, Carcinoma de células de los islotes, Tumores de células de los islotes, Leucemia mielomonocítica juvenil, Sarcoma de Kaposi, Sarcoma de Kaposi, Cáncer de riñón, Tumor de Klatzkin, Tumor de Krukenberg, Cáncer de laringe, Cáncer de laringe, Lentigo maligno melanoma, Leukemia, Cáncer de labio y de la cavidad oral, Liposarcoma, Cáncer de pulmón, Luteoma, Linfangioma, Linfangiosarcoma, Linfopitelioma, Leucemia linfóide, Linfoma, Macroglobulinemia, Histiocitoma fibroso maligno, Histiocitoma fibroso maligno, Histiocitoma fibroso maligno de hueso, Glioma maligno, Mesotelioma maligno, Tumor maligno de la vaina del nervio periférico, Tumor rabdoide maligno, Tumor tritón maligno, Linfoma MALT, Linfoma de células del manto, Leucemia de mastocitos, Tumor de células germinales mediastinales, Tumor mediastinal, Cáncer medular de tiroides, Meduloblastoma, Meduloepitelioma, Melanoma, Meningioma, Carcinoma de células de Merkel, Mesotelioma, Cáncer de cuello escamoso metastásico con tumor primario oculto, Carcinoma urotelial metastásico, Tumor mulleriano mixto, Leucemia monocítica, Cáncer de boca, Cáncer mucinoso, Síndrome de neoplasia endocrina múltiple, Mieloma múltiple, Micosis fungoides, Enfermedad mielodisplásica, Síndromes mielodisplásicos, Leucemia mieloide, Sarcoma mieloide, Enfermedad mieloproliferativa, Mixoma, Cáncer de la cavidad nasal, Cáncer nasofaríngeo, Carcinoma nasofaríngeo, Neoplasia, Neurinoma, Neuroblastoma, Neuroblastoma, Neurofibroma, Neuroma, Melanoma nodular, Linfoma no de Hodgkin, Cáncer de piel no melanoma, Cáncer de pulmón no microcítico, Oncología ocular, Oligoastrocitoma, Oligodendroglioma, Oncocitoma, Meningioma de la vaina del nervio óptico, Cáncer oral, Cáncer oral, Cáncer orofaríngeo, Osteosarcoma, Osteosarcoma, Cáncer de ovarios, Cáncer de ovarios, Cáncer de ovarios epitelial, Tumor de células germinales ováricas, Tumor ovárico de bajo potencial maligno, Enfermedad de Paget de mama, Tumor de Pancoast, Cáncer de páncreas, Cáncer de páncreas, Cáncer papilar de tiroides, Papilomatosis, Paraganglioma, Cáncer de seno paranasal, Cáncer paratiroideo, Cáncer de pene, Tumor de células epiteloides perivascuales, Cáncer de faringe, Feocromocitoma, Tumor del parénquima pineal de diferenciación intermedia, Pineoblastoma, Pituicitoma, Adenoma pituitario, Tumor pituitario, Neoplasia de células plasmáticas, Blastoma pleuropulmonar, Poliembrioma, Linfoma linfoblástico de precursores T, Linfoma primario del sistema nervioso central, Linfoma de derrame primario, Cáncer hepatocelular primario, Cáncer de hígado primario, Cáncer peritoneal primario, Tumor neuroectodérmico primitivo, Cáncer de próstata, Pseudomixoma peritoneal, Cáncer de recto, Carcinoma de células renales, Carcinoma de las vías respiratorias que involucra el gen NUT en el cromosoma 15, Retinoblastoma, Rabdomioma, Rabdomiosarcoma, Transformación de Richter, Teratoma sacrococcígeo, Cáncer de las glándulas salivales, Sarcoma, Schwannomatosis, Carcinoma de las glándulas sebáceas, Neoplasia secundaria, Seminoma, Tumor seroso, Tumor de células de Sertoli-Leydig, Tumor del estroma del cordón sexual, Síndrome de Sezary, Carcinoma de células de anillo de sello, Cáncer de piel, Tumor de células pequeñas redondas y azules, Carcinoma microcítico, Cáncer de pulmón microcítico, Linfoma de células pequeñas, Cáncer del intestino delgado, Sarcoma de tejidos blandos, Somatostatina, Verruga del hollín, Tumor de la médula espinal, Tumor espinal, Linfoma esplénico de la zona marginal, Carcinoma de células escamosas, Cáncer de estómago, Melanoma de diseminación superficial, Tumor neuroectodérmico primitivo supratentorial, Tumor de superficie epitelial-estromal, Sarcoma sinovial, Leucemia linfoblástica aguda de linfocitos T, Leucemia de linfocitos T granulares grandes, Leucemia de linfocitos T, Linfoma de linfocitos T, Leucemia prolinfocítica de linfocitos T, Teratoma, Cáncer linfático terminal, Cáncer de testículos, Tecoma, Cáncer de garganta, Carcinoma tímico, Timoma, Cáncer de tiroides, Cáncer de células transicionales de la pelvis renal y el uréter, Carcinoma de células transicionales, Cáncer de uraco, Cáncer uretral, Neoplasia urogenital, Sarcoma uterino, Melanoma uveal, Cáncer de vagina, Síndrome de Verner Morrison, Carcinoma verrucoso, Glioma de la vía visual, Cáncer de vulva, Macroglobulinemia de Waldenstrom, Tumor de Warthin, Tumor de Wilms y combinaciones de los mismos. En la Tabla 6 se proporcionan ejemplos no limitantes de variantes de secuencia específicas asociadas con el cáncer.

Tabla 6

Gen	Mutación Aminoácido Nucleótido	% de mutación COSMIC en melanoma	Relevancia clínica
BRAF	V600E 1799T>A	50 % (80-90 % entre mutaciones de BRAF)	Aumentan la sensibilidad a los inhibidores de BRAF Aumentan la sensibilidad a los inhibidores de MEK
BRAF	V600E 1799_1800delTGinsAA	50 % (80-90 % entre mutaciones de BRAF)	Aumentan la sensibilidad a los inhibidores de BRAF Aumentan la sensibilidad a los inhibidores de MEK
BRAF	V600R	50 % (<5 % entre mutaciones de BRAF)	Aumenta la sensibilidad a BRAF
	1798_1799delGTinsAG		los inhibidores responden a los inhibidores de MEK
BRAF	V600M 1798G>A	50 % (<1 % entre mutaciones de BRAF)	Aumentan la sensibilidad a los inhibidores de BRAF Aumentan la sensibilidad a los inhibidores de MEK
BRAF	V600K 1798_1799delGTinsAA	50 % (5 % entre mutaciones de BRAF)	Aumentan la sensibilidad a los inhibidores de BRAF Aumentan la sensibilidad a los inhibidores de MEK

BRAF	V600G 1799T>G	50 % (<1 % entre mutaciones de BRAF)	Aumentan la sensibilidad a los inhibidores de BRAF Responden a los inhibidores de MEK
BRAF	V600D 1799-800delTGinsAT	50 % (<5 % entre mutaciones de BRAF)	Aumentan la sensibilidad a los inhibidores de BRAF Aumentan la sensibilidad a los inhibidores de MEK
BRAF	L597V 1789C>G	50 % (1 % entre mutaciones de BRAF)	Responden a los inhibidores de BRAF Responden a los inhibidores de MEK
BRAF	L597S 1789_1790delCTinsTC	50 % (<1 % entre mutaciones de BRAF)	Responden a los inhibidores de BRAF Responden a los inhibidores de MEK
BRAF	L597Q 1790T>A	50 % (<1 % entre mutaciones de BRAF)	Responden a los inhibidores de BRAF Responden a los inhibidores de MEK
BRAF	L597R 1790T>G	50 % (<1 % entre mutaciones de BRAF)	Responden a los inhibidores de BRAF Responden a los inhibidores de MEK
BRAF	D594N 1799_1780delTGinsGA	50 % (<1 % entre mutaciones de BRAF)	No a los inhibidores de BRAF Pueden responder a los inhibidores de MEK
BRAF	D594H 1780C>G	50 % (<1 % entre mutaciones de BRAF)	No a los inhibidores de BRAF Pueden responder a los inhibidores de MEK
BRAF	D594E 1782T>A	50 % (<1 % entre mutaciones de BRAF)	No a los inhibidores de BRAF Pueden responder a los inhibidores de MEK
BRAF	D594E 1782T>G	50 % (<1 % entre mutaciones de BRAF)	No a los inhibidores de BRAF Pueden responder a los inhibidores de MEK
BRAF	D594G 1781A>G	50 % (<1 % entre mutaciones de BRAF)	No a los inhibidores de BRAF Pueden responder a los inhibidores de MEK
BRAF	D594V 1781A>T	50 % (<1 % entre mutaciones de BRAF)	No a los inhibidores de BRAF Pueden responder a los inhibidores de MEK
NRAS	G12D 35G>A	13 -25 % (4 % entre mutaciones de NRAS)	quimioterapia citotóxica
NRAS	G13R 37G>C	13 -25 % (2 % entre mutaciones de NRAS)	quimioterapia citotóxica No a los inhibidores de BRAF Pueden responder a los inhibidores de MEK
NRAS	G13D 38G>A	13 -25 % (2 % entre mutaciones de NRAS)	quimioterapia citotóxica No a los inhibidores de BRAF Pueden responder a los inhibidores de MEK
NRAS	G13V 38G>T	13 -25 % (2 % entre mutaciones de NRAS)	quimioterapia citotóxica No a los inhibidores de BRAF Pueden responder a los inhibidores de MEK
NRAS	Q61K 181C>A	13 -25 % (34 % entre mutaciones de NRAS)	Responden a los inhibidores de MEK
NRAS	Q61L 182A>T	13 -25 % (8 % entre mutaciones de NRAS)	Responden a los inhibidores de MEK
NRAS	Q61R 182A>G	13 -25 % (35 % entre mutaciones de NRAS)	Responden a los inhibidores de MEK
NRAS	Q61H 183A>C	13 -25 % (2 % entre mutaciones de NRAS)	Responden a los inhibidores de MEK
NRAS	Q61H 183A>T	13 -25 % (2 % entre mutaciones de NRAS)	Responden a los inhibidores de MEK
CTNNB1	S37F 110C>T	~ 2-3% en el melanoma uveal primario (46% entre las mutaciones de CTNNB1)	
CNA11	Q209L 626A>T	34% en melanoma uveal primario (92% entre mutaciones de CNA11)	
CNA11	Q209P 626A>C	34% en melanoma uveal primario (1 % entre mutaciones de CNA11)	
GNAQ	Q209L 626A>T	~ 50 % en melanoma uveal primario (~33 % entre mutaciones de GNAQ)	Sensible a los inhibidores de MEK
GNAQ	Q209P 626A>C	~ 50 % en melanoma uveal primario (~64 % entre mutaciones de GNAQ)	Sensible a los inhibidores de MEK Sensible a los inhibidores de MEK

GNAQ	Q209R 626A>G	~ 50 % en melanoma uveal primario (~2 % entre mutaciones de CNAQ)	
KIT	K624E 1924A>G	~ 20% entre melanomas malignos mutantes de KIT	Aumenta la sensibilidad a inhibidor de KIT
KIT	D816H 2446G>C	~ 5 % entre melanomas malignos mutantes de KIT	Aumenta la sensibilidad a inhibidor de KIT
KIT	L567P 1727T>C	~ 25 % entre melanomas malignos mutantes de KIT	Aumenta la sensibilidad a inhibidor de KIT
KIT	V559A 1676T>C	~ 20% entre melanomas malignos mutantes de KIT	Aumenta la sensibilidad a inhibidor de KIT
KIT	V559D 1676T>A	~ 5 % entre melanomas malignos mutantes de KIT	Aumenta la sensibilidad a inhibidor de KIT
KIT	W557R 1669T>C	~ 10 % entre melanomas malignos mutantes de KIT	Aumenta la sensibilidad a inhibidor de KIT
KIT	W557R 1669T>A	~ 10 % entre melanomas malignos mutantes de KIT	Aumenta la sensibilidad a inhibidor de KIT

Además, los métodos y composiciones descritos en el presente documento pueden ser útiles para descubrir mutaciones nuevas y raras que están asociadas con uno o más tipos de cáncer, etapas o características del cáncer. Por ejemplo, las poblaciones de individuos que comparten una característica en análisis (por ejemplo, una enfermedad particular, tipo de cáncer, etapa del cáncer, etc.) pueden someterse a un método de detección de variantes de secuencia de acuerdo con la divulgación para identificar variantes de secuencia o tipos de variantes de secuencia (por ejemplo, mutaciones en genes particulares o partes de genes). Las variantes de secuencia identificadas como que ocurren con una frecuencia estadísticamente significativamente mayor entre el grupo de individuos que comparten la característica que entre individuos sin la característica, se les pueden asignar un grado de asociación con esa característica. Las variantes de secuencia o los tipos de variantes de secuencia así identificadas se pueden usar, por lo tanto, para diagnosticar o tratar a los individuos que se ha descubierto que las albergan.

Otras aplicaciones terapéuticas incluyen el uso en diagnósticos fetales no invasivos. El ADN fetal se puede encontrar en la sangre de una mujer embarazada. Los métodos y composiciones descritos en el presente documento se pueden usar para identificar variantes de secuencia en el ADN fetal circulante y, por lo tanto, se pueden usar para diagnosticar una o más enfermedades genéticas en el feto, como las asociadas con una o más variantes genéticas causales. Los ejemplos no limitantes de variantes genéticas causales se describen en el presente documento, e incluyen trisomías, fibrosis quística, anemia de células falciformes y enfermedad de Tay-Saks. En esta realización, la madre puede proporcionar una muestra de control y una muestra de sangre que se utilizará para la comparación. La muestra de control puede ser cualquier tejido adecuado, y normalmente se procesará para extraer ADN celular, que a continuación, se puede secuenciar para proporcionar una secuencia de referencia. Las secuencias de ADNcf correspondientes al ADN genómico fetal se pueden identificar como variantes de secuencia con respecto a la referencia materna. El padre también puede proporcionar una muestra de referencia para ayudar a identificar secuencias fetales y variantes de secuencia.

Otras aplicaciones terapéuticas adicionales incluyen la detección de polinucleótidos exógenos, tal como los patógenos (por ejemplo, bacterias, virus, hongos y microbios), cuya información puede informar un diagnóstico y selección de tratamiento. Por ejemplo, algunos subtipos de VIH se correlacionan con la resistencia a fármacos (véase, por ejemplo, hivdb.stanford.edu/pages/genotype-rx). De manera similar, También se pueden realizar tipificaciones de HCV, subtipos y mutaciones de isotipos utilizando los métodos y composiciones de la presente divulgación. Por otra parte, Cuando un subtipo de HPV se correlaciona con un riesgo de cáncer cervical, tal diagnóstico puede proporcionar una evaluación adicional del riesgo de cáncer. Otros ejemplos no limitantes de virus que pueden detectarse incluyen el hepadnavirus del virus de la hepatitis B (HBV), virus de la hepatitis de la marmota, virus de la hepatitis de la Ardilla de tierra (Hepadnaviridae), virus de la hepatitis B de pato, virus de la hepatitis B de garza, Herpesvirus del virus del herpes simple (HSV) tipos 1 y 2, virus Varicela zóster, citomegalovirus (CMV), citomegalovirus humano (CMVH), citomegalovirus de ratón (MCMV), citomegalovirus del cerdo de guinea (GPCMV), virus Epstein Barr (EBV), virus del herpes humano 6 (variantes A y B del HHV), virus del herpes humano 7 (HHV-7), virus del herpes humano 8 (HHV-8), Virus del herpes asociado al sarcoma de Kaposi (KSHV), Virus B de virus vaccinia Poxvirus, virus variola, virus de la viruela, virus de la viruela del mono, virus de la viruela de vaca, virus de la viruela del camello, virus de la ectromelia, virus de la viruela del ratón, virus de la viruela del conejo, virus de la viruela del mapache, virus del molusco contagioso, virus orf, virus de los nodos del lechero, virus de la estomatitis papular bovina, virus de la viruela ovina, virus de la viruela caprina, virus de la enfermedad de la piel gruesa, virus de la viruela aviar, virus de la viruela del canario, virus de la viruela de la paloma, virus de la viruela del gorrión, virus del mixoma, virus del fibroma de liebre, virus del fibroma de conejo, virus del fibroma de ardilla, virus suínpox, virus tanapox, Virus Yabapox, Virus del dengue Flavivirus, virus de la hepatitis C (HCV), Virus de la hepatitis GB (GBV-A, GBV-B y GBV-C), virus del Nilo Occidental, virus de la fiebre amarilla, Virus de la encefalitis de San Luis, virus de la encefalitis japonesa, Virus de Powassan, virus de la encefalitis transmitida por garrapatas, virus de la enfermedad de la selva de Kyasanur, Togavirus, Virus de la encefalitis equina venezolana (VEE), virus chikungunya, Virus del río ross, Virus mayaro, Virus Sindbis, virus de la rubeola, virus

retrovirus de inmunodeficiencia humana (VIH) tipos 1 y 2, virus de leucemia de linfocitos T humanos (HTLV) tipos 1, 2 y 5, virus de tumor mamario de ratón (MMTV), virus del sarcoma de Rous (RSV), lentivirus, coronavirus, virus del síndrome respiratorio agudo grave (SARS), virus filovirus del Ébola, Virus de Marburg, Metapneumovirus (MPV) tal como el metapneumovirus humano (HMPV), Virus rabdovirus de rabia, virus de la estomatitis vesicular, Bunyavirus, Virus de la fiebre hemorrágica crimea-congo, virus de la fiebre del Valle del Rift, Virus de la crosse, Virus Hantaan, Ortomixovirus, virus de la influenza (tipos A, B, y C), Paramixovirus, virus parainfluenza (PIV tipos 1, 2 y 3), virus sincitial respiratorio (tipos A y B), virus del sarampión, virus de las paperas, Arenavirus, virus de la coriomeningitis linfocítica, Virus junin, Virus machupo, Virus guaranarito, Virus lassa, Virus ampari, Virus flexal, Virus lppy, Virus mobala, Virus mopeia, Virus latino, Virus paraná, Virus Pichinde, Virus Punta Toro (PTV), Virus Tacaribe y virus Tamiami.

Los ejemplos de patógenos bacterianos que pueden detectarse por los métodos de la divulgación incluyen, sin limitación, Los ejemplos específicos de patógenos bacterianos incluyen, sin limitación, uno o más de (o cualquier combinación de) *Acinetobacter baumannii*, especies de *Actinobacillus*, *Actinomycetes*, especies de *Actinomyces* (tales como *Actinomyces israelii* y *Actinomyces naeslundii*), especies de *Aeromonas* (tal como *Aeromonas hydrophila*, *Aeromonas veronii* biovar *sobria* (*Aeromonas sobria*) y *Aeromonas caviae*), *Anaplasma phagocytophilum*, *Alcaligenes xylosoxidans*, *Acinetobacter baumannii*, *Actinobacillus actinomycetemcomitans*, especies de *Bacillus* (tal como *Bacillus anthracis*, *Bacillus cereus*, *Bacillus subtilis*, *Bacillus thuringiensis* y *Bacillus stearothermophilus*), especies de *Bacteroides* (tal como *Bacteroides fragilis*), especies de *Bartonella* (tal como *Bartonella bacilliformis* y *Bartonella henselae*, especies de *Bifidobacterium*, especies de *Bordetella* (tal como *Bordetella pertussis*, *Bordetella parapertussis* y *Bordetella bronchiseptica*), especies de *Borrelia* (tal como *Borrelia recurrentis* y *Borrelia burgdorferi*), especies de *Brucella* (tal como *Brucella abortus*, *Brucella canis*, *Brucella melintensis* y *Brucella suis*), especies de *Burkholderia* (tal como *Burkholderia pseudomallei* y *Burkholderia cepacia*), especies de *Campylobacter* (tal como *Campylobacter jejuni*, *Campylobacter coli*, *Campylobacter lari* y *Campylobacter fetus*), especies de *Capnocytophaga*, *Cardiobacterium hominis*, *Chlamydia trachomatis*, *Chlamydophila pneumoniae*, *Chlamydophila psittaci*, especies de *Citrobacter*, *Coxiella burnetii*, especies de *Corynebacterium* (tal como, *Corynebacterium diphtheriae*, *Corynebacterium jeikeum* y *Corynebacterium*), especies de *Clostridium* (tal como *Clostridium perfringens*, *Clostridium difficile*, *Clostridium botulinum* y *Clostridium tetani*), *Eikenella corrodens*, especies de *Enterobacter* (tal como *Enterobacter aerogenes*, *Enterobacter agglomerans*, *Enterobacter cloacae* y *Escherichia coli*, incluyendo *Escherichia coli* oportunistas, tal como *E. coli* enterotoxigénica, *E. coli* enteroinvasiva, *E. coli* enteropatógena, *E. coli* enterohemorrágica, *E. coli* enteroagregativa y *E. coli* uropatógena), especies de *Enterococcus* (tal como *Enterococcus faecalis* y *Enterococcus faecium*), especies de *Ehrlichia* (tal como *Ehrlichia chaffeensis* y *Ehrlichia canis*), *Erysipelothrix rhusiopathiae*, especies de *Eubacterium*, *Francisella tularensis*, *Fusobacterium nucleatum*, *Gardnerella vaginalis*, *Gemella morbillorum*, especies de *Elaemophilus* (tal como *Elaemophilus influenzae*, *Elaemophilus ducreyi*, *Elaemophilus aegyptius*, *Elaemophilus parainfluenzae*, *Elaemophilus haemolyticus* y *Elaemophilus parahaemolyticus*, Especies de *Elleicobacter* (tal como *Elleicobacter pylori*, *Elleicobacter cinaedi* y *Elleicobacter fennelliae*), *Kingella kingii*, especies de *Klebsiella* (tal como *Klebsiella pneumoniae*, *Klebsiella granulomatis* y *Klebsiella oxytoca*), especies de *Lactobacillus*, *Listeria monocytogenes*, *Leptospira interrogans*, *Legionella pneumophila*, *Leptospira interrogans*, especies de *Peptostreptococcus*, *Moraxella catarrhalis*, especies de *Morganella*, especies de *Mobiluncus*, especies de *Micrococcus*, especies de *Mycobacterium* (tal como *Mycobacterium leprae*, *Mycobacterium tuberculosis*, *Mycobacterium intracellulare*, *Mycobacterium avium*, *Mycobacterium bovis* y *Mycobacterium marinum*), especies de *Mycoplasma* (tal como *Mycoplasma pneumoniae*, *Mycoplasma hominis* y *Mycoplasma genitalium*), especies de *Nocardia* (tal como *Nocardia asteroides*, *Nocardia cyriacigeorgica* y *Nocardia brasiliensis*), especies de *Neisseria* (tal como *Neisseria gonorrhoeae* y *Neisseria meningitidis*), *Pasteurella multocida*, *Plesiomonas shigelloides*. Especies de *Prevotella*, especies de *Porphyromonas*, *Prevotella melaminogenica*, especies de *Proteus* (tal como *Proteus vulgaris* y *Proteus mirabilis*), especies de *Providencia* (tal como *Providencia alcalifaciens*, *Providencia rettgeri* y *Providencia stuartii*), *Pseudomonas aeruginosa*, *Propionibacterium acnes*, *Rhodococcus equi*, especies de *Rickettsia* (tal como *Rickettsia rickettsii*, *Rickettsia akari* y *Rickettsia prowazekii*, *Orientia tsutsugamushi* (antes: *Rickettsia tsutsugamushi*) y *Rickettsia typhi*), especies de *Rhodococcus*, *Serratia marcescens*, *Stenotrophomonas maltophilia*, especies de *Salmonella* (tal como *Salmonella enterica*, *Salmonella typhi*, *Salmonella paratyphi*, *Salmonella enteritidis*, *Salmonella choleraesuis*, y *Salmonella typhimurium*), especies de *Serratia* (tal como *Serratia marcescens* y *Serratia liquefaciens*), especies de *Shigella* (tal como *Shigella dysenteriae*, *Shigella flexneri*, *Shigella boydii* y *Shigella sonnei*), especies de *Staphylococcus* (tal como *Staphylococcus aureus*, *Staphylococcus epidermidis*, *Staphylococcus hemolyticus*, *Staphylococcus saprophyticus*), especies de *Streptococcus* (tal como *Streptococcus pneumoniae* (por ejemplo, el serotipo 4 de *Streptococcus pneumoniae* resistente al cloranfenicol, serotipo 6B de *Streptococcus pneumoniae* resistente a la espectinomicina, serotipo 9V de *Streptococcus pneumoniae* resistente a la estreptomina, serotipo 14 de *Streptococcus pneumoniae* resistente a la eritromicina, serotipo 14 de *Streptococcus pneumoniae* resistente a la optoquina, serotipo 18C de *Streptococcus pneumoniae* resistente a la rifampicina, serotipo 19F de *Streptococcus pneumoniae* resistente a la tetraciclina, serotipo 19F de *Streptococcus pneumoniae* resistente a la penicilina y serotipo resistente 23F de *Streptococcus pneumoniae* resistente a trimetoprima, serotipo 4 de *Streptococcus pneumoniae* resistente al cloranfenicol, serotipo 6B de *Streptococcus pneumoniae* resistente a la espectinomicina, serotipo 9V de *Streptococcus pneumoniae* resistente a la estreptomina, serotipo 14 de *Streptococcus pneumoniae* resistente a la optoquina, serotipo 18C de *Streptococcus pneumoniae* resistente a la rifampicina, serotipo 19F de *Streptococcus pneumoniae* resistente a la penicilina y serotipo resistente 23F de *Streptococcus pneumoniae* resistente a trimetoprima), *Streptococcus agalactiae*, *Streptococcus mutans*, *Streptococcus pyogenes*, *Streptococci del grupo A*, *Streptococcus pyogenes*, *Streptococci del grupo B*, *Streptococcus agalactiae*, *Streptococci* de grupo C., *Streptococcus*

anginosus, *Streptococcus equismilis*, *Streptococci* del grupo D, *Streptococcus bovis*, *streptococci* del grupo F y *Streptococcus anginosus* *streptococci* del grupo G), *Spirillum minus*, *Streptobacillus moniliformi*, especies de *Treponema* (tal como *Treponema carateum*, *Treponema petenue*, *Treponema pallidum* y *Treponema endemicum*, *Tropheryma whippelii*, *Ureaplasma urealyticum*, especies de *Veillonella*, especies de *Vibrio* (tal como *Vibrio cholerae*, *Vibrio parahemolyticus*, *Vibrio vulnificus*, *Vibrio parahaemolyticus*, *Vibrio vulnificus*, *Vibrio alginolyticus*, *Vibrio mimicus*, *Vibrio hollisae*, *Vibrio fluvialis*, *Vibrio metchnikovii*, *Vibrio damsela* y *Vibrio fumisii*), especies de *Yersinia* (tal como *Yersinia enterocolitica*, *Yersinia pestis*, y *Yersinia pseudotuberculosis*) y *Xanthomonas maltophilia* entre otros.

En algunas realizaciones, los métodos y composiciones de la divulgación se utilizan para monitorizar receptores de trasplante de órganos. Normalmente, los polinucleótidos de las células donantes se encontrarán en circulación en un fondo de polinucleótidos de las células del receptor. El nivel de ADN circulante del donante generalmente será estable si el órgano es bien aceptado, y el rápido aumento del ADN del donante (por ejemplo, como una frecuencia en una muestra dada) se puede usar como un signo temprano del rechazo del trasplante. El tratamiento se puede administrar en esta etapa para prevenir el fracaso del trasplante. Se ha demostrado que el rechazo del órgano del donante da como resultado un aumento del ADN del donante en la sangre; véase Snyder et al., PNAS 108(15):6629 (2011). La presente divulgación proporciona mejoras significativas en la sensibilidad sobre técnicas anteriores en esta área. En esta realización, se puede usar una muestra de control del receptor (por ejemplo, frotis de mejilla, etc.) y una muestra de control del donante para la comparación. La muestra del receptor se puede usar para proporcionar esa secuencia de referencia, mientras que las secuencias correspondientes al genoma del donante se pueden identificar como variantes de secuencia con respecto a esa referencia. La monitorización puede comprender obtener muestras (por ejemplo, muestras de sangre) del receptor durante un período de tiempo. Se pueden usar muestras tempranas (por ejemplo, dentro de las primeras semanas) para establecer unos datos de referencia para la fracción de ADNcf del donante. Las muestras posteriores se pueden comparar con los datos de referencia. En algunas realizaciones, un aumento en la fracción de ADNcf del donante de aproximadamente el 10% o al menos aproximadamente, 20 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 %, 95 %, 100 %, 250 %, 500 %, 1000% o más puede servir como una indicación de que un receptor está en el proceso de rechazar el tejido del donante.

Ejemplos

Los siguientes ejemplos se dan con el propósito de ilustrar varias realizaciones de la invención y no pretenden limitar la presente invención de ninguna manera. Los presentes ejemplos, junto con los métodos descritos en el presente documento son actualmente representativos de realizaciones preferidas, son ejemplos y no pretenden ser limitaciones en el alcance de la invención.

Ejemplo 1: Preparación de la biblioteca de secuenciación repetitiva en tándem para la detección de mutaciones

Comenzando con >10 ng de fragmentos de ADN de ~150 pb en 12 µl de agua o Tris-HCl 10 mM, pH 8,0, se agregaron 2 µl de la mezcla de tampón de CircLigase 10X y la mezcla se calentó a 95 °C durante 2 minutos y se enfrió en hielo durante 5 minutos. A esto se agregaron 4 µl de Betaina 5M, 1 µl de MnCl₂ 50 mM y 1 µl de CircLigase II. La reacción se incubó a 60 °C durante al menos, 12 horas. A continuación, se agregaron y mezclaron 2 µl de la mezcla de cebador de RCA (50 nM cada uno, hasta una concentración final de 5 nM). La mezcla se calentó a 95 °C durante 2 min y después se enfrió a 42 °C durante 2 h. El producto de unión circular se purificó con el kit de purificación de nucleótidos oligo Zymo. De acuerdo con las instrucciones del fabricante, se agregaron 28 µl de agua a 22 µl del producto de unión circular hasta un volumen total de 50 µl. Esto se mezcló con 100 µl de tampón de unión Oligo y 400 µl de etanol. Esto se centrifugó durante 30 segundos a > 10.000xg, y se descartó el flujo continuo. Se agregaron 750 µl de tampón de lavado de ADN, a continuación se centrifugó durante 30 segundos a > 10.000xg, se descartó el flujo continuo y se centrifugó durante otro minuto a la velocidad máxima. La columna se movió a un nuevo tubo Eppendorf y se eluyó con 17 µl de agua (el volumen final eluido fue de aproximadamente 15 µl).

La amplificación por círculo rodante se realizó en un volumen de aproximadamente 50 µl. En la muestra de elución de 15 µl, se agregaron 5 µl de tampón RepliPHI 10X (Epicentre), 1 µl de dNTP 25 mM, 2 µl de DTT 100 mM, 1 µl de 100U/µl de RepliPHI Phi29 y 26 µl de agua. La mezcla de reacción se incubó a 30 °C durante 1 hora. Los productos de RCA se purificaron agregando 80 µl de perlas de Ampure, siguiendo las instrucciones del fabricante para los pasos de lavado restantes. Para la elución, se añadieron 22,5 µl de tampón de elución y las perlas se incubaron a 65 °C durante 5 minutos. Después de girar brevemente, se devolvió el tubo a los imanes.

Aproximadamente 20 µl de producto eluido de la reacción de RCA se mezclaron con 25 µl de mezcla 2X Phusion Master, 2,5 µl de DMSO y 0,5 µl de 10 µM de cada mezcla de cebadores B2B. La amplificación utilizó el siguiente programa de PCR: 95 °C durante 1 minuto, 5 ciclos de extensión (95 °C durante 15 segundos, 55 °C durante 15 segundos, 72 °C durante 1 minuto, 13 -18 ciclos de replicación (95 °C durante 15 segundos, 68 °C durante 15 segundos, 72 °C durante 1 minuto) y 72 °C durante 7 minutos de extensión final. El tamaño del producto de la PCR se comprobó ejecutando un E-gel. Si el intervalo era de 100-500 pb, se realizó una purificación de perla Ampure de 0.6X para enriquecer 300-500 pb y tomar 1-2 ng para otra ronda de PCR con cebadores de adaptadores de biblioteca de ARN pequeños. Si el intervalo de tamaño del producto era >1000 pb, los productos se purificaron con perlas de Amure 1.6X y se tomaron 2-3 ng para la preparación de la biblioteca de amplicones Nextera XT para enriquecer tamaños en el intervalo de 400-1000 pb por una purificación 0,6X de perlas de Ampure.

Para llevar a cabo bioinformática en los datos de secuenciación, se obtuvieron los archivos FASTQ de una ejecución MiSeq. Las secuencias se alinearon en archivos FASTQ para hacer referencia a secuencias genómicas que contienen secuencias dirigidas (por ejemplo, KRAS y EGFR) utilizando BWA. Se encontraron las regiones y las longitudes de las unidades de repetición y su posición de referencia para cada secuencia (ambas lecturas) utilizando los resultados de alineación. Se encontraron variantes en todos los loci usando los resultados de alineación y la información de unidades de repetición de cada secuencia. Se combinaron los resultados de dos lecturas. Se calculó la frecuencia normalizada de variantes y el nivel de ruido. Se aplicaron múltiples criterios adicionales en identificaciones de variante de variantes confirmadas, incluyendo qscore > 30 y valor de p < 0,0001. Las variantes confirmadas que pasaron estos criterios se indicaron como variantes verdaderas (mutaciones). El proceso se puede automatizar mediante lenguajes informáticos (por ejemplo, python).

Ejemplo 2: Creación de una biblioteca de secuenciación repetitiva en tándem para la detección de variantes de secuencia

Se utilizaron 10 ng de fragmentos de ADN con una longitud promedio de 150 pb en un volumen de 12 µl para la construcción de la biblioteca de secuenciación repetitiva en tándem. El ADN se procesó previamente con polinucleótido quinasa T4 (New England Biolabs) para agregar un grupo fosfato en el extremo 5' y dejar un grupo hidroxilo en el extremo 3'. Para los fragmentos de ADN generados a partir de DNase I o fragmentación enzimática o extraídos de suero o plasma, se omitió la etapa de procesamiento del extremo. El ADN se mezcló con 2 µl de tampón CircLigase 10X (Epicentre CL9021K). La mezcla se calentó a 95 °C durante 2 min y después se enfrió en hielo durante 5 minutos, a continuación se agregaron 4 µl de Betaina, 1 µl de MnCl₂ 50 mM y 1 µl de CircLigase II (Epicentre CL9021K). La reacción de unión se llevó a cabo a 60 °C durante al menos, 12 horas. Se añadió 1 µl de cada mezcla de cebador de RCA a 200 nM (hasta una concentración final de 10 nM) a los productos de unión y se mezcló, se calentó a 96 °C durante 1 minuto, se enfrió a 42 °C y se incubó a 42 °C durante 2 horas.

El producto de unión circular con cebadores de RCA hibridados se purificó con el kit de purificación de nucleótidos oligo Zymo (Zymo Research, D4060). Para hacer esto, los 21 µl de producto se diluyeron hasta 50 µl con 28 µl de agua y 1 µl de ARN transportador (Sigma-Aldrich, R5636, diluido hasta 200 ng/µl con 1X de tampón TE). La muestra diluida se mezcló con 100 µl de tampón de unión Oligo y 400 µl de etanol al 100%. La mezcla se cargó en la columna y se centrifugó durante 30 segundos a >10.000 xg. Se descartó el flujo continuo. La columna se lavó con 750 µl de tampón de lavado de ADN mediante centrifugación durante 30 segundos a >10.000 xg, descartando el flujo continuo y centrifugando durante 1 minuto más a la velocidad máxima. La columna se movió a un nuevo tubo Eppendorf de 1,5 ml y el ADN se eluyó con 17 µl de tampón de elución (Tris-Cl 10 mM, pH 8,0, volumen eluido final de aproximadamente 15 µl).

Se agregaron 5 µl de tampón RepliPHI 10X, 2 µl de dNTP 25 mM, 2 µl de DTT 100 mM, 1 µl de 100 U/µl de RepliPHI Phi29 y 25 µl de agua (Epicentre, RH040210) a la muestra eluida de 15 µl de la columna hasta un volumen de reacción total de 50 µl. La mezcla de reacción se incubó a 30 °C durante 2 horas. Los productos de RCA se purificaron agregando 80 µl de perlas de Ampure XP (Beckman Coulter, A63881). Se siguieron las instrucciones del fabricante para las etapas de lavado. Los productos de RCA se eluyeron después de 5 minutos de incubación a 65 °C en 22,5 µl de tampón de elución. El tubo se centrifugó brevemente antes de volver a los imanes.

Aproximadamente 20 µl de producto eluido de la reacción de RCA se mezclaron con 25 µl de mezcla 2X Phusion Master (New England Biolabs M0531S), 2,5 µl de agua, 2,5 µl de DMSO y 0,5 µl de mezcla de cebadores B2B (10 µM de cada). La amplificación se realizó con el siguiente programa de termociclado: 95 °C durante 2 minutos, 5 ciclos de extensión (95 °C durante 30 segundos, 55 °C durante 15 segundos, 72 °C durante 1 minuto, 18 ciclos de replicación (95 °C durante 15 segundos, 68 °C durante 15 segundos, 72 °C durante 1 minuto) y 72 °C durante 7 minutos de extensión final. El tamaño del producto de la PCR se verificó mediante electroforesis. Una vez que se confirmaron los productos de PCR largos por electroforesis, los productos de PCR se mezclaron con 30 µl de perlas de Ampure (volumen de 0,6X) para purificación para enriquecer productos de PCR de >500 pb. Los productos purificados se cuantificaron con la plataforma de cuantificación Qubit 2.0 (Invitrogen). Se usó aproximadamente 1 ng de ADN purificado para la preparación de la biblioteca de amplicones Nextera XT (Illumina FC-131-1024). Los elementos de la biblioteca con un tamaño de inserto de >500 pb se enriquecieron por purificación con cuentas de Ampure 0,6X.

La concentración y la distribución del tamaño de las bibliotecas amplificadas se analizaron utilizando el kit de alta sensibilidad de ADN Agilent para el bioanalizador 2100 (Agilent Technologies Inc., Santa Clara, CA). La secuenciación se realizó utilizando Illumina MiSeq con un kit de secuenciación MiSeq de 2-250 pb. De acuerdo con el manual de MiSeq, Se cargó una biblioteca desnaturalizada de 12 pM en la ejecución de la secuenciación.

En una variación de este procedimiento, se usaron adaptadores de Illumina en la preparación de bibliotecas en lugar de la preparación de Nextera. Para hacer esto, se utilizó aproximadamente 1 ng de ADN purificado de manera similar para la amplificación por PCR con un par de cebadores que contienen la parte universal de los cebadores B2B y la secuencia del Adaptador de Illumina (P5 y P7; 5'CAAGCAGAAGACGGCATACGA3' y 5'ACACTCTTTCCCTACACGACGCTCTTCCGATCT3'). Utilizando Phusion Master Mix, se realizaron 12 ciclos de etapas de replicación (95 °C durante 30 segundos, 55 °C durante 15 segundos, 72 °C durante 60 segundos). El

propósito de esta etapa de amplificación fue agregar adaptadores de Illumina para la secuenciación de amplicones. Los amplicones >500 pb de longitud se enriquecieron con perlas de Ampure 0,6X. La concentración y la distribución del tamaño de la biblioteca de amplicones se analizaron utilizando el kit de alta sensibilidad de ADN Agilent para el bioanalizador 2100 (Agilent technologies Inc., Santa Clara, CA). La secuenciación se realizó utilizando Illumina MiSeq con un kit de secuenciación MiSeq de 2x250 pb. La parte universal de los cebadores B2B también sirvió como secuencias de cebadores de secuenciación y se agregó un cebador de secuenciación personalizado si el kit de Illumina no contenía el cebador. Se cargó una biblioteca desnaturalizada de 12 pM en la ejecución de la secuenciación.

La cobertura de la región diana en un análisis de ejemplo se ilustra en la FIG. 33. La tabla 3 a continuación describe los resultados para el análisis de las regiones dirigidas.

La Tabla 1 proporciona ejemplos de cebadores de RCA útiles en los métodos de la divulgación. La Tabla 2 proporciona ejemplos de cebadores B2B útiles en los métodos de la divulgación.

Tabla 1

Nombre del gen	Nombre del cebador A de RCA	Secuencia del cebador A de RCA	Nombre del cebador B de RCA	Secuencia del cebador B de RCA
PIK3CA	PIK3CA-E1a	GCTTTGAGCTGTTCTT TGTCATT	PIK3CA-E1b	AAAGCAATTTCTACACGA GATCC
PIK3CA	PIK3CA-E2a	TTTAATTGTGTGGAAG ATCCAATC	PIK3CA-E2b	ATTAAACAGCATGCATTG AACTG
EGFR	EGFR-E1a	CTTTCTCACCTTCTGG GATCC	EGFR-E1b	AAATTCCCGTCGCTATCA AG
EGFR	EGFR-E2a	CCATCACGTAGGCTTC CTG	EGFR-E2b	ATGGCCAGCGTGGACAAC TGTCCGGGAACACAAAG AC
EGFR	EGFR-E3a	GACATAGTCCAGGAG GCAGC	EGFR-E3b	TGGCAGCCAGGAACGTAC AACACCGCAGCATGTCAA G
EGFR	EGFR-E4a	AAGCGACGGTCCTCC AAG	EGFR-E4b	AGTACGTTCCCTGGCTG CC
EGFR	EGFR-E5a	ATCCACTTGATAGGC ACCTTG	EGFR-E5b	AAAGTGGATGGCATTGGAA TC
EGFR	EGFR-E6a	TCTCGCTGGCAGGGA TTC	EGFR-E6b	AGTTCCGTGAGTTGATCA TCG
EGFR	EGFR-E7a	AACTTTGGGCGACTAT CTGC	EGFR-E7b	ACTTCTACCGTGCCCTGA TG
EGFR	EGFR-E8a	TTGGAGTCTGTAGGA CTTGGC	EGFR-E8b	CACAGCAGGGCTTCTTCA G
EGFR	EGFR-E9a	CTGCTGTGGGATGAG GTACTC	EGFR-E9b	CATGGGCAACTTCTCTGT TTC
EGFR	EGFR-E10a	CATGGAATGCTTGTAC CACATC	EGFR-E10b	AAACTGATGGGACCCACT CC
BRAF	BRAF-E1a	CAGTTTGAACAGTTGT CTGGATC	BRAF-E1b	AGGAGATATCAAGAGGA TGGATTG
PTEN	PTEN-E1a	TGTTTCTGCTAACGAT CTCTTTG	PTEN-E1b	TCTTGCAGAAAGACTTGA AGG
PTEN	PTEN-E2a	CAGGAAATCCCATAG CAATAATG	PTEN-E2b	GGATTCAAAGCATAAAA ACCATTAC
PTEN	PTEN-E3a	GCTTTGAATCCAAAA ACCTTAAAAAC	PTEN-E3b	TATGTTGTATAACTTAAA CCCGATAGAC
PTEN	PTEN-E4a	TACAGTACATTTCATAC CTACCTCTGC	PTEN-E4b	TTGAAGACCATAAACCAC CAC
PTEN	PTEN-E5a	AAAGGATATTGTGCA ACTGTGG	PTEN-E5b	AAGTAAGGACCAGAGAC AAAAAGG
PTEN	PTEN-E6a	CCATAGAAATCTAGG GCCTCTTG	PTEN-E6b	GGATTATAGACCAGTGGC ACTG
PTEN	PTEN-E7a	CCAGATGATTCTTTAA CAGGTAGC	PTEN-E7b	

PTEN	PTEN-E9a	GAACCTGTCTTCCCGT CGTG	PTEN-E9b	CATGTACTTTGAGTTCCC TCAGC
PTEN	PTEN-E12a	TCTGGTCCTGGTATGA AGAATG	PTEN-E12b	CAGGACCAGAGGAAACC TCAG
PTEN	PTEN-E13a	GCTCTATACTGCAAAT GCTATCG	PTEN-E13b	CGTGCAGATAATGACAAG GAATATC
PTEN	PTEN-E14a	TTGGAGAAAAGTATC GGTTGG	PTEN-E14b	GGTCAGTTAAATTAAACA TTTTGTGG
PTEN	PTEN-E15a	TGGTGTACAGAAGTT GAACTGC	PTEN-E15b	GATGTTAGTGACAATGAA CCTGATC
KRAS	KRAS-E1a	AAGAGTGCCTTGACG ATACAGC	KRAS-E1b	TCTTGCCTACGCCACCAG CAGGCCCTTCTGTCTTGA AC
TP53	TP53-E1a	CCTGACTCAGACTGA CATCTCC	TP53-E1b	GAACATCTCGAAGCGCTC AC
TP53	TP53-E2a	ATGTTCCGAGAGCTG AATGAG	TP53-E2b	TTATGGTATAAGTTGGTG
TP53	TP53-E3a	TTAAAGGACCAGACC AGCTTC	TP53-E3b	TTCTGAAG
TP53	TP53-E4a	CTTGGGACCTCTTATC AAGTGG	TP53-E4b	AGAGGTCCCAAGACTTAG TACCTG
TP53	TP53-E5a	AAGCAAGCAGGACAA GAAGC	TP53-E5b	GCTTGCTTACCTCGCTTA GTG
TP53	TP53-E6a	GGGACGGAACAGCTT TGAG	TP53-E6b	TTCCGTCCCAGTAGATTA CCAC
TP53	TP53-E7a	CAACTACATGTGTAA CAGTTCCTGC	TP53-E7b	CATGTAGTTGTAGTGGAT GGTGG
TP53	TP53-E8a	GTGGAGTATTTGGAT GACAGAAAC	TP53-E8b	ATACTCCACACGCAAATT TCC
TP53	TP53-E9a	TGCTCAGATAGCGAT GGTGAG	TP53-E9b	CTATCTGAGCAGCGCTCA TG
TP53	TP53-E10a	CTGTGCAGCTGTGGGT TGA	TP53-E10b	GCAGGTCTTGGCCAGTTG TGTCCCAGAATGCAAGAA GC
TP53	TP53-E12a	AAGTCTGTGACTTGCA CGGTC	TP53-E12b	GATGACAGGGGCCAGGA G
TP53	TP53-E13a	CCTGTCATCTTCTGTC CCTTC	TP53-E13b	TGGGTCTTCAGTGAACCA TTG
TP53	TP53-E14a	AAGACCCAGGTCCAG ATGAAG	TP53-E14b	GAGCAGAAAGTCAGTCCC ATG
TP53	TP53-E15a	CTGCTCTTGTCTTTCA GACTTCC	TP53-E15b	GCTCGACGCTAGGATCTG AC
TP53	TP53-E16a	CTCTGAGTCAGGAAA CATTTTCAG	TP53-E16b	

Tabla 2

Nombre del gen	Cebador A B2B	Secuencia del cebador A	Cebador B B2B	Secuencia del cebador B
PIK3CA	PIK3CA-BX1a	GTTCAGAGTTCTACAGTCC GACGATCGCTTTGAGCTGT TCTTTGTCATT	PIK3CA-BX1b	CCTTGGCACCCCGAGAAATTC AAAAGCAATTTCTACACGA GATCC
PIK3CA	PIK3CA-BX2a	GTTCAGAGTTCTACAGTCC GACGATCTTTAATTGTGTG GAAGATCCAATC	PIK3CA-BX2b	CCTTGGCACCCCGAGAAATTC AATTAACAGCATGCAATTGA ACTG
EGFR	EGFR-BX1a	GTTCAGAGTTCTACAGTCC GACGATCCTTTCTCAGCTT CTGGGATCC	EGFR-BX1b	CCTTGGCACCCCGAGAAATTC AAAAATTCCTCGTCTATCAA G
EGFR	EGFR-BX2a	GTTCAGAGTTCTACAGTCC GACGATCCCATCACGTAG GCTTCTGT	EGFR-BX2b	CCTTGGCACCCCGAGAAATTC AATGGCCAGCTGGACAAC CCTTGGCACCCCGAGAAATTC
EGFR	EGFR-BX3a	GTTCAGAGTTCTACAGTCC GACGATCGACATAGTCCA GGAGGCAGC	EGFR-BX3b	ATGTCCGGGAACACAAAGA C
EGFR	EGFR-BX4a	GTTCAGAGTTCTACAGTCC GACGATCAAGCGACGGTC CTCCAAG	EGFR-BX4b	CCTTGGCACCCCGAGAAATTC ATGGCAGCAGGAACGTAC CCTTGGCACCCCGAGAAATTC
EGFR	EGFR-BX5a	GTTCAGAGTTCTACAGTCC GACGATCAGTACGTTCTCTG GCTGCC	EGFR-BX5b	AAACACCGCAGCATGTCAA G
EGFR	EGFR-BX6a	GTTCAGAGTTCTACAGTCC GACGATCATCCACTTGATA GGCACCTTG	EGFR-BX6b	CCTTGGCACCCCGAGAAATTC AAAGTGGATGGCATTTGGAA TC
EGFR	EGFR-BX7a	GTTCAGAGTTCTACAGTCC GACGATCTCTCGCTGGCA GGGATTC	EGFR-BX7b	CCTTGGCACCCCGAGAAATTC ACCTGGAGAAAGGAGAACG C
EGFR	EGFR-BX9a	GTTCAGAGTTCTACAGTCC GACGATCAACTTTGGGOG ACTATCTGC	EGFR-BX9b	CCTTGGCACCCCGAGAAATTC AAGTTCCTGTGAGTTGATCAT CG
EGFR	EGFR-BX10a	GTTCAGAGTTCTACAGTCC GACGATCTTGGAGTCTGTA GGACTTGGC	EGFR-BX10b	CCTTGGCACCCCGAGAAATTC AACTTCTACCTGCTGCTGAT G
EGFR	EGFR-BX11a	GTTCAGAGTTCTACAGTCC GACGATCTGCTGTGGGA TGAGGTACTC	EGFR-BX11b	CCTTGGCACCCCGAGAAATTC ACACAGCAGGGCTTCTTCAG

EGFR	EGFR-BX12a	GTTACAGAGTTCTACAGTCC GACGATCCATGGAATGCT TGTAACACATC	EGFR-BX12b	CCTTGGCACCCCGAGAAATTCC ACATGGGCAACTTCTCTGTT TC
BRAF	BRAF-BX1a	GTTACAGAGTTCTACAGTCC GACGATCCAGTTGAACA GTTGTCTGATC	BRAF-BX1b	CCTTGGCACCCCGAGAAATTCC AAACTGTGATGGACCCACT CC
PTEN	PTEN-BX1a	GTTACAGAGTTCTACAGTCC GACGATCTGTCTCTGCTAA CGATCTCTTG	PTEN-BX1b	CCTTGGCACCCCGAGAAATTCC AAGGAGATATCAAGAGGAT GGATTTC
PTEN	PTEN-BX2a	GTTACAGAGTTCTACAGTCC GACGATCCAGGAAATCCC ATAGCAATAATG	PTEN-BX2b	CCTTGGCACCCCGAGAAATTCC ATCCTGCAGAAAGACTTGA AGG
PTEN	PTEN-BX3a	GTTACAGAGTTCTACAGTCC GACGATCGCTTGAATCCA AAACCTTAAAC	PTEN-BX3b	CCTTGGCACCCCGAGAAATTCC AGGATTCAAGCATAAAA CCATTAC
PTEN	PTEN-BX4a	GTTACAGAGTTCTACAGTCC GACGATCTACAGTACATT ATACCTACCTCTGC	PTEN-BX4b	CCTTGGCACCCCGAGAAATTCC ATATGTGTGATAACTTAAAC CCGATAGAC
PTEN	PTEN-BX5a	GTTACAGAGTTCTACAGTCC GACGATCAAGGATAATG TGCAACTGTGG	PTEN-BX5b	CCTTGGCACCCCGAGAAATTCC AC
PTEN	PTEN-BX6a	GTTACAGAGTTCTACAGTCC GACGATCCATAGAAATC TAGGGCTCTTG	PTEN-BX6b	CCTTGGCACCCCGAGAAATTCC AAAGTAAAGGACCAGAGACA AAAAGG
PTEN	PTEN-BX7a	GTTACAGAGTTCTACAGTCC GACGATCCAGATGTTCT TTAACAGGTAGC	PTEN-BX7b	CCTTGGCACCCCGAGAAATTCC AGGATTATAGACCAGTGGC ACTG
PTEN	PTEN-BX9a	GTTACAGAGTTCTACAGTCC GACGATCGAACTTGTCTTC CCGTCTGTG	PTEN-BX9b	CCTTGGCACCCCGAGAAATTCC ACAATGTAATTTGAGTTCCCT CAGC
PTEN	PTEN-BX12a	GTTACAGAGTTCTACAGTCC GACGATCTCTGGTCTGGT ATGAAGAATG	PTEN-BX12b	CCTTGGCACCCCGAGAAATTCC ACAGGACCAGAGGAAACCT CAG
PTEN	PTEN-BX13a	GTTACAGAGTTCTACAGTCC GACGATCGCTCTATACTGC AAATGCTATCG	PTEN-BX13b	CCTTGGCACCCCGAGAAATTCC ACGTGCAGATAATGACAAG GAATATC

PTEN	PTEN-BX14a	GTTACAGAGTTCTACAGTCC GACGATCTTGGAGAAAAG TATCGGTGG	PTEN-BX14b	CCTTGGCACCCGAGAAATTC AGGTCAGTTAAATTTAAACAT TTTGTGG
PTEN	PTEN-BX15a	GTTACAGAGTTCTACAGTCC GACGATCTGGTGTACAG AAGTTGAAGTGC	PTEN-BX15b	CCTTGGCACCCGAGAAATTC AGATGTTAGTGACAATGAA CCTGATC
KRAS	KRAS-BX1a	GTTACAGAGTTCTACAGTCC GACGATCAAGAGTGCCCT GACGATACAGC	KRAS-BX1b	CCTTGGCACCCGAGAAATTC ATCTTGCCTACGCCACCAG CCTTGGCACCCGAGAAATTC
TP53	TP53-BX1a	GTTACAGAGTTCTACAGTCC GACGATCCCTGACTCAGA CTGACATCTCC	TP53-BX1b	ACAGGCCCTTCTGTCTTGAA C CCTTGGCACCCGAGAAATTC
TP53	TP53-BX2a	GTTACAGAGTTCTACAGTCC GACGATCAATGTTCCGAGA GCTGAATGAG	TP53-BX2b	AGAACAATCTCGAAGCGCTC AC CCTTGGCACCCGAGAAATTC
TP53	TP53-BX3a	GACGATCTTAAAGGACCA GACCAGCTTC	TP53-BX3b	ATTATGGTATAAGTTGGTGT TCTGAAG
TP53	TP53-BX4a	GTTACAGAGTTCTACAGTCC GACGATCCTTGGGACCTCT TATCAAGTGG	TP53-BX4b	CCTTGGCACCCGAGAAATTC AAGAGTCCCAAGACTTAG TACCTG
TP53	TP53-BX5a	GTTACAGAGTTCTACAGTCC GACGATCAAGCAAGCAGG ACAAGAAGC	TP53-BX5b	CCTTGGCACCCGAGAAATTC AGCTTGCTTACCTCGCTTAG TG
TP53	TP53-BX6a	GTTACAGAGTTCTACAGTCC GACGATCGGGACGGAACA GCTTTGAG	TP53-BX6b	CCTTGGCACCCGAGAAATTC AATCCGTCCAGTAGATTAC CAC
TP53	TP53-BX7a	GTTACAGAGTTCTACAGTCC GACGATCCAATCATGT GTAACAGTCTGC	TP53-BX7b	CCTTGGCACCCGAGAAATTC ACATGTAGTTGTAGTGATG GTGG
TP53	TP53-BX8a	GTTACAGAGTTCTACAGTCC GACGATCGTGGAGTATT GGATGACAGAAAC	TP53-BX8b	CCTTGGCACCCGAGAAATTC AATACTCCACACGCCAAATT CC
TP53	TP53-BX9a	GTTACAGAGTTCTACAGTCC GACGATCTGCTCAGATAG CGATGGTGAG	TP53-BX9b	CCTTGGCACCCGAGAAATTC ACTATCTGAGCAGCGGCTCAT G

TP53	TP53-BX10a	GTTCAGAGTTCTACAGTCC GACGATCCTGTGACAGCTGT GGTTGA	TP53-BX10b	CCTTGGCACCCGAGAAATTC AGCAGGTCCTTGGCCAGTTG
TP53	TP53-BX12a	GTTCAGAGTTCTACAGTCC GACGATCAAGTCTGTGAC TTGCACGGTC	TP53-BX12b	CCTTGGCACCCGAGAAATTC ATGTCCCAGAATGCAAGAA GC
TP53	TP53-BX13a	GTTCAGAGTTCTACAGTCC GACGATCCTGTCTATCTTC TGTCCTTC	TP53-BX13b	CCTTGGCACCCGAGAAATTC AGATGACAGGGGCCAGGAG CCTTGGCACCCGAGAAATTC
TP53	TP53-BX14a	GTTCAGAGTTCTACAGTCC GACGATCAAGACCCAGGT CCAGATGAAG	TP53-BX14b	ATGGGTCTTCAGTGAACCCAT TG
TP53	TP53-BX15a	GTTCAGAGTTCTACAGTCC GACGATCCTGCTCTGTCT TTCAGACTTC	TP53-BX15b	CCTTGGCACCCGAGAAATTC AGAGCAGAAAGTCAGTCCC ATG
TP53	TP53-BX16a	GTTCAGAGTTCTACAGTCC GACGATCCTCTGAGTCAG GAAACATTTTCAG	TP53-BX16b	CCTTGGCACCCGAGAAATTC AGCTCGACGCTAGGATCTGA C

Tabla 3

	Resultados
lecturas	1.5M
% base diana w/lx	97,8 %
% específico	63,4 %
% de duplicados	18,2 %
cobertura promedio	74.5x
d.e de la cobertura	0,21

Ejemplo 3: Fragmentación del ADN genómico para la construcción de bibliotecas de secuenciación

- 5 Se procesó 1 µl de ADN genómico utilizando un kit de fragmentasa de ADNcd NEBNext (New England Biolabs) siguiendo el protocolo del fabricante. El tiempo de incubación se extendió a 45 minutos a 37 °C. La reacción de fragmentación se detuvo agregando 5 µl de EDTA 0,5 M, pH 8,0, y se purificó agregando 2X volúmenes de perlas Ampure XP (Beckman Coulter, A63881) de acuerdo con el protocolo del fabricante. El ADN fragmentado se analizó en un bioanalizador con un kit de ADN de alta sensibilidad (Agilent). El intervalo de tamaño del ADN fragmentado era normalmente de aproximadamente 100 pb a aproximadamente 200 pb con un pico de aproximadamente 150 pb.

Ejemplo 4: Procedimientos de Preparación de Bibliotecas

- 15 En este ejemplo, se usó un kit de preparación de bibliotecas KAPA (KK8230) con fines ilustrativos.
- Para las etapas que involucran la purificación de perlas, las perlas AMPure XP (n. ° de cat. A63881) se equilibraron a temperatura ambiente y se resuspendieron completamente antes de mezclarlas con la muestra. Después de mezclar bien con la muestra en un mezclador vorticial, se incubó a temperatura ambiente durante 15 minutos para permitir que el ADN se uniera a las perlas. A continuación, se colocaron perlas sobre un soporte magnético hasta que el líquido estuvo claro. Las perlas se lavaron dos veces con 200 µl de etanol al 80% y se secaron a temperatura ambiente durante 15 minutos.
- 25 Para realizar la reacción de reparación en los extremos, se mezclaron hasta 50 µl (2-10 ng) de ADN libre de células con 20 µl de la mezcla maestra de reparación en los extremos (8 µl de agua, 7 µl de tampón de reparación en los extremos 10X KAPA y 5 µl de mezcla de enzimas de reparación en los extremos KAPA) y se incubó durante 30 minutos a 20 °C. A continuación se agregaron 120 µl de perlas AMPure XP a la reacción de reparación en los extremos de 70 µl. La muestra se purificó, a continuación, como se describe anteriormente.
- 30 Para realizar reacciones de adición de colas A, las perlas secas que contenían el fragmento de ADN reparado en el extremo se mezclaron con la mezcla maestra de adición de colas A (42 µl de agua, 5 µl de tampón de adición de colas A KAPA 10X y enzima de adición de colas A KAPA). La reacción se incubó a 30 °C durante 30 minutos. Después de agregar 90 µl de solución PEG (PEG 8000 al 20%, NaCl 2,5 M), la mezcla se lavó de acuerdo con el protocolo de purificación de perlas anterior. Este paso de adición de colas A se saltó para reacciones de unión de extremos romos.
- 35 Para la unión del enlazador, se utilizaron dos oligos que tienen las siguientes secuencias (5' a 3') para formar un polinucleótido dúplex adaptador:
- /5Phos/CCATTTCATTACCTCTTCTCCGCACCCGACATAGAT*T y
/5Phos/ATCTATGTCGGGTGCGGAGAAAGAGGTAATGAAATGG*T. Las perlas secas que contenían un extremo reparado (para la unión roma) o una cola (para la unión basada en enlazadores), se mezclaron con 45 µl de la mezcla maestra de la unión (30 µl de agua, 10 µl de 5x de tampón de unión KAPA y 5 µl de ligasa de ADN KAPA T4), y 5 µl de agua (para la unión de extremo romo) o 5 µl de una mezcla molar igual de oligonucleótidos enlazadores (para unión basada en enlazadores). Las perlas se resuspendieron completamente y se incubaron a 20 °C durante 15 minutos. Después de agregar 50 µl de solución PEG (véase arriba), la mezcla se lavó de acuerdo con el protocolo de purificación de perlas anterior.
- 45 La amplificación de desplazamiento múltiple (MDA) se realizó con los kits de amplificación de ADN Illustra Genomiphi V2. Las perlas secas que contienen la cadena del fragmento unido se resuspendieron en 9 µl de tampón que contiene hexámero aleatorio y se calentaron durante 3 minutos a 95 °C, seguido de un enfriamiento rápido en hielo. Después de agregar 1 µl de mezcla de enzimas, la muestra enfriada en hielo se incubó a 30 °C durante 90 minutos. La reacción se paró, a continuación, calentando a 65 °C durante 10 minutos. Después de agregar 30 µl de solución PEG (véase arriba), la mezcla se lavó de acuerdo con el protocolo de purificación descrito anteriormente y se resuspendió en 200 µl de TE (con una incubación a 65 °C durante 5 minutos). Si se desea, el producto purificado se podría cuantificar con PCR cuantitativa, PCR digital en gotas (ddPCR), o someterse a la secuenciación de nueva generación (NGS).
- 55 Después de la MDA, se sonicaron cadenas de fragmentos unidos largos (por ejemplo, >2 kb) a ~ 300 pb usando un Covaris S220 en un volumen total de 130 µl. El protocolo del fabricante indicaba una potencia máxima de 140 W, un factor de actividad del 10%, 200 ciclos por ráfaga y 80 segundos de tiempo de tratamiento. La longitud del fragmento

de ~300 pb se seleccionó para aumentar la posibilidad de mantener un fragmento de ADN libre de células original intacto. Se puede usar un protocolo convencional de preparación de bibliotecas para colocar adaptadores en fragmentos de ADN sonificados para la secuenciación, si se desea. Se devolvió varias composiciones de lectura a partir de ejecuciones de secuenciación de pares extremos en secuenciadores Illumina (ya sea HiSeq o MiSeq). Aquellas en las que la unión (ya sea una auto-unión o una unión de adaptador en el caso en que se incluyeron adaptadores en la etapa de unión) fue interna a la lectura (flanqueada 5' y 3' por secuencia no adaptadora) se usaron para codificar secuencias de código de barras de interés.

Ejemplo 5: Circularización y Amplificación

Esto proporciona una ilustración de ejemplo de un procedimiento de circularización y amplificación (también denominado procedimiento de "Nebula"). El procedimiento utilizó los siguientes suministros: Máquina de PCR (por ejemplo, termociclador MJ research PTC-200 Peltier); CircLigase II, ligasa de ADNcs Epicentre n.º de cat. CL9025K; Exonucleasa (p.ej. Exol, NEB Biolabs n.º de cat. M0293S; ExoIII, NEB Biolabs n.º de cat. M0206S); Polinucleótido Quinasa de T4 (NEB Biolabs n.º de cat. M0201S); kit de amplificación del genoma completo (p.ej. GE Healthcare, Illustra, Ready-To-Go, Genomiphi, V3 DNA amplification kit); GlycoBlue (p.ej. Ambion n.º de cat. AM9515); Micro centrífuga (p.ej. Eppendorf 5415D); perlas de purificación de ADN (p.ej. Agencourt, AMPure XP, Beckman Coulter n.º de cat. A63881); Soporte magnético (por ejemplo, el MagnaRack™ Invitrogen n.º de cat. CS15000); Fluorómetro Qubit® 2.0 (Invitrogen, n.º de cat. Q32866); kit de ensayo HS de sondas moleculares de ADNcd (Life Technology n.º de cat. 032854); y un Bioanalizador (Agilent 2100), y reactivos de ADN de alta sensibilidad (n.º de cat. 5067-4626).

Para la amplificación de fragmentos de ADN que carecen de un fosfato 5' terminal (por ejemplo, ADN libre de células), el primer paso fue la reparación en los extremos y la formación de cadenas sencillas. El ADN se desnaturalizó a 96 °C durante 30 segundos (por ejemplo, en una máquina de PCR). Se preparó una reacción de polinucleótido quinasa (PNK) combinando 40 µl de ADN y 5 µl de tampón de reacción 10X PNK, seguido de incubación a 37 °C durante 30 minutos. Se añadieron ATP 1 mM y la enzima PNK a la reacción, y se incubaron a 37 °C durante 45 minutos. Se realizó un intercambio de tampón precipitando y resuspendiendo el ADN. Se combinaron 50 µl de ADN de la reacción PNK, 5 µl de acetato de sodio 0.5M pH 5,2, 1 µl de GlycoBlue, 1 µl de oligo (100 ng/µl) y 150 µl de etanol al 100%. La mezcla se incubó a -80 °C durante 30 minutos y se centrifugó a 16k rpm durante 5 minutos para sedimentar el ADN. El sedimento de ADN se lavó con 500 µl de etanol al 70%, se secó al aire durante 5 minutos a temperatura ambiente y el ADN se suspendió en 12 µl de Tris-Cl 10 mM, pH 8,0.

El ADN resuspendido se circularizó a continuación, por unión. El ADN se desnaturalizó a 96 °C durante 30 segundos, la muestra se enfrió en hielo durante 2 minutos, y se añadió una mezcla de ligasa (2 µl de tampón CircLigase 10X, 4 µl de Betaína 5M, 1 µl de MnCl₂ 50 mM, 1 µl de CircLigase II). La reacción de unión se incubó a 60 °C durante 16 horas en una máquina de PCR. Los polinucleótidos no unidos se degradaron por digestión con exonucleasa. Para ello, el ADN se desnaturalizó a 80 °C durante 45 segundos y se añadió a cada tubo 1 µl de la mezcla de exonucleasa (Exol 20U/µl: Exo III 100U/µl = 1:2). Esto se mezcló pipeteando arriba y abajo 5 veces, y se giró brevemente. La mezcla de digestión se incubó a 37 °C durante 45 minutos. El volumen se llevó a 50 µl con 30 µl de agua, y se realizó un intercambio adicional de tampón mediante precipitación y resuspensión como se describe anteriormente.

Para realizar la amplificación del genoma completo (WGA), el ADN purificado se desnaturalizó primero a 65 °C durante 5 minutos. Se añadieron 10 µl de tampón de desnaturalización del kit GE WGA a 10 µl de ADN purificado. El ADN se enfrió en un bloque frío o hielo durante 2 minutos. Se agregaron 20 µl de ADN al taponador GenomiPhi V3 (WGA) Ready-To-Go. La reacción de WGA se incubó a 30 °C durante 1,5 horas, seguido de inactivación por calor a 65 °C durante 10 minutos.

La muestra se purificó utilizando perlas magnéticas AmpureXP (1,6X). Las perlas se agitaron en vortex y se colocaron alícuotas de 80 µl en tubos de 1,5 ml. A continuación, se combinaron 30 µl de agua, 20 µl de ADN amplificado y los 80 µl de perlas, y se incubaron a temperatura ambiente durante 3 minutos. Los tubos se colocaron en un soporte magnético durante 2 minutos y la solución transparente se pipeteó. Las perlas se lavaron dos veces con etanol al 80%. El ADN se eluyó agregando 200 µl de Tris-Cl 10 mM pH 8,0. La mezcla de perlas de ADN se incubó a 65 °C durante 5 minutos. Los tubos se colocaron de nuevo en el soporte magnético durante 2 minutos. Se transfirieron 195 µl de ADN a un nuevo tubo. Se usó 1 µl para la cuantificación por Qubit. Por último, se sometieron a sonicación 130 µl de producto de WGA utilizando Covaris S220 para alcanzar un tamaño de alrededor de 400 pb.

Ejemplo 6: Circularización y Amplificación con purificación adicional

Esto proporciona una ilustración de ejemplo de un procedimiento de circularización y amplificación (también denominado procedimiento de "Nebula") con una etapa de extracción con fenol cloroformo.

La etapa 1 fue la eliminación del ARN competitivo (del ARN portador en la extracción) y del ARN natural (co-purificado) para una reacción de circLigase. El ARN se eliminó agregando 1 µl de RNasa A (10 mg/ml) (Qiagen 1007885) a 50 µl de ADNcf (2-10 ng) e incubando a 37 °C durante 30 minutos en una máquina de PCR (MJ research PTC-200 Peltier Thermal Cyclic).

La etapa 2 fue un intercambio de tampón con sal y precipitación con etanol. Esta etapa fue útil para limpiar y concentrar la entrada para la unión con una recuperación cercana al 100% (mientras que las columnas generalmente solo recuperan el 30%). La mezcla coprecipitante de etanol (50 µl de ADN del tratamiento con ARNasa, 5 µl de acetato de sodio 0,5 M pH 5,2, 1 µl de GlycoBlue (Ambion AM9515), 1 µl de oligo portador (100 ng/µl), 150 µl de etanol al 100 %) se incubaron a -80 ° C durante 30 minutos y se centrifugaron a 16k rpm (Eppendorf 5415D) durante 5 minutos para precipitar el ADN. El uso de un oligo portador no específico de 20 meros (se usó un cebador de PCR), aumentó ligeramente el rendimiento y la estabilidad de la recuperación de la precipitación. El sedimento de ADN se lavó con 500 µl de etanol al 70%. El sedimento de ADN se secó al aire durante 5 minutos a temperatura ambiente y se resuspendió con 13 µl de Tris-Cl 10 mM, pH 8,0.

La etapa 3 fue la circularización. Se desnaturalizaron 12 µl de ADNcf a 96 ° C durante 30 segundos y se enfriaron en un bloque de hielo durante 2 minutos. La adición de la mezcla de unión (12 µl de ADNcf, 2 µl de tampón de Circligase 10X, 4 µl de betaína 5M, 1 µl de MnCl₂ 50 µM, 1 µl de Circligase II (Epicentre n.° CL9025K) se configuró en un bloque frío y se realizó una ligación Realizado a 60 ° C durante 16 horas.

La etapa 4 fue la digestión con exonucleasa. La mezcla de ADN de unión se incubó a 80 ° C durante 45 segundos en una máquina de PCR, seguido de un tratamiento con exonucleasa. Se añadió 1 µl de mezcla de exonucleasa (Exol 20U/µl: ExoIII 100U/µl = 1:2) a cada tubo y las reacciones se incubaron a 37 ° C durante 30 minutos. Para fines de control de calidad, no fue necesaria la eliminación de la plantilla lineal.

La etapa 5 fue la extracción con fenol cloroformo y el intercambio de tampón con sal y precipitación con etanol. El fenol/etanol fue útil para lograr más del 80% de eficiencia de unión (la cantidad de producto circularizado fue aproximadamente igual a la cantidad de polinucleótidos de entrada). Se agregaron 180 µl de Tris 10 mM a 20 µl de ADN del tratamiento con exonucleasa para obtener un volumen de 200 µl, y se utilizaron 200 µl de fenol para extraer el ADN. Se recogió la capa acuosa y se recuperó el ADN mediante precipitación con etanol. La mezcla de coprecipitante de etanol (200 µl de solución de ADN después de la extracción con fenol, 20 µl de acetato de sodio 0.5M pH 5,2, 1 µl de GlycoBlue, 1 µl de oligo portador (100 ng/µl), 600 µl de etanol al 100%) se incubó a -80 ° C durante 30 minutos y se centrifugó a 16k rpm durante 5 minutos para precipitar el ADN. El sedimento de ADN se lavó con 500 µl de etanol al 70%. El sedimento de ADN se secó al aire durante 5 minutos a temperatura ambiente y se resuspendió con 11 µl de Tris-C1 10 mM, pH 8,0.

La etapa 6 fue la amplificación del genoma completo. Se incubaron 10 µl de ADN purificado a 65 ° C durante 5 minutos en un bloque de calentamiento, y se agregaron 10 µl de tampón de desnaturalización de GE Healthcare, Ready-To-Go, Genomiphi, kit de amplificación de ADN V3. Después de enfriar el ADN durante 5 minutos a temperatura ambiente, se sometieron 20 µl de ADN al taponador Genomiphi V3 Ready-To-Go. (WGA). La reacción de amplificación se incubó a 30 ° C durante 1,5 horas, y la reacción se terminó por inactivación por calor a 65 ° C durante 10 minutos.

El paso 7 fue la purificación de las perlas utilizando perlas magnéticas AmpureXP (1.6X). Esto se hizo como en el ejemplo anterior.

La etapa 8 fue la sonicación como en el ejemplo anterior. El ADN estaba listo a continuación, para la PCR cuantitativa, ddPCR o la construcción de la biblioteca de secuenciación.

Ejemplo 7: Análisis de la eficacia de la unión y las tasas específicas

El ADNcf que se circularizó y se sometió a la aplicación del genoma completo como en los ejemplos anteriores se analizó mediante PCR cuantitativa (qPCR). Los resultados de la curva de amplificación de qPCR para una diana de muestra (utilizando cebadores KRAS) se muestran en la FIG. 18. Como se muestra en la FIG. 18A, la amplificación de qPCR de 1/10 * de la entrada de ADNcf dio un Ct promedio (umbral de ciclo) de 31,75, y 1/10 * del producto de unión de la misma muestra dio un Ct promedio de 31,927, lo que indica una alta eficacia de unión de aproximadamente 88%. La eficacia de la unión puede variar desde aproximadamente o aproximadamente más del 70%, 80 %, 90 %, el 95 % o más, tal como aproximadamente el 100%. El ADN lineal que no se circularizó se elimina en algunos ejemplos, de manera que casi todo el ADN puede amplificarse a partir de formas circulares. Cada muestra se ejecutó tres veces, por duplicado. Como se muestra en la figura 18B, las curvas de amplificación de 10 ng de producto de WGA y ADN genómico de referencia (ADNg) (12878, 10 ng) prácticamente se superponen entre sí. El Ct promedio para la muestra de WGA fue de 26,655, mientras que el de la muestra de ADNg fue de 26,605, lo que indica una alta tasa específica de más del 96%. El número de KRAS en una cantidad dada de ADN amplificado fue comparable con el ADNg no amplificado, lo que indica un proceso de amplificación sin sesgo. Cada muestra se probó tres veces, por duplicado. Como punto de contraste, también se probó el protocolo de circularización provisto por Lou et al. (PNAS, 2013, 110 (49)). Usando el método de Lou, que carecía de las etapas de precipitación y purificación de los ejemplos descritos anteriormente, solo el 10-30% del ADN de entrada lineal se convirtió en ADN circular. Una recuperación tan baja presenta un desafío para la secuenciación cadena abajo y la detección de variantes.

Ejemplo 8: Análisis de ADN circularizado mediante ddPCR

La PCR digital en gotas (ddPCR) se usó para evaluar la preservación de frecuencia de alelo y el sesgo en productos

de amplificación del genoma completo generados a partir de polinucleótidos circularizados. En general, la ddPCR se refiere a un ensayo de PCR digital que mide cantidades absolutas contando las moléculas de ácido nucleico encapsuladas en particiones discretas, definidas volumétricamente, de gotas de agua en aceite que admiten la amplificación por PCR (Hinson et al, 2011, Anal. Chem. 83:8604-8610; Pinheiro et al, 2012, Anal. Chem. 84: 1003-1011). Una sola reacción de ddPCR puede comprender al menos 20.000 gotitas divididas por pocillo. La PCR digital en gotas se puede realizar utilizando cualquier plataforma que realice un ensayo de PCR digital que mida cantidades absolutas contando las moléculas de ácido nucleico encapsuladas en particiones de gotitas de agua en aceite discretas, definidas volumétricamente, que admiten la amplificación por PCR. Una estrategia típica para la PCR digital en gotas se puede resumir de la siguiente manera: una muestra se diluye y se divide en miles o millones de cámaras de reacción separadas (gotitas de agua en aceite) para que cada una contenga una o ninguna copia de la molécula de ácido nucleico de interés. El número de gotitas "positivas" detectadas, que contienen el amplicón diana (es decir, la molécula de ácido nucleico de interés), en comparación con el número de gotitas "negativas", que no contienen el amplicón diana (molécula de ácido nucleico de interés), puede utilizarse para determinar el número de copias de la molécula de ácido nucleico de interés que estaban en la muestra original. Los ejemplos de sistemas de PCR digital en gotas incluyen el Sistema de PCR digital en gotas QX100™ de Bio-Rad, que divide las muestras que contienen la plantilla de ácido nucleico en 20.000 gotitas de tamaño nanolitro; y el sistema de PCR digital RainDrop™ de RainDance, que divide las muestras que contienen la plantilla de ácido nucleico en 1.000.000 a 10.000.000 gotitas de tamaño picolitro. Se proporcionan ejemplos adicionales de métodos para ddPCR en el documento WO2013181276A1.

En este ejemplo, El ADN genómico V600E de BRAF (ADNg) de una línea celular de melanoma se mezcló con el genoma de referencia 12878 en proporciones específicas (0%, 0,67 %, 2,0 %, 6,67 %, 20% o 100%) y se fragmentó para generar fragmentos de un tamaño similar a los encontrados en el ADNcf (en este caso, alrededor de 150 pb). Las muestras de ADN mixtas (10 ng) se circularizaron y amplificaron de acuerdo con el Ejemplo 2. Se sometieron 40 ng de ADN amplificado a ddPCR para V600E de BRAF y de tipo silvestre. Las frecuencias alélicas de mutación observadas se ilustran gráficamente y se tabulan en la FIG. 19. Tal como se muestra, la frecuencia alélica de mutación observada con amplificación (fila central de la tabla de la Figura 19) refleja la frecuencia alélica mutante de entrada (fila superior), así como el resultado de la ddPCR de 100 ng de ADN genómico sin amplificación (fila inferior). La frecuencia alélica por salida de ddPCR se calcula como el número de mutaciones de BRAF que contienen gotitas sobre la suma de las gotitas que contienen tanto mutantes como de tipo silvestre. El ADN con amplificación se indica como un círculo abierto, y sin amplificación se indica como un círculo lleno pequeño relleno. Con la excepción de una pequeña desviación del 0,67%, los dos conjuntos de datos se superponen completamente. Esto demuestra la conservación de la representación real de la frecuencia alélica mutante, sustancialmente sin sesgo.

Ejemplo 9: Detección de variantes de secuencia sobre el fondo

Se circularizaron 10 ng de ADNg sonificado (150 pb, ADN de referencia de Multi-Gene Multiplex, Horizon) y se amplificaron como se describe en el Ejemplo 2, y seguido de sonicación. El ADN fragmentado se sometió a continuación, a la construcción de la biblioteca de secuenciación de Rubicon. Después de la secuenciación de captura, se trazaron las variantes dentro de 50 pb a partir de puntos calientes de referencia. Los resultados para la detección de variantes, donde identificar una variante requiere detección en dos polinucleótidos diferentes que se distinguen por uniones diferentes, se muestran en la FIG. 20. Los siete puntos calientes de referencia esperados (KIT D816V, EGFR G719S, EGFR T790M, EGFR L858R, KRAS G13D, KRAS G12D, NRAS Q61K) se trazaron en la posición 0. También se confirmaron otras dos variantes, ilustradas como el triángulo abierto y el diamante en la Fig. 20.

A efectos de comparación, el ADNg se sonicó como se indica anteriormente, pero 10 ng del ADNg sonificado se sometieron directamente a la construcción de la biblioteca de secuenciación de Rubicon de acuerdo con la práctica común, sin circularización y sin requerir confirmación de una secuencia de variantes en dos polinucleótidos diferentes. Después de la secuenciación de captura, se trazaron las variantes dentro de 50 pb a partir de puntos calientes de referencia, con los resultados en la FIG.21. Los siete puntos calientes de referencia esperados (KIT D816V, EGFR G719S, EGFR T790M, EGFR L858R, KRAS G13D, KRAS G12D, NRAS Q61K) se trazaron en la posición 0. No se esperaban variantes en otras posiciones y probablemente se deban a errores de secuenciación. A diferencia de los resultados del método empleado para generar la FIG. 20, los resultados en la FIG. 21 indican que los métodos de secuenciación convencionales tienen una tasa de error aleatorio mucho más alta que puede enmascarar la verdadera señal de mutación cuando la frecuencia alélica es baja (por ejemplo, inferior al 5%).

Los resultados de un análisis separado de la sensibilidad y el ruido de fondo detectados por los métodos de secuenciación con y sin requerir la detección en dos polinucleótidos diferentes se ilustran en las FIG. 16-17. Como ilustran estas figuras, el requisito de validación reduce en gran medida el ruido de fondo y aumenta la sensibilidad.

Ejemplo 10: Análisis de la composición de GC y distribución de tamaño

Se circularizaron 10 ng de ADNg sonificado (150 pb, ADN de referencia de Multi-Gene Multiplex, Horizon) y se amplificaron en forma circular y se amplificaron como en el Ejemplo 5, se secuenciaron y analizaron con la variante que identifica el filtro de verificación de dos polinucleótidos (izquierda). El número de secuencias con un intervalo de porcentajes de CG se tabularon y se representaron gráficamente, como se muestra en la FIG. 22. Como se muestra en el gráfico del extremo izquierdo, las secuencias para las muestras preparadas de acuerdo con el Ejemplo 5 se parecen mucho a la distribución teórica, excepto el pico central (corresponde al contenido global de GC del genoma

subyacente). Por el contrario, cuando se utilizó la misma cantidad de ADN_g directamente para construir una biblioteca de secuenciación sin amplificación utilizando un kit de construcción de la biblioteca de secuenciación de Rubicon, la diferencia entre el resultado de la secuenciación y la distribución teórica es muy evidente (véase la gráfica central). El pico central de esta secuenciación directa de Rubicon es más alto que la distribución teórica. Newman *et al.* (2014; Nature Medicine, (20):548-54) indicaron que la distribución del contenido de GC de secuenciación de ADN_{cf} fue similar con la distribución teórica cuando se usaron 32ng de ADN_{cf}. Esto se ilustra en el gráfico del extremo derecho.

La distribución del tamaño del ADN se evaluó para el ADN_{cf} que se había circularizado, amplificado y secuenciado como en el Ejemplo 5. Como se muestra en la FIG. 23, el pico de la distribución de las longitudes de los fragmentos indicados por los resultados de la secuenciación es de aproximadamente 150-180 pb, que se asemeja al patrón de distribución típico de ADN_{cf}.

Ejemplo 11: Evaluación de la uniformidad de amplificación

Los resultados de qPCR de 10 productos circularizados y amplificados según el Ejemplo 5 se compararon con el ADN de referencia no amplificado (ADN_g de la línea celular 12878, Coriell Institute). Se utilizaron 10 ng de ADN de referencia genómico o producto de amplificación para cada reacción de qPCR en tiempo real, y se generaron relaciones por cuantificación relativa del producto de amplificación sobre referencia genómica. Como se muestra en la FIG. 24, la proporción de cada PCR está dentro de un cambio de 2 veces, lo que sugiere que el número de copias de estas dianas en el conjunto de ADN amplificado es muy similar al ADN de referencia no amplificado. Los 10 pares de cebadores de PCR de 6 genes (BRAF, cKIT, EGFR, KRAS, NRAS, PI3KCA) se diseñaron y validaron previamente.

Ejemplo 12: Cuantificación del rendimiento de amplificación de fragmentos de ADN

Se aisló ADN_{cf} de cuatro pacientes (paciente 1-4) y de un control sano. El ADN genómico (ADN_g, Multi-Gene Multiplex Horizon) se sonizó a fragmentos de aproximadamente 150 pb. El ADN se circularizó y amplificó con cebadores aleatorios. La Tabla 7 muestra la cantidad de entrada de ADN en la reacción de amplificación y la cantidad de ADN producida por amplificación. Se obtuvo una amplificación significativa incluso para la muestra más pequeña (0,4 ng) y todas las muestras se amplificaron al menos 600 veces.

Tabla 7

Tipo de Muestra	Entrada (ng)	Rendimiento ng
ADN _g	10	6100
ADN _g	4	6880
ADN _g	2	5700
ADN _g	1	5760
ADN _g	0,5	3280
ADN _{cf} Control sano	0,4	6240
ADN _{cf} paciente1	4	13800
ADN _{cf} paciente2	3	9480
ADN _{cf} paciente3	3	7840
ADN _{cf} paciente4	1	3180

Ejemplo 13: Detección de mutaciones de baja frecuencia a partir de ADN_{cf} de pacientes con cáncer

En la etapa 1, se circularizó el ADN_{cf}. La mezcla de unión circular se preparó en un tubo de PCR a temperatura ambiente. Se pipetaron 4 ng-10 ng de ADN_{cf} en un volumen de 12 µl en el tubo de PCR. El ADN se desnaturalizó a 96 °C durante 30 segundos, a continuación, se enfriaron los tubos de PCR en hielo durante 2 minutos. Se añadió a cada tubo una mezcla de unión (2 µl de 10X de tampón CircLigase, 4 µl de betaína 5 M, 1 µl de MnCl₂ 50 mM, 1 µl de CircLigase II) y la reacción transcurrió a 60 °C durante 16 horas en una máquina de PCR.

En la etapa 2, las reacciones de unión se trataron para eliminar el ADN lineal no unido. Se añadió 1 µl de mezcla de exonucleasa (NEB M0206S, M0293S; ExoI 20u/µl; ExoIII 100u/µl = 1:2) a cada tubo, se mezcló e incubó a 37 °C durante 30 minutos en una máquina de PCR.

En la etapa 3, la reacción de unión se purificó para el intercambio de tampón. El producto de unión se purificó con Oligo Clean & Concentrator (Zymo Research). La mezcla de unión (30 µl de Tris 10 mM, 100 µl de tampón de unión Oligo, 400 µl de etanol al 100%) se añadió a la reacción de unión después del tratamiento con Exonucleasa, se mezcló y se centrifugó brevemente. Las columnas de Zymo-spin se cargaron y se giraron a más de 10.000 xg durante 30 segundos. Las columnas se lavaron con 750 µl de tampón de lavado de ADN y se centrifugaron a 14.000 xg durante 1 minuto. El ADN se eluyó con 15 µl de Tris 10 mM por centrifugación a más de 10.000 xg durante 30 segundos.

En la etapa 4, el ADN se amplificó por cebado aleatorio. La amplificación del genoma completo (WGA) se realizó con el kit de amplificación de ADN Ready-To-Go Genomiphi V3 (GE Healthcare). 10 µl de unión purificada se mezclaron con 10 µl de tampón de desnaturalización 2x, se incubaron a 95 °C durante 3 minutos, a continuación, se enfriaron a

4 ° C en hielo. Se agregaron 20 µl de ADN desnaturalizado a la premezcla de WGA, las muestras se incubaron a 30 °C durante 1,5 horas, seguido de inactivación a 65 °C durante 10 minutos.

5 En la etapa 5, los productos de amplificación se mejoraron utilizando Agencourt AMPure XP Purification (1.6X) (Beckman Coulter). Se añadieron 30 µl de Tris 10 mM y 80 µl de perlas de AMPure a 20 µl de reacción de WGA. La mezcla se incubó a temperatura ambiente durante 2 minutos. Los tubos se colocaron en un soporte magnético y se incubaron durante 2 minutos. Los sobrenadantes se eliminaron y se desecharon. Las muestras se lavaron con 200 µl de etanol (80%) dos veces, se secaron al aire durante 5 minutos y el ADN se eluyó con 200 µl de Tris 10 mM, pH 8,0.

10 En la etapa 6, el ADN de WGA se fragmentó. Se sonicaron 130 µl de producto de WGA utilizando un sonicador Covaris S220 para obtener un tamaño de fragmento de aproximadamente 400 pb. Los ajustes de Covaris S220 fueron los siguientes: Potencia incidente máxima = 140W, factor de actividad = 10%, ciclos por ráfaga = 200, tiempo de tratamiento = 55 segundos.

15 En la etapa 7, las muestras fueron cuantificadas por qPCR. Se usaron 1/10 de entrada de unión y el producto de unión para las reacciones de qPCR con tres repeticiones para medir la eficacia de la unión. Se usaron 10 ng del producto de WGA fragmentado junto con 10 ng de ADN de referencia (línea celular 12878) para que la qPCR midiera la velocidad específica. Las reacciones comprendieron 5 µl de mezcla maestra 2x (mezcla maestra de PCR universal rápida de TaqMan (2x), Applied Biosystems; tinte Evagreen, Biotium), 0,5 µl de cebador (5 µM), 1,2 µl de H₂O, 10 µl de ADN. La amplificación procedió de acuerdo con el siguiente programa: 95 °C 2 minutos; y 40 ciclos de [95 °C, 10 segundos; 60 °C, 20 segundos].

25 En la etapa 8, se construyeron bibliotecas de secuenciación. Se prepararon bibliotecas de secuenciación a partir de 500-1000 ng de ADN amplificado sonificado utilizando el kit KAPA Hyper Prep (KK8500) o el kit de preparación de la biblioteca KAPA con la biblioteca de PCR estándar (KK8200). Las uniones del adaptador (con la concentración final del adaptador 1µM) se prepararon según el protocolo del fabricante. Se añadieron 30 µl (0,3x) de una solución al 20% de PEG 8000/2,5 M NaCl de lavado del adaptador unido del producto unido a 100 µl del producto unido resuspendido. Las perlas se mezclaron completamente con el producto unido y se incubaron a temperatura ambiente durante 15 minutos. Las perlas se capturaron en un imán hasta que el líquido estuvo claro. Luego se sometieron 130 µl de sobrenadante a selección por tamaño usando perlas Ampure XP. Las muestras se transfirieron a una nueva placa seguido de una adición de 20 µl de perlas Ampure XP (0.5x). El producto unido se capturó en ese momento en las perlas y se lavó dos veces con 200 µl de etanol al 80%. El producto unido se resuspendió y se eluyó en 20 µl de tampón EB. Después de la selección por tamaño y la purificación, se agregaron 20 µl de producto unido a 25 µl de premezcla 2x de KAPA HiFi Hotstart y 5 µl de 10 µM de cebadores P5 + P7 (5 'CAAGCAGAAGACGGCATACGA3'), 5'ACACTCTTTCCCTACACGACGCTCTTCCGATCT3 ') para amplificar la biblioteca utilizando el siguiente programa de ciclos: 98 °C, 45 segundos; 5 ciclos de (98 °C, 15 segundos; 60 °C, 30 segundos; 72 °C durante 30 segundos); 72 °C, 60 segundos. La biblioteca amplificada se diluyó 20x antes de cargarla en un analizador de fragmentos o bioanalizador (chip de alta sensibilidad) para la cuantificación. Se realizó una selección de tamaño adicional a través del selector de tamaño de gel (preparación Blue Pippin de Sage Science).

40 En la etapa 9, la biblioteca de secuenciación se enriqueció mediante el enriquecimiento de la captura de la sonda utilizando sondas de xGEN Pan-Cancer Panel v1.5, 127908597 (IDT). En la etapa 10, la biblioteca se secuenció en un HiSeq 2500, con una profundidad promedio de 1000x.

45 En la etapa 11, los datos de secuenciación fueron analizados para realizar identificaciones de variante. La identificación de la variante incluyó una etapa que requiere que se produzca una diferencia de secuencia en dos polinucleótidos diferentes (por ejemplo, identificados por uniones diferentes) para contarse como una variante. Se detectaron varias mutaciones somáticas y también se indicaron en una base de datos pública (COSMIC (Catálogo de mutaciones somáticas en cáncer)). Entre las mutaciones identificadas se encontraba V600M de BRAF con una frecuencia alélica de 0,05%, que demuestra la alta sensibilidad de este sistema incluso cuando la entrada es baja. Los resultados para la detección de varias mutaciones, incluida su frecuencia en la muestra, se muestran en la Tabla 8.

Tabla 8

Tipo de cáncer	Gen mutado	Mutación	Frecuencia en la muestra
Melanoma	FGFR2	V191I	0,20 %
Melanoma	NRAS	Q61K	0,20 %
Melanoma	CTNNB1	S37F	0,10 %
Melanoma	BRAF	V600E	0,10 %
Melanoma	BRAF	V600M	0,05 %
Cáncer de mama	PIK3CA	H1407L	53 %
Cáncer de mama	LIFR	S679L	0,60 %
Cáncer de mama	KRAS	G12D	0,60 %
Cáncer de mama	ATM	R2993*	0,30 %
Cáncer de mama	ATRX	S618F	0,10 %
Próstata	AR	T346A	6,40 %

Próstata	SPOP	F133L	3,30 %
Próstata	BRAF	V600M	1,10 %
Próstata	NSD1	R1557C	1,00 %
Próstata	SF3B1	K700E	1,60 %
Pancreático	KRAS	G12V	1,70 %
Pancreático	TP53	7578176C>T (variante de donante de corte y empalme)	0,5 %

Ejemplo 14: Detección precisa de mutaciones a partir de ADN de referencia múltiplex de muestra FFPE

El ADN se extrajo de la muestra Horizon FFPE-multiplex (HD200) siguiendo el protocolo del fabricante (kit de ADN Covaris truXTRAC™ FFPE). 130 µl de ADNg FFPE se sonicaron utilizando un sonicador Covaris S220 para obtener un tamaño de fragmento de aproximadamente 150 pb (configuración de Covaris S220: Potencia incidente máxima = 175W, factor de actividad = 10%, ciclos por ráfaga = 200, tiempo de tratamiento = 430 segundos). 50 ng de ADN en un volumen de 11 µl se desnaturalizaron a 95 °C durante 30 segundos. Se agregaron 10,5 µl de H₂O, 2,5 µl de tampón de ligasa (NEB B0202S) y 1 µl de polinucleótido quinasa de T4 (NEB M0201S). Las reacciones se incubaron a 37 °C durante 30 minutos para fosforilar.

Las muestras se unieron, a continuación se purificaron con Oligo Clean & Concentrator (Zymo Research). La mezcla de unión (30 µl de Tris 10 mM, 100 µl de tampón de unión Oligo, 400 µl de etanol al 100%) se añadió a la reacción de unión después del tratamiento con Exonucleasa, se mezcló con un vórtice y se giró brevemente. Las muestras se cargaron en una columna Zymo-spin y se giraron a más de 10.000 xg durante 30 segundos. Las columnas se lavaron con 750 µl de tampón de lavado de ADN y se centrifugaron a 14.000xg durante 1 minuto. El ADN se eluyó con 15 µl de Tris 10 mM por centrifugación a más de 10.000xg durante 30 segundos.

Las muestras se procesaron y analizaron posteriormente de acuerdo con los pasos 5-11 en el Ejemplo 13. Los resultados se resumen en la Tabla 9. La representación de nueve mutaciones en el ADN convencional de la mutación múltiple de Horizon se mantuvo aproximadamente por este proceso, mientras que la cantidad de ADN aumentó al menos 600 veces.

Tabla 9

Gen	Mutaciones	Frecuencia Alélica	Frecuencia Alélica Medida
BRAF	p.V600E	10,5 %	8 %
KIT	p.D816V	10 %	9,4 %
EGFR	P.L858R	3 %	4 %
EGFR	p.T790M	1 %	1 %
EGFR	p.G719S	24,5 %	14 %
KRAS	p.G13D	15 %	9,1 %
KRAS	p.G12D	6 %	5,4 %
NRAS	p.Q61K	12,5 %	8,3 %
PIK3CA	p.H1047R	17,5 %	14 %

Ejemplo 15: Detección de mutaciones de baja frecuencia de multímeros de ADNg de línea celular de mutación de cáncer

En este ejemplo, el ADN genómico sonificado se unió para formar multímeros, que a continuación se sometieron a amplificación, fragmentación y análisis. La FIG. 25 ilustra un ejemplo de este proceso.

El ADNg de una línea celular de melanoma SK-mel-28 (ATCC) que contiene la mutación V600E de BRAF se mezcló con el ADNg de referencia (12878 Coriell Institute) para lograr un 1% de V600E de BRAF. El ADN se sonicó como en el Ejemplo 14 para obtener un tamaño de fragmento de aproximadamente 150 pb. Se desnaturalizaron 100 ng de ADN en un volumen de 11 µl a 95 °C durante 30 segundos. Se agregaron 10,5 µl de H₂O, 2,5 µl de tampón de ligasa (NEB B0202S) y 1 µl de polinucleótido quinasa de T4 (NEB M0201S), seguido de incubación a 37 °C durante 30 minutos para fosforilar el ADN.

Las muestras se purificaron con Oligo Clean & Concentrator (Zymo Research). Esto incluyó la adición de la mezcla de unión (25 µl de Tris 10 mM, 100 µl de tampón de unión Oligo, 400 µl de etanol al 100%) a la reacción de unión después del tratamiento con Exonucleasa. Esto se mezcló con un vórtice y se giró brevemente. Se cargó una columna de Zymo-spin y se centrifugó a más de 10.000 xg durante 30 segundos, se lavó con 750 µl de tampón de lavado de ADN y se centrifugó a 14.000xg durante 1 minuto. El ADN se eluyó con 15 µl de Tris 10 mM por centrifugación a más de 10.000 xg durante 30 segundos.

Para unir, se mezclaron 6 ng de ADN en 4 µl de volumen con 0,45 µl de tampón de reparación en los extremos 10x (Enzymatics), 0,05 µl de dNTP 25 mM, 0,5 ATP 10 mM, mezcla de enzimas de reparación en los extremos (Enzymatics) y ligasa de T4 2000 unidades/µl. La reacción se incubó a 25 °C durante 30 minutos y a continuación a

75 °C durante 20 minutos.

La amplificación del genoma completo se realizó con el kit de amplificación de ADN Ready-To-Go Genomiphi V3 (GE Healthcare). Se mezclaron 8 µl de H₂O y 10 µl de unión purificada con 10 µl de tampón de desnaturalización 2x. El ADN se desnaturalizó a 95 °C durante 3 minutos y a continuación, se enfrió a 4 °C en hielo. Se añadieron 20 µl de ADN desnaturalizado a la premezcla de WGA y se incubaron a 30 °C durante 1,5 horas, seguido de inactivación a 65 °C durante 10 minutos.

La reacción de amplificación se mejoró a continuación utilizando la purificación Agencourt AMPure XP (1.6X) (Beckman Coulter). Se añadieron 30 µl de Tris 10 mM y 80 µl de perlas de AMPure a 20 µl de reacción de WGA. Esto se incubó a temperatura ambiente durante 2 minutos. El tubo se colocó en un soporte magnético y se incubaron durante 2 minutos. El sobrenadante se eliminó y se desechó. Las perlas se lavaron con 200 µl de etanol (80%) dos veces, a continuación se secaron al aire durante 5 minutos. El ADN se eluyó con 200 µl de Tris 10 mM, pH 8,0. A continuación, se fragmentaron 130 µl de producto de WGA utilizando el sonicador Covaris S220 para obtener un tamaño de fragmento de aproximadamente 400 pb (configuración de Covaris S220: Potencia incidente máxima = 140W, factor de actividad = 10%, ciclos por ráfaga = 200, tiempo de tratamiento = 55 segundos).

Las mutaciones se detectaron por ddPCR utilizando ensayos de detección de mutaciones de ddPCR BioRad Prime PCR. La reacción de ddPCR de detección de mutación se montó en un tubo de PCR a temperatura ambiente (80 ng de ADN amplificado, 10 µl de supermezcla de ddPCR 2x para sondas, 1 µl de diana 20x (V600E de BRAF, BioRad) (9 µM)/sonda (FAM; 5 µM), 1 µl de 20x cebadores de tipo silvestre (9 µM)/sonda (HEX; 5 µM), 8 µl de muestra de ADN (50 ng). La reacción se mezcló pipeteando arriba y abajo 5 veces, y luego se transfirió al cartucho del generador de gotas. Las gotas se generaron utilizando el generador de gotas QX200, se transfirieron a una placa de PCR de 96 pocillos y se amplificaron utilizando el siguiente programa de PCR: 95 °C, 10 minutos; 40 ciclos de [94 °C, 30 segundos, 55 °C 1 minuto]; 98 °C, 10 minutos. La placa de reacción de PCR se transfirió a un lector de gotas QX200 para cuantificar el resultado. Basándose en el ADN de entrada, la frecuencia esperada de la mutación Y600E BRAF fue del 1%. Por este procedimiento de unión y amplificación, esta frecuencia se mantuvo aproximadamente (1,41% según el análisis de ddPCR) mientras que la cantidad de ADN aumentó aproximadamente 200 veces.

Aunque las realizaciones preferidas de la presente invención se han mostrado y descrito en el presente documento, será obvio para los expertos en la materia que tales realizaciones se proporcionan solo a modo de ejemplo. A los expertos en la materia se les ocurrirán numerosas variaciones, cambios y sustituciones sin apartarse de la invención como se define en las reivindicaciones. Debería entenderse que se pueden emplear varias alternativas a las realizaciones de la invención descritas en el presente documento para llevar a la práctica la invención.

REIVINDICACIONES

1. Un método para identificar una variante de secuencia en una muestra de ácido nucleico que comprende una pluralidad de polinucleótidos libres de células, teniendo cada polinucleótido libre de células de la pluralidad un extremo 5' y un extremo 3', comprendiendo el método:
 - (a) circularizar polinucleótidos libres de células individuales de dicha pluralidad para formar una pluralidad de polinucleótidos circulares, cada uno de los cuales tiene una unión entre el extremo 5' y el extremo 3';
 - (b) amplificar los polinucleótidos circulares de (a);
 - (c) secuenciar los polinucleótidos amplificados para producir una pluralidad de lecturas de secuenciación;
 - (d) identificar diferencias de secuencia entre lecturas de secuenciación y una secuencia de referencia; y
 - (e) identificar una diferencia de secuencia como la variante de la secuencia solamente cuando la diferencia de secuencia se produce en al menos dos polinucleótidos circulares que tienen uniones diferentes.
2. El método de la reivindicación 1, en donde la pluralidad de polinucleótidos libres de células son de cadena sencilla.
3. El método de la reivindicación 1, en donde la circularización se efectúa sometiendo la pluralidad de polinucleótidos libres de células a una reacción de unión.
4. El método de la reivindicación 1, en donde un polinucleótido circular individual tiene una unión que es única entre los polinucleótidos circularizados.
5. El método de la reivindicación 1, en donde la secuencia de referencia es una secuencia consenso formada al alinear las lecturas de secuenciación entre sí.
6. El método de la reivindicación 1, en donde la circularización comprende la etapa de unir un polinucleótido adaptador al extremo 5', al extremo 3', o tanto al extremo 5' como al extremo 3' de un polinucleótido libre de células en la pluralidad de polinucleótidos libres de células.
7. El método de la reivindicación 1, en donde la amplificación se efectúa utilizando una polimerasa que tiene actividad de desplazamiento de cadena.
8. El método de la reivindicación 1, en donde la amplificación comprende someter los polinucleótidos circulares a una mezcla de reacción de amplificación que comprende cebadores aleatorios.
9. El método de la reivindicación 8, en donde los polinucleótidos amplificados se someten a la etapa de secuenciación sin enriquecimiento.
10. El método de la reivindicación 8, que comprende adicionalmente el enriquecimiento de uno o más polinucleótidos diana entre los polinucleótidos amplificados realizando una etapa de enriquecimiento antes de la secuenciación, en donde, opcionalmente, la etapa de enriquecimiento comprende hibridar polinucleótidos amplificados con una pluralidad de sondas unidas a un sustrato o amplificar una secuencia diana que comprende la secuencia A y la secuencia B orientadas en una dirección de 5' a 3' en una mezcla de reacción de amplificación que comprende:
 - (a) los polinucleótidos amplificados;
 - (b) un primer cebador que comprende la secuencia A', en donde el primer cebador hibrida específicamente con la secuencia A de la secuencia diana mediante complementariedad de secuencia entre la secuencia A y la secuencia A';
 - (c) un segundo cebador que comprende la secuencia B, en donde el segundo cebador hibrida específicamente con la secuencia B' presente en un polinucleótido complementario que comprende un complemento de la secuencia diana mediante complementariedad de secuencia entre B y B'; y
 - (d) una polimerasa que extiende el primer cebador y el segundo cebador para producir polinucleótidos amplificados;en donde la distancia entre el extremo 5' de la secuencia A y el extremo 3' de la secuencia B de la secuencia diana es de 75 nt o menos.
11. El método de la reivindicación 1, en donde los polinucleótidos libres de células comprenden ADN tumoral circulante.
12. El método de la reivindicación 1, en donde amplificar los polinucleótidos circulares de (a) forma una pluralidad de concatémeros.
13. El método de la reivindicación 1, en donde la variante de secuencia es una variante de secuencia rara que ocurre a una frecuencia de menos de aproximadamente un 5 %, 4 %, 3 %, 2 %, 1 %, 0,5 % o 0,1 %.
14. El método de la reivindicación 1, en donde dicha variante de secuencia es indicativa de cáncer.
15. El método de la reivindicación 1, en donde la muestra de ácidos nucleicos comprende menos de 50 ng, 45 ng, 40

ng, 35 ng, 30 ng, 25 ng, 20 ng, 15 ng, 10 ng, 5 ng, 4 ng, 3 ng, 2 ng, o 1 ng de polinucleótidos libres de células.

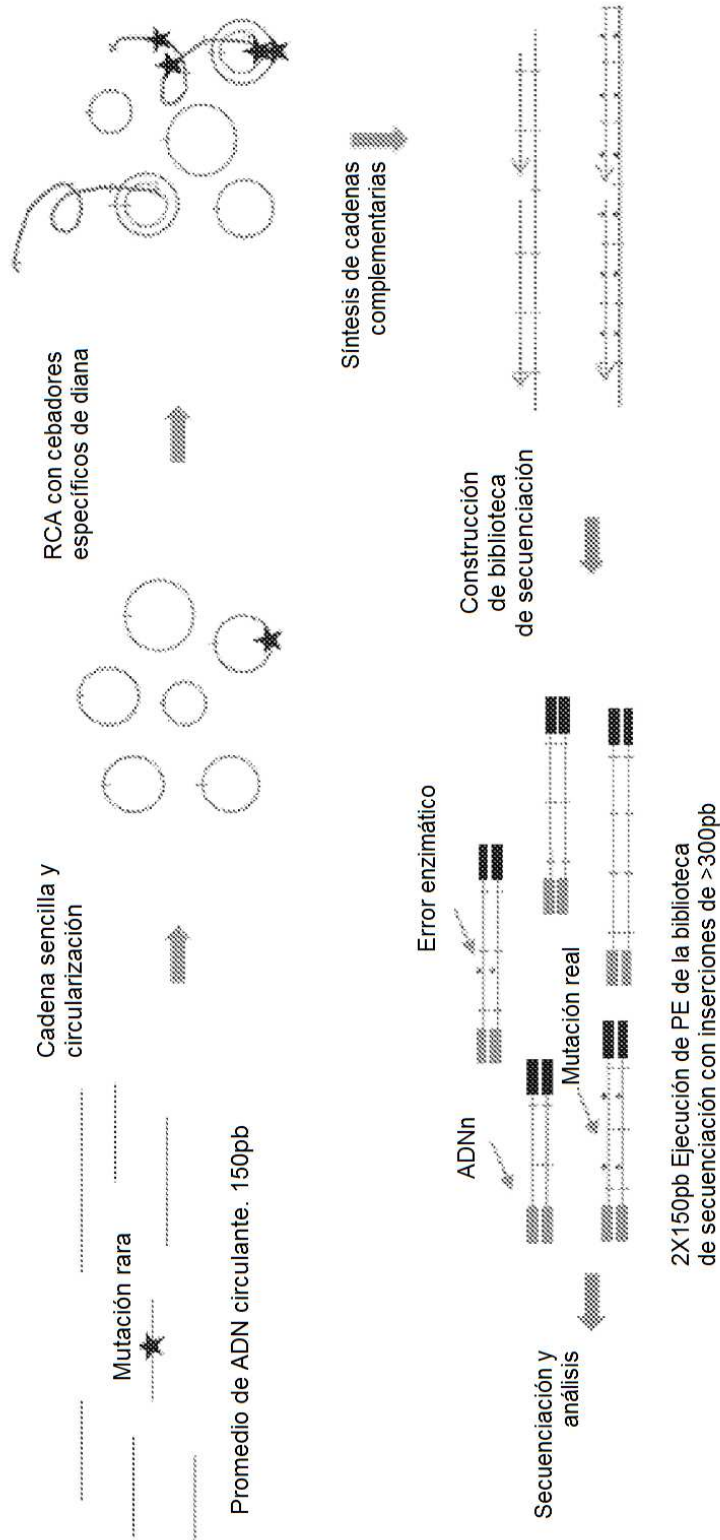


FIG. 1

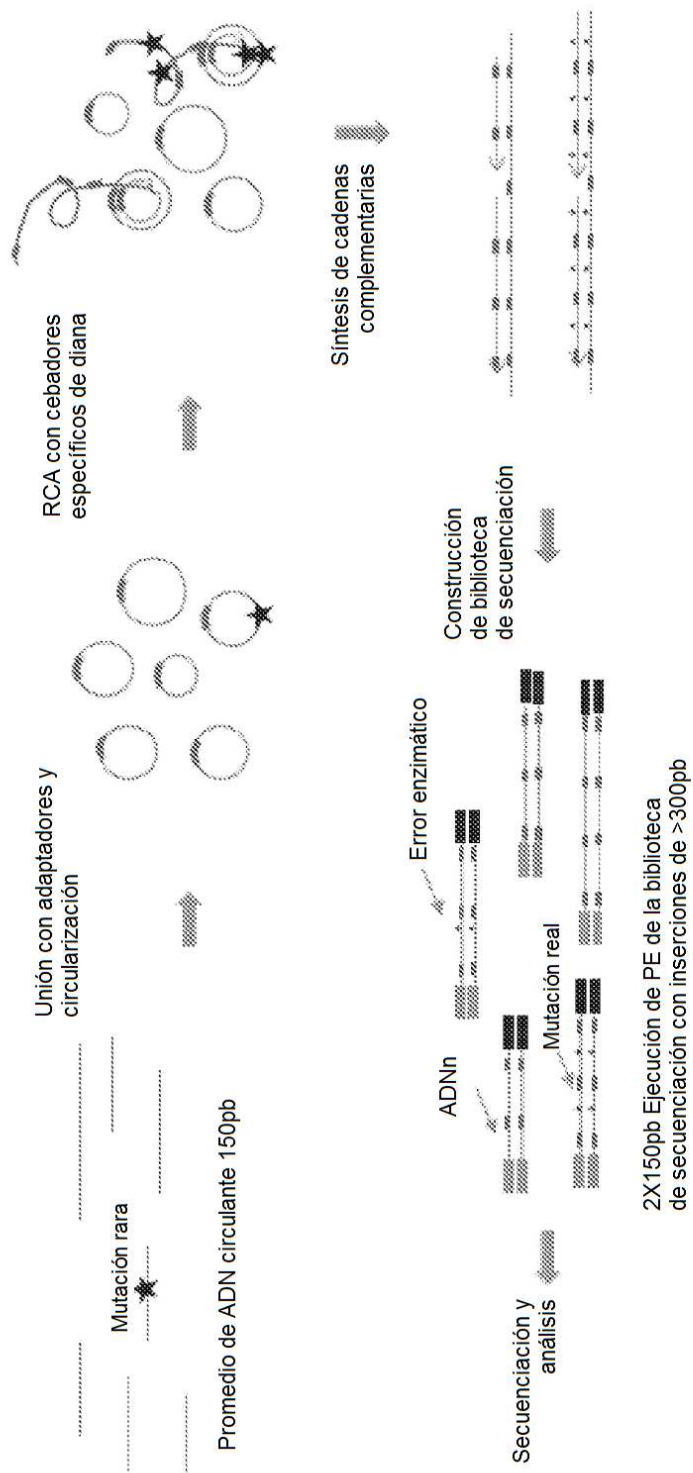


FIG. 2

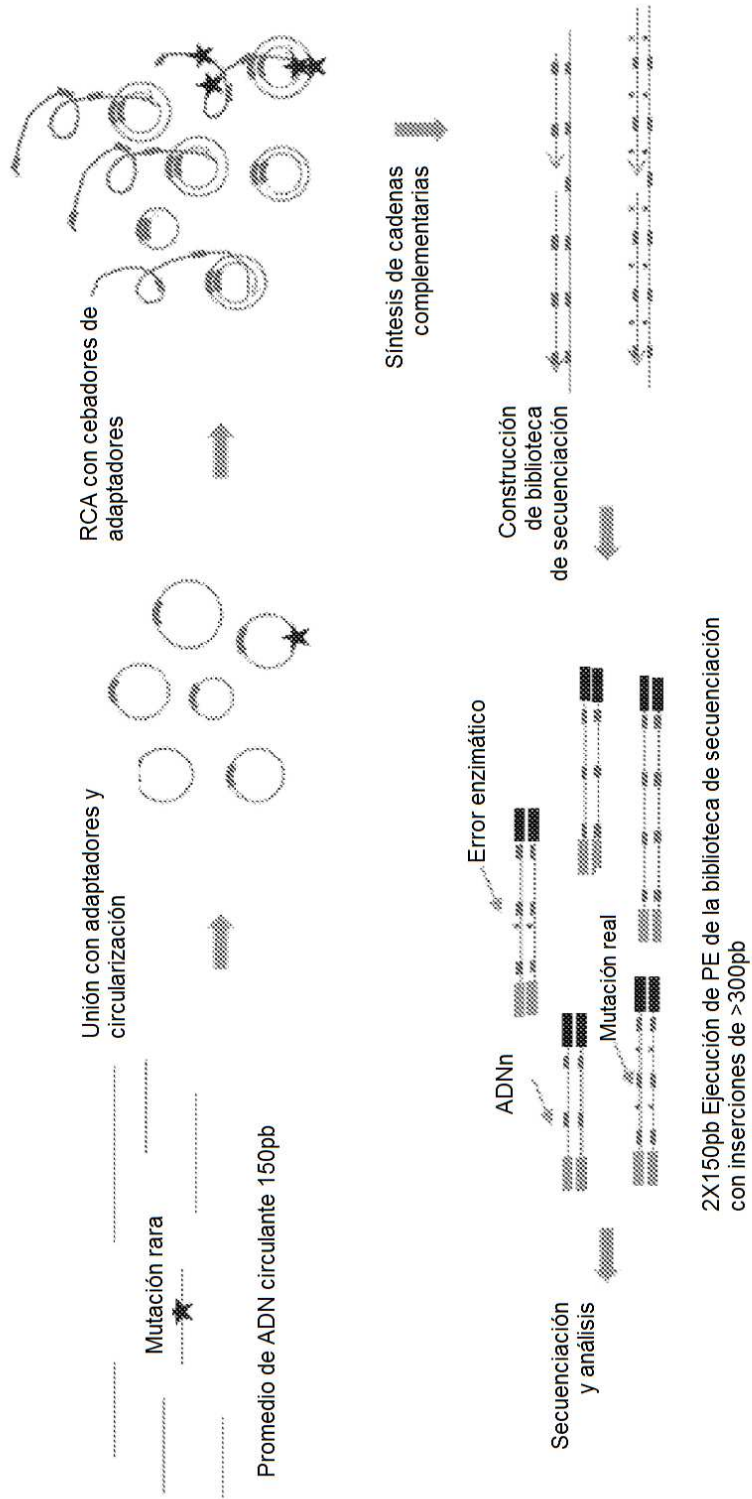


FIG. 3

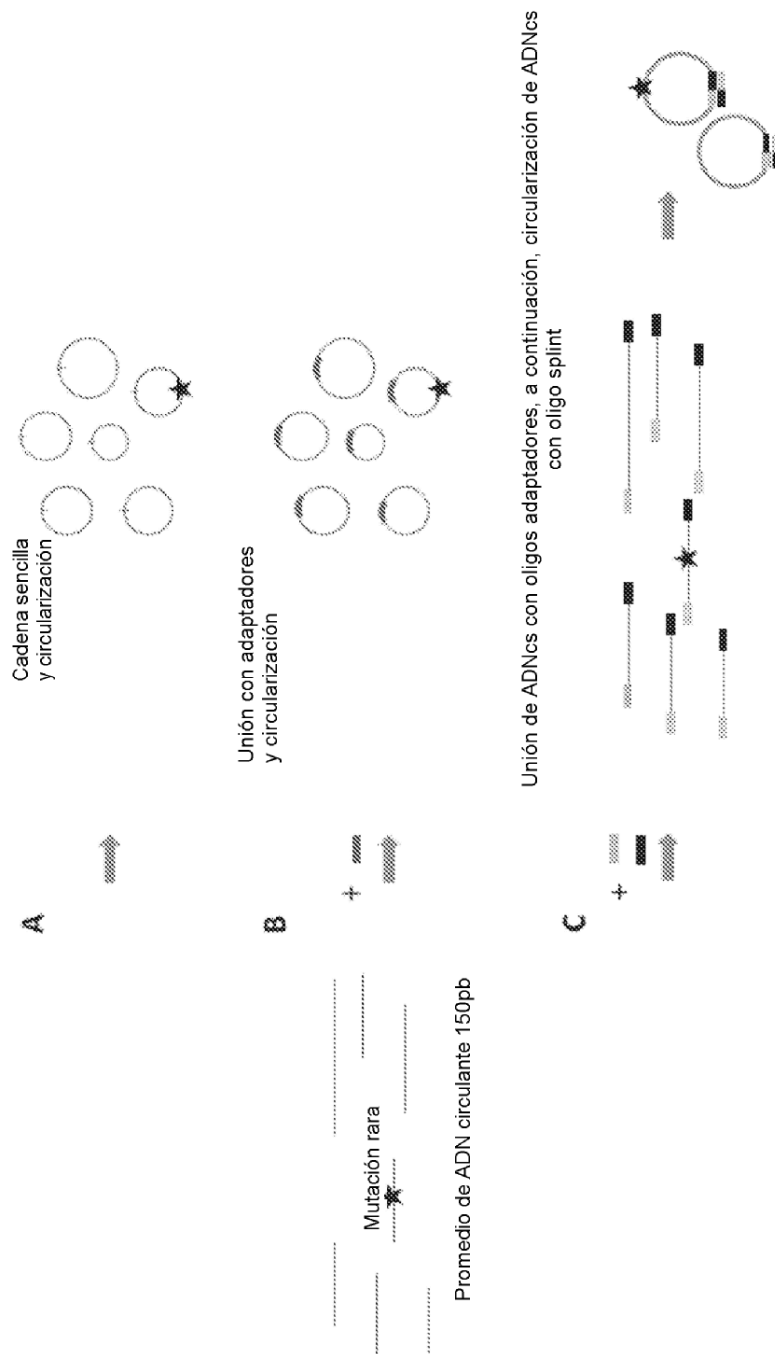


FIG. 4

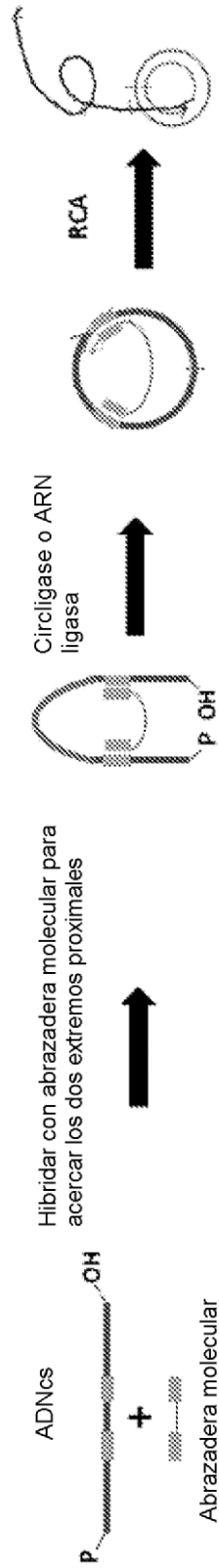


FIG. 5

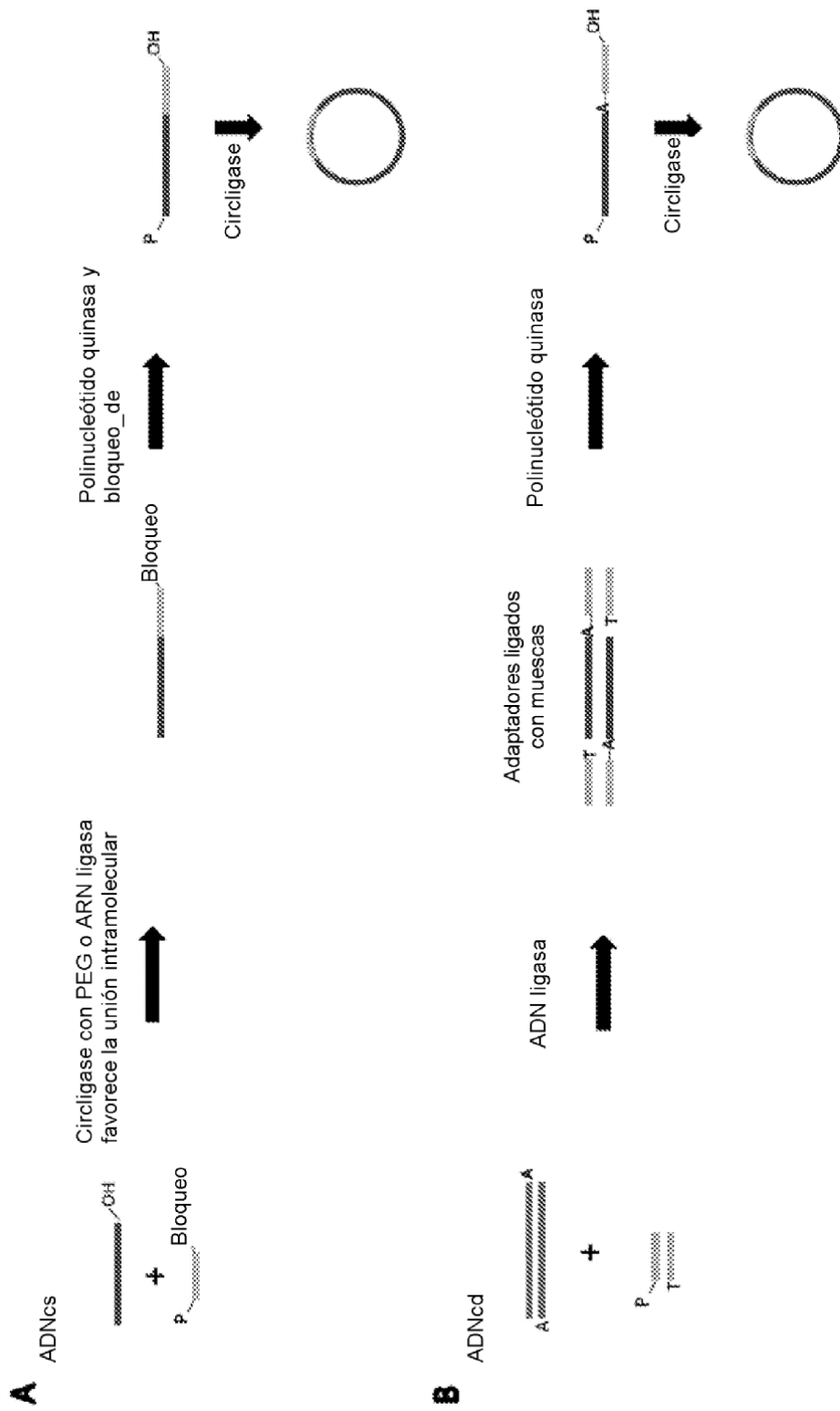


FIG. 6

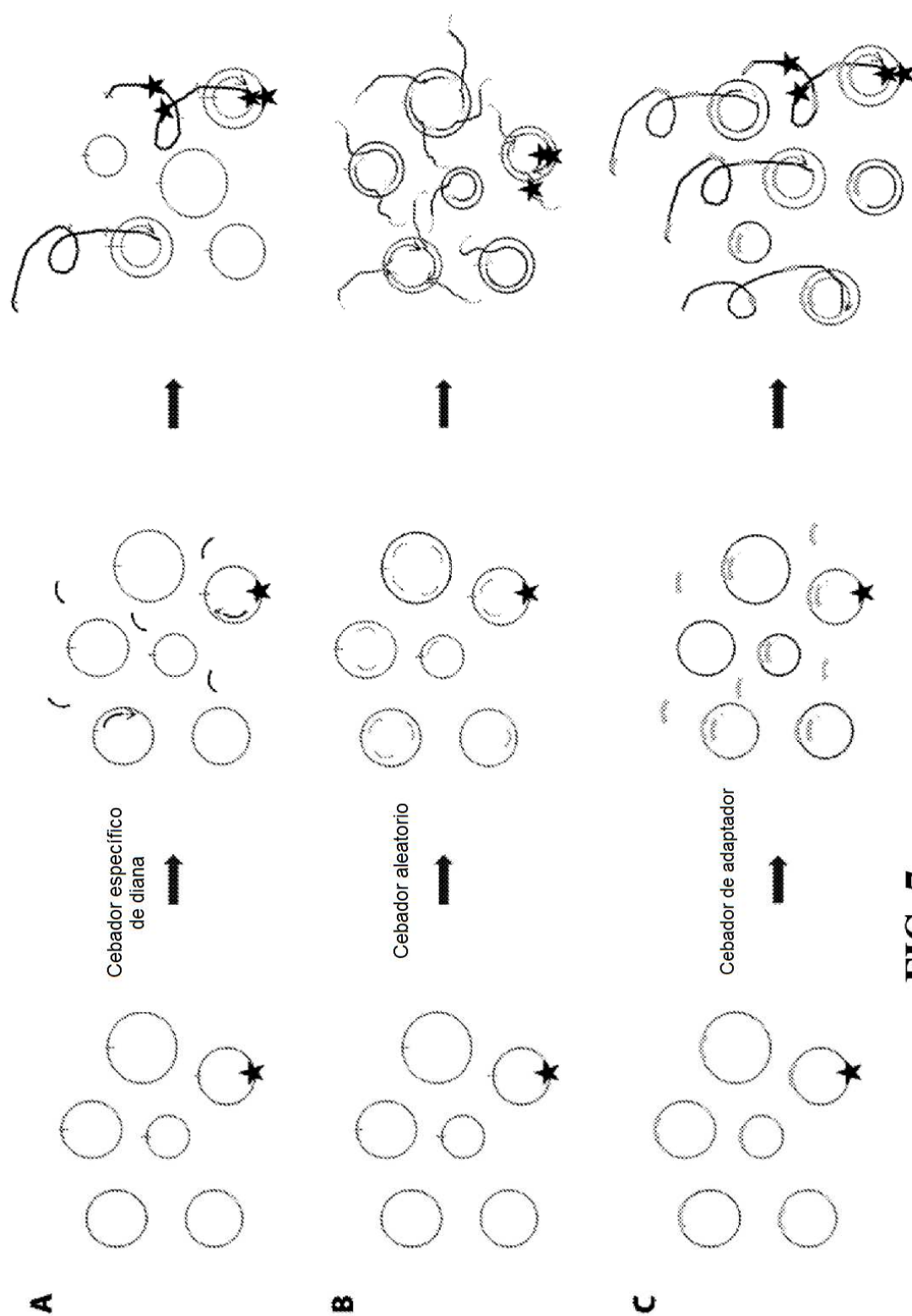


FIG. 7

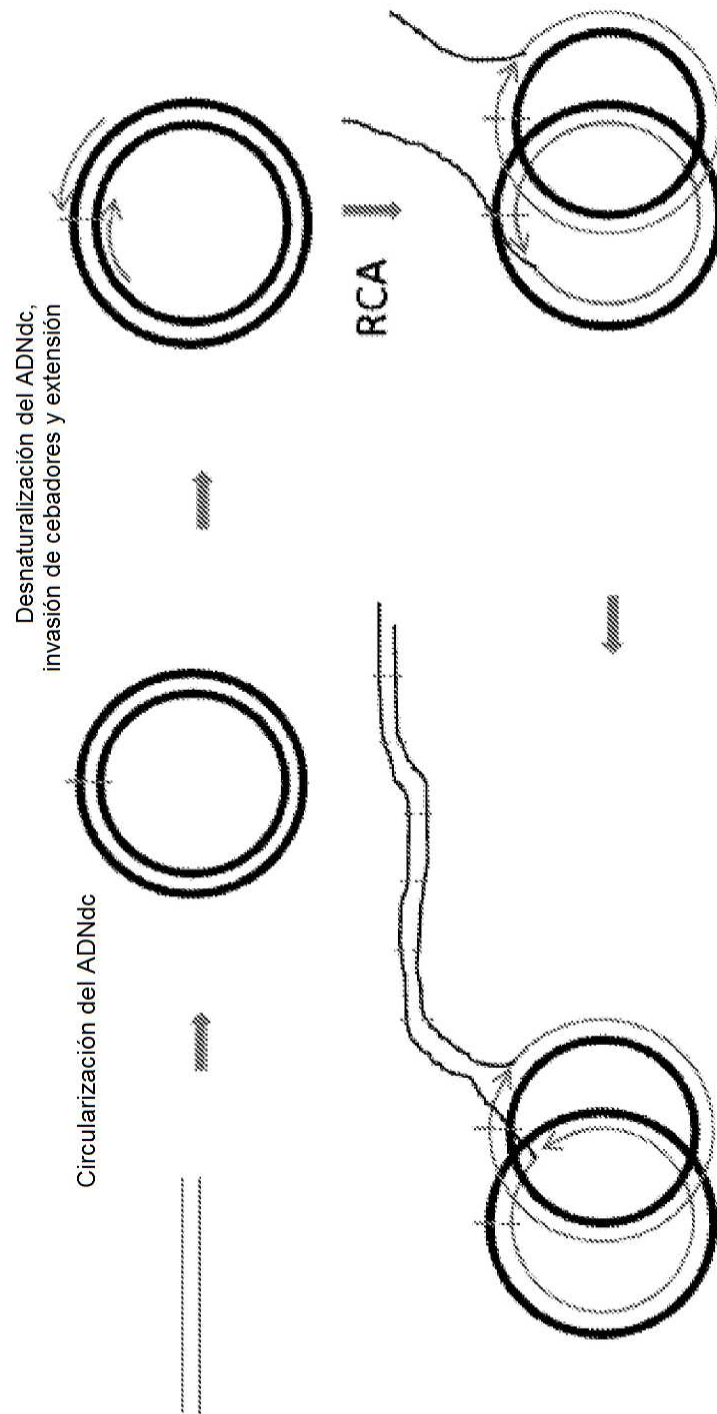


FIG. 8

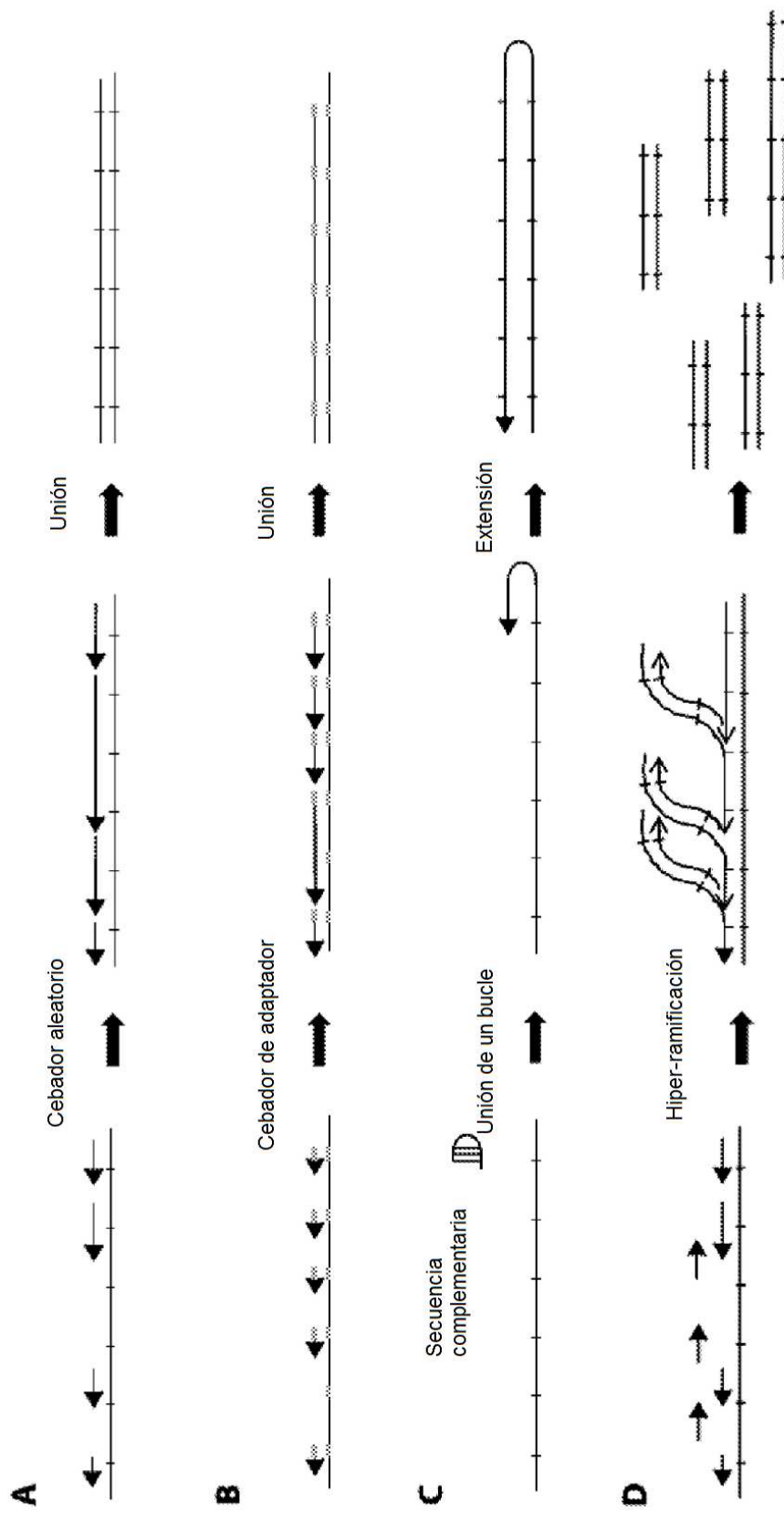


FIG. 9

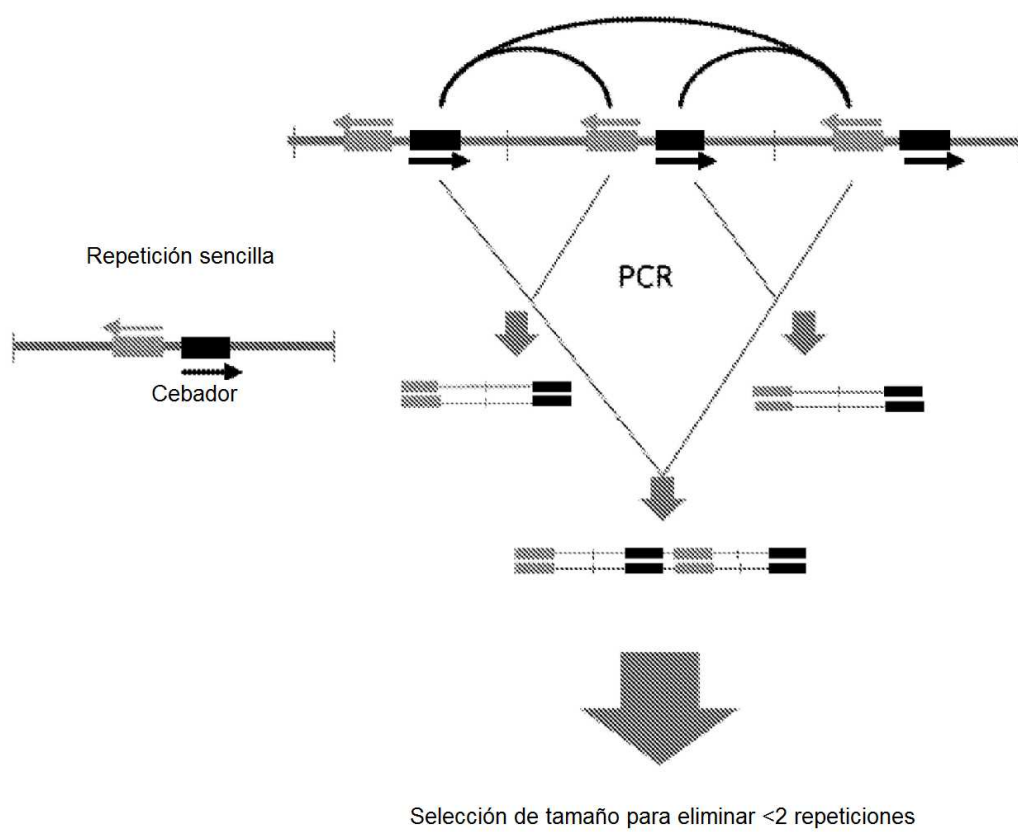


FIG. 10

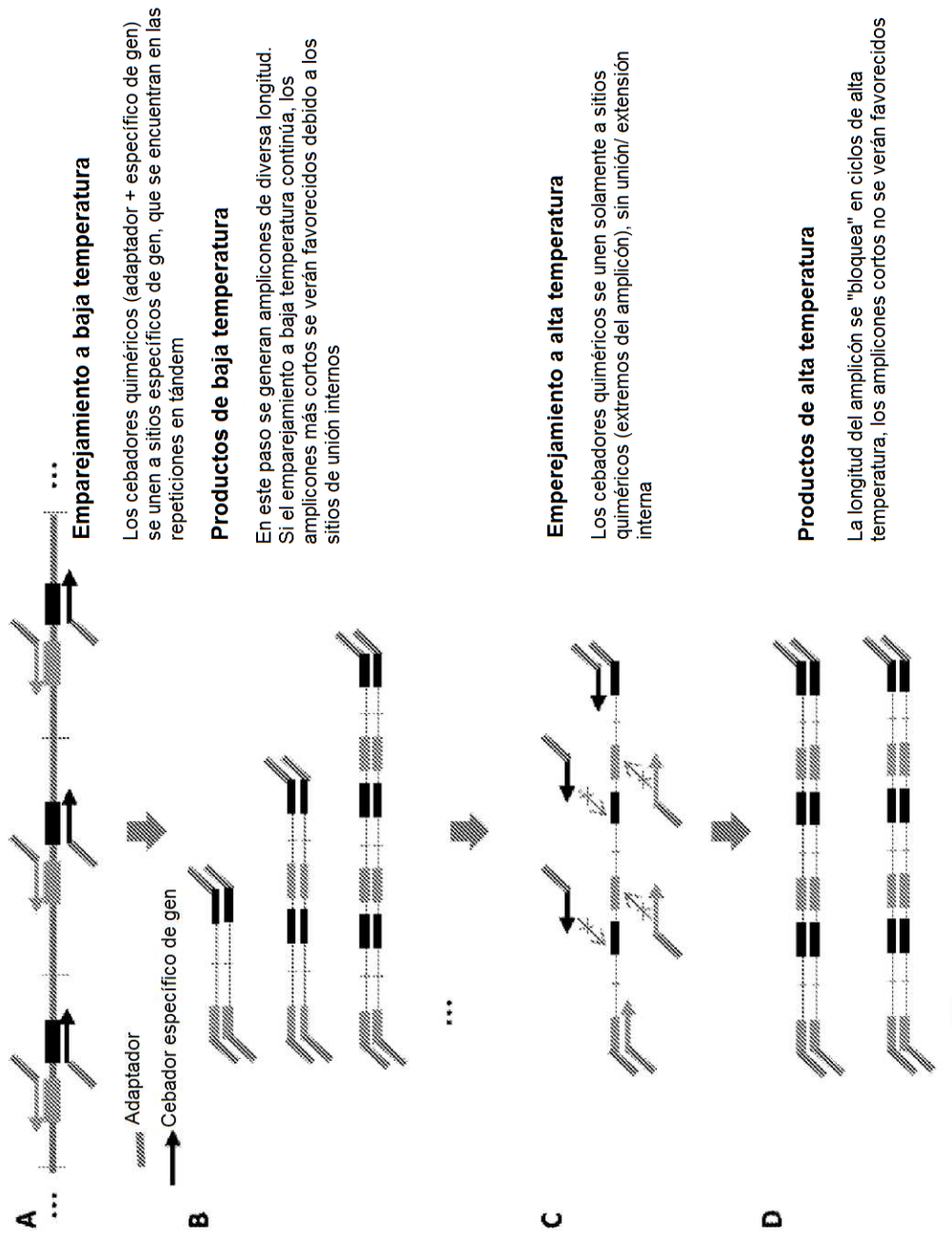


FIG. 11

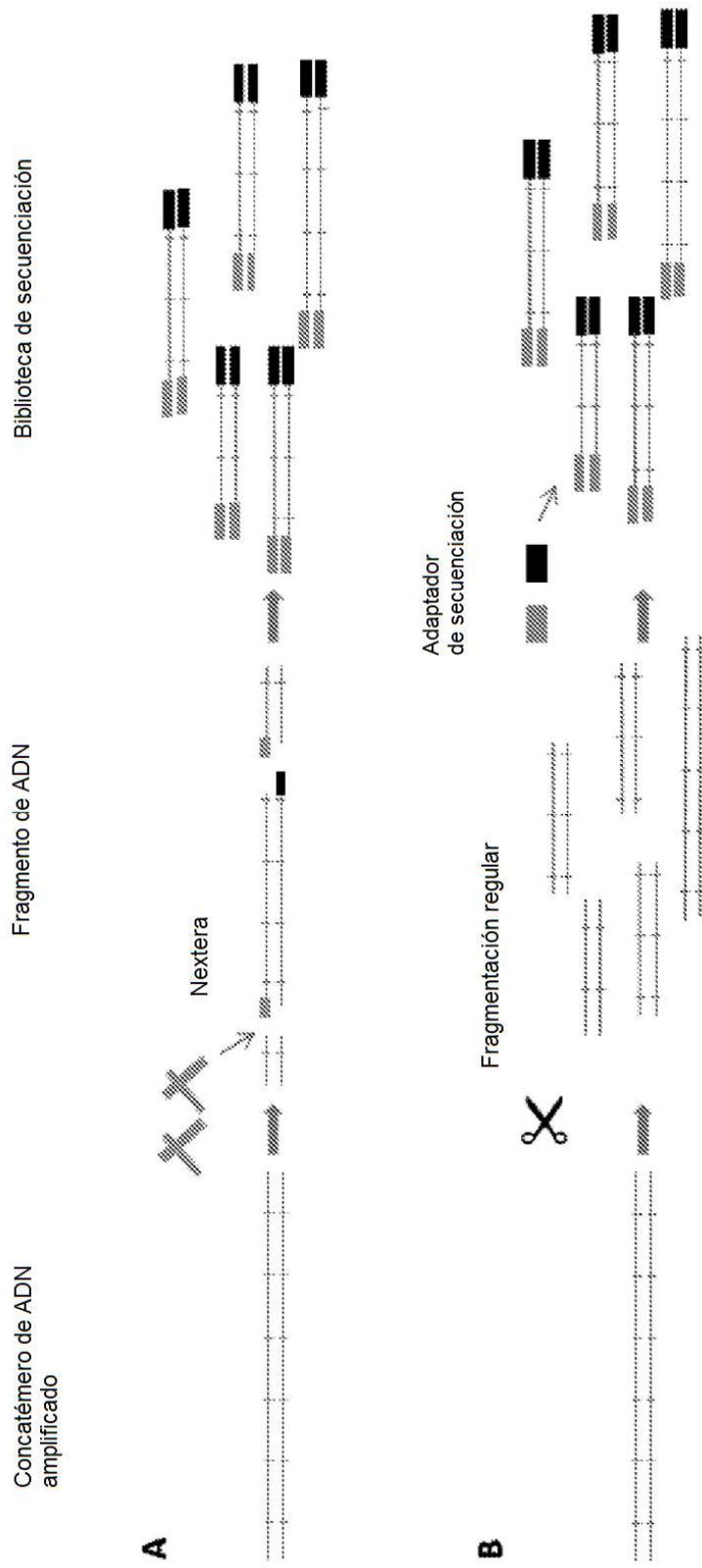


FIG. 12

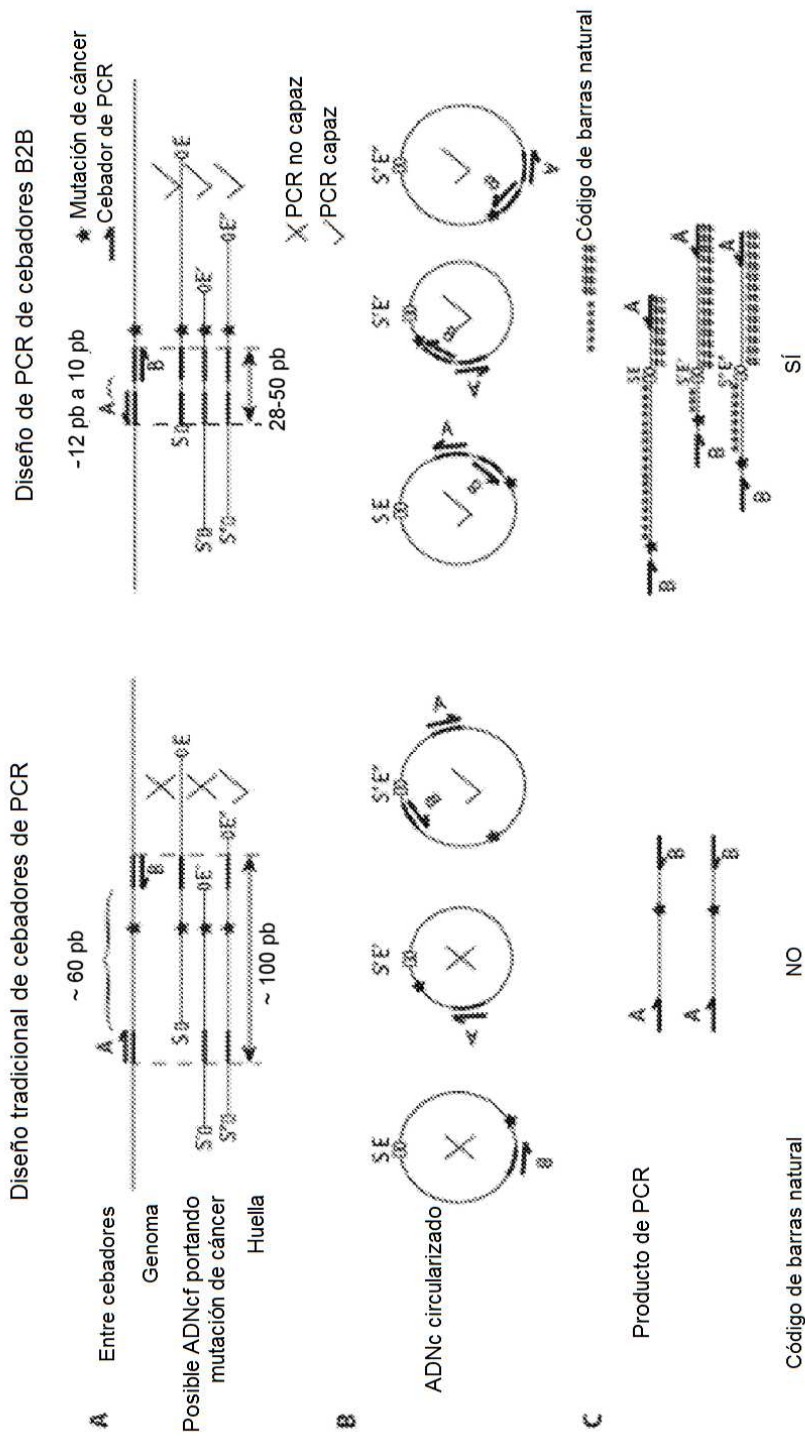


FIG. 13

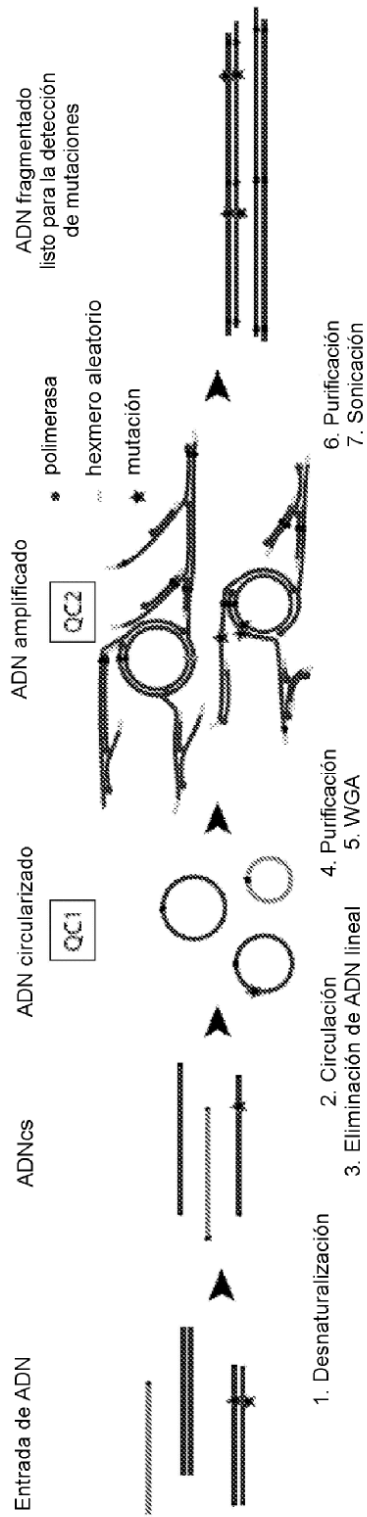


FIG. 14

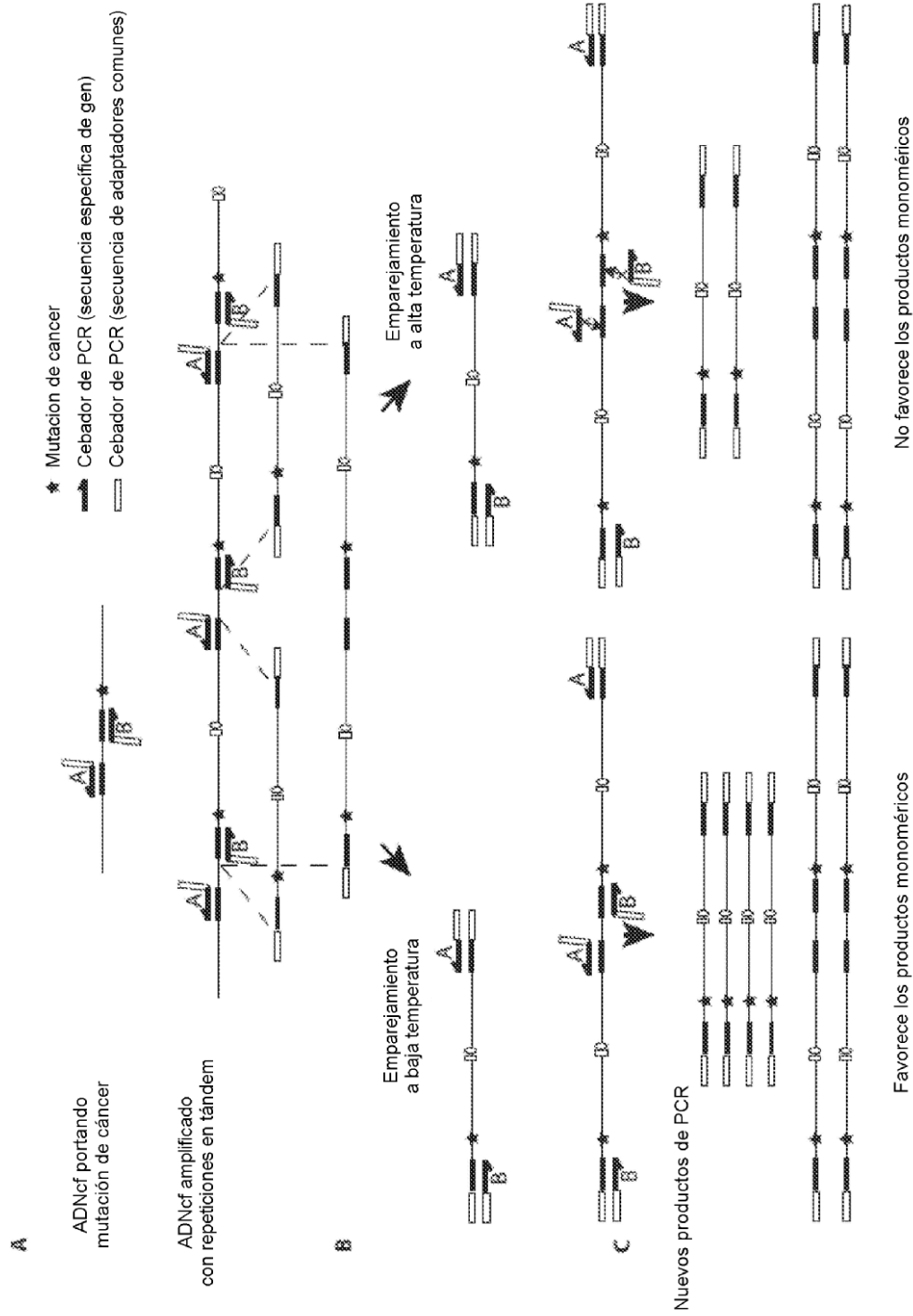


FIG. 15

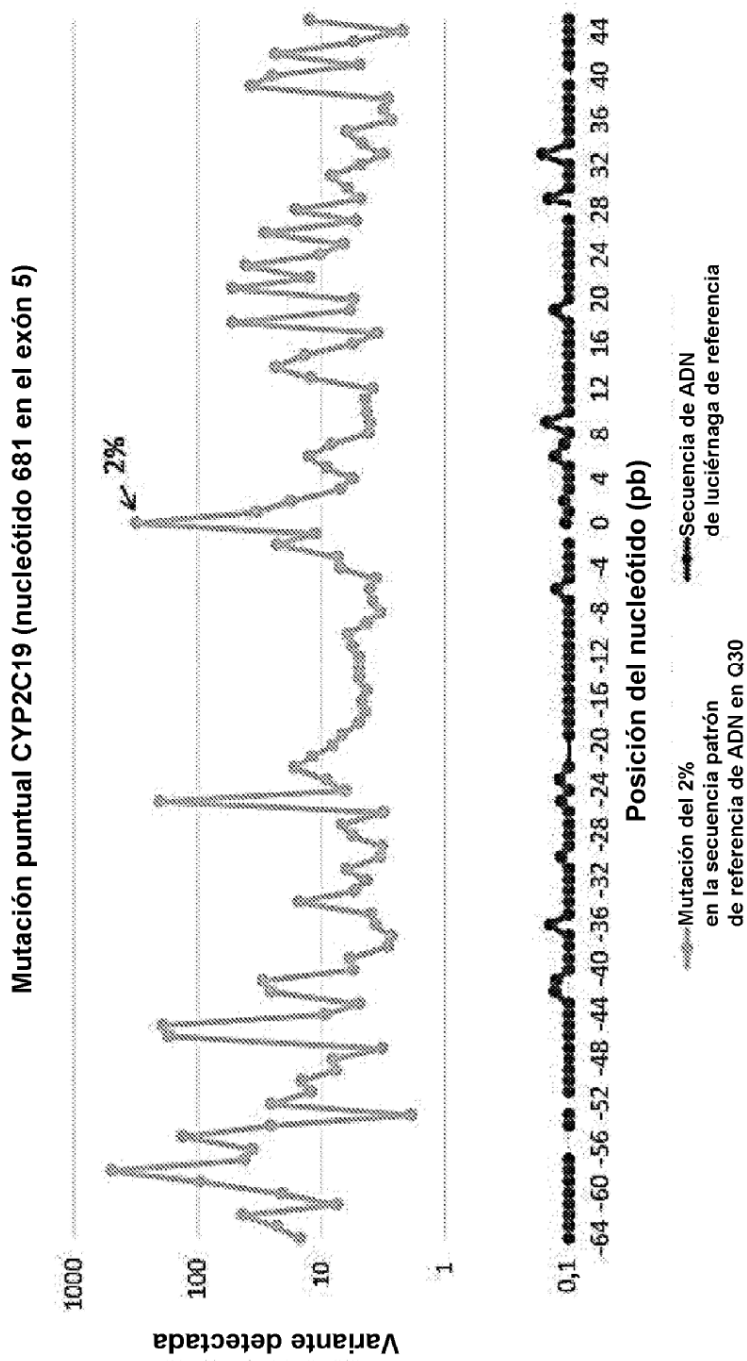


FIG. 16

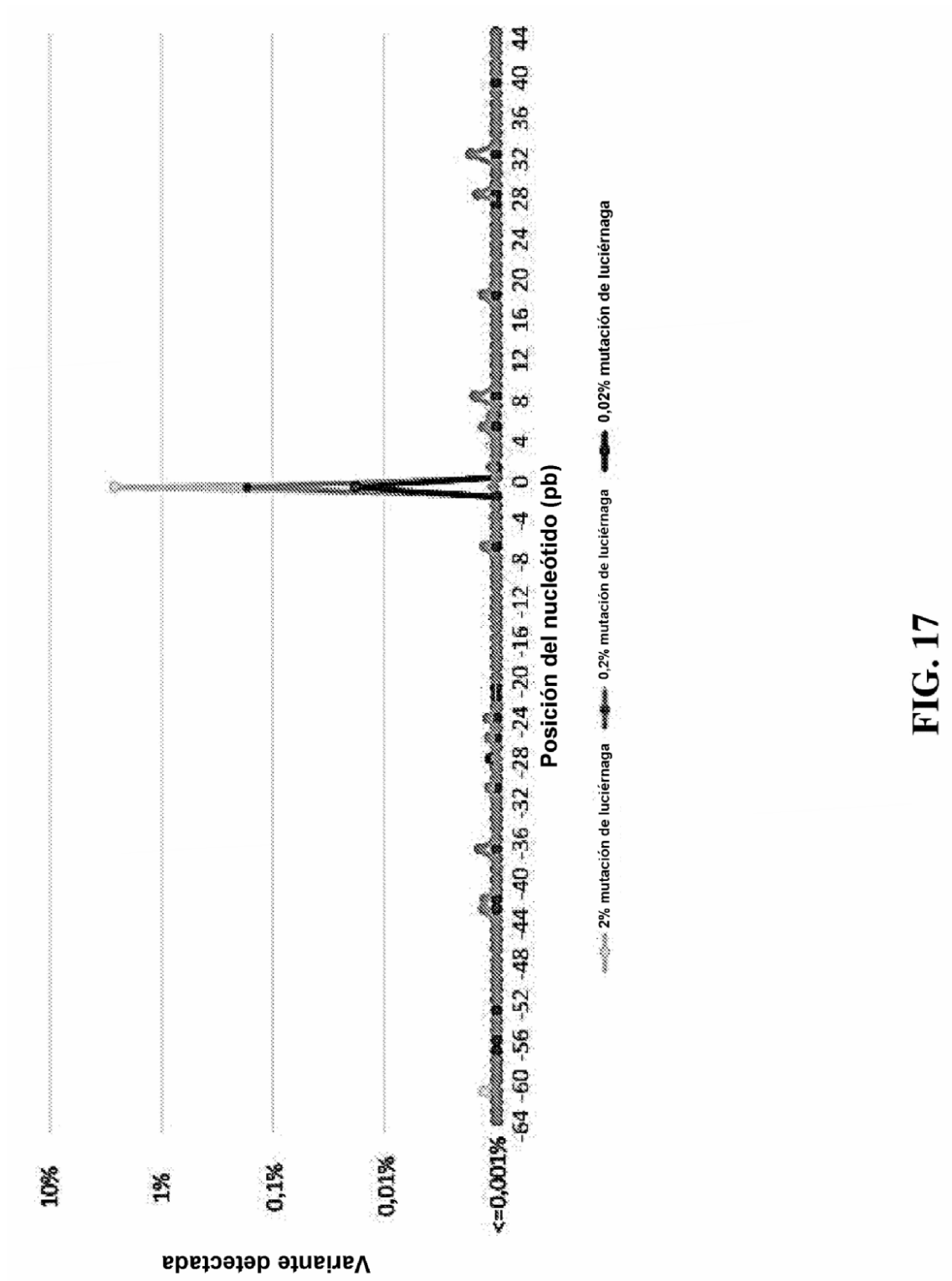


FIG. 17

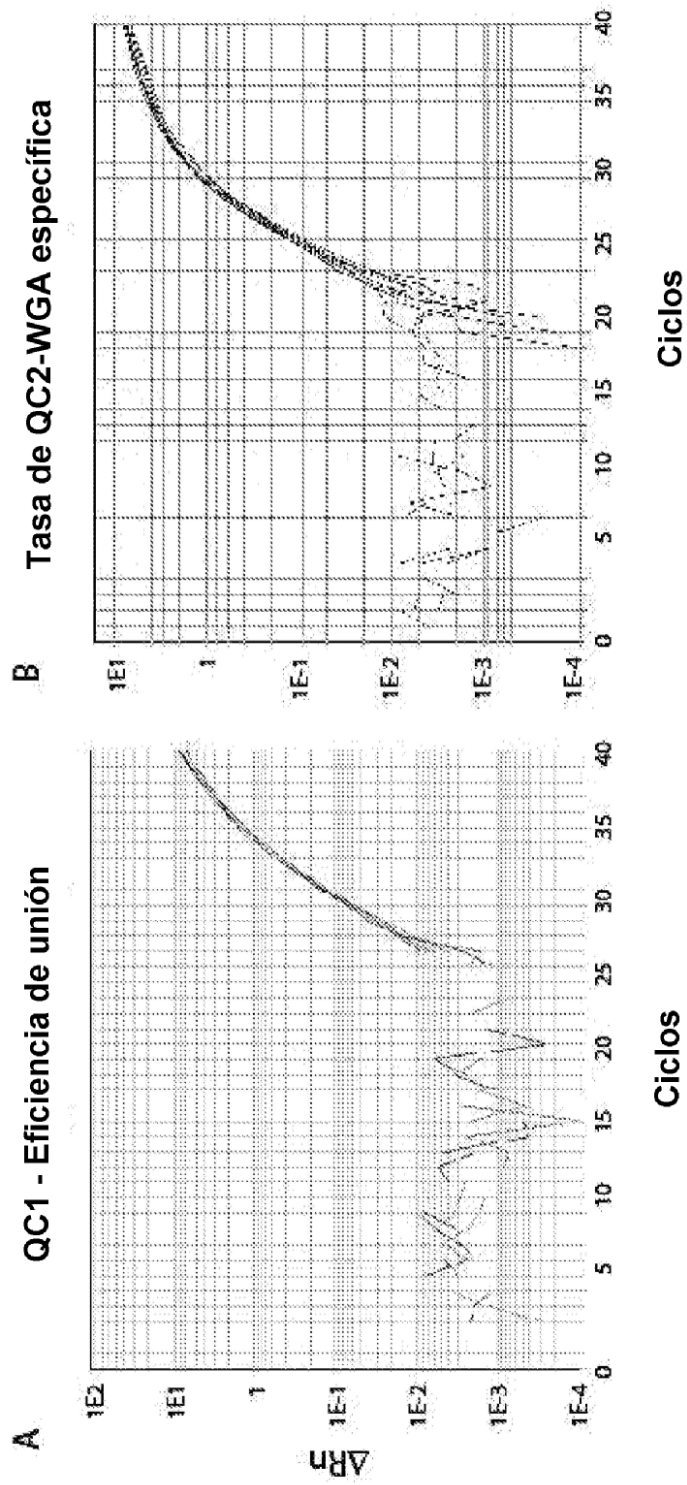
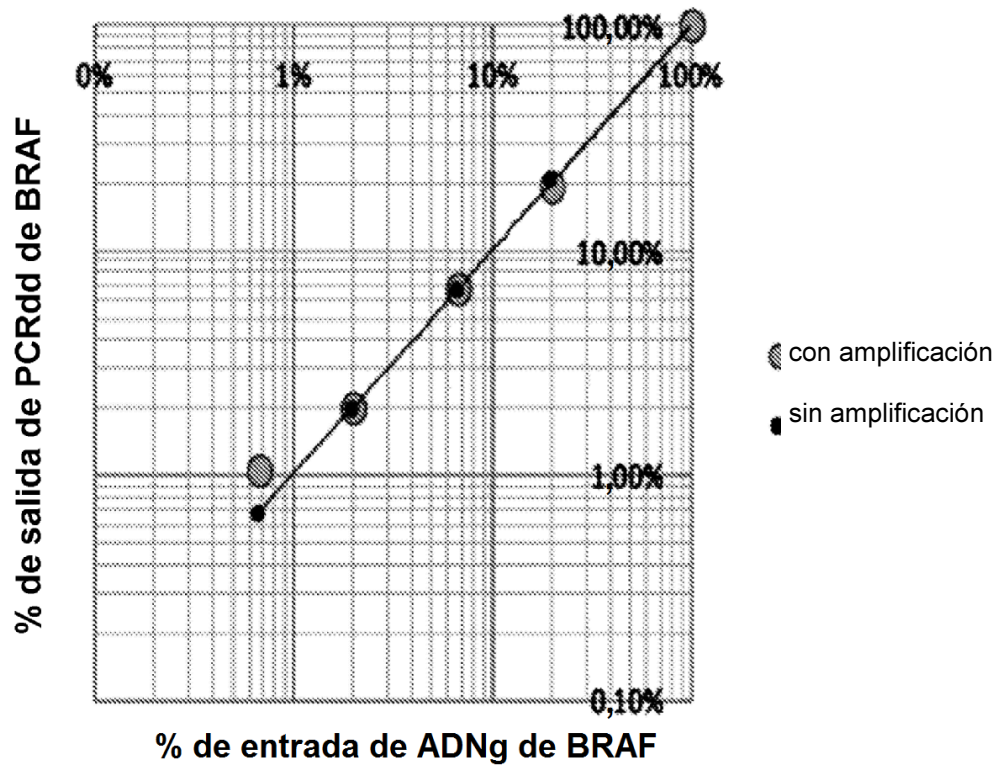
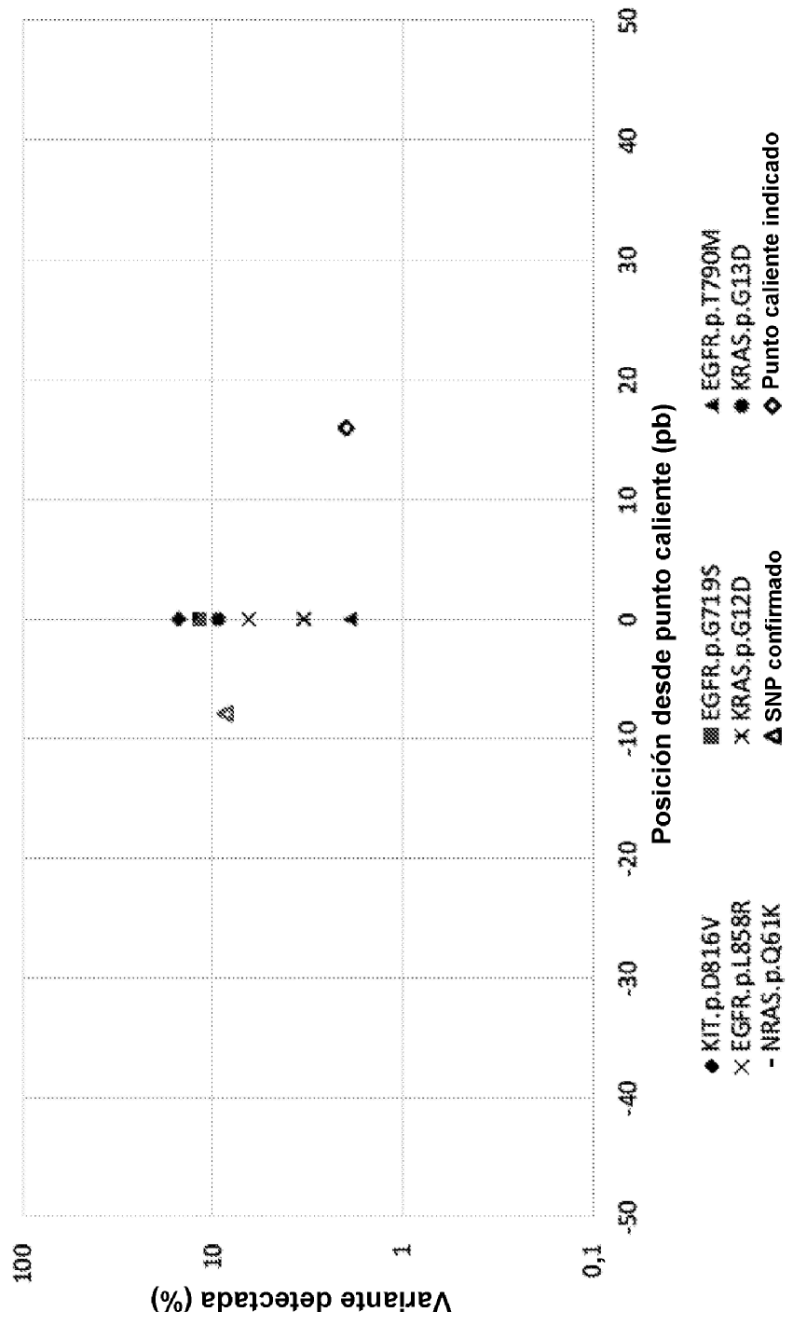


FIG. 18



% esperado	0%	0,67%	2,00%	6,67%	20,00%	100%
con amplificación	0%	1,07%	2,04%	6,77%	19,41%	100%
sin amplificación 1		0,66%	1,97%	6,67%	20,26%	

FIG. 19



Mapeo de variantes de luciérnaga Q60, +/- 50 pb desde puntos calientes de referencia

FIG. 20

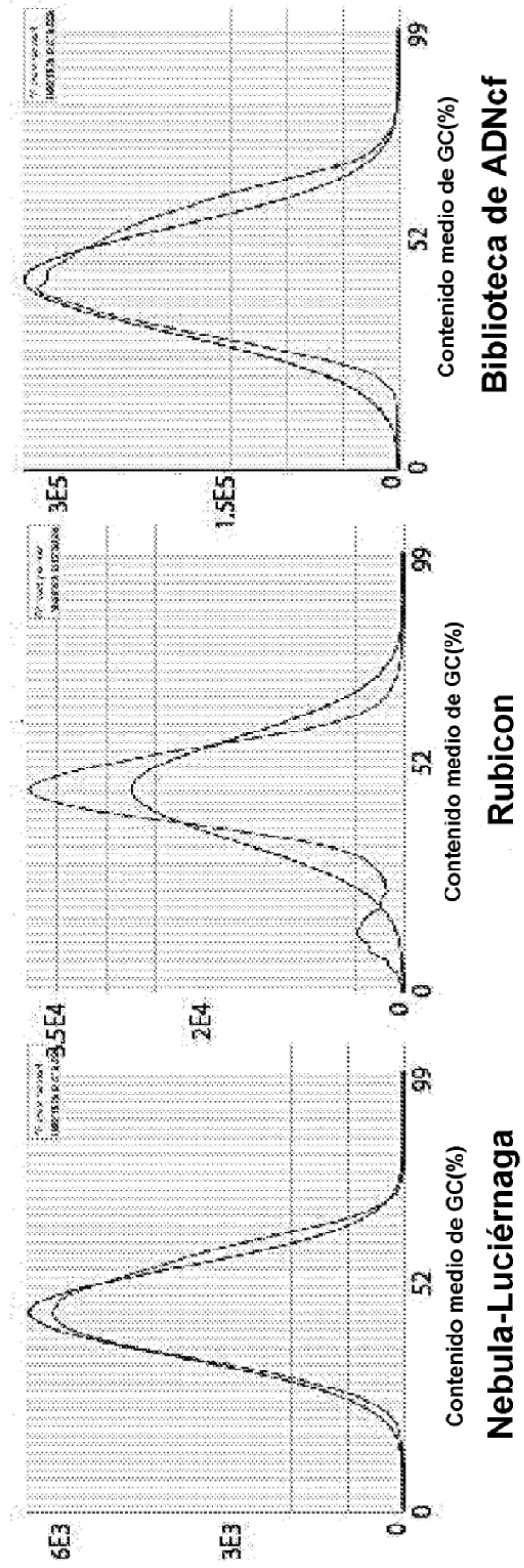


FIG. 22

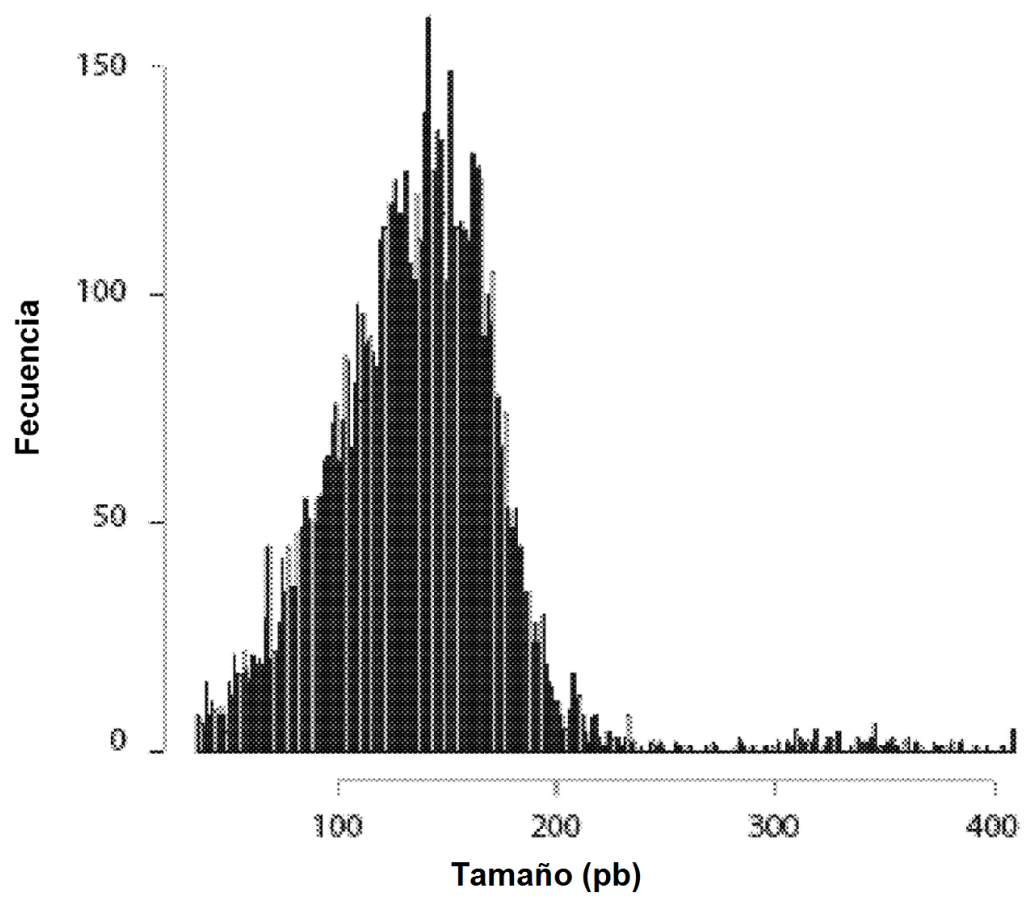


FIG. 23

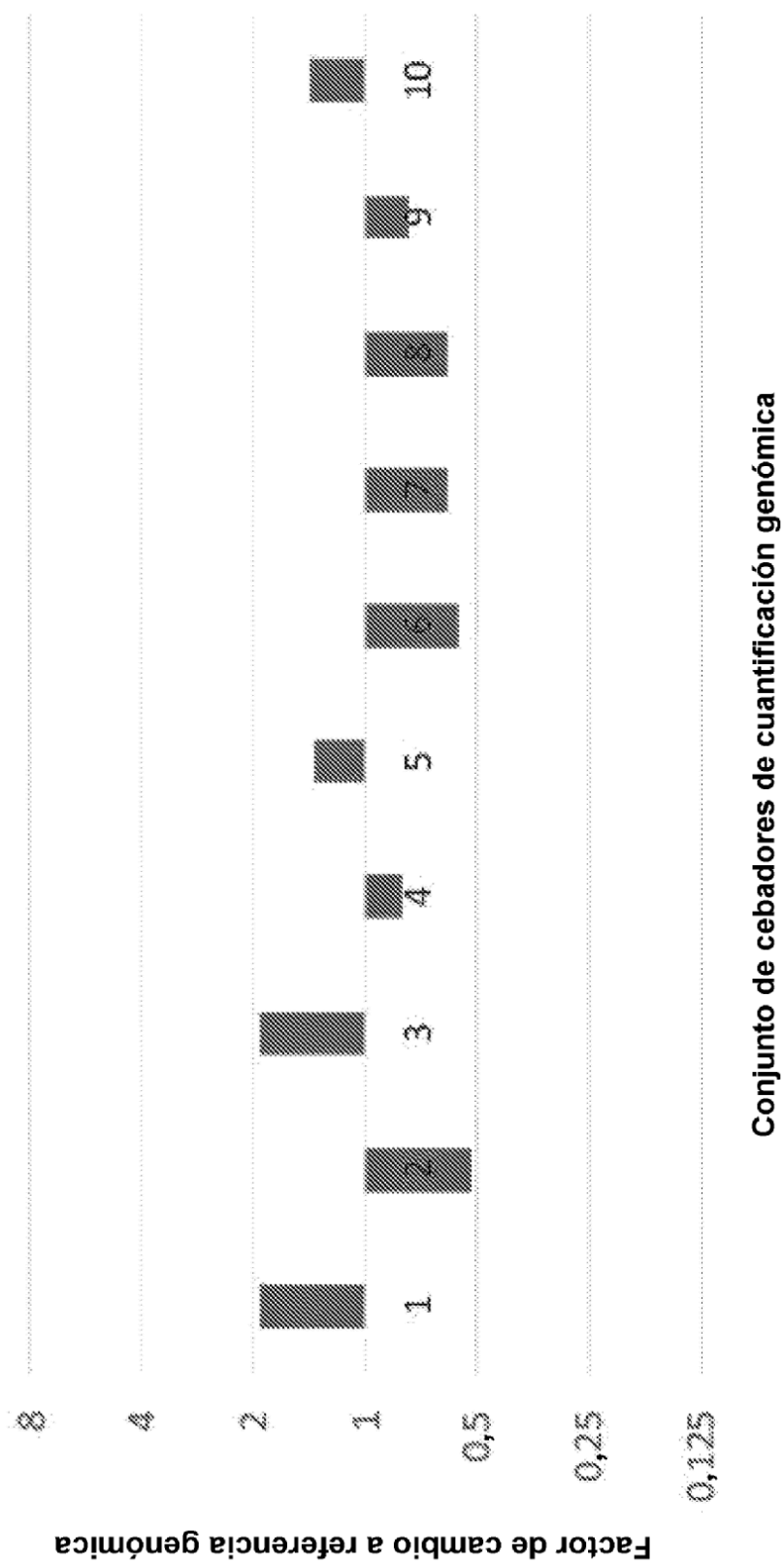


FIG. 24

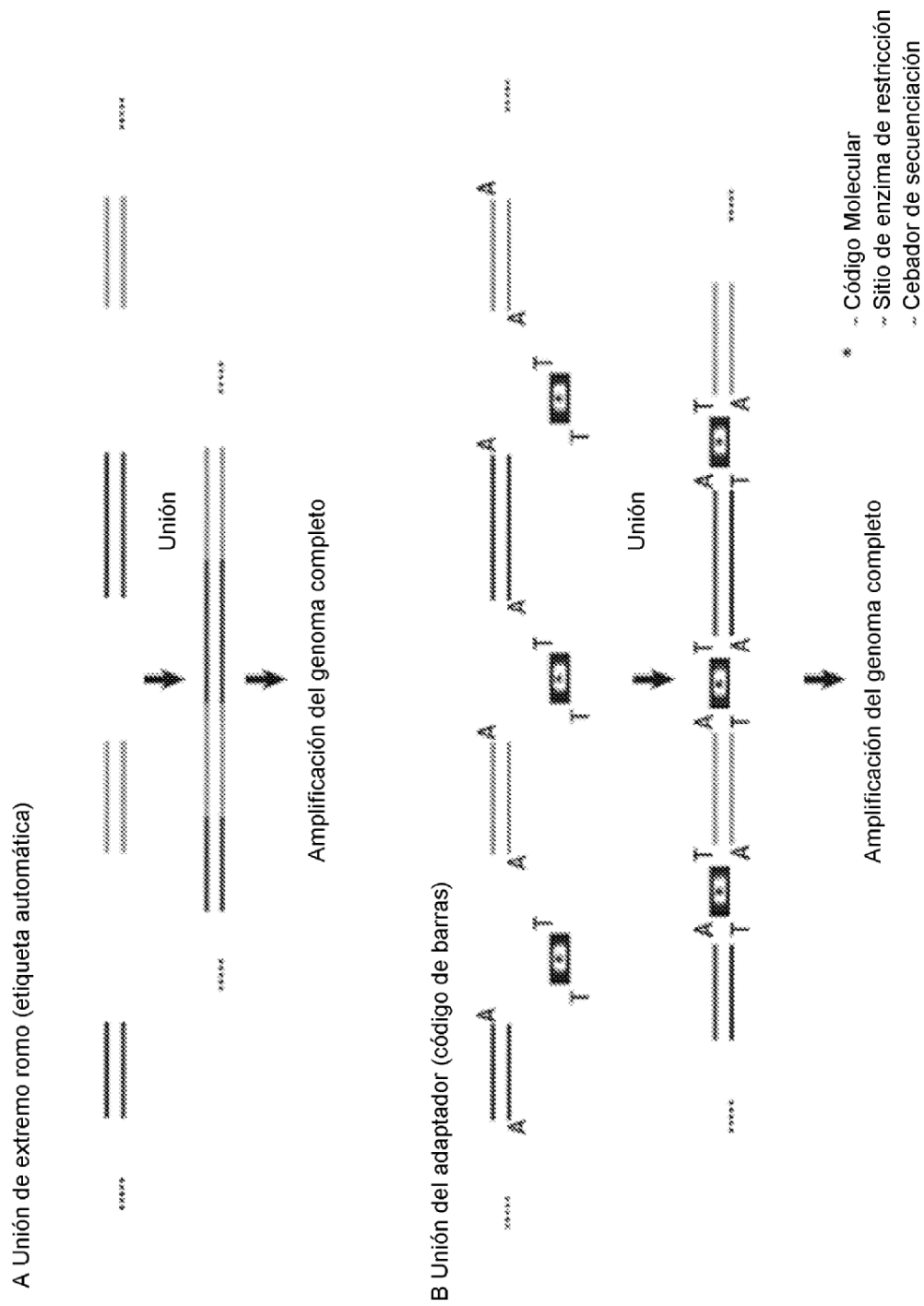


FIG. 25

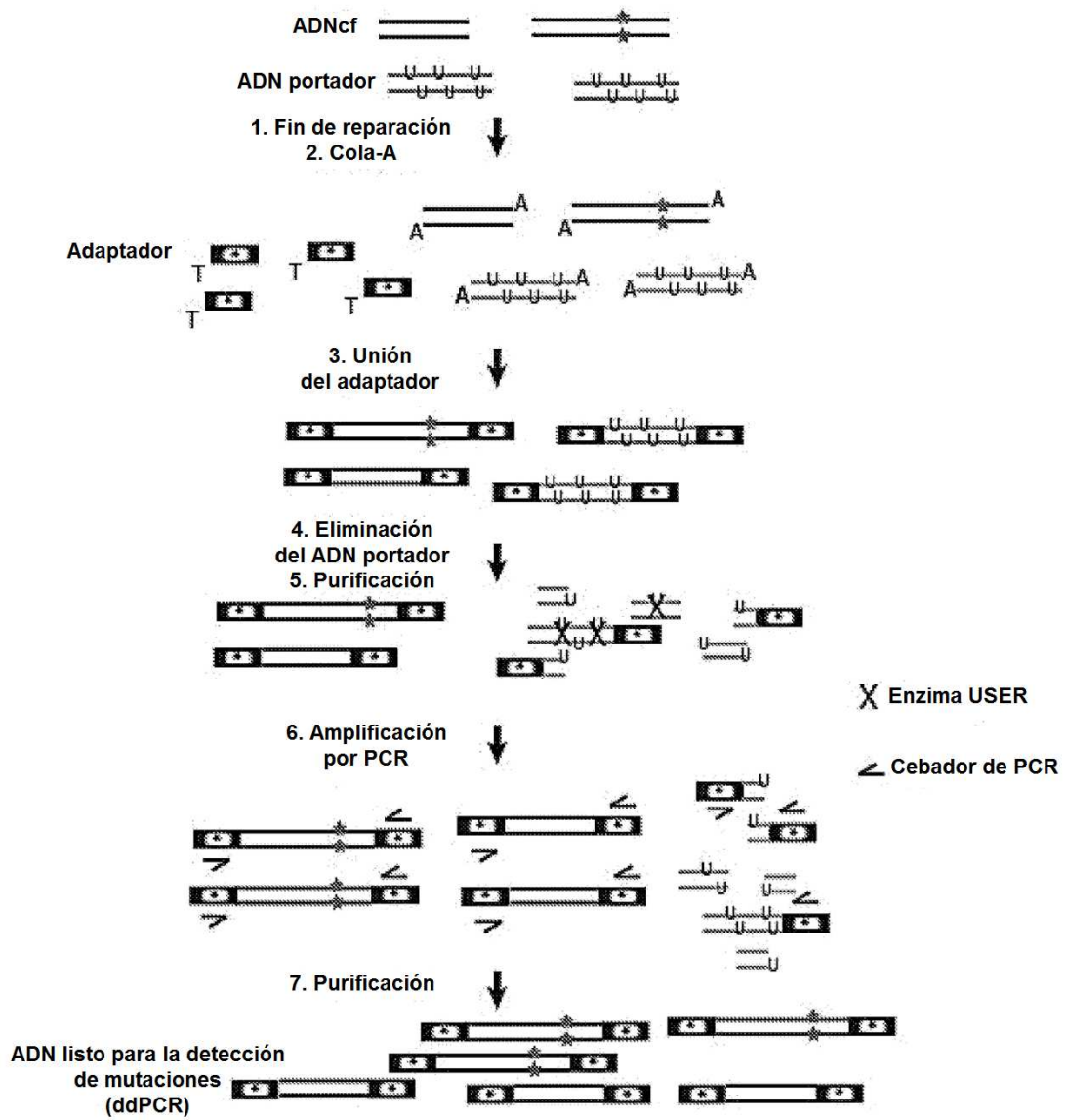


FIG. 26

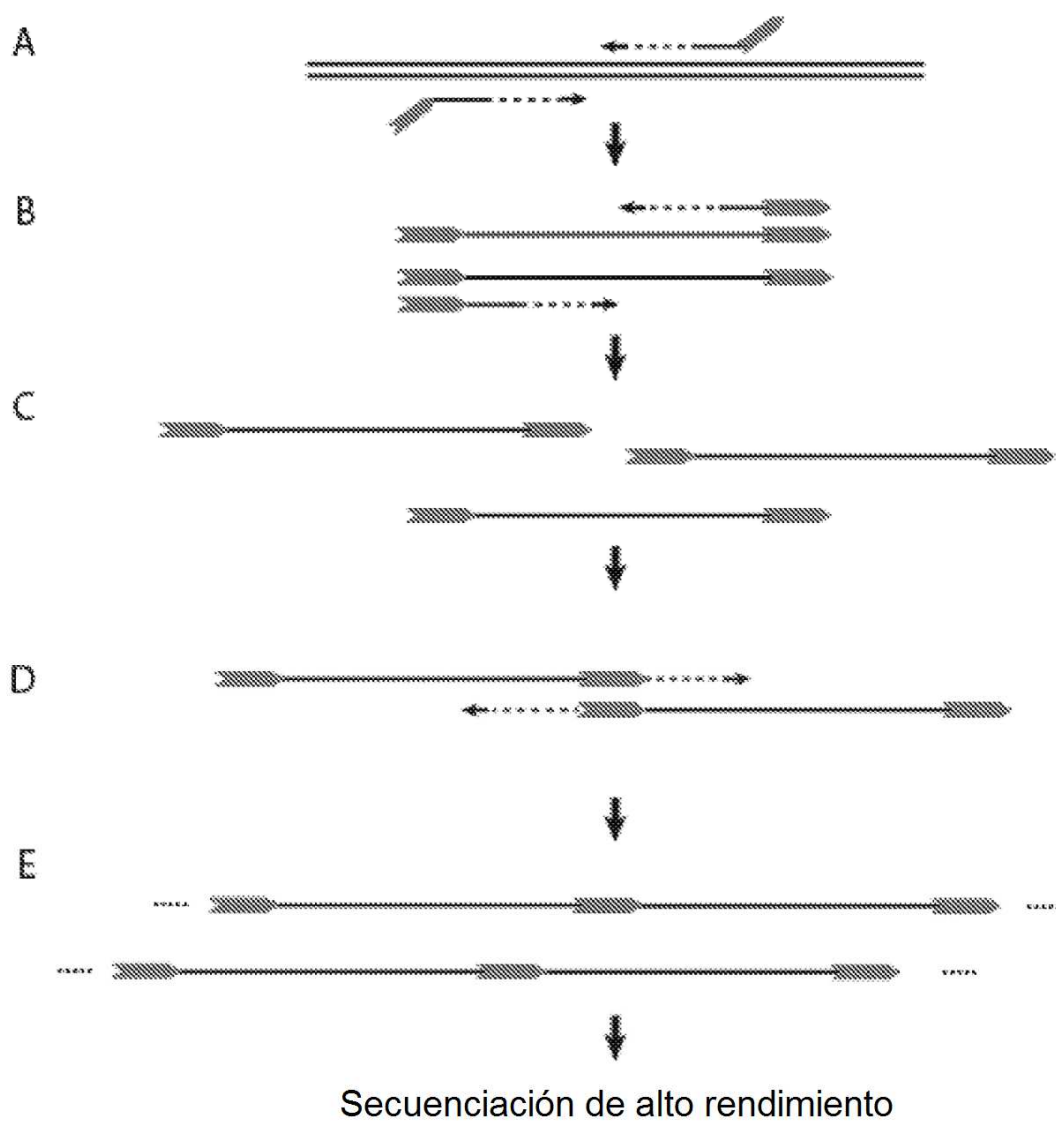


FIG. 27

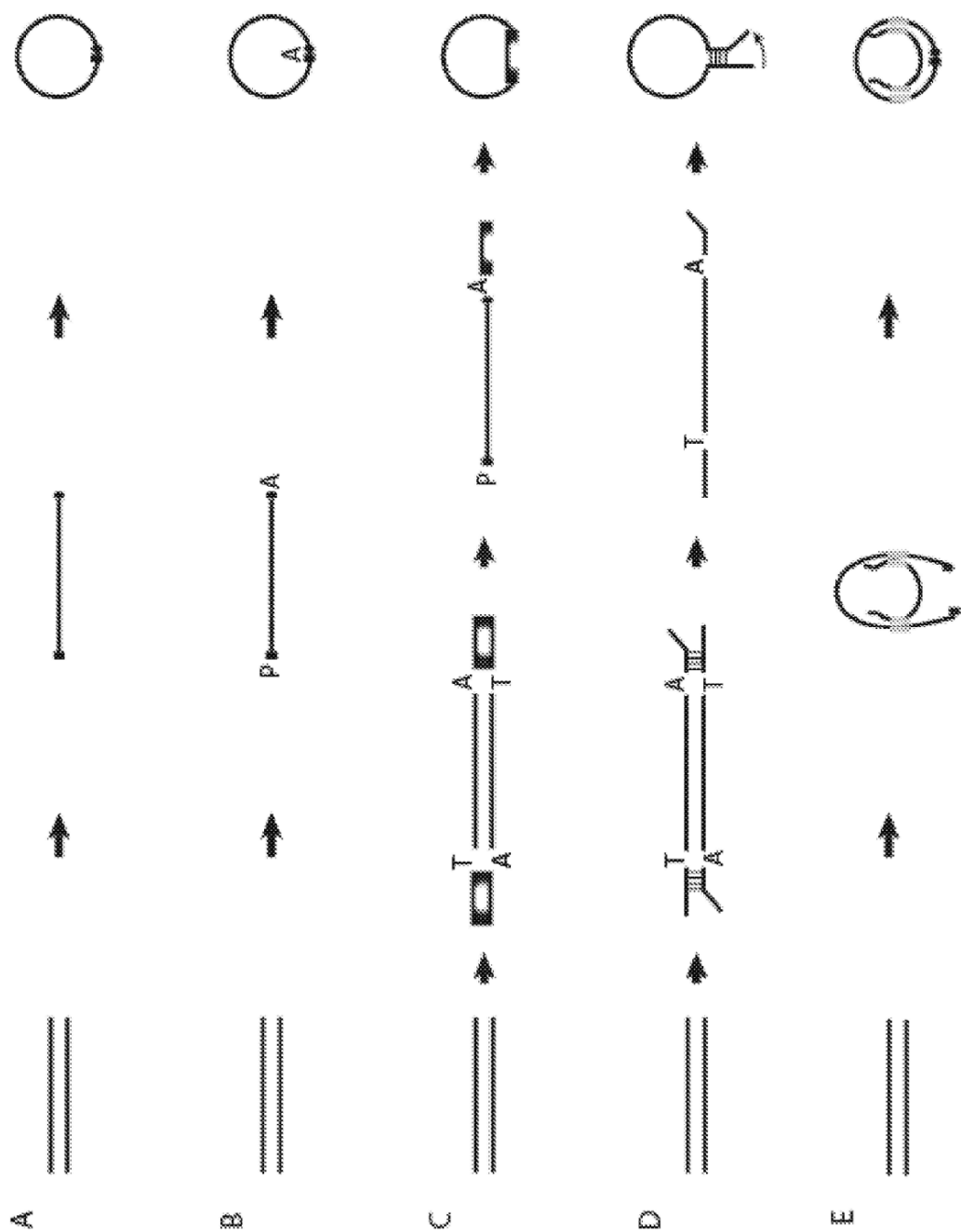


FIG. 28

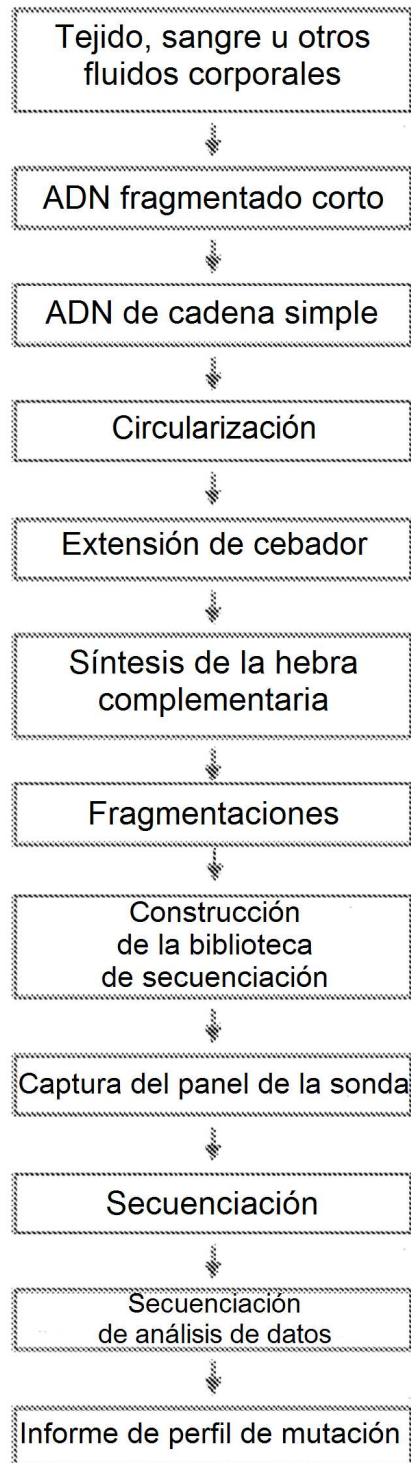


FIG. 29

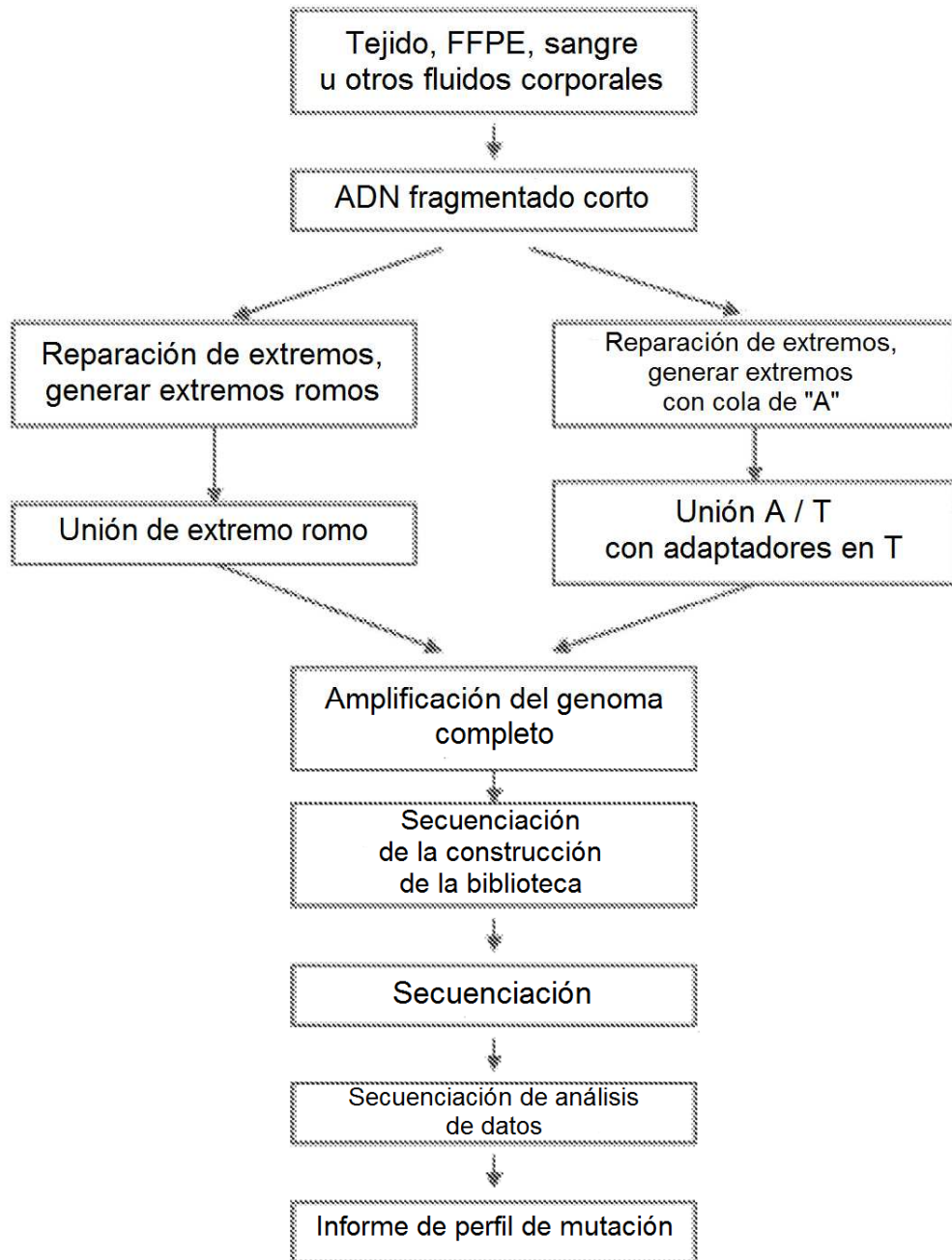


FIG. 30

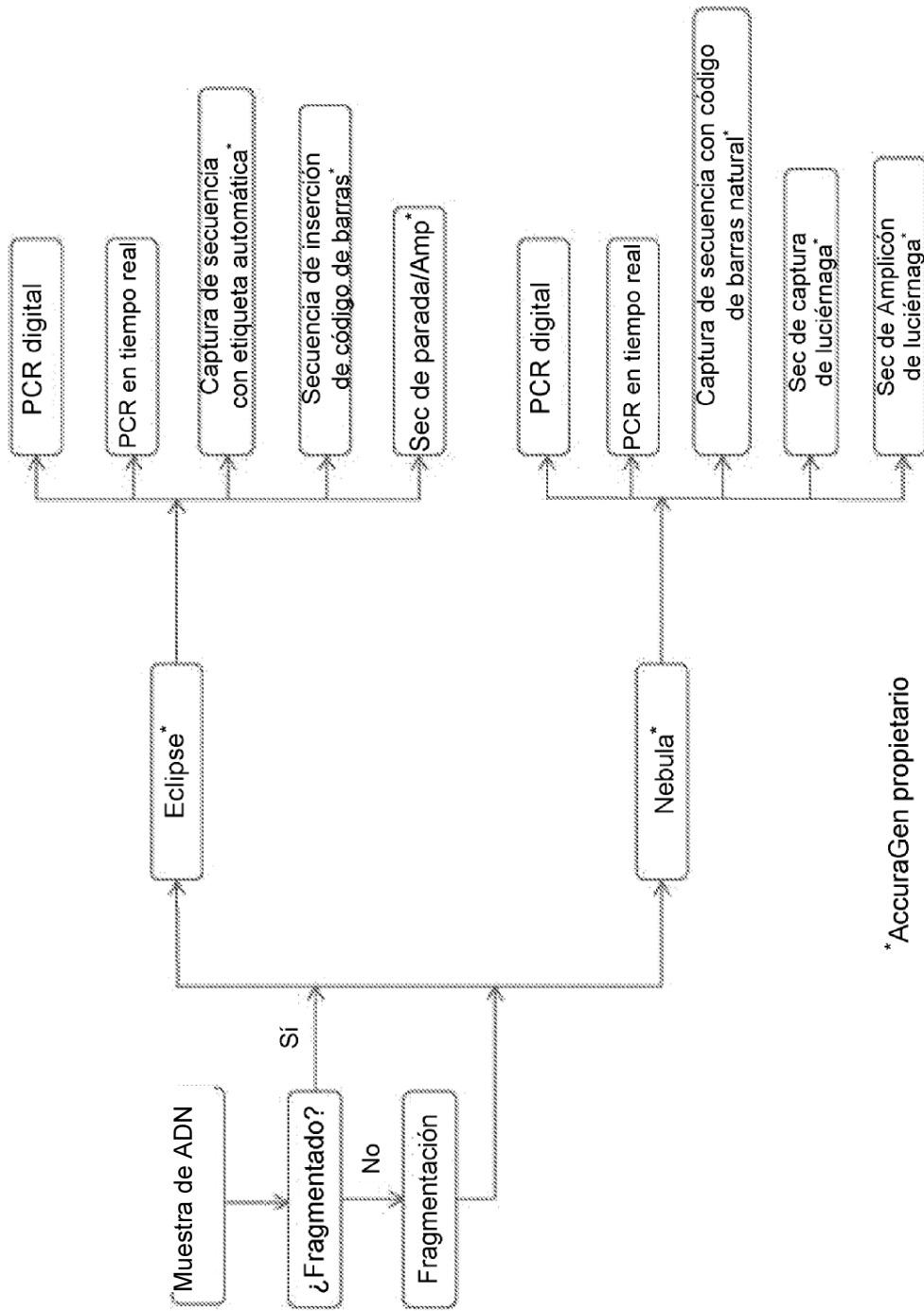


FIG. 31

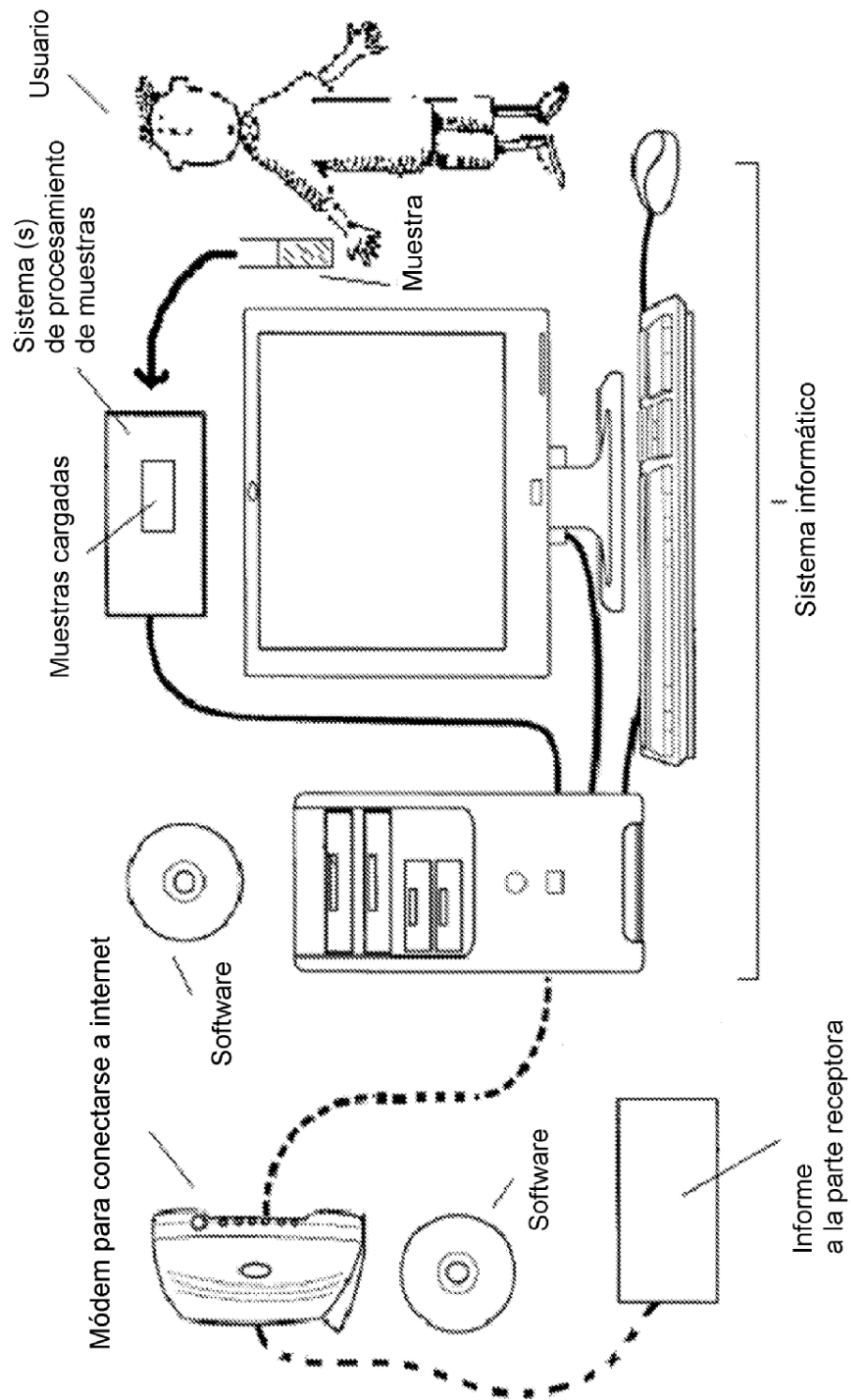


FIG. 32

Cobertura de la región diana

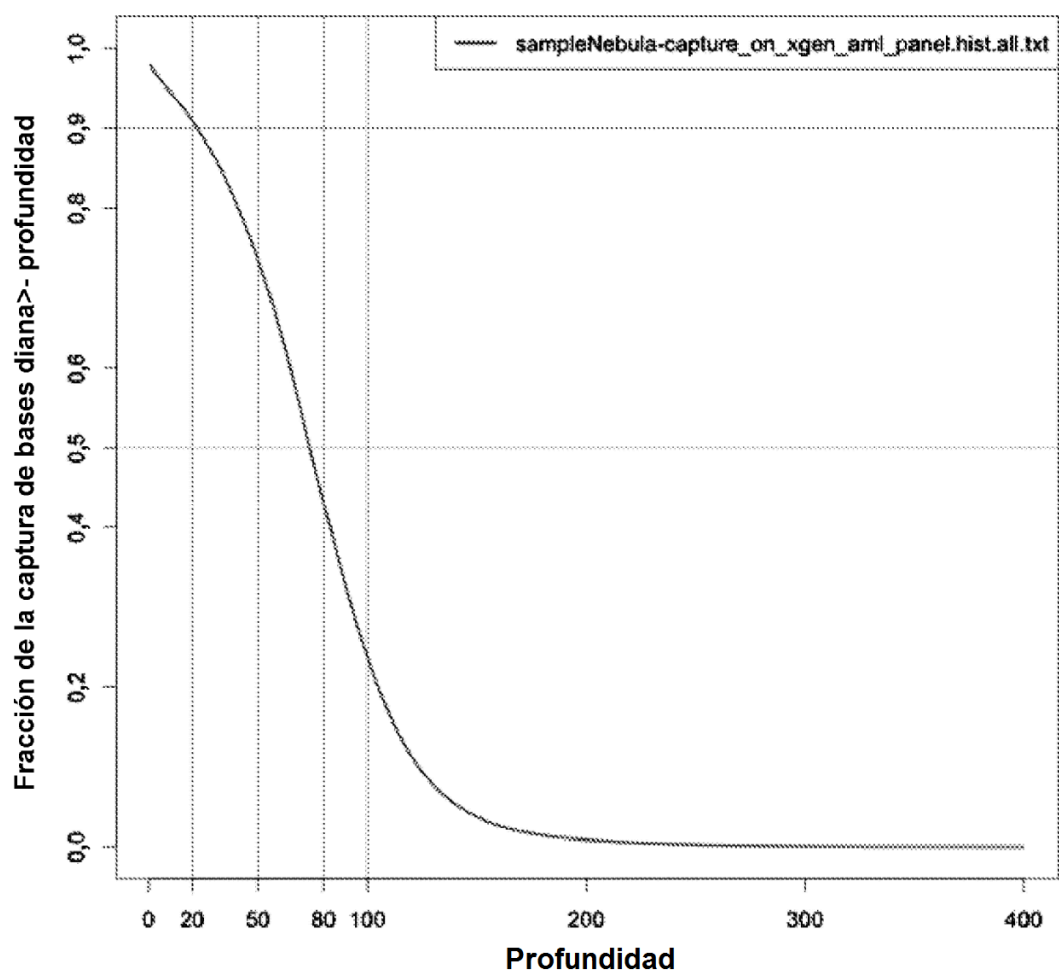


FIG. 33