

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 713 153**

51 Int. Cl.:

C12Q 1/68 (2008.01)

C12N 9/10 (2006.01)

C12N 15/10 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **26.06.2015 PCT/US2015/038050**

87 Fecha y número de publicación internacional: **07.01.2016 WO16003814**

96 Fecha de presentación y número de la solicitud europea: **26.06.2015 E 15734543 (0)**

97 Fecha y número de publicación de la concesión europea: **26.12.2018 EP 3161152**

54 Título: **Métodos y composiciones que utilizan la transposición unilateral**

30 Prioridad:

30.06.2014 US 201462019209 P

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

20.05.2019

73 Titular/es:

**ILLUMINA, INC. (100.0%)
5200 Illumina Way
San Diego, CA 92122, US**

72 Inventor/es:

**STEEMERS, FRANK J.;
FISHER, JEFFREY S.;
GUNDERSON, KEVIN L.;
AMINI, SASAN y
GLOECKNER, CHRISTIAN**

74 Agente/Representante:

ELZABURU, S.L.P

ES 2 713 153 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 y composiciones que utilizan la transposición unilateral

Campo de la invención

5 Las realizaciones proporcionadas en este documento se refieren a métodos y composiciones para una secuenciación de próxima generación. Algunas realizaciones incluyen la preparación de una biblioteca de plantillas a partir de un ácido nucleico diana mediante transposición unilateral, también conocida como transposición de un solo lado, secuenciación de la biblioteca de plantillas y captura de información de contigüidad.

Antecedentes de la invención

10 Hay disponibles varias tecnologías de secuenciación de próxima generación para la determinación rápida y económica de la secuencia completa de un genoma. Por lo general, se prepara una biblioteca de ácidos nucleicos plantilla a partir de una muestra de ADN genómica diana antes de la secuenciación. La preparación de la muestra generalmente incluye una etapa de fragmentación del ADN que rompe las cadenas de ADN más grandes en fragmentos de ADN más pequeños que son más susceptibles de tratar con las tecnologías de secuenciación de próxima generación. A menudo, los adaptadores se unen a los extremos de los fragmentos de ADN, que se pueden lograr mediante la reparación del extremo de ADN, seguido de la unión del adaptador o, más recientemente, mediante un sistema de transposomas. El uso de los transposomas, que es un complejo de una transposasa y de ácidos nucleicos del transposón, permite la fragmentación genómica simultánea y la unión del adaptador de fragmentos, de tal modo que se simplifica la preparación de la biblioteca. Sin embargo, la fragmentación del ADN genómico puede conducir a una pérdida en la información con respecto a las moléculas individuales de ácidos nucleicos con respecto a la contigüidad, eliminación gradual y el haplotipo. Por lo tanto, existe una necesidad para los métodos alternativos de preparación de la biblioteca.

25 El documento US 2010/120098A1 describe métodos y composiciones de generación de bibliotecas de secuenciación mediante Tn5. El método consiste en la generación de fragmentos de ADN marcados con 5' y 3', implicando el método la fragmentación y el marcado de 5' del ADN diana de doble cadena in vitro, luego usando una DNA polimerasa para generar fragmentos de ADN de una sola cadena de marcados con 5' y 3'. El método implica mellar ambas hebras del ADN diana.

Leschziner et al., P.N.A.S. USA Vol. 95 7345-7350 (1998) describe la expresión y purificación de Tn552 transposasa.

Resumen de la invención

30 En algunas realizaciones descritas en este documento se presentan métodos para realizar transposiciones unilaterales. Los inventores de la presente solicitud han encontrado sorprendentemente que al realizar una transposición unilateral, el ADN diana bicatenario se mella en una sola hebra y el ADN diana, después de dicha transposición, permanece intacto incluso después de que los transposomas se eliminan. Por lo tanto, la contigüidad del ADN diana se mantiene siempre después del evento de transposición. En algunas realizaciones, se pueden utilizar métodos para la transposición unilateral para capturar información de contigüidad. En algunas realizaciones, se pueden utilizar métodos para la transposición unilateral para preparar una biblioteca de secuenciación. En algunas realizaciones, se pueden utilizar métodos para la transposición unilateral para determinar la información de eliminación gradual o la información del haplotipo.

40 En algunas realizaciones, se configura un dímero de transposoma para mellar solamente una cadena del ADN diana bicatenario y se transfiere solamente una cadena transferida del transposón de un monómero de transposoma al ADN diana mellado. En algunas realizaciones, una unidad monomérica de un dímero de transposoma es incapaz de transposición, dando como resultado una transposición unilateral. En algunas realizaciones, una transposasa de un dímero de transposoma puede formar el complejo de transposoma al unirse al transposón, pero incapaz de mellar el ADN diana.

45 En algunas realizaciones, el transposón es funcional tal que los transposomas se forman poniéndose en contacto con los transposones a las transposasas y la secuencia del transposón se puede transferir al ácido nucleico diana. En algunas realizaciones, el transposón no es funcional tal que los transposomas se forman poniéndose en contacto con los transposones con las transposasas, pero la secuencia del transposón no puede transferirse al ácido nucleico diana. En algunas realizaciones, el extremo 3' de la hebra transferida comprende un ácido nucleico del extremo 3' que es incapaz de un ataque nucleofílico en el extremo 5' del ácido nucleico diana. En algunas realizaciones, el extremo 3' de la secuencia de reconocimiento se bloquea. En algunas realizaciones, el extremo 3' de la secuencia bloqueada del reconocimiento comprende un nucleótido dideoxi del extremo 3', un grupo amina, un grupo alquilo, un grupo arilo, un grupo tiol, un grupo sulfato, un nucleótido inverso, un grupo azido o una biotina.

55 En algunas realizaciones, la transposasa es capaz de formar transposomas pero incapaz de mellar el ADN diana. En algunas realizaciones, la transposasa comprende una o más modificaciones de aminoácidos, de modo que es capaz de formar un transposoma pero incapaz de mellar un ADN diana.

5 En algunas realizaciones, el complejo de transposoma se configura de tal manera que el transposoma es incapaz de formar un dímero eficientemente. En algunas realizaciones, el complejo de transposoma se configura de tal manera que el transposoma es incapaz totalmente de formar un dímero. En la invención, el monómero transposoma forma una mella en una cadena del ADN bicatenario solamente y transfiere la cadena transferida del transposón al ADN mellado diana.

10 En algunas realizaciones, la transposición unilateral se realiza explotando la resistencia diferencial a la transposición por las dos hebras de ADN diana. En algunas realizaciones, una hebra de ADN diana comprende bases modificadas o enlaces fosfodiéster modificados que son resistentes a la transposición. Exponer un ADN diana con resistencia diferencial a la transposición por las dos hebras de ADN diana a los transposomas da como resultado una transposición unilateral. En algunas realizaciones, el ácido nucleico diana es un ADNc de doble cadena en el cual una hebra del ADNc comprende bases modificadas y/o enlaces de fosfodiéster modificados de tal manera que esa hebra es resistente a la transposición. En algunas realizaciones, el ácido nucleico diana es un ADN genómico bicatenario en el que se modifica una hebra de manera tal que esa hebra es, parcialmente o totalmente, resistente a la transposición.

15 En la FIG. 14 se muestra un esquema ejemplar de transposición unilateral. En algunas realizaciones, comenzando con una plantilla de ácido nucleico monocatenario (línea continua), se sintetiza una hebra complementaria (línea de puntos) que tiene una resistencia diferencial a la transposición respecto a la cadena de la plantilla original. Luego, usando incluso complejos de transposón normales (por ejemplo, activos y desbloqueados), se produce una transposición unilateral. En un ejemplo, la cadena recién sintetizada tiene una mayor resistencia frente a la transposición. En otro ejemplo, la plantilla original es altamente resistente, permitiendo la cadena sintetizada la transposición en sí mismo. En esta realización, la cadena menos resistente forma los elementos de la biblioteca, que se mantienen en contigüidad por la cadena más resistente.

20 El solicitante descubrió sorprendentemente que después de llevar a cabo la transposición unilateral, el ácido nucleico diana de doble cadena permanece intacto sin perder la información de contigüidad incluso después de retirar la transposasa de la transposoma. En algunas realizaciones, las transposasas se eliminan del ácido nucleico diana transpuesto después de la transposición por tratamiento con SDS, urea, proteasa o calor. En consecuencia, una transposición unilateral puede ser ventajosa para determinar la información de la secuencia, la información de contigüidad, la información de fases y la información de haplotipo. La información de contigüidad puede proporcionar una resolución extensa del poder del haplotipo. La formación de haplotipos permite la eliminación gradual de alelos raros y de variantes estructurales tales como los re-ensamblamientos génicos y la duplicación de genes.

25 En algunas realizaciones, la transposición unilateral puede acoplarse con una codificación de barras combinatoria en la que los primeros conjuntos de códigos de barra se adjuntan a través de la transposición unilateral y el segundo conjunto de códigos de barras se adjunta mediante la amplificación posterior.

30 En algunas realizaciones, los primeros conjuntos de códigos de barras se introducen en el ácido nucleico diana durante la transposición para generar el ácido nucleico diana transpuesto que comprende el primer conjunto de códigos de barras. Los ácidos nucleicos diana transpuestos se agrupan para generar un primer grupo de ácidos nucleicos dianas transpuestos. Un segundo conjunto de códigos de barras se introduce en el primer grupo de ácidos nucleicos diana transpuestos para generar un ácido nucleico diana que comprende primeros y segundos conjuntos de códigos de barras. El segundo conjunto de códigos de barras se puede introducir ya sea mediante amplificación posterior, unión o transposición adicional. En algunas realizaciones, el primer y segundo conjunto de códigos de barras es diferente. El ácido nucleico diana comprende los primeros y segundos juegos de códigos de barras; se agrupan para generar un segundo grupo de ácido nucleico diana transpuesto. Opcionalmente, se pueden repetir las etapas de introducir códigos de barras adicionales y agruparlos para generar una biblioteca de ácidos nucleicos diana con código de barra.

35 En algunas realizaciones, la transposición unilateral se puede utilizar para determinar la información de la secuencia o la información de contigüidad del ácido nucleico a partir de células individuales. El ácido nucleico puede ser un ácido nucleico genómico o ADNc generado a partir del ARNm de la célula individual. En algunas realizaciones, se puede introducir un primer conjunto de códigos de barras en el ácido nucleico a partir de células individuales que sirven como identificadores de la célula individual. En algunas realizaciones, después de introducir el primer conjunto de códigos de barras al ácido nucleico a partir de células individuales, el ácido nucleico con código de barra puede agruparse y procesarse posteriormente mediante amplificación, unión o transposición adicional con o sin introducir otros códigos de barras.

40 Un primer aspecto de la invención proporciona un método para preparar una biblioteca de secuenciación a partir de un ácido nucleico diana de doble cadena que comprende: (a) proporcionar una pluralidad de los transposomas, comprendiendo cada transposoma una transposasa y un ácido nucleico transposón, en el cual el transposoma se configura para mellar y transferir el transposón a solamente una hebra del ácido nucleico de la diana; y (b) poner en contacto el ácido nucleico diana con los transposomas de tal manera que el ácido nucleico diana es mellado en una pluralidad de sitios del ácido nucleico diana y los ácidos nucleicos del transposón se unen al ácido nucleico diana mellado, obteniendo así una biblioteca de ácidos nucleicos modificados para la secuenciación.

- Algunas realizaciones incluyen un método para preparar una biblioteca de secuenciación a partir de un ácido nucleico diana de doble cadena que comprende: (a) proporcionar una pluralidad de los transposomas, comprendiendo cada transposoma una transposasa y un ácido nucleico transposón, en el cual el transposoma se configura para mellar y transferir el transposón a solamente una hebra del ácido nucleico de la diana; (b) poner en contacto el ácido nucleico diana con los transposomas de tal manera que el ácido nucleico diana se mella en una pluralidad de sitios del ácido nucleico diana y los ácidos nucleicos del transposón se unen al ácido nucleico diana mellado; y (c) hibridar cebadores a los ácidos nucleicos del transposón y extender los cebadores hibridados, obteniendo así una biblioteca de ácidos nucleicos modificados para la secuenciación. En la FIG. 13 se muestran esquemas ejemplares de la preparación de la biblioteca utilizando una transposición unilateral.
- Un segundo aspecto de la invención proporciona un método para capturar información de contigüidad de un ADN diana. El método incluye (a) proporcionar una pluralidad de los transposomas, comprendiendo cada monómero de transposoma una transposasa y un ácido nucleico de transposón en el que el transposoma es configurado para mellar sólo una hebra del ácido nucleico diana bicatenario; (b) poner en contacto el ADN diana con los transposomas de tal manera que el ADN diana es mellado en una pluralidad de sitios del ácido nucleico diana; (c) añadir o insertar una o más secuencias de reconocimiento a la secuencia de ADN diana para generar un ADN diana tratado; (d) secuenciar el ADN diana tratado; y (e) capturar la información de contigüidad identificando las secuencias de ADN diana o las secuencias de reconocimiento que tienen una propiedad compartida.
- Un tercer aspecto proporciona un método para capturar información de contigüidad de un ADN diana. El método incluye (a) proporcionar una pluralidad de transposomas, comprendiendo cada monómero de transposoma una transposasa y un ácido nucleico de transposón que comprende una secuencia de reconocimiento, donde el transposoma se configura para mellar solamente una cadena del ácido nucleico diana bicatenario; (b) insertar los ácidos nucleicos del transposón en hebras del ácido nucleico diana, que comprenden: (i) poner en contacto el ácido nucleico diana con los transposomas de tal manera que el ácido nucleico diana se mella en una pluralidad de sitios y los ácidos nucleicos del transposón individual se unen a las cadenas melladas en un lado de los sitios mellados, y (ii) unir los ácidos nucleicos del transposón individuales unidos a las cadenas melladas en el otro lado de los sitios mellados, obteniéndose así un ácido nucleico modificado; (c) amplificar el ácido nucleico modificado, obteniéndose así una pluralidad de ácidos nucleicos que comprenden secuencias de reconocimiento insertadas; (d) secuenciar el ADN diana tratado; y (e) capturar la información de contigüidad identificando las secuencias de ADN diana o las secuencias de reconocimiento que tienen una propiedad compartida.
- Algunas realizaciones también incluyen la captura de los ácidos nucleicos modificados en una superficie.
- En algunas realizaciones, los transposomas que se ponen en contacto con los ácidos nucleicos diana (b) se unen a una superficie, capturando así los ácidos nucleicos modificados en la superficie.
- Algunas realizaciones también incluyen la secuenciación de los ácidos nucleicos capturados en la superficie.
- En algunas realizaciones, la proximidad de la información de la secuencia obtenida a partir de dos ácidos nucleicos capturados en una representación lineal de la secuencia de ácidos nucleicos diana es indicativa de la proximidad de los ácidos nucleicos capturados en la superficie.
- En algunas realizaciones, los ácidos nucleicos capturados más próximos entre sí en la superficie comprenden secuencias más próximas en la representación de la secuencia de ácidos nucleicos de la diana comparada a los ácidos nucleicos capturados más próximos.
- En algunas realizaciones, la representación de la secuencia de ácido nucleico diana comprende una representación de haplotipo. En algunas realizaciones, la representación de las secuencias del ácido nucleico diana comprende lecturas cortas ordenadas.
- En algunas realizaciones, la transposasa comprende una actividad de transposasa unilateral.
- En algunas realizaciones, la transposasa comprende una subunidad monomérica que carece de actividad de transposasa. En algunas realizaciones, la transposasa comprende subunidades monoméricas unidas covalentemente. En algunas realizaciones, la estructura cuaternaria de la transposasa es monomérica. En algunas realizaciones, la transposasa carece de la capacidad de formar dímeros.
- En algunas realizaciones, la transposasa se selecciona del grupo que consiste en Mu, Mu E392Q, Tn5, Tn5 hiperactivo (Goryshin and Reznikoff, J. Biol. Chem., 273:7367 (1998)), Transposasa EZ-Tn5™ (Epicentre Biotechnologies, Madison, Wisconsin), variantes de la transposasa Tn5, RAG, Tn7, Tn10, Vibhar, y Tn552. Se describen variantes de las transposasas Tn5, como las que tienen sustituciones de aminoácidos, inserciones, deleciones y/o fusiones con otras proteínas o péptidos en las patentes de EE.UU. : 5.925.545; 5.965.443; 7.083.980; 7.608.434; y solicitud de patente de EE.UU. 14/686.961. En algunas realizaciones, la transposasa Tn5 comprende una o más sustituciones en las posiciones 54, 56, 372, 212, 214, 251 y 338 con respecto a la proteína natural, tal como se describe en la solicitud de patente de los Estados Unidos 14/686.961. En algunas realizaciones, la proteína natural Tn5 o su variante puede incluir además un polipéptido de fusión. En algunas realizaciones, el dominio polipeptídico fusionado a la transposasa puede comprender, por ejemplo, el factor de elongación TS.

5 En algunas realizaciones, el ácido nucleico del transposón se bloquea. En algunas realizaciones, el extremo 3' de la cadena transferida del transposón se bloquea. En algunas realizaciones, el extremo 3' del ácido nucleico del transposón bloqueado se selecciona del grupo que consiste en un grupo dideoxi, un grupo espaciador, un grupo amina, un grupo azido, un grupo fosfato, un grupo alquilo, un nucleótido inverso y un grupo biotina. En algunas realizaciones, la secuencia del transposón se puede alterar por la sustitución, la adición o la eliminación de bases de la secuencia del transposón.

10 En algunas realizaciones, la pluralidad de los transposomas se prepara poniendo en contacto las transposasas con los ácidos nucleicos funcionales del transposón y ácidos nucleicos del transposón no funcionales. En algunas realizaciones, el transposón no funcional comprende el transposón bloqueado. In some embodiments, the ratio of transposon nucleic acids comprising non-functional transposon nucleic acids to functional transposon nucleic acids is greater than or equal to 1:1. In some embodiments, the ratio of transposon nucleic acids comprising non-functional transposon nucleic acids to functional transposon nucleic acids can be. En algunas realizaciones, la proporción de ácidos nucleicos de transposón que comprenden ácidos nucleicos de transposón no funcionales frente a ácidos nucleicos de transposón funcionales es mayor o igual a 1:1. En algunas realizaciones, la proporción de ácidos nucleicos de transposón que comprenden los ácidos nucleicos de transposón no funcionales frente a los ácidos nucleicos de transposón funcionales pueden ser 1:2, 1:3, 1:5, 1: 10, 1:20, 1:30, 1:40, 1:50, 1:75, 1:100, 2:1, 3:1, 4:1, 5:1, 6:1, 7:1, 8:1, 9:1, 10:1, 20:1, 30:1, 40:1, 50:1, 60:1, 70:1, 80:1, 90:1, o 100:1.

20 Algunas realizaciones también incluyen la amplificación de ácidos nucleicos extendidos. En algunas realizaciones, la amplificación de los ácidos nucleicos extendidos se realiza con cebadores de amplificación con cola que comprenden una secuencia seleccionada del grupo que consiste en un sitio de anclaje, un sitio de cebado de secuenciación, un sitio de cebado de amplificación y un marcador indicador.

Algunas realizaciones también incluyen la amplificación de los ácidos nucleicos capturados. En algunas realizaciones, la amplificación de los ácidos nucleicos capturados comprende una amplificado del puente.

25 En algunas realizaciones, la superficie comprende una pluralidad de sondas de captura. En algunas realizaciones, las sondas de captura comprenden ácidos nucleicos. Algunas realizaciones también incluyen la hibridación de los ácidos nucleicos modificados con las sondas de captura.

30 En algunas realizaciones, los ácidos nucleicos modificados y las sondas de captura comprenden, cada una, una fracción de afinidad. En algunas realizaciones, las fracciones de afinidad pueden ser miembros de un par de enlace. En algunos casos, los ácidos nucleicos modificados pueden comprender un primer miembro de un par de enlace y la sonda de captura puede abarcar un segundo miembro del par de enlace. En algunos casos, las sondas de captura pueden ser inmovilizadas en una superficie sólida y el ácido nucleico modificado puede comprender un primer miembro de un par de enlace y la sonda de captura puede comprender un segundo miembro del par de enlace. En tales casos, la unión de los primeros y segundos miembros del par de enlace inmoviliza el ácido nucleico modificado a la superficie sólida. Los ejemplos del par de enlace incluyen, entre otros, a la biotina-avidina, biotina-estreptavidina, biotina-neutravidina, ligando-receptor, hormona-receptor, lecitina-glicoproteína y antígeno-anticuerpo.

35 Algunas realizaciones también incluyen la unión de la fracción de afinidad de los ácidos nucleicos modificados con la fracción de afinidad de las sondas de captura.

40 En algunas realizaciones, el ácido nucleico del transposón consiste en una secuencia seleccionada del grupo que consiste en un sitio del ancla, un código de barras, un sitio del cebador de secuencia, un sitio del cebador de amplificación, un índice molecular único, y un marcador indicador.

En algunas realizaciones, al menos un transposoma comprende dos ácidos nucleicos transposón.

En algunas realizaciones, los dos ácidos nucleicos del transposón tienen diversas secuencias.

En algunas realizaciones, la pluralidad de los transposomas comprende al menos dos ácidos nucleicos de transposón diferentes.

45 En algunas realizaciones, el ácido nucleico diana se selecciona del grupo consistente en ADN y ARN. . En algunas realizaciones, el ácido nucleico diana se selecciona del grupo consistente en ADN genómico y ADN. En algunas realizaciones, el ácido nucleico diana es el ADN genómico.

En algunas realizaciones, la superficie está en un sustrato seleccionado del grupo que consiste en perlas, láminas, células de flujo, canales, barritas y pocillos.

50 En algunas realizaciones, la superficie comprende al menos alrededor de 10.000 ácidos nucleicos capturados por mm². En algunas realizaciones, la superficie comprende al menos alrededor de 100.000 ácidos nucleicos capturados por mm². En algunas realizaciones, la superficie comprende por lo menos alrededor de 1.000.000, 1.500.000, 2.000.000, 3.000.000, 5.000.000, 10.000.000, 15.000.000, 20.000.000, 30.000.000, 40.000.000, 50.000.000, 60.000.000, 70.000.000, 80.000.000, 90.000.000, 100.000.000, 150.000.000, 200.000.000, 300.000.000, 350.000.000, 400.000.000, 450.000.000, 500.000.000, 550.000.000, 600.000.000, 650.000.000, 700.000.000,

55

750.000.000, 800.000.000, 850.000.000, 900.000.000, 950.000.000, 1.000.000.000, 1.200.000.000, 1.300.000.000, 1.400.000.000, 1.500.000.000, 1.600.000.000, 1.700.000.000, 1.800.000.000, 1.900.000.000, 2000.000.000, 3.000.000.000, 4.000.000.000, 5.000.000.000, 6.000.000.000, 7.000.000.000, 8.000.000.000, 9.000.000.000, 10.000.000.000 o más de ácidos nucleicos capturados por mm².

5 También se describe una biblioteca de secuenciación preparada por cualquiera de los métodos anteriores.

Algunas de las realizaciones de los métodos proporcionados en este documento incluyen un método de preparación de una biblioteca de secuenciación con códigos de barras de un ácido nucleico diana de doble cadena que comprende: (a) proporcionar una pluralidad de los transposomas, comprendiendo cada transposoma una transposasa y un ácido nucleico de transposón que comprende un código de barras; y (b) insertar los ácidos nucleicos del transposón en hebras del ácido nucleico diana, que comprende: (i) poner en contacto el ácido nucleico diana con los transposomas de tal manera que el ácido nucleico diana se mella en una pluralidad de sitios y los ácidos nucleicos de transposón individuales se unen a las cadenas melladas en un lado de los sitios mellados, y (ii) unir los ácidos nucleicos de transposón individuales unidos a las cadenas melladas en el otro lado de los sitios mellados, obteniéndose así un ácido nucleico modificado.

10

15 Algunas realizaciones también incluyen (c) capturar el ácido nucleico diana modificado en una superficie.

Algunas realizaciones incluyen un método de preparación de una biblioteca de secuenciación con códigos de barras a partir de un ácido nucleico diana de doble cadena que comprende: (a) proporcionar una pluralidad de los transposomas, comprendiendo cada transposoma una transposasa y un ácido nucleico de transposón que comprende un código de barras; y (b) insertar los ácidos nucleicos del transposón en hebras del ácido nucleico diana, que comprende: (i) poner en contacto el ácido nucleico diana con los transposomas de tal manera que el ácido nucleico diana se mella en una pluralidad de sitios y los ácidos nucleicos del transposón individual se unen a las cadenas melladas en un lado de los sitios mellados, y (ii) unir los ácidos nucleicos del transposón individuales unidos a las cadenas melladas en el otro lado de los sitios mellados, obteniéndose así un ácido nucleico modificado; (c) amplificar el ácido nucleico modificado, obteniéndose así una pluralidad de ácidos nucleicos que comprenden códigos de barras insertados.

20

25

Algunas realizaciones también incluyen la captura del ácido nucleico diana modificado en una superficie.

En algunas realizaciones, los transposomas que se ponen en contacto con los ácidos nucleicos diana en (b) se unen a una superficie, capturándose así los ácidos nucleicos modificados en la superficie.

Algunas realizaciones también incluyen la secuenciación de los ácidos nucleicos capturados.

30 En algunas realizaciones, la proximidad de la información de la secuencia obtenida a partir de dos ácidos nucleicos capturados en una representación lineal de la secuencia de ácidos nucleicos diana es indicativa de la proximidad de los ácidos nucleicos capturados en la superficie.

En algunas realizaciones, los ácidos nucleicos capturados más próximos entre sí en la superficie comprenden secuencias más próximas en la representación de la secuencia de ácidos nucleicos diana comparado con los ácidos nucleicos capturados más próximos.

35

En algunas realizaciones, la representación de la secuencia de ácido nucleico diana comprende una representación de haplotipo.

En algunas realizaciones, el código de barras de, al menos, un ácido nucleico de transposón es diferente.

En algunas realizaciones, los códigos de barras de los ácidos nucleicos de transposón no son los mismos.

40 Algunas realizaciones también incluyen alinear las secuencias de ácidos nucleicos de acuerdo con la presencia de códigos de barras comunes en las secuencias para generar una representación del ácido nucleico diana.

En algunas realizaciones, la transposasa comprende una actividad de transposasa unilateral.

En algunas realizaciones, la transposasa comprende una subunidad monomérica que carece de actividad de transposasa.

45 En algunas realizaciones, la transposasa comprende subunidades monoméricas unidas covalentemente. En algunas realizaciones, la estructura cuaternaria de la transposasa es monomérica. En algunas realizaciones, la transposasa carece de la capacidad de formar dímeros.

En algunas realizaciones, la transposasa se selecciona del grupo que consiste en Mu, Mu E392Q, Tn5, Tn5 hiperactivo, EZ-Tn5™, variantes de Tn5, RAG, Tn7, Tn10, Tn552, y transposasa Vibhar.

50 En algunas realizaciones, el ácido nucleico del transposón se bloquea.

En algunas realizaciones, el extremo 3' del ácido nucleico de transposón bloqueado se selecciona del grupo que consiste en un grupo dideoxi, un grupo espaciador, un grupo amina, un grupo azido, un grupo alquilo, un grupo arilo, un nucleótido inverso, un grupo tiofosfato y un grupo biotina.

5 En algunas realizaciones, la pluralidad de los transposomas se prepara poniendo en contacto las transposasas con ácidos nucleicos de transposón no funcionales y ácidos nucleicos de transposón funcionales. En algunas realizaciones, los transposones no funcionales comprenden el extremo 3' bloqueado. En algunas realizaciones, la pluralidad de los transposomas se prepara poniendo en contacto las transposasas con ácidos nucleicos de transposón bloqueados y ácidos nucleicos de transposón no bloqueados. En algunas realizaciones, la proporción de ácidos nucleicos de transposón que comprenden ácidos nucleicos de transposón bloqueados frente a ácidos nucleicos de transposón no bloqueados es mayor o igual a 1:1. En algunas realizaciones, la proporción de ácidos nucleicos de transposón que comprenden los ácidos nucleicos de transposón frente a los ácidos nucleicos de transposón no-bloqueados puede ser 1:2, 1:3, 1:5, 1: 10, 1:20, 1:30, 1:40, 1:50, 1:75, 1:100, 2:1, 3:1, 4:1, 5:1, 6:1, 7:1, 8:1, 9:1, 10:1, 20:1, 30:1, 40:1, 50:1, 60:1, 70:1, 80:1, 90:1, o 100:1.

15 Algunas realizaciones también incluyen la fijación de adaptadores de amplificación al ácido nucleico diana. En algunas realizaciones, los adaptadores de amplificación comprenden una secuencia seleccionada del grupo que consiste en un sitio de anclaje, un sitio de cebado de secuenciación, un sitio de cebado de amplificación y un marcador indicador.

Algunas realizaciones también incluyen la amplificación de los ácidos nucleicos capturados. En algunas realizaciones, la amplificación de los ácidos nucleicos capturados comprende la amplificación del puente.

20 En algunas realizaciones, la superficie comprende una pluralidad de sondas de captura. En algunas realizaciones, las sondas de captura comprenden ácidos nucleicos. En algunas realizaciones, las sondas de captura comprenden una fracción de afinidad. En algunas realizaciones, la fracción de afinidad se selecciona del grupo que consiste en biotina, avidina, estreptavidina y una recombinasa.

25 En algunas realizaciones, el ácido nucleico del transposón comprende una secuencia seleccionada del grupo que consiste en un sitio ancla, un sitio cebador de secuencia, un sitio cebador de amplificación, un índice molecular único y un marcador indicador.

En algunas realizaciones, al menos un transposoma comprende dos ácidos nucleicos de transposón. En algunas realizaciones, los dos ácidos nucleicos de transposón tienen diversas secuencias.

30 En algunas realizaciones, la pluralidad de los transposomas comprende al menos dos ácidos nucleicos de transposón diferentes.

En algunas realizaciones, el ácido nucleico diana se selecciona del grupo consistente en fragmentos de ADN de ADN genómico y ADNc. . En algunas realizaciones, el ácido nucleico diana se selecciona del grupo consistente en ADN genómico y ADNc En algunas realizaciones, el ácido nucleico diana es el ADN genómico.

35 En algunas realizaciones, la superficie está en un substrato seleccionado del grupo que consiste en una perla, una lámina, una célula de flujo, un canal, una barrilita y un pocillo.

40 En algunas realizaciones, la superficie comprende al menos alrededor de 10.000 ácidos nucleicos capturados por mm². En algunas realizaciones, la superficie comprende al menos alrededor de 100.000 ácidos nucleicos capturados por mm². En algunas realizaciones, la superficie comprende al menos alrededor de 1.000.000 de ácidos nucleicos capturados por mm². En algunas realizaciones, la superficie comprende por lo menos alrededor de 1.000.000, 1.500.000, 2.000.000, 3.000.000, 5.000.000, 10.000.000, 15.000.000, 20.000.000, 30.000.000, 40.000.000, 50.000.000, 60.000.000, 70.000.000, 80.000.000, 90.000.000, 100.000.000, 150.000.000, 200.000.000, 300.000.000, 350.000.000, 400.000.000, 450.000.000, 500.000.000, 550.000.000, 600.000.000, 650.000.000, 700.000.000, 750.000.000, 800.000.000, 850.000.000, 900.000.000, 950.000.000, 1.000.000.000, 1.200.000.000, 1.300.000.000, 1.400.000.000, 1.500.000.000, 1.600.000.000, 1.700.000.000, 1.800.000.000, 1.900.000.000, 2.000.000.000, 3.000.000.000, 4.000.000.000, 5.000.000.000, 6.000.000.000, 7.000.000.000, 8.000.000.000, 9.000.000.000, 10.000.000.000, o más ácidos nucleicos capturados por mm².

También se describe en este documento una biblioteca de secuenciación que comprende códigos de barras preparados por cualquiera de los métodos anteriores.

50 En algunas realizaciones, después del tratamiento con transposasa o después de una amplificación posterior, se pueden agregar o insertar una o más secuencias de reconocimiento en el ácido nucleico diana mellado. La una o más secuencias de reconocimiento pueden incluir, entre otras, un código de barras, una cebado o una secuencia de ADN adaptadora en el sitio de mellado que marque el fragmento de ácido nucleico diana como único con respecto a la relación adyacente, compartimental o distancia espacial.

55 Después de ser marcadas, las moléculas de ácido nucleico de la pistola se pueden secuenciar utilizando la plataforma de secuenciación descrita anteriormente y la información de contigüidad se captura mediante la

identificación de secuencias de reconocimiento que tienen una propiedad compartida. En algunas realizaciones, la propiedad compartida es una secuencia de códigos de barras idéntica o complementaria. Por ejemplo, las secuencias de lectura de origen adyacente pueden identificarse mediante secuencias de códigos de barras compartidas; o pueden definirse lecturas por compartimentos basados en códigos de barras específicos del compartimento compartido derivados del mismo segmento de ADN diana. En otras realizaciones, la propiedad compartida es una ubicación física compartida o restringida, que puede ser indicada por una o más coordenadas x, y en una célula de flujo. Una ubicación física "restringida" puede referirse a una ubicación física cercana, idéntica o casi idéntica o a un conjunto de dos o más ubicaciones físicas cuyas coordenadas físicas relativas están correlacionadas con las coordenadas de secuencia relativas en la secuencia de ácido nucleico diana a partir de la cual se derivaron los fragmentos de ácido nucleico. Por ejemplo, en los métodos relacionados con la contigüidad a largo alcance, la transposición in situ en el ADN genómico extendido de HMW en la superficie de una célula de flujo de secuenciación se realiza mediante secuencias adaptadoras para obtener relaciones espaciales a distancia mediante la identificación de las ubicaciones físicas restringidas (es decir, las coordenadas relativas en las que se inmovilizan las plantillas de secuenciación unidas físicamente) de las secuencias adaptadoras, fragmentos de ADN híbridos, o una combinación de los mismos. Los métodos se pueden utilizar para capturar información de contigüidad de corto alcance, de medio alcance y de largo alcance.

En algunas realizaciones, una transposición unilateral se puede combinar con la codificación de barras combinatoria. Un uso de los elementos transpuestos de un solo lado es uno que permitiría la codificación de barras combinatoria, sin la necesidad de ningún mecanismo adicional para mantener juntos los elementos de la biblioteca relacionados durante el proceso. En la FIG. 15 se muestra un esquema ejemplar de la combinación de la transposición unilateral con la codificación de barras combinatoria.

Breve descripción de los dibujos

La figura 1 muestra una realización ilustrativa en la que se pone en contacto un ácido nucleico diana con una población de transposomas que comprende un ácido nucleico transposón.

La figura 2 muestra una realización ilustrativa en la que una población de transposomas comprende diferentes ácidos nucleicos de transposón que se ponen en contacto con un ácido nucleico diana y los diferentes ácidos nucleicos de transposón se unen a una hebra del ácido nucleico diana en diferentes sitios de mella.

La figura 3 representa una realización ilustrativa en la que un ácido nucleico diana modificado se amplifica mediante amplificación lineal para obtener ciertos productos de amplificación.

La figura 4 representa una realización ilustrativa en la que un transposoma comprende una transposasa dimérica, y un ácido nucleico de transposón comprende dos elementos de transposón que comprenden elementos de mosaico (ME) en los que uno de los ME se bloquea con un grupo dideoxi en un extremo de 3'.

La figura 5 representa una realización ilustrativa en la que una población de los transposomas que comprende diferentes códigos de barras entra en contacto con un ácido nucleico diana; los ácidos nucleicos de transposón se adhieren a un lado de los sitios de mella y el otro extremo no unido de los ácidos nucleicos de transposón se une al otro lado del sitio mella por ligación; y el ácido nucleico diana modificado es amplificado por la amplificación del genoma entero (WGA).

La figura 6 muestra los resultados de tratar ADN genómico sin transposomas (Amplicon), o los transposomas que comprenden transposasa y (1) ácido nucleico de transposón bloqueado con un grupo de biotina de 3' (Bio 3'); (2) ácido nucleico de transposón bloqueado con un grupo espaciador 3' (espaciador 3'); o (3) ácido nucleico de transposón no bloqueado (TDE1).

La figura 7 muestra las gráficas del número de coberturas nominales, y la media de la longitud de la lectura sintética para lecturas de 500 bp con una frecuencia de inserción de 100 bp, y para lecturas de 300 bp con una frecuencia de inserción de 50 bp.

La figura 8A y 8B muestra una transposición unilateral con transposones de adaptadores en forma de y.

La figura 9 muestra una fotografía de un gel de agarosa cargado con muestras de reacciones de transposición unilaterales.

La figura 10 muestra fotografías de geles de agarosa cargados con muestras a partir de reacciones de transposición que se realizan con las variantes n+1 y n-1 de un transposón.

La figura 11 muestra un diagrama de una reacción de transposición realizada con una mezcla de transposomas activos e inactivos.

La figura 12 muestra un esquema ejemplar de la mella del ácido nucleico diana y la unión del adaptador de

oligonucleótidos.

La figura 13 muestra esquemas ejemplares de preparación de bibliotecas utilizando una transposición unilateral.

5 La figura 14 muestra un esquema ejemplar de transposición unilateral que explota la resistencia diferencial a la transposición por dos hebras de un ADN.

10 La figura 15 muestra un esquema ejemplar de transposición unilateral junto con la codificación de barras combinatorias. Utilizando la transposición unilateral, los productos unilaterales pueden mantener la contigüidad sin necesidad de un mecanismo externo. Las moléculas únicas, pero indistinguibles, (A, B y C) se encuentran juntas. Se dividen aleatoriamente en reacciones separadas, en las que se añaden códigos de barras modulares. Aunque el número de reacciones separadas en cada paso es menor que el número de moléculas, el camino a través de las reacciones tiende a ser único para cada molécula, lo que resulta en una combinación única de código de barras para cada una.

Descripción detallada

15 Las realizaciones proporcionadas en este documento se refieren a los métodos para secuenciación de próxima generación. Algunas realizaciones incluyen la preparación de una biblioteca de plantillas a partir de un ácido nucleico diana mediante la transposición unilateral y la secuenciación de la biblioteca de plantillas. En algunas realizaciones, la transposición unilateral incluye una transposasa que mella una cadena de un ácido nucleico de doble cadena, y que fija un ácido nucleico de transposón a la cadena mellada en un extremo del sitio de mella. Ventajosamente, la transposición unilateral no fragmenta el ácido nucleico diana de doble cadena en comparación con la transposición de doble cara (p. ej., Nextera™). Por lo tanto, la contigüidad, haplotipo, y/o la información de fase se puede retener para ciertos ácidos nucleicos diana, como el ADN genómico.

20 Algunas realizaciones de los métodos proporcionados en este documento incluyen transposomas con una actividad de transposasa unilateral, el uso de los transposomas para preparar una biblioteca de secuenciación y secuenciar dichas bibliotecas. En algunas realizaciones, un transposoma puede incluir una transposasa con una actividad de transposasa unilateral. En algunas realizaciones, un transposoma puede incluir un ácido nucleico de transposón que puede tener un grupo de bloqueo que inhibe la inserción del transposón en ambas hebras de un ácido nucleico diana de doble cadena. Las transposasas también incluyen integrasas de retrotransposones y transposasas de retrovirus. Las transposasas ejemplares incluyen, entre otras a Mu, Tn10, Tn5 y Tn5 hiperactivo (Goryshin and Reznikoff, J. Biol. Chem., 273:7367 (1998)). Las realizaciones de transposasas útiles con algunos de los métodos proporcionados en este documento incluyen los descritos en la publicación de solicitud de patente de EE.UU. No. 2010/0120098. Más realizaciones de transposasas y elementos del transposón incluyen una transposasa hiperactiva Tn5 y un elemento de transposasa tipo Tn5 (Goryshin and Reznikoff, J. Biol. Chem., 273:7367 (1998)), una transposasa y un elemento de transposasa MU que comprende las secuencias finales R1 y R2 (Mizuuchi, Cell, 35: 785, (1983) y Savilahti, et al., EMBO J., 14: 4893, 15 (1995)). Ejemplos de elementos de transposasa que forman un complejo con una transposasa hiperactiva Tn5 (p. ej., transposasa EZ-Tn5™, Epicentre Biotechnologies, Madison, Wisconsin) se exponen en los documentos WO 2012/061832; U.S. 2012/0208724, U.S. 2012/0208705 y WO 2014018423. Más realizaciones de transposasas y de secuencias de transposón útiles con algunos de los métodos y de las composiciones proporcionadas en este documento incluyen Staphylococcus aureus Tn552 (Colegio et al., J. Bacteriol., 183: 2384-8 (2001); Kirby et al., Mol. Microbiol., 43: 173-86 (2002)), Ty1 (Devine & Boeke, Nucleic Acids Res., 22: 3765-72 (1994) y el documento WO 95/23875), Transposon Tn7 (Craig, Science 271: 1512 (1996); Craig, Curr Top Microbiol Immunol., 204:27-48 (1996)), Tn/O e IS10 (Kleckner et al., Curr Top Microbiol Immunol., 204:49-82 (1996)), transposasa Mariner (Lampe et al., EMBO J., 15: 5470-9, (1996)), Tel (Plasterk, Curr. Top Microbiol. Immunol., 204: 125-43, (1996)), Elemento P (Gloor, Methods Mol. Biol., 260: 97-114, (2004)), Tn3 (Ichikawa & Ohtsubo, J Biol. Chem. 265: 18829-32, (1990)), secuencias de inserción bacterianas (Ohtsubo & Sekine, Curro Top. Microbiol. Immunol. 204: 1-26, (1996)), retrovirus (Brown, et al., Proc Natl Acad Sci USA, 86:2525-9, (1989)), y retrotransposon de levadura (Boeke & Corces, Annu Rev Microbiol. 43:403-34, (1989)). Más ejemplos incluyen IS5, Tn10, Tn903, IS911, y las versiones de ingeniería de las enzimas de la familia de la transposasa (Zhang et al., PLoS Genet. 5:e1000689. Epub 2009 Oct 16; y Wilson et al. Microbiol. Methods 71:332-5 (2007)). Más ejemplos incluyen transposasas Mua (Véase p.ej., Rasila TS, et al., (2012) PLoS ONE 7(5): e37922. doi:10.1371/journal.pone.0037922). Los ejemplos de transposasas útiles con algunas realizaciones de los métodos y de las composiciones proporcionados en este documento se describen en Leschziner, A.E., et al., (1998) P.N.A.S. 95:7345-7350; y Haapa S., et al., (1999) N.A. Res. 27:2777-2784. Se describen variantes de las transposasas Tn5, tales como las que tienen sustituciones de aminoácidos, inserciones, deleciones y/o fusiones con otras proteínas o péptidos en las patentes de EE.UU.: 5.925.545; 5.965.443; 7.083.980; 7.608.434; y la solicitud de patente de EE.UU. 14/686.961. En algunas realizaciones, la transposasa Tn5 comprende una o más sustituciones en las posiciones 54, 56, 372, 212, 214, 251 y 338 con respecto a la proteína natural, tal como se describe en la solicitud de patente de los Estados Unidos 14/686,961. En algunas realizaciones, la proteína natural Tn5 o su variante puede incluir además un polipéptido de fusión. En algunas realizaciones, el dominio polipeptídico fusionado a la transposasa puede comprender, por ejemplo, el factor de elongación TS.

60 En algunas realizaciones, se pone en contacto un ácido nucleico diana de doble cadena con una pluralidad de

transposomas tal que las cadenas del ácido nucleico diana son melladas y los ácidos nucleicos de transposón se unen a las cadenas del ácido nucleico diana mellado en un lado de los sitios de la mella para obtener un ácido nucleico diana modificado. En algunas realizaciones, el transposoma se pone en contacto con el ácido nucleico diana en la solución. En esta realización, el ácido nucleico diana modificado puede ser producido en solución y, posteriormente, ser capturado en una superficie. Alternativamente, el contacto entre el transposoma y el ácido nucleico diana puede ocurrir en una superficie. El transposoma o el ácido nucleico diana se puede adherir a la superficie antes de que se realice el contacto. El ácido nucleico diana modificado que resulta del contacto entre el transposoma y el ácido nucleico diana en la superficie puede permanecer capturado en la superficie o el ácido nucleico diana modificado se puede liberar de la superficie.

- 5
- 10 En algunas realizaciones, el ácido nucleico capturado es secuenciado. En algunas realizaciones, la proximidad de la información de la secuencia obtenida a partir de dos ácidos nucleicos capturados en una representación lineal de la secuencia de ácidos nucleicos diana es indicativa de la proximidad de los ácidos nucleicos capturados en la superficie. En algunas realizaciones, los ácidos nucleicos capturados con una mayor cercanía entre sí en la superficie comprenden secuencias con una mayor cercanía en la representación de la secuencia del ácido nucleico diana comparado con los ácidos nucleicos capturados más próximos. En algunas realizaciones, la representación de la secuencia del ácido nucleico diana comprende un haplotipo o una representación de ensamblaje.
- 15

Algunas realizaciones de los métodos y de las composiciones proporcionados en este documento también incluyen el uso de la transposición unilateral en el ensamblaje *de novo* de fragmentos secuenciados de un ácido nucleico diana. En algunas realizaciones, los puntos de referencia se insertan en un ácido nucleico diana y se pueden utilizar en el ensamblaje de fragmentos secuenciados de un ácido nucleico diana para generar una representación de la secuencia del ácido nucleico diana. En algunas realizaciones, los fragmentos superpuestos pueden incluir puntos de referencia insertados comunes. El uso de puntos de referencia es particularmente ventajoso con los ácidos nucleicos diana que comprenden secuencias altamente repetitivas. Además, en algunas realizaciones, no se requiere ninguna secuencia de referencia.

- 20
- 25 En algunas realizaciones, los puntos de referencia se insertan en un ácido nucleico diana poniendo en contacto el ácido nucleico diana con una población de transposomas con una actividad de transposasa unilateral, y ácidos nucleicos de transposón que comprenden diferentes códigos de barras. En algunas realizaciones, los ácidos nucleicos de transposón se insertan en cadenas sencillas del ácido nucleico diana mediante la transposición de un lado y, después, la unión. En algunas realizaciones, la transposasa mella una cadena del ácido nucleico diana, y el ácido nucleico del transposón se une a una cadena del ácido nucleico diana mellada en el sitio mellado, y el otro extremo del ácido nucleico del transposón se une al ácido nucleico diana mellado en el otro lado del lado mellado, obteniéndose así un ácido nucleico diana de doble cadena modificado que tiene una inserción en la cadena que comprende un bucle. En algunas realizaciones, el ácido nucleico modificado se puede amplificar y secuenciar. En algunas realizaciones, el ácido nucleico modificado se puede adherir a una superficie. En algunas realizaciones, la unión se puede hacer a través de una proteína de unión de una sola cadena, o una proteína que una bucles de una sola cadena, tal como una recombinasa.
- 30
- 35

En algunas realizaciones, el ácido nucleico diana se modifica sin transposición. En algunas realizaciones, el ácido nucleico diana se puede utilizar aleatoriamente con el mellado por endonucleasa, p. ej., mellado por endonucleasa de New England BioLabs, MA, EE.UU., o endonucleasas de restricción. Las endonucleasas de restricción ejemplares incluyen, entre otras, EcoRI, EcoRII, BamHI, Hind III, TaqI, NotI. Otros ejemplos de endonucleasas de restricción se pueden encontrar en el catálogo New England BioLabs. Opcionalmente, se pueden ampliar huecos con las enzimas que tienen actividad de exonucleasa en 3' o 5', por ejemplo, con exonucleasa I o exonucleasa II, o exonucleasa III. Los adaptadores de oligonucleótidos pueden ser ligados al extremo mellado del ácido nucleico diana. En algunas realizaciones, los adaptadores de oligonucleótidos pueden incluir un sitio de unión de cebado, tal como un sitio de cebado de secuenciación, y un sitio de cebado de amplificación; las secuencias adicionales también pueden incluir un sitio de división, un índice molecular único, un sitio de anclaje, un marcador indicador y un código de barras. Por lo tanto, el mellado del ácido nucleico diana y la unión de uno o más adaptadores mantienen intacto el ácido nucleico diana sin fragmentación. En la figura 12, se muestra un esquema ejemplar del mellado del ácido nucleico diana y la unión del adaptador de oligonucleótidos.

- 40
- 45
- 50 Como se usa en este documento, un "ácido nucleico" incluye, al menos, dos monómeros de nucleótido unidos. Los ejemplos incluyen, entre otros, al ADN, tal como el ADN genómico o el ADNc; ARN, tal como ARNm, ARNs o ARNr; o un híbrido de ADN y ARN. Como se desprende de los ejemplos a continuación y en otras partes del presente documento, un ácido nucleico puede tener una estructura de ácido nucleico natural o una estructura analógica de ácido nucleico no natural. Un ácido nucleico puede contener enlaces de fosfodiéster; sin embargo, en algunas realizaciones, los ácidos nucleicos pueden tener otros tipos de estructuras, comprendiendo, por ejemplo, fosforamida, fosforotioato, fosforoditioato, O-metilfosforoamidita y estructuras de ácido nucleico peptídicas y enlaces. Los ácidos nucleicos pueden tener estructuras positivas; estructuras no iónicas, y estructuras no basadas en ribosa. Los ácidos nucleicos también pueden contener uno o más azúcares carbocíclicos. Los ácidos nucleicos utilizados en los métodos o las composiciones de este documento pueden tener una sola cadena o, alternativamente, dos cadenas, según se especifique. En algunas realizaciones, un ácido nucleico puede contener porciones de secuencia bicatenaria y de una sola cadena, por ejemplo, según queda demostrado por los adaptadores bifurcados. Un ácido nucleico puede contener cualquier combinación de desoxirribonucleótidos y ribonucleótidos, y cualquier combinación
- 55
- 60

- de bases, incluyendo uracilo, adenina, timina, citosina, guanina, inosina, xantina, hipoxantina, isocitosina, isoguanina y análogos de base, tales como nitropirrol (incluyendo 3-nitropirrol) y nitroindol (incluyendo 5-nitroindol), etc. En algunas realizaciones, el ácido nucleico puede incluir al menos una base promiscua. Una base promiscua puede emparejarse por bases con más de un tipo diverso de base y puede ser útil, por ejemplo, cuando esté incluida en los cebadores o los insertos del oligonucleótido que se utilizan para la hibridación al azar en muestras complejas del ácido nucleico, tales como las muestras genómicas de ADN. Un ejemplo de una base promiscua incluye la inosina, que puede emparejarse con adenina, timina o citosina. Otros ejemplos incluyen hipoxantina, 5-nitroindol, 5-nitroindol acílico, 4-nitropirazol, 4-nitroimidazol y 3-nitropirrol. Pueden ser utilizadas bases promiscuas que pueden emparejarse mediante bases con, al menos, dos, tres, cuatro o más tipos de bases.
- Como se usa en este documento, la "secuencia de nucleótidos" incluye el orden y el tipo de monómeros de nucleótido en un polímero de ácido nucleico. Una secuencia de nucleótidos es una entidad de una molécula de ácido nucleico y puede representarse en cualquiera de una variedad de formatos incluyendo, por ejemplo, una representación, imagen, medio electrónico, serie de símbolos, serie de números, serie de letras, serie de colores, etc. La información puede representarse, por ejemplo, a una resolución de un solo nucleótido, con una resolución más alta (p. ej., indicando una estructura molecular para las subunidades de nucleótidos) o a una resolución más baja (p. ej., indicando regiones cromosómicas, como los bloques de haplotipo). Una serie de letras "A," "T," "G," y "C" es una representación de secuencia muy conocida para el ADN que puede correlacionarse, a una resolución de un solo nucleótido, con la secuencia real de una molécula de ADN. Una representación similar se utiliza para el ARN, excepto que "T" se sustituye por "U" en la serie.
- Como se usa en este documento, el término "diferente", cuando se utiliza en referencia a los ácidos nucleicos, significa que los ácidos nucleicos tienen secuencias de nucleótidos que no son las mismas entre sí. Dos o más ácidos nucleicos pueden tener secuencias de nucleótidos que sean diferentes a lo largo de toda su longitud. Alternativamente, dos o más ácidos nucleicos pueden tener secuencias de nucleótidos diferentes a lo largo de una parte sustancial de su longitud. Por ejemplo, dos o más ácidos nucleicos pueden tener porciones de secuencias de nucleótidos diana que sean diferentes para las dos o más moléculas, mientras que también retienen una porción de secuencia universal que es la misma en las dos, o más, moléculas. Pueden aparecer secuencias universales en los extremos de un ácido nucleico, o flanqueando una región de ácido nucleico que deba ser copiada, detectada o amplificada.
- Como se usa en este documento, el "haplotipo" incluye un conjunto de alelos en más de un locus heredado por un individuo de uno de sus progenitores. Un haplotipo puede incluir dos o más loci de todo o parte de un cromosoma. Los alelos incluyen, por ejemplo, polimorfismos de nucleótido único (SNP), repeticiones en tándem cortas (STR), secuencias de genes, inserciones cromosómicas, deleciones cromosómicas, etc. El término "alelos escalonados" se refiere a la distribución de los alelos particulares a partir de un cromosoma particular, o de su porción. En consecuencia, la "fase" de dos alelos puede referirse a una caracterización o representación de la ubicación relativa de dos o más alelos en uno o más cromosomas.
- Como se usa en este documento, una "mella" en un ácido nucleico significa una región de un ácido nucleico bicatenario donde sólo una de las dos cadenas contiene una estructura con una disposición dividida. Por tanto, el "mellado" se refiere al acto de romper la estructura covalente de una sola cadena de ácido nucleico dentro de una región de un ácido nucleico bicatenario. La región es, generalmente, solamente una porción del ácido nucleico bicatenario. La porción puede incluir, por ejemplo, como mucho, 5 pares de bases, 10 pares de bases, 25 pares de bases, 50 pares de bases, 100 pares de bases, 200 pares de bases, 300 pares de bases, 400 pares de bases, 500 pares de bases, 1000 pares de bases. Las regiones pueden incluir porciones más grandes o porciones más pequeñas de un ácido nucleico bicatenario. Por ejemplo, alternativamente o además de los límites superiores ilustrados más arriba, el límite inferior de una porción de un ácido nucleico que se mella puede tener opcionalmente, al menos, 500 pares de bases, 400 pares de bases, 300 pares de bases, 200 pares de bases, 100 pares de bases, 50 pares de bases, 25 pares de bases, 10 pares de bases o más pequeños. Los valores enumerados en los intervalos anteriores pueden definir el tamaño máximo o mínimo para todos los miembros de una población de regiones de ácidos nucleicos o, alternativamente, pueden referirse a un promedio para la población de ácidos nucleicos que tienen las regiones. Se entenderá que un ácido nucleico bicatenario puede ser mellado en ambas cadenas, una primera mella realizándose en una primera región y una segunda mella realizándose en una segunda región. En general, la conectividad efectiva de las dos regiones melladas se puede mantener en condiciones en las que el ácido nucleico permanece en una forma hibridada y bicatenaria. En cambio, la separación de ambas cadenas en la misma región de un ácido nucleico bicatenario puede dar como resultado la pérdida de conectividad efectiva entre las regiones del ácido nucleico que flanquean el sitio de división.
- Como se usa en este documento, el término "superficie" pretende dar a entender una parte o una capa de un soporte sólido o material en gel que está en contacto directo con un fluido circundante, tal como un fluido gaseoso o fluido líquido. La superficie puede estar en contacto con otro material, como un gas, líquido, gel, polímero, polímero orgánico, segunda superficie de un material similar o diferente, metal o revestimiento. La superficie o sus regiones pueden ser sustancialmente planas. La superficie puede tener características superficiales tales como pocillos, huecos, canales, crestas, regiones levantadas, varillas, palitos o similares. En el caso de un sustrato poroso, una superficie puede ubicarse en un poro donde un fluido está en contacto con el sustrato. Por ejemplo, una superficie puede darse en los poros de un gel donde las mitades unidas interactúan con el fluido que entra en los poros. Por lo

tanto, una fracción que está "sobre" una superficie puede estar localizada en un poro de un material poroso, tal como un gel.

Como se usa en este documento, el término "soporte sólido" se refiere a un sustrato rígido que es insoluble en un líquido acuoso. El sustrato puede ser no poroso o poroso. Opcionalmente, el sustrato puede ser capaz de absorber un líquido (por ejemplo, debido a la porosidad) pero, por lo general, será lo suficientemente rígido para que el sustrato no se hinche sustancialmente al absorber el líquido y no se contraiga, sustancialmente, cuando el líquido se retire por secado. Un soporte sólido no poroso es generalmente impermeable a líquidos o gases. Los soportes sólidos ejemplares incluyen, entre otros, vidrio y vidrio modificado o funcionalizado, plásticos (incluyendo acrílicos, poliestireno y copolímeros de estireno y otros materiales, polipropileno, polietileno, polibutileno, poliuretanos, Teflon™, olefinas cíclicas, polímidas, etc.), nilón, cerámica, resinas, Zeonor, sílice o materiales a base de sílice, incluyendo silicio y silicio modificado, carbón, metales, vidrios inorgánicos, paquetes de fibra óptica y polímeros. Los soportes sólidos particularmente útiles para algunas realizaciones se localizan dentro de un aparato de célula de flujo.

Como se utiliza en este documento, la expresión "material en gel" pretende dar a entender un sustrato semirrígido que es permeable a líquidos y gases. Típicamente, el material en gel puede hincharse cuando se absorbe el líquido y puede contraerse cuando el líquido se quita secando. Los geles ejemplares incluyen, entre otros, aquellos que tienen una estructura coloidal, como la agarosa; estructura en malla polimérica, tal como gelatina; o estructura polimérica reticulada, tal como poliacrilamida, SFA (véase, por ejemplo, publicación de solicitud de patente de EE.UU. N.º. 2011/0059865 A1) o PAZAM (véase, por ejemplo, Solicitud de patente provisional de EE.UU. N.º. Ser. 61/753.833).

Como se usa en este documento, el término "unido" pretende dar a entender algo conectado por fuerzas que impiden su separación por difusión. El término puede incluir conexiones que sean de naturaleza covalente o no covalente. Por ejemplo, un ácido nucleico puede estar covalentemente unido a una superficie a través de uno o más enlaces covalentes que creen una cadena de enlaces. La unión no covalente se da cuando al menos uno de los enlaces entre dos cosas (por ejemplo, entre un ácido nucleico y una superficie) no es un enlace covalente. Ejemplos de enlaces no covalentes incluyen, por ejemplo, enlaces de hidrógeno, enlaces iónicos, fuerzas de van der Waals, enlaces hidrófobos o similares.

Como se usa en este documento, el término "información de contigüidad" se refiere a una relación espacial entre dos o más fragmentos de ADN basados en información compartida. El aspecto compartido de la información puede ser con respecto a las relaciones espaciales adyacentes, compartimentarias y a distancia. La información sobre estas relaciones a su vez facilita el ensamblaje jerárquico o mapeo de las lecturas de secuencia derivadas de los fragmentos de ADN. Esta información de contigüidad mejora la eficiencia y exactitud de dicho ensamblaje o mapeo debido a que los métodos tradicionales de ensamblaje o mapeo, utilizados junto con la secuenciación convencional con pistola, no tienen en cuenta los orígenes genómicos relativos o las coordenadas de las lecturas de la secuencia individual, según se relacionan con la relación espacial entre los dos fragmentos de ADN, o más, de los que se derivaron las lecturas de la secuencia individual. Por lo tanto, de acuerdo con las realizaciones descritas en este documento, los métodos de captura de información de contigüidad pueden lograrse mediante métodos de contigüidad de corto alcance para determinar las relaciones espaciales adyacentes, métodos de contigüidad de rango medio para determinar relaciones espaciales compartimentales, o métodos de contigüidad de largo alcance para determinar las relaciones espaciales a distancia. Estos métodos facilitan la exactitud y la calidad del ensamblaje o mapeo de la secuencia de ADN, y pueden utilizarse con cualquier método de secuenciación, como los descritos anteriormente.

En algunas realizaciones, este paso da como resultado la generación de una biblioteca de moléculas de ácido nucleico de pistola derivadas de la secuencia de ADN diana. En una realización alternativa, la fragmentación o inserción puede lograrse incluso mediante un enfoque con adaptador Y, como se describe a continuación. Las una o más moléculas de transposasa pueden ser transposasas libres solubles o se pueden asociar a una secuencia de reconocimiento unida a la superficie.

Como se usa en este documento, el término "código de barras" se refiere a una secuencia de ácidos nucleicos que es única y totalmente independiente de la secuencia de ácidos nucleicos diana. Generalmente, un código de barras puede incluir una o más secuencias de nucleótidos que se pueden utilizar para identificar uno o más ácidos nucleicos en particular. El código de barras puede ser una secuencia artificial, o puede ser una secuencia natural generada durante la transposición, como secuencias de ADN genómico de flanco idéntico (códigos g) al final de los fragmentos de ADN anteriormente yuxtapuestos. Un código de barras puede comprender, al menos, alrededor de 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 o más nucleótidos consecutivos. En algunas realizaciones, un código de barras comprende, al menos, aproximadamente 10, 20, 30, 40, 50, 60, 70, 80, 90, 100 o más nucleótidos consecutivos. En algunas realizaciones, al menos una parte de los códigos de barras en una población de ácidos nucleicos que comprenden códigos de barras es diferente. En algunas realizaciones, al menos alrededor del 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 99% de los códigos de barras son diferentes. En otras de dichas realizaciones, todos los códigos de barras son diferentes. La diversidad de diferentes códigos de barras en una población de ácidos nucleicos que comprenden códigos de barras puede ser generada aleatoriamente o no generada aleatoriamente.

En algunas realizaciones, una secuencia de transposón comprende al menos un código de barras. En algunas realizaciones, como en las que los transposomas comprenden dos secuencias transposón no contiguas, la primera secuencia transposón comprende un primer código de barras, y la segunda secuencia transposón comprende un segundo código de barras. En algunas realizaciones, una secuencia de transposón comprende un código de barras que comprende una primera secuencia de códigos de barras y una segunda secuencia de códigos de barras. En algunas de las realizaciones anteriores, la primera secuencia de códigos de barras se puede identificar o designar para emparejarse con la segunda secuencia de códigos de barras. Por ejemplo, se puede saber que una primera secuencia de códigos de barras conocida se empareja con una segunda secuencia de códigos de barras conocida utilizando una tabla de referencia que comprende una pluralidad de primeras y segundas secuencias de código de barras que se sabe que se emparejan entre sí.

En otro ejemplo, la primera secuencia de códigos de barras puede comprender la misma secuencia que la segunda secuencia de códigos de barras. En otro ejemplo, la primera secuencia de códigos de barras puede comprender el complemento inverso de la segunda secuencia de códigos de barras. En algunas realizaciones, la primera secuencia de códigos de barras y la segunda secuencia de códigos de barras son diferentes. Las secuencias de códigos de barras primera y segunda pueden constar de un bi-código.

En algunas realizaciones de las composiciones y los métodos descritos en este documento, los códigos de barras se utilizan en la preparación de ácidos nucleicos plantilla. Como debererá entenderse, el gran número de códigos de barras disponibles permite que cada molécula de ácido nucleico plantilla comprenda una identificación única. La identificación única de cada molécula en una mezcla de ácidos nucleicos plantilla se puede utilizar en varias aplicaciones. Por ejemplo, se pueden aplicar moléculas de identificación única para identificar moléculas de ácidos nucleicos individuales, en muestras con múltiples cromosomas, en genomas, en células, en tipos celulares, en estados de enfermedades celulares y en especies, por ejemplo, en secuenciación de haplotipos, en discriminación parental del alelo, en la secuenciación metagenómica y en la secuenciación de la muestra de un genoma.

En algunas realizaciones, se puede insertar una pluralidad de códigos de barras únicos a lo largo del ácido nucleico diana durante la transposición. En algunas realizaciones, cada código de barras incluye una primera secuencia de códigos de barras y una segunda secuencia de códigos de barras, teniendo un sitio de fragmentación dispuesto en el mismo. La primera secuencia de códigos de barras y la segunda secuencia de códigos de barras se pueden identificar o designar para emparejarse entre sí. El emparejamiento puede ser informativo para que un primer código de barras esté asociado a un segundo código de barras. Ventajosamente, las secuencias de códigos de barras emparejadas se pueden utilizar para ensamblar los datos de secuenciación de la biblioteca de ácidos nucleicos plantilla. Por ejemplo, la identificación de una primera plantilla de ácido nucleico que comprende una primera secuencia de códigos de barras y una segunda plantilla de ácido nucleico que comprende una segunda secuencia de códigos de barras que se combina con la primera indica que la primera y segunda plantilla de ácidos nucleicos representan secuencias adyacentes entre sí en una representación de secuencia del ácido nucleico diana. Estos métodos se pueden utilizar para ensamblar una representación de secuencia de un ácido nucleico diana de novo, sin el requisito de un genoma de referencia.

Como se usa en este documento, la expresión "al menos una porción" y/o sus equivalentes gramaticales pueden referirse a cualquier fracción de una cantidad entera. Por ejemplo, "al menos una porción" puede referirse al menos al 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 99%, 99,9% o 100% de una cantidad total.

Como se usa en este documento, el término "aproximadamente" significa +/- 10%.

Ácidos nucleicos diana

Las realizaciones de los métodos proporcionados en este documento incluyen un ácido nucleico diana. El ácido nucleico diana es un ácido nucleico de doble cadena. En algunas realizaciones, un ácido nucleico diana incluye ADN genómico, o ADNc. En algunas realizaciones, se utiliza ADN mitocondrial o de cloroplasto. En algunas realizaciones, los ácidos nucleicos diana incluyen ARN o derivados de los mismos como ARNm o ADNc. Algunas realizaciones descritas en este documento pueden utilizar una sola especie de ácido nucleico diana, presente en una copia (es decir, una sola molécula) o, alternativamente, presente en múltiples copias (es decir, un conjunto de moléculas de ácido nucleico que tienen la misma secuencia). Otras realizaciones pueden utilizar una pluralidad de diferentes especies de ácidos nucleicos diana (por ejemplo, moléculas de ácido nucleico que tienen diferentes secuencias de nucleótidos que están presentes en la pluralidad). Por lo tanto, una pluralidad de ácidos nucleicos diana puede incluir una pluralidad de los mismos ácidos nucleicos diana, una pluralidad de diferentes ácidos nucleicos diana donde algunos ácidos nucleicos diana son los mismos, o una pluralidad de ácidos nucleicos diana donde todos los ácidos nucleicos diana son diferentes. Los ácidos nucleicos diana pueden prepararse a partir de moléculas de ácido nucleico obtenidas de un solo organismo o de poblaciones de moléculas de ácido nucleico obtenidas de fuentes que incluyan más de un organismo. Un ácido nucleico diana puede provenir de una sola célula; de múltiples células, tejidos o fluidos corporales de un solo organismo; de células, tejidos o fluidos corporales de varios organismos de la misma especie; o de múltiples especies, como con muestras metagenómicas, tales como de muestras ambientales. Las fuentes de moléculas de ácido nucleico incluyen, entre otras, organelos, células, tejidos, órganos u organismos.

En algunas realizaciones, se pone en contacto un ácido nucleico diana con un transposoma de tal manera que el ácido nucleico transposón se inserta o se adhiere al ácido nucleico diana para proporcionar un ácido nucleico modificado. En algunas realizaciones, los ácidos nucleicos modificados pueden estar más manipulados, por ejemplo, pueden extenderse, amplificarse y unirse.

5 Transposomas

Los métodos indicados en este documento utilizan los transposomas. En algunas realizaciones, un transposoma incluye una transposasa ligada a uno o más ácidos nucleicos de transposón. En algunas realizaciones, el transposoma comprende una actividad de transposasa unilateral que incluye mellar una cadena de un ácido nucleico bicatenario, y unir un ácido nucleico de transposón a la cadena mellada en un lado del sitio de mella.

10 En algunas realizaciones, los transposomas con actividad de transposasa unilateral incluyen los transposomas que comprenden ciertos tipos de transposasas que tienen actividad de transposasa unilateral. En algunas realizaciones, una transposasa natural tiene actividad unilateral de transposasa o se modifica para tener actividad unilateral de transposasa. Los ejemplos de transposasas con actividad unilateral de transposasa o que se pueden modificar por tener actividad unilateral de transposasa incluyen Mu, Mu E392Q, Tn5, RAG, Tn5 hiperactivo, variantes Tn5, Vibhar, y Tn552 (Leschziner, A.E., et al., (1998) P.N.A.S. 95:7345-7350; y Haapa S., et al., (1999) N.A. Res. 27:2777-2784).
 15 Más ejemplos de transposasas con actividad unilateral de transposasa o que se pueden modificar para tener actividad unilateral de transposasa se enumeran en este documento. En algunas realizaciones, una transposoma que tiene actividad unilateral de transposasa comprende un solo monómero y un ácido nucleico de transposón. En algunas realizaciones, una transposasa se puede modificar para carecer de la capacidad de formar un dímero. En
 20 algunas realizaciones, un transposoma que tiene actividad de transposasa unilateral comprende un dímero en el que uno de los monómeros carece de actividad de transposasa. En algunas realizaciones, las subunidades de monómero del dímero se pueden asociar covalentemente.

En algunas realizaciones, un transposoma que tiene actividad de transposasa unilateral consiste en un ácido nucleico de transposón bloqueado. En algunas realizaciones, un ácido nucleico bloqueado del transposón se
 25 bloquea de ser unido a una cadena de un ácido nucleico bicatenario mellado. El ácido nucleico bloqueado del transposón puede incluir grupos bloqueantes en el extremo 3' del ácido nucleico del transposón que inhibe la unión del ácido nucleico del transposón a otros ácidos nucleicos. En algunas realizaciones, los grupos bloqueantes pueden incluir un grupo didesoxi, un grupo espaciador y un grupo biotina. En algunas realizaciones, una población de transposomas que tiene actividad unilateral de transposasa se puede preparar poniendo en contacto las
 30 transposasas con ácidos nucleicos bloqueados del transposón y ácidos nucleicos de transposón no bloqueados. Los ácidos nucleicos de transposón no bloqueados incluyen ácidos nucleicos de transposón que carecen de grupos bloqueantes. En algunas realizaciones, se obtiene una población que incluye transposomas que comprende dímeros de transposasa que comprenden un ácido nucleico bloqueado del transposón y un ácido nucleico no bloqueado del transposón. En algunas realizaciones, los dímeros de transposasa comprenden dos ácidos nucleicos bloqueados del transposón. En algunas realizaciones, los dímeros de transposasa comprenden dos ácidos nucleicos no bloqueados del transposón. En algunas realizaciones, la proporción de los diferentes tipos de dímeros en una población puede ser manipulada poniendo en contacto las transposasas con varias proporciones de ácidos nucleicos de transposón bloqueados frente a ácidos nucleicos de transposón no bloqueados. En algunas realizaciones, la proporción de ácidos nucleicos de transposón bloqueados frente a ácidos nucleicos de transposón no bloqueados es mayor o igual
 40 a 1:1, 5:1, 10:1, 50:1, 100:1, 200:1, 500:1, 1000:1, o cualquier intervalo entre las proporciones anteriores.

Otros ácidos nucleicos de transposón útiles son los que son más cortos o más largos que el transposón estándar. Por ejemplo, la cadena transferida en el extremo de 3' se puede hacer más corta (p. ej., quitando una o más bases) para inhibir la reacción de transferencia que ocurra. Del mismo modo, el extremo 3' puede ser más largo para dar lugar a tal inhibición.

45 Los métodos para la transposición unilateral también se pueden utilizar para regular el tamaño de inserto de la biblioteca después de la transposición. Una ventaja de este enfoque es que la longitud del tamaño del inserto puede determinarse por la relación entre el complejo activo y el no activo que, en muchas situaciones, es más fácil de controlar que el tiempo de incubación o concentración de transposoma y ácido nucleico.

Los dímeros de transposoma que tienen una subunidad monomérica de transposoma activa y una subunidad de transposoma no activa se puede preparar y agregar a los ácidos nucleicos antes o después de que se inicie una
 50 reacción de transposición. Por ejemplo, la reacción se puede iniciar mediante la adición de Mg^{2+} . En realizaciones particulares, una población que incluye una mezcla de tres especies de dímero de transposomas (es decir, pueden formarse activos: dímeros activos, inactivos: dímeros inactivos, y activos: dímeros inactivos). Pueden prepararse poblaciones que tienen diferentes actividades alterando la proporción de subunidades de transposoma activas y no activas que se combinan para formar una mezcla. La relación se puede seleccionar para influir en el tamaño medio de inserción que se produce cuando una muestra de ácido nucleico se trata con la mezcla. Se puede lograr un control similar del tamaño de inserto utilizando una mezcla de especies de transposoma que tienen sólo las dos especies de transposoma de activo: activo e inactivo: inactivo. Como se demuestra en el diagrama de la FIG. 11, las especies inactivas: que son capaces de unirse al ADN diana, pero incapaces de transponer la diana, actúan como
 55 espaciadores. En otras palabras, los inactivos: los dímeros inactivos compiten por sitios que, de otro modo, estarían
 60

unidos a los activos: dímeros activos y transpuestos. La titulación rutinaria de la cantidad de inactivos: dímero inactivo que es añadido en una mezcla de reacción de la transposasa se puede utilizar para controlar tamaños de fragmento medios producidos por la mezcla. Estos métodos tienen varias ventajas incluyendo, por ejemplo, ser relativamente independiente del tiempo y controlable. Las reacciones de transposición convencionales (p. ej., métodos de preparación de muestras Nextera® de Illumina, Inc. (San Diego, CA)) requieren un control cuidadoso del período de tiempo de reacción para lograr una reacción de transposición que produzca fragmentos de un tamaño promedio deseado. Los métodos actuales de transposición unilateral, por otro lado, se pueden llevar a cabo como se establece anteriormente para ser menos sensibles al tiempo. Más concretamente, se puede seleccionar la proporción de subunidades de monómero activo a inactivo para determinar el tamaño de los fragmentos de la biblioteca.

En algunas realizaciones, un transposoma con actividad de transposasa unilateral se puede unir a una superficie. El transposoma puede unirse a través de la transposasa o a través del ácido nucleico transposón. Por ejemplo, una transposasa puede ser covalentemente o no covalentemente unida a una superficie. Alternativamente o además, un ácido nucleico del transposón puede ser covalentemente o no covalentemente unido a la superficie. Se establecen uniones útiles, superficies y métodos asociados para su preparación y uso con más detalle en este documento y en la solicitud de patente de EE.UU. N°. Ser. 13/790.220.

En algunas realizaciones, una transposasa incluye una enzima que es capaz de formar un complejo funcional con un ácido nucleico de transposón que comprende un elemento del transposón o el elemento de transposasa, y catalizar la inserción o la transposición del ácido nucleico del transposón en un ácido nucleico diana para proporcionar un ácido nucleico modificado. Por ejemplo, en una reacción de transposición *in vitro*, la inserción de ácidos nucleicos de transposón en un ADN diana para proporcionar un ADN modificado. En algunas realizaciones, una transposasa incluye una enzima que es capaz de formar un complejo funcional con un ácido nucleico de transposón que comprende un elemento del transposón o el elemento de transposasa, y catalizar la transposición unilateral en un ácido nucleico diana para proporcionar un ácido nucleico modificado.

En algunas realizaciones, la inserción o la unión de los ácidos nucleicos de transposón por una transposasa pueden ser en un sitio al azar o substancialmente aleatorio en un ácido nucleico diana. Las transposasas también incluyen integrasas de retrotransposones y transposasas de retrovirus. Las realizaciones de transposasas útiles con algunos de los métodos proporcionados en este documento incluyen las descritas en el documento U.S. 2010/0120098. Otras realizaciones de transposasas y los elementos del transposón incluyen una transposasa hiperactiva Tn5 y un elemento de transposasa tipo Tn5 (Goryshin and Reznikoff, J. Biol. Chem., 273:7367 (1998)), transposasa MuA y un elemento de transposasa Mu que comprende las secuencias finales R1 y R2 (Mizuuchi, Cell, 35: 785, (1983) y Savilahti, et al., EMBO J., 14: 4893, 15 (1995)). Ejemplos de elementos de transposasa que forman un complejo con una transposasa hiperactiva Tn5 (p. ej., EZ-Tn5™ transposasa, Epicentre Biotechnologies, Madison, Wisconsin) se exponen en los documentos WO 2012/061832; U.S. 2012/0208724, U.S. 2012/0208705 y WO 2014018423. Más realizaciones de transposasas y de ácidos nucleicos de transposón útiles con algunos de los métodos y de las composiciones proporcionadas en este documento incluyen Staphylococcus aureus Tn552 (Colegio et al., J. Bacteriol., 183: 2384-8 (2001); Kirby et al., Mol. Microbiol., 43: 173-86 (2002)), Ty1 (Devine & Boeke, Nucleic Acids Res., 22: 3765-72 (1994) y el documento WO 95/23875), Transposón Tn7 (Craig, Science 271: 1512 (1996); Craig, Curr Top Microbiol Immunol., 204:27-48 (1996)), Tn/O y IS10 (Kleckner et al., Curr Top Microbiol Immunol., 204:49-82 (1996)), transposasa Mariner (Lampe et al., EMBO J., 15: 5470-9, (1996)), Tel (Plasterk, Curro Topics Microbiol. Immunol., 204: 125-43, (1996)), Elemento P (Gloor, Methods Mol. Biol., 260: 97-114, (2004)), transposasa Mos-1 (Richardson et al., EMBO Journal 25:1324-1334 (2006)), Tn3 (Ichikawa & Ohtsubo, J Biol. Chem. 265: 18829-32, (1990)), secuencias de inserción bacteriana (Ohtsubo & Sekine, Curro Top. Microbiol. Immunol. 204: 1-26, (1996)), retrovirus (Brown, et al., Proc Natl Acad Sci USA, 86:2525-9, (1989)), y retrotransposón de levadura (Boeke & Corces, Annu Rev Microbiol. 43:403-34, (1989)). Más ejemplos incluyen IS5, Tn10, Tn903, IS911, y las versiones modificadas de las enzimas de la familia de la transposasa (Zhang et al., PLoS Genet. 5:e1000689. Epub 2009 Oct 16; y Wilson et al. Microbiol. Methods 71:332-5 (2007)). Más ejemplos incluyen transposasas Mua (véase, p. ej., Rasila TS, et al., (2012) PLoS ONE 7(5): e37922. doi:10.1371/journal.pone.0037922).

En algunas realizaciones, un ácido nucleico transposón comprende un ácido nucleico bicatenario. Un elemento transposón incluye una molécula de ácido nucleico, o parte de la misma, que incluye las secuencias de nucleótidos que forman un transposoma con una enzima transposasa o integrasa. En algunas realizaciones, un elemento transposón es capaz de formar un complejo funcional con la transposasa en una reacción de transposición. Se proporcionan en este documento ejemplos de los elementos del transposón, e incluyen el extremo de transposón del extremo externo de 19 bp ("OE") el extremo de transposón del extremo interno ("IE"), o el extremo de transposón del "extremo del mosaico" ("ME") reconocido, por ejemplo, por una transposasa Tn5 natural o mutante, o el extremo de transposón R1 y de R2 (véase, por ejemplo, el documento US 2010/0120098). Los elementos transposón pueden comprender cualquier ácido nucleico o análogo de ácido nucleico adecuado para formar un complejo funcional con la enzima transposasa o integrasa en una reacción de transposición *in vitro*. Por ejemplo, el extremo del transposón puede abarcar el ADN, el ARN, las bases modificadas, las bases no naturales, la estructura modificada, y puede incluir mellas en una o ambas cadenas.

En algunas realizaciones, un ácido nucleico del transposón puede incluir un elemento del transposón y secuencias adicionales. En algunas realizaciones, las secuencias adicionales pueden insertarse o unirse a un ácido nucleico

diana en una reacción de transposición. Las secuencias adicionales pueden incluir un sitio de enlace de cebado, tal como un sitio de cebado de secuenciación, y un sitio de cebado de amplificación; las secuencias adicionales también pueden incluir un sitio de división, un índice molecular único, un sitio de anclaje, un marcador indicador y un código de barras.

- 5 En algunas realizaciones, un sitio de enlace de cebado puede incluir secuencias para que los cebadores de secuenciación se hibriden a un ácido nucleico en una reacción de secuenciación. En algunas realizaciones, un sitio de enlace de cebado puede incluir secuencias para que los cebadores se hibriden a un ácido nucleico en una reacción de amplificación u otra reacción de extensión.

10 En algunas realizaciones, un sitio de división puede incluir un sitio en un ácido nucleico de transposón que puede ser fragmentado. Por ejemplo, un ácido nucleico del transposón que comprende un sitio de división se puede insertar en un ácido nucleico diana y el ácido nucleico modificado puede, entonces, ser fragmentado en el sitio de división insertado. En algunas realizaciones, un sitio de división incluye una secuencia de reconocimiento de la enzima de restricción y/o un sitio de división de la enzima de restricción. En algunas realizaciones, un sitio de división puede incluir, al menos, un ribonucleótido en un ácido nucleico que puede abarcar, por el contrario, los desoxirribonucleótidos y puede ser dividido con un RNAsa. Pueden ser utilizados agentes químicos de división capaces de dividir selectivamente el enlace fosfodiéster entre un desoxirribonucleótido y un ribonucleótido incluyendo, por ejemplo, iones metálicos tales como iones de metal de tierra rara (p. ej. La^{3+} , particularmente Tm^{3+} , Yb^{3+} o Lu^{3+} , $\text{Fe}(3)$ o $\text{Cu}(3)$), o la exposición a un pH elevado. En algunas realizaciones, un sitio de división puede incluir una o más secuencias de reconocimiento para una nickasa, es decir, una endonucleasa de mellar que rompe una cadena de una región particular de un ácido nucleico bicatenario. Por lo tanto, el sitio de fragmentación puede incluir una primera secuencia de reconocimiento de nickasa y, opcionalmente, una segunda secuencia de reconocimiento de nickasa. La primera y la segunda secuencias de reconocimiento de nickasa pueden ser iguales que las otras o diferentes entre sí. En algunas realizaciones, un sitio de división puede incluir uno o más análogos de nucleótido que comprenden un sitio no básico y permite la división en el sitio de fragmentación en presencia de ciertos agentes químicos, tales como poliamina, N,N'- dimetiletildiamina (DMED) (Véase, por ejemplo, el documento, U.S. 2010/0022403).

30 En algunas realizaciones, un sitio no básico puede ser creado por la modificación de un nucleótido de uracilo dentro del sitio de división, por ejemplo, usando una enzima de uracilo ADN glicosilasa (UDG). La cadena de polinucleótidos, incluido el sitio no básico, puede ser dividido entonces en el sitio no básico por tratamiento con endonucleasa (p. ej. endonucleasa Endo IV, Liasa AP, glicosilasa FPG/Liasa AP, Endo VIII glicosilasa/Liasa AP), calor o álcali. También pueden generarse sitios no básicos en análogos de nucleótidos distintos de desoxiuridina y dividirse de una manera análoga por tratamiento con endonucleasa, calor o álcali. Por ejemplo, la 8-oxo-guanina se puede convertir en un sitio no básico por exposición a la glucosilasa FPG. La desoxiinosina se puede convertir en un sitio no básico por exposición a Alka glicosilasa. Los sitios no básicos así generados pueden entonces ser divididos, típicamente por el tratamiento con una endonucleasa conveniente, tal como la Endo IV o Liasa AP (véase p. ej., el documento U.S. 2011/0014657). En otro ejemplo, un sitio de división puede incluir un acoplamiento diol que permita la división por tratamiento con peryodato (p. ej., peryodato de sodio). En otro ejemplo, un sitio de división puede incluir un grupo disulfuro que permita la división con un agente reductor químico, p. ej. clorhidrato de tris (2-carboxietil)-fosfato (TCEP). En algunas realizaciones, un sitio de división puede incluir un resto de fotodivisión. La división fotoquímica se puede llevar a cabo por cualquiera de una variedad de métodos que utilicen energía lumínica para romper los enlaces covalentes. Un sitio para la división fotoquímica se puede proporcionar por una fracción química no nucleotídica en un ácido nucleico, tal como la fosforamidita [4-(4,4'-dimetoxitritiloxi)butiramidometil]-1-(2-nitrofenil)-etil]-2-cianoetil-(N,N-diisopropil)-fosforamidita) (Glen Research, Sterling, Va., USA, No. Cat 10-4913-XX).

45 En algunas realizaciones, un ácido nucleico de transposón puede incluir un sitio de anclaje. En algunas realizaciones, un sitio de anclaje puede incluir secuencias que se pueden enlazar específicamente a sondas de captura. En algunas realizaciones, el sitio de anclaje comprende secuencias que son complementarias y/o sustancialmente complementarias para capturar sondas que comprenden ácidos nucleicos. En algunas realizaciones, un sitio de anclaje puede incluir un ligando o receptor que una sonda de captura que comprenda un receptor o ligando correspondiente. En otras palabras, un sitio de anclaje y una sonda de captura pueden abarcar un par ligando/receptor. En algunas realizaciones, se puede asociar un ligando o receptor al sitio de anclaje de un ácido nucleico de transposón a través de un nucleótido modificado. Ejemplos de ligandos y receptores incluyen biotina o polyHis que se pueden unir a estreptavidina o níquel, respectivamente. Otros ejemplos incluyen pares de ligandos y sus receptores conocidos en la técnica, por ejemplo, avidina-biotina, estreptavidina-biotina y derivados de biotina, estreptavidina o avidina, incluyendo, entre otros, 2-iminobiotina, destiobiotina, NeutrAvidin (Molecular Probes, Eugene, Oreg.), CaptAvidin (Molecular Probes), y similares; proteínas de unión/péptidos, incluyendo la proteína de unión a maltosa-maltosa (MBP), proteína/péptido de unión calcio-calcio (CBP); antígeno-anticuerpo, incluyendo los marcadores de epítipo, incluyendo c-MYC, HA, VSV-G, HSV, V5 y FLAG Tag™, y sus anticuerpos anti-epítipo correspondientes; haptenos, por ejemplo, dinitrofenilo y digoxigenina, y sus anticuerpos correspondientes; aptámeros y sus dianas correspondientes; los marcadores Poly-His (p. ej., Penta-His y hexa-His) y sus parejas de unión, incluidos los materiales de cromatografía de afinidad de iones metálicos inmovilizados (IMAC) correspondientes y los anticuerpos anti-Poly-His; fluoróforos y anticuerpos anti-fluoróforo; cadenas de ácido nucleico y sus cadenas complementarias; y similares.

En algunas realizaciones, un ácido nucleico de transposón puede incluir un marcador indicador. Los marcadores indicadores útiles incluyen cualquiera de una variedad de marcadores identificables, marcadores o grupos conocidos en la técnica. En ciertas realizaciones, un marcador indicador puede emitir una señal. Entre los ejemplos de señales se incluyen las que son fluorescentes, quimioluminiscentes, bioluminiscentes, fosforescentes, radiactivas, calorimétricas o electro-quimioluminiscentes. Los marcadores indicadores ejemplares incluyen fluoróforos, radioisótopos, cromógenos, enzimas, antígenos, incluyendo marcadores de epítipo, nanocristales semiconductores tales como puntos cuánticos, metales pesados, tintes, grupos fosforescentes, grupos quimioluminiscentes, fracciones de detección electroquímica, proteínas de unión, fósforos, quelatos de tierras raras, quelatos metálicos de transición, tintes cercanos al infrarrojo, marcadores electro-quimioluminiscentes y marcadores indicadores compatibles con el espectrómetro de masas, tales como marcadores de masas, marcadores de carga e isótopos. Más marcadores indicadores que se pueden utilizar con los métodos y las composiciones descritas en este documento incluyen marcadores espectrales, tales como tintes fluorescentes (p. ej., isotiocianato de fluoresceína, rojo de Tejas, Rodamina y similares); radiomarcadores (p. ej., ^3H , ^{125}I , ^{35}S , ^{14}C , ^{32}P , ^{33}P , etc.); enzimas (p. ej., peroxidasa de rábano picante, fosfatasa alcalina, etc.); marcadores colorimétricos espectrales, tales como oro coloidal o vidrio coloreado o plástico (p. ej., poliestireno, polipropileno, látex, etc.); perlas; marcadores magnéticos; marcadores eléctricos; marcadores térmicos; y marcadores de masas.

En algunas realizaciones, un ácido nucleico de transposón puede incluir un código de barras. En algunas realizaciones, una población de transposomas puede incluir ácidos nucleicos de transposón que comprenden el mismo código de barras, uno o más códigos de barras diferentes, o cada ácido nucleico de transposón puede incluir un código de barras diferente. En algunas realizaciones, se puede utilizar un código de barras insertado o unido a un ácido nucleico diana para identificar un ácido nucleico diana. En algunas realizaciones, se puede utilizar un código de barras para identificar un evento de inserción en un ácido nucleico diana. En algunas realizaciones, cada transposoma, en una población de transposomas, incluye un ácido nucleico de transposón con un código de barras diferente que se puede utilizar para identificar un sitio de inserción en el ácido nucleico diana. En algunas realizaciones, se puede utilizar un código de barras para identificar el sitio de inserción después de la fragmentación en un sitio de división, por ejemplo, donde un código de barras se extiende por un sitio de división. Los códigos de barras ilustrativos y los métodos para su preparación y uso se establecen en la pub. int. No. WO 2012/061832; y los documentos US 2012/0208724, US 2012/0208705 y PCT/US2013/031023. En algunas realizaciones, los códigos de barras insertados en un ácido nucleico diana pueden ser útiles como hitos en la alineación posterior de secuencias fragmentadas para obtener una representación de secuencia del ácido nucleico diana. En algunas realizaciones, los fragmentos que incluyen códigos de barras comunes pueden identificarse como si tuvieran secuencias superpuestas.

En algunas realizaciones, un ácido nucleico del transposón puede incluir dos elementos del transposón que están ligados entre sí. Un enlazante se puede incluir en el inserto de tal manera que un primer elemento de transposón es contiguo a un segundo elemento de transposón. Un inserto particularmente útil es uno que forma un complejo "en bucle" como se establece en la publicación internacional No. WO 2012/061832; y los documentos US 2012/0208724, US 2012/0208705 y PCT/US2013/031023. En tales estructuras, un solo inserto que tiene elementos de transposón contiguos se une a dos subunidades de transposasa formando un complejo "en bucle". En algunas realizaciones, el ácido nucleico del transposón puede incluir un grupo de bloqueo.

40 **Sustratos**

Algunas realizaciones de los métodos proporcionados en este documento incluyen el uso de un sustrato que tiene una superficie. Los sustratos útiles incluyen, por ejemplo, soportes sólidos y geles. En algunas realizaciones, la superficie se une a los ácidos nucleicos. En algunas realizaciones, la superficie comprende una pluralidad de sondas de captura que unen los ácidos nucleicos a la superficie a través de la complementariedad de Watson-Crick. En algunas realizaciones, las sondas de captura se unen a los marcadores de anclaje. En algunas realizaciones, las sondas de captura y los marcadores de anclaje comprenden los ácidos nucleicos. En algunas realizaciones, las sondas de captura y los marcadores de anclaje comprenden grupos de moléculas pequeñas que se unen específicamente entre sí, tal como un receptor o ligando como se proporciona en este documento, por ejemplo, biotina, avidina, HisD, níquel, anticuerpos y antígenos.

Los sustratos pueden tener dos o tres dimensiones y pueden ser una superficie plana (p. ej., un portaobjetos de vidrio) o pueden ser moldeados. Los materiales útiles incluyen vidrio (p. ej., vidrio de poro controlado (CPG)), cuarzo, plástico (como poliestireno (poliestireno reticulado poco reticulado y muy reticulado), policarbonato, polipropileno y poli(metilmacrilato)), copolímero acrílico, poliamida, silicio, metal (p. ej., oro derivatizado con alcanotiolato), celulosa, nylon, látex, dextrano, matriz de gel (p. ej., gel de sílice), poliácroleino o materiales compuestos. Los soportes sólidos tridimensionales convenientes incluyen, por ejemplo, esferas, micropartículas, granos, membranas, láminas, placas, microchips trabajados a máquina, tubos (p. ej., tubos capilares), micropocillos, dispositivos microfluidos, canales, filtros, o cualquier otra estructura adecuada para anclar un ácido nucleico u otra sonda de captura. Los soportes sólidos pueden incluir matrices micro planas o matrices capaces de disponer de regiones que incluyan poblaciones de ácidos nucleicos o cebadores u otras sondas de captura. Algunos ejemplos son las láminas de CPG y poliestireno derivatizadas con nucleósidos; láminas magnéticas derivatizadas; poliestireno injertado con polietilenglicol, y similares.

Se pueden utilizar varias composiciones y métodos asociados de preparación y uso de esas composiciones para unir, anclar o inmovilizar sondas de captura, tales como ácidos nucleicos a una superficie de un sustrato. La adhesión se puede lograr a través de la unión directa o indirecta a la superficie. La unión puede ser por acoplamiento covalente (véase p. ej., Joos et al. (1997) *Analytical Biochemistry*, 247:96-101; Oroskar et al. (1996) *Clin. Chem.*, 42:1547-1555; y Khandjian (1986) *Mol. Bio. Rep.*, 11:107-11). Una adhesión preferida es la unión directa de la amina de un nucleótido terminal de un ácido nucleico a un epóxido integrado en la superficie. La unión también puede ser a través de la unión no covalente. Por ejemplo, la biotina-estreptavidina (Taylor et al. (1991) *Phys. D: Appl. Phys.*, 24:1443) y la digoxigenina con anti-digoxigenina (Smith et al., *Science*, 253: 1122 (1992)) son herramientas comunes para anclar ácidos nucleicos a las superficies. La fijación de un ácido nucleico a una superficie puede ser a través de una estructura intermedia tal como una perla, una partícula o un gel. La fijación de los ácidos nucleicos a una matriz a través de un gel es ejemplificada por células de flujo disponibles comercialmente en Illumina Inc. (San Diego, CA) o descrita en los documentos US 2010/10111768; U.S. 2012/0270305; y WO 05/065814.

En algunas realizaciones, un sustrato puede tener una superficie continua o monolítica. Por lo tanto, los fragmentos de ácido nucleico pueden unirse en lugares espacialmente aleatorios donde la distancia entre los fragmentos vecinos más cercanos (o los clústeres vecinos más cercanos derivados de los fragmentos) será variable. Las matrices resultantes pueden tener un patrón espacial variable o aleatorio de propiedades. En algunas realizaciones, un sustrato utilizado en un método establecido en este documento puede incluir una matriz de sondeos de captura que están presentes en un patrón repetido. En algunas de estas realizaciones, las sondas de captura proporcionan las ubicaciones a las que los ácidos nucleicos se pueden unir. En algunas realizaciones, los patrones de repetición son patrones hexagonales, patrones rectilíneos, patrones cuadrículares, patrones con simetría reflexiva, patrones con simetría rotacional, o similares. Las sondas de captura a las que los ácidos nucleicos modificados se adhieren pueden tener cada una un área que es aproximadamente 1 mm², 500 μm², 100 μm², 25 μm², 10 μm², 5 μm², 1 μm², 500 nm², o 100 nm², o más pequeña que esto, o un intervalo definido por cualquiera de los dos valores precedentes. Alternativamente o además, cada entidad puede tener un área que es aproximadamente 100 nm², 250 nm², 500 nm², 1 μm², 2.5 μm², 5 μm², 10 μm², 100 μm², o 500 μm², o más grande que esto o un intervalo definido por cualesquiera dos de los valores precedentes. Un grupo o una colonia de ácidos nucleicos que son el resultado de la amplificación de fragmentos en una matriz (ya sea en patrones o espacialmente aleatorios) puede tener, de forma similar, un área que esté en un intervalo por encima o entre un límite superior e inferior seleccionado de los ejemplificados anteriormente.

En algunas realizaciones, la densidad de entidades tales como ácidos nucleicos, sondas de captura o ácidos nucleicos capturados en una superficie puede ser, al menos, 1000 entidades/mm², 10000 entidades/mm², 100000 entidades/mm², 1000000 entidades/mm², o cualquier intervalo entre los valores anteriores. En algunas realizaciones, la densidad de entidades tales como ácidos nucleicos, sondas de captura o ácidos nucleicos capturados en una superficie puede ser, al menos, 1000 entidades/μm², 10000 entidades/μm², 100000 entidades/μm², 1000,000 entidades/μm², 2.000.000 entidades/μm², 3000.000 entidades/μm², 4000.000 entidades/μm², 5000.000 entidades/μm², 6000.000 entidades/μm², 7000.000 entidades/μm², 8000.000 entidades/μm², 9000.000 entidades/μm², 10.000.000 entidades/μm², 20.000.000 entidades/μm², 50.000.000 entidades/μm², 100.000.000 entidades/μm², o cualquier intervalo entre los valores anteriores.

Varias plataformas de secuenciación disponibles comercialmente utilizan sustratos que tienen pocillos que proporcionan una barrera a la difusión de reactivos de detección (p. ej., pirofosfato en plataformas disponibles a partir de 454 Lifesciences (una filial de Roche, Basilea, Suiza) o protones en plataformas disponibles de Ion Torrent (una subsidiaria de Life Technologies, Carlsbad, California)) durante las etapas de detección de secuencia.

Algunas de las realizaciones proporcionadas en este documento incluyen la amplificación de porciones de un ácido nucleico diana, ácido nucleico modificado, o fragmentos de los mismos. Se puede utilizar cualquier metodología de amplificación adecuada conocida en la técnica. En algunas realizaciones, los fragmentos de ácido nucleico se amplifican en un sustrato, o sobre éste. Por ejemplo, en algunas realizaciones, los fragmentos de ácido nucleico se amplifican utilizando metodologías de amplificación de puentes como se ejemplifican en las descripciones de Pat. de EE.UU. No. 5.641.658; publicación de patente de EE.UU. No. 2002/0055100; Pat. de EE.UU. No. 7.115.400; publicación de patente de EE.UU. No. 2004/0096853; publicación de patente de EE.UU. No. 2004/0002090; publicación de patente de EE.UU. No. 2007/0128624; y publicación de patente de EE.UU. No. 2008/0009420.

Los métodos de amplificación de puentes permiten que los productos de amplificación se inmovilicen en un sustrato, o sobre éste, para formar matrices compuestas de grupos (o "colonias") de moléculas de ácido nucleico inmovilizadas. Cada grupo o colonia en tal matriz se forma a partir de una pluralidad de cadenas inmovilizadas idénticas del polinucleótido y de una pluralidad de cadenas complementarias inmovilizadas idénticas del polinucleótido. Se puede hacer referencia a las matrices así formadas en este documento como "matrices agrupadas". Los productos de las reacciones de amplificación en fase sólida son las estructuras denominadas "en puente" cuando se forman por pares hibridados de cadenas de polinucleótidos inmovilizadas y cadenas complementarias inmovilizadas, ambas cadenas inmovilizadas en el soporte sólido en el extremo 5', preferiblemente a través de una adhesión covalente. Las metodologías de amplificación en puente son ejemplos de métodos en los que se utiliza una plantilla de ácido nucleico inmovilizado para producir amplicones inmovilizados. También se pueden utilizar otras metodologías adecuadas para producir amplicones inmovilizados a partir de fragmentos de

ácido nucleico inmovilizados producidos de acuerdo con los métodos proporcionados en este documento. Por ejemplo, se pueden formar uno o más grupos o colonias a través de PCR en fase sólida, MDA en fase sólida, RCA en fase sólida, etc., ya sea que uno o ambos cebadores de cada par de cebadores de amplificación se inmovilicen.

5 Se apreciará que cualquiera de las metodologías de amplificación descritas en este documento o generalmente conocidas en la técnica se pueden utilizar con cebadores universales o específicos de diana para amplificar fragmentos de ADN inmovilizados. Los métodos adecuados para la amplificación incluyen, entre otros, la reacción en cadena de la polimerasa (PCR), la amplificación por desplazamiento de cadena (SDA), la amplificación mediada por transcripción (TMA) y la amplificación basada en secuencias de ácido nucleico (NASBA), por ejemplo, como se describe en la patente estadounidense no. 8.003.354. Los métodos de amplificación anteriores se pueden emplear para amplificar uno o más ácidos nucleicos de interés. Por ejemplo, pueden ser utilizados PCR, multiplex PCR, SDA, TMA, NASBA y similares para amplificar fragmentos de ácido nucleico inmovilizados. En algunas realizaciones, se incluyen en la reacción de la amplificación cebadores dirigidos específicamente a los ácidos nucleicos de interés.

15 Otros métodos apropiados para la amplificación de ácidos nucleicos pueden incluir la extensión y la unión de oligonucleótidos, la amplificación del círculo rodante (RCA) (Lizardi et al., Nat. Genet. 19:225-232 (1998)) y análisis de unión de oligonucleótidos (OLA) (Véase p.ej., las patentes de EE. UU. Nos. 7.582.420, 5.185.243, 5.679.524 y 5.573.907; EP 0320308; EP 0336731; EP 0439182; WO 90101069; WO 89/12696; y WO 89109835). It will be appreciated that these amplification methodologies can be designed to amplify immobilized nucleic acid fragments. Por ejemplo, en algunas realizaciones, el método de amplificación puede incluir la amplificación de la sonda de unión o las reacciones de análisis de unión de oligonucleótidos (OLA) que contienen cebadores dirigidos específicamente al ácido nucleico de interés. En algunas realizaciones, el método de amplificación puede incluir una reacción de extensión-unión de cebadores que contiene cebadores dirigidos específicamente al ácido nucleico de interés. Como ejemplo no limitante de cebadores de extensión y unión que pueden diseñarse específicamente para amplificar un ácido nucleico de interés, la amplificación puede incluir cebadores utilizadas para el ensayo Goldengate® (Illumina, Inc., San Diego, CA) o uno o más ensayos establecidos en Pat. de EE.UU. No. 7.582.420 y 7.611.869.

25 Una técnica de amplificación isotérmica se puede utilizar en un método de la presente descripción. Los métodos de amplificación isotérmica ejemplar incluyen, entre otros, la amplificación por desplazamiento múltiple (MDA), como se ilustra, por ejemplo, en Dean et al., Proc. Natl. Acad. Sci. USA 99:5261-66 (2002) o la amplificación del ácido nucleico por desplazamiento de cadena isotérmica como se ilustra, por ejemplo, en la Pat. de EE.UU. No. 6.214.587. Otros métodos no basados en PCR que se pueden utilizar en la presente descripción incluyen, por ejemplo, la amplificación por desplazamiento de cadena (SDA) que se describe, por ejemplo, en Walker et al., Molecular Methods for Virus Detection, Academic Press, Inc., 1995; Pat. de EE.UU. Nos. 5.455.166, y 5.130.238, y Walker et al., Nucl. Acids Res. 20:1691-96 (1992) o la amplificación por desplazamiento de cadena hiper-ramificada que se describe, por ejemplo, en Lage et al., Genome Research 13:294-307 (2003).

35 La descripción adicional de las reacciones de amplificación, las condiciones y los componentes se establecen en la patente de EE.UU. No. 7.670.810. Otras técnicas útiles de amplificación isotérmica incluyen técnicas de amplificación facilitadas por recombinasa, como las vendidas comercialmente como los kits TwistAmp™ por TwistDx (Cambridge, Reino Unido). Los componentes útiles del reactivo de amplificación y las condiciones de reacción facilitadas por recombinasa se establecen en los documentos US 5.223.414 y US 7.399.590. La amplificación dependiente de helicasa también se puede utilizar, por ejemplo, como se describe en Xu et al. EMBO Rep 5:795-800 (2004).

45 En algunas realizaciones, puede ser deseable realizar una etapa de re-siembra. Por ejemplo, los fragmentos de ácidos nucleicos modificados pueden capturarse en localizaciones dentro de una región de una superficie, replicarse en uno o más ciclos de un proceso de amplificación, los fragmentos originales y/o los amplicones de los mismos pueden ser liberados a partir de las localizaciones, los ácidos nucleicos liberados se pueden capturar en otras localizaciones en la misma región, y los ácidos nucleicos nuevamente capturados pueden ser amplificados. En un ejemplo específico, se puede llevar a cabo un solo ciclo de amplificación de puente para un fragmento que se sembró en una superficie y, en lugar de lavar el fragmento de plantilla original en su liberación de la superficie, el fragmento plantilla puede volverse a sembrar la superficie en una nueva localización próxima a la localización donde se había sembrado originalmente. Las sucesivas rondas de amplificación del puente permitirán el crecimiento del grupo tanto en la localización original de la semilla como en la localización de la re-siembra. Usando estos métodos se pueden crear colonias replicadas en una región de una superficie para proporcionar réplicas técnicas. El análisis de las secuencias para las réplicas técnicas puede proporcionar la ventaja de comprobar los errores. Por ejemplo, se pueden identificar como errores de amplificación variantes de secuencia observadas que aparecen en sólo un subconjunto de grupos próximos (que se identifican como réplicas técnicas), mientras que las variantes de secuencia, que se producen en todos los grupos que se identifican como réplicas técnicas para un fragmento en particular, es más probable que sean variantes verdaderas.

Secuenciación de ácidos nucleicos

60 Algunas realizaciones de los métodos descritos en este documento pueden incluir una etapa de secuenciación de fragmentos derivados de un ácido nucleico diana. Un ejemplo es la secuenciación por síntesis (SBS, por sus siglas en inglés). En la SBS, se supervisa la extensión de un cebador de ácido nucleico a lo largo de una plantilla de ácido

nucleico (por ejemplo, un fragmento de un ácido nucleico diana o su amplicón) para determinar la secuencia de nucleótidos en la plantilla. El cebador puede hibridarse en un sitio de cebado que esté presente en un inserto, como se indica anteriormente. El proceso químico subyacente puede ser la polimerización (por ejemplo, como cuando se cataliza mediante una enzima como la polimerasa). En una realización de SBS particular basada en polimerasa, se agregan nucleótidos marcados fluorescentemente a un cebador (con lo que se amplía el cebador) en una forma dependiente de la plantilla, de modo que se puede utilizar la detección del orden y el tipo de nucleótidos agregados al cebador para determinar la secuencia de la plantilla. Pueden ser sometidos varios fragmentos diferentes de ácido nucleico, que han sido adheridos en diferentes localizaciones de una matriz usando las etapas establecidas en este documento, a una técnica SBS en condiciones en las que pueden ser distinguidos los eventos que ocurren para diferentes plantillas a razón de su localización en la matriz.

En algunas realizaciones, las células de flujo proporcionan un formato conveniente para albergar una matriz de fragmentos de ácido nucleico que es producida por los métodos de la presente descripción y que se somete a una SBS u otra técnica de detección que implique la administración repetida de reactivos en ciclos. Como se usa en este documento, la expresión "célula de flujo" incluye una cámara que tiene una superficie a través de la cual puede hacerse circular uno o más reactivos fluidos. Generalmente, una célula de flujo tendrá una abertura de entrada y una abertura de salida para facilitar el flujo de fluido. Ejemplos de células de flujo y sistemas de fluidos relacionados y plataformas de detección que se pueden utilizar fácilmente en los métodos de la presente descripción se describen, por ejemplo, en Bentley et al., *Nature* 456:53-59 (2008), los documentos WO 04/018497; US 7.057.026; WO 91/06678; WO 071123744; US 7.329.492; US 7.211.414; US 7.315.019; US 7.405.281, y US 2008/0108082. En particular, se presenta un gel en la superficie interior de una célula de flujo y el gel proporciona un sustrato al cual se adhiere una o más de las composiciones expuestas en este documento y/o donde se presentan una o más de las etapas del método establecidas en este documento.

En algunas realizaciones, para iniciar un primer ciclo de SBS, se pueden hacer fluir uno o más nucleótidos marcados, la ADN polimerasa, etc., hacia dentro de una célula de flujo que contiene una matriz de fragmentos de ácido nucleico, o a través de ésta. Se pueden detectar los sitios de una matriz donde la extensión de cebado (p. ej. mediante hibridación del cebador a un sitio de cebado ubicado en un inserto unido a un fragmento de ácido nucleico) hacen que se incorpore un nucleótido marcado. Opcionalmente, los nucleótidos pueden incluir además una propiedad de terminación reversible que termine después en la extensión del cebador, una vez que se haya agregado un nucleótido a un cebador. Por ejemplo, se puede agregar un análogo de nucleótido que tiene una fracción reversible del terminador a un cebador tal que la extensión subsecuente no pueda ocurrir hasta que un agente de desbloqueo sea suministrado para retirar la fracción. Por lo tanto, para las realizaciones que utilizan la terminación reversible, se puede administrar un reactivo de desbloqueo a la célula de flujo (antes o después de que ocurra la detección). Pueden llevarse a cabo lavados entre las diferentes etapas de administración. El ciclo se puede repetir "n" veces para extender el cebador por n nucleótidos, detectando así una secuencia de longitud "n". Se describen procedimientos ejemplares de SBS, sistemas fluidos y plataformas de detección que se pueden adaptar fácilmente para su uso con una matriz producida por los métodos de la presente descripción, por ejemplo, en Bentley et al., *Nature* 456:53-59 (2008), los documentos WO 04/018497; US 7.057.026; WO 91/06678; WO 071123744; US 7.329.492; US 7.211.414; US 7.315.019; US 7.405.281, y US 2008/0108082.

En algunas realizaciones, se pueden utilizar otros procedimientos de secuenciación que utilizan reacciones cíclicas, como la pirosecuenciación. La pirosecuenciación detecta la liberación de pirofosfato inorgánico (PPi), ya que los nucleótidos particulares se incorporan en una cadena de ácido nucleico nascente (Ronaghi, et al., *Analytical Biochemistry* 242(1), 84-9 (1996); Ronaghi, *Genome Res.* 11(1),3-11 (2001); Ronaghi et al. *Science* 281(5375), 363 (1998); documentos US 6.210.891; US 6.258.568 y US. 6.274.320). En la pirosecuenciación, el PPi liberado puede ser detectado al ser convertido a trifosfato de adenosina (ATP) por la ATP sulfúrilasa y el nivel de ATP generado se puede detectar por medio de los fotones producidos por luciferasa. Así, la reacción de secuenciación se puede monitorear a través de un sistema de detección de luminiscencia. Las fuentes de radiación de excitación utilizadas para los sistemas de detección basados en fluorescencia no son necesarias para los procedimientos de pirosecuenciación. Se describen sistemas de fluidos, detectores y procedimientos útiles que se pueden utilizar para la aplicación de la pirosecuenciación en los métodos de la presente descripción, por ejemplo, en los documentos WO 2012058096, US 2005/0191698, US 7.595.883, y US 7.244.559. También son útiles reacciones de secuenciación por unión, incluyendo, por ejemplo, las descritas en Shendure et al. *Science* 309:1728-1732 (2005); los documentos US 5.599.675; y US 5.750.341. Algunas realizaciones pueden incluir procedimientos de secuenciación por hibridación como se describe, por ejemplo, en Bains et al., *Journal of Theoretical Biology* 135(3),303-7 (1988); Drmanac et al., *Nature Biotechnology* 16, 54-58 (1998); Fodor et al., *Science* 251(4995), 767-773 (1995); y el documento WO 1989110977.

En algunas realizaciones, en los procedimientos de secuenciación por unión y secuenciación por hibridación, los fragmentos de ácido nucleico diana (o sus amplicones), que están presentes en los sitios de una matriz, se someten a ciclos repetidos de administración y detección de oligonucleótidos. Los sistemas fluidos para los métodos SBS establecidos en este documento o en las referencias citadas en este documento pueden ser fácilmente adaptados para la administración de reactivos para los procedimientos de secuenciación por unión o secuenciación por hibridación. Típicamente, los oligonucleótidos se marcan con fluorescencia y pueden detectarse utilizando detectores de fluorescencia similares a los descritos con respecto a los procedimientos de SBS descritos en este documento o expuestos en las referencias citadas en este documento.

Algunas realizaciones pueden utilizar métodos que implican la supervisión a tiempo real de la actividad de la ADN polimerasa. Por ejemplo, pueden detectarse incorporaciones de nucleótidos a través de interacciones por transferencia de energía de resonancia fluorescente (FRET) entre una polimerasa que lleva un fluoróforo y nucleótidos marcados con γ -fosfato, o con guías de ondas zero-mode (ZMWs). Se describen técnicas y reactivos para una secuenciación basada en FRET, por ejemplo, en Levene et al. *Science* 299, 682-686 (2003); Lundquist et al. *Opt. Lett.* 33, 1026-1028 (2008); y Korlach et al. *Proc. Natl. Acad. Sci. USA* 105, 1176-1181 (2008).

Algunas realizaciones SBS incluyen la detección de un protón liberado en la incorporación de un nucleótido en un producto de extensión. Por ejemplo, la secuenciación basada en la detección de protones liberados puede utilizar un detector eléctrico y técnicas asociadas que están disponibles comercialmente en Ion Torrent (Guilford, CT, una filial de Life Technologies) o métodos y sistemas de secuenciación descritos en los documentos US 2009/10026082 A1; US 2009/10127589 A1; US 2010/10137143; o US 2010/10282617.

En algunas realizaciones, una etapa de secuenciación de los métodos actuales puede incluir una técnica de secuenciación de nanoporos, como las descritas en Deamer & Akeson *Trends Biotechnol.* 18, 147- 151 (2000); Deamer & Branton, *Acc. Chem. Res.* 35:817-825 (2002); y Li et al., *Nat. Mater.* 2:611-615 (2003). En tales realizaciones, el fragmento de ácido nucleico diana pasa a través de un nanoporo. El nanoporo puede ser un poro sintético o una proteína de membrana biológica, tal como una α -hemolisina. A medida que el ácido nucleico diana pasa a través del nanoporo, cada par de bases se puede identificar midiendo las fluctuaciones en la conductancia eléctrica del poro. (documento Pat. de EE.UU. No. 7.001.792; Soni & Meller *Clin. Chem.* 53, 1996-2001 (2007); Healy, *Nanomed.* 2:459- 481 (2007); y Cockroft et al., *J. Am. Chem. Soc.* 130:818-820 (2008)). En algunas realizaciones, la localización de los nanoporos individuales es similar a un sitio o entidad en las matrices ilustradas en este documento. La proximidad de los nanoporos entre sí puede correlacionarse con la proximidad de las secuencias de fragmentos que leen, por ejemplo, para facilitar el ensamblaje de esos fragmentos en la secuencia más grande de la que se derivaron.

En algunas realizaciones, las etapas de secuenciación descritas en este documento pueden llevarse a cabo de forma ventajosa en formatos multiplex, de modo que se manipulan simultáneamente múltiples ácidos nucleicos diana diferentes. En realizaciones particulares, los diferentes ácidos nucleicos diana pueden tratarse en un recipiente de reacción común o en una superficie de un sustrato particular. Esto permite la administración conveniente de los reactivos de secuenciación, la eliminación de reactivos no reaccionados y la detección de eventos de incorporación de manera multiplex. En las realizaciones que utilizan ácidos nucleicos diana unidos a una superficie, o sus fragmentos, los ácidos nucleicos diana, o fragmentos, pueden estar en un formato de matriz. En un formato de matriz, se pueden acoplar típicamente fragmentos de ácidos nucleicos diana a una superficie de una manera espacialmente distinguible, por ejemplo, utilizando técnicas de adhesión establecidas en este documento. La matriz puede incluir una sola copia de un fragmento de ácido nucleico diana en cada sitio (también conocido como una entidad) o pueden estar presentes varias copias con la misma secuencia en cada sitio o entidad. Pueden ser producidas múltiples copias por métodos de amplificación, como la amplificación de puente o PCR de emulsión.

Preparación y secuenciación de ácidos nucleicos

Algunas realizaciones de los métodos proporcionados en este documento incluyen la preparación de una biblioteca de secuenciación de un ácido nucleico diana. Algunas realizaciones también incluyen la secuenciación de la biblioteca preparada. En algunas realizaciones, En algunas realizaciones, una población de los transposomas, comprendiendo cada transposoma una transposasa y un ácido nucleico de transposón, se pone en contacto con un ácido nucleico diana. El contacto puede ocurrir en un sustrato, o sobre es éste, o alternativamente, en

solución. El transposoma puede abarcar la actividad unilateral de transposasa de modo que el ácido nucleico diana sea mellado en una variedad de sitios y los ácidos nucleicos individuales de transposón se unan a las cadenas melladas en un lado de los sitios mellados. En algunas realizaciones, un cebador puede ser hibridado a cada uno de los ácidos nucleicos de transposón adheridos y extenderse para obtener una población de ácidos nucleicos modificados de una sola cadena. En algunas realizaciones, los ácidos nucleicos extendidos pueden ser amplificados. En algunas realizaciones, los ácidos nucleicos extendidos y/o amplificados, a saber, los ácidos nucleicos modificados, se pueden capturar en una superficie para la secuenciación. Algunas realizaciones también incluyen la secuenciación de los ácidos nucleicos capturados.

La FIG. 1 muestra una realización ilustrativa en la que se pone en contacto un ácido nucleico diana con una población de transposomas que comprende un ácido nucleico transposón. El ácido nucleico diana se mella en una pluralidad de sitios, y el ácido nucleico de transposón se une a una cadena de ácido nucleico diana mellado en un lado del sitio de mella. Los cebadores se hibridan a los ácidos nucleicos de transposón adjuntos para proporcionar una población de ácidos nucleicos extendidos. En algunas realizaciones, los ácidos nucleicos extendidos pueden ser amplificados. En algunas realizaciones, los ácidos nucleicos extendidos proporcionan plantillas para una biblioteca de secuenciación.

En algunas realizaciones, el transposoma con actividad de transposasa unilateral comprende una transposasa que tiene una actividad de transposasa unilateral. En algunas realizaciones, el transposoma comprende un ácido nucleico de transposón bloqueado. Se proporcionan en este documento transposomas útiles con métodos y

composiciones de bibliotecas de preparación y secuenciación a partir del ácido nucleico diana. En algunas realizaciones, el ácido nucleico del transposón consiste en un sitio de anclaje, un código de barras, un sitio del cebador de secuenciación, un sitio del cebador de amplificación y/o un marcador indicador. La figura 2 muestra una realización ilustrativa en la que una población de transposomas comprende diferentes ácidos nucleicos de transposón que se ponen en contacto con un ácido nucleico diana y los diferentes ácidos nucleicos de transposón se unen a una hebra del ácido nucleico diana en diferentes sitios de mella. En algunas realizaciones, los diferentes ácidos nucleicos de transposón pueden incluir diferentes sitios de anclaje, códigos de barras, sitios de cebador de secuenciación, sitios de cebador de amplificación y/o marcadores indicadores.

En algunas realizaciones, los ácidos nucleicos extendidos se amplifican. En algunas realizaciones, la amplificación es con cebadores de amplificación de cola. Un cebador de cola puede incluir secuencias del extremo adicionales, de modo que las secuencias adicionales se incluyan en los productos de amplificación. En algunas realizaciones, los cebadores de amplificación pueden incluir un sitio de anclaje, un sitio del cebador de secuencia, un sitio del cebador de amplificación y un marcador indicador. La figura 3 representa una realización ilustrativa en la que un ácido nucleico diana modificado se amplifica mediante amplificación lineal para obtener ciertos productos de amplificación.

La figura 4 representa una realización ilustrativa en la que un transposoma comprende una transposasa dimérica, y un ácido nucleico de transposón comprende dos elementos transposón que comprenden elementos de mosaico (ME, por sus siglas en inglés) en los que uno de los ME se bloquea con un grupo dideoxi en un extremo 3'. En algunas realizaciones, el ácido nucleico transposón comprende un enlazador divisible entre los dos elementos del transposón. El ácido nucleico del transposón puede ser dividido, y el fragmento no-bloqueado del ácido nucleico del transposón se puede adherir a una cadena del ácido nucleico diana mellado en un sitio de mella.

En algunas realizaciones, los ácidos nucleicos modificados se capturan en una superficie. En algunas realizaciones, la superficie comprende una pluralidad de sondas de captura. En algunas realizaciones, las sondas de captura comprenden ácidos nucleicos. En algunas realizaciones, las sondas de captura se hibridan específicamente a los ácidos nucleicos modificados. En algunas realizaciones, las sondas de captura comprenden una fracción de afinidad que se une a una fracción de afinidad de los ácidos nucleicos modificados. En algunas realizaciones, los ácidos nucleicos capturados se amplifican, por ejemplo, por la amplificación del puente. En algunas realizaciones, los ácidos nucleicos de captura se secuencian en la superficie.

Algunas realizaciones de los métodos proporcionados en este documento también incluyen preparar una biblioteca de secuenciación que comprende códigos de barras. Algunas realizaciones también incluyen secuenciación de estas bibliotecas. En algunas realizaciones, los códigos de barras proporcionan puntos de referencia útiles en la alineación de los fragmentos secuenciados del ácido nucleico diana. En algunas realizaciones, los ácidos nucleicos de transposón se insertan en cadenas sencillas del ácido nucleico diana por la transposición y la unión unilateral. El ácido nucleico diana modificado se amplifica y los fragmentos se secuencian. Los fragmentos superpuestos pueden incluir inserciones comunes que son útiles en la alineación de los fragmentos secuenciados para generar una representación de secuencia del ácido nucleico diana. La figura 5 representa una realización ilustrativa en la que una población de los transposomas que comprende diferentes códigos de barras entra en contacto con un ácido nucleico diana; los ácidos nucleicos de transposón se adhieren a un lado de los sitios de mella y el otro extremo no unido de los ácidos nucleicos de transposón se une al otro lado del sitio mella por ligación; y el ácido nucleico diana modificado es amplificado por la amplificación del genoma entero (WGA).

En algunas realizaciones, una población de transposomas con actividad de transposasa unilateral se pone en contacto con un ácido nucleico diana. Los transposomas comprenden ácidos nucleicos de transposón que se insertan en cadenas del ácido nucleico diana. Se describen en este documento transposomas útiles con tales realizaciones. En algunas realizaciones, se insertan ácidos nucleicos de transposón en cadenas monocatenarias y bicatenarias del ácido nucleico diana poniendo en contacto el ácido nucleico diana con los transposomas, de tal manera que el ácido nucleico diana se mella en una variedad de sitios y los ácidos nucleicos de transposón monocatenario se unen a las cadenas melladas en un lado de los sitios mellados, y ligando los ácidos nucleicos de transposón monocatenarios unidos a las cadenas melladas en el otro lado de los sitios de mellado. En algunas realizaciones, la ligasa puede incluir una ligasa que une extremos no homólogos. En algunas realizaciones, la ligasa puede incluir la ligasa IV. En algunas realizaciones, el ácido nucleico modificado se amplifica. En algunas realizaciones, los ácidos nucleicos modificados se capturan en una superficie. En algunas realizaciones, los ácidos nucleicos modificados se secuencian. En algunas realizaciones, las secuencias de los ácidos nucleicos modificados se alinean según la presencia de códigos de barras comunes en secuencias superpuestas. Algunas realizaciones incluyen una biblioteca de secuenciación que incluye códigos de barras preparados por un método proporcionado en este documento.

Obtención de la información del haplotipo

Los ácidos nucleicos diana, tales como el ADN genómico, pueden incluir más de un solo haplotipo. Por ejemplo, el ADN genómico humano, contiene dos conjuntos de moléculas de ADN, cada una con una combinación diferente de secuencias maternas y paternas. Algunas de las realizaciones proporcionadas en este documento son útiles para obtener información de la secuencia a partir de los fragmentos de una sola molécula de ácido nucleico o de sus copias.

En algunas realizaciones, se mantiene la proximidad física de ciertos fragmentos en el sustrato. En algunas realizaciones, las secuencias de fragmentos que tienen una mayor cercanía entre sí en la secuencia del ácido nucleico diana lineal tienen una proximidad física más cercana entre sí en la superficie en comparación con las secuencias de fragmentos que son menos próximos entre sí en la secuencia del ácido nucleico diana lineal. La proximidad física de ciertos fragmentos puede ser retenida por una variedad de métodos.

En algunas realizaciones, la transposición unilateral no fragmenta un ácido nucleico diana. En algunas realizaciones, se puede poner en contacto un ácido nucleico diana con los transposomas que tienen una actividad de transposasa unilateral para obtener un ácido nucleico modificado. En algunas realizaciones, el ácido nucleico modificado puede ser puesto en contacto con una superficie. En algunas realizaciones, los ácidos nucleicos de transposón incluyen marcadores de anclaje tal que las secuencias modificadas se pueden capturar en una superficie que comprende las sondas de captura. En algunas realizaciones, el ácido nucleico modificado puede ser fragmentado mientras se pone en contacto con la superficie. En algunas realizaciones, el ácido nucleico modificado puede ser fragmentado en un lugar próximo a la superficie. En algunas realizaciones, el ácido nucleico modificado se puede secuenciar en la superficie.

En algunas realizaciones, los métodos para obtener la información del haplotipo incluyen la comparación de secuencias complementarias determinadas para localizaciones próximas en la superficie para identificar errores de secuencia. En algunas realizaciones, la proximidad relativa de cualquiera de las dos especies de fragmentos en la superficie puede proporcionar información útil para la alineación de la información de secuencia obtenida a partir de los dos fragmentos. Específicamente, la distancia entre los grupos en la superficie, derivada de cualesquiera dos fragmentos dados, se puede correlacionar positivamente con la probabilidad de que los dos grupos sean de la misma molécula de polinucleótido diana, según lo descrito más detalladamente en el documento WO 2012/025250.

Como ejemplo, en algunas realizaciones, los fragmentos derivados de una molécula de ácido nucleico larga capturada en la superficie de una célula de flujo aparecen en una línea a través de la superficie de la célula de flujo (p. ej., si el ácido nucleico fuera estirado antes de la fragmentación o de la amplificación) o en una nube en la superficie. Además, se puede generar entonces un mapa físico del ácido nucleico inmovilizado. El mapa físico correlaciona así la relación física de grupos después de que el ácido nucleico inmovilizado se amplifique. En concreto, el mapa físico se utiliza para calcular la probabilidad de que los datos de secuencia obtenidos de cualquiera de los dos grupos estén unidos, como se describe en el documento WO 2012/025250.

En algunas realizaciones, el mapa físico se genera por imágenes de la superficie para establecer la localización de las moléculas de ácido nucleico inmovilizadas a través de la superficie. En algunas realizaciones, el ácido nucleico inmovilizado se visualiza agregando un agente de visualización al soporte sólido y detectando una señal del agente de visualización. En algunas realizaciones, el agente de visualización es un marcador detectable. Los marcadores detectables convenientes, incluyen, entre otros, protones, haptenos, radionucleidos, enzimas, marcadores fluorescentes, marcadores quimioluminiscentes y/o agentes cromogénicos. Por ejemplo, en algunas realizaciones, el agente de visualización es un colorante intercalante o un agente de unión de ADN no intercalante. Puede ser utilizado cualquier tinte de intercalación adecuado o

agente de unión de ADN no intercalante, como se conocen en la técnica, incluyendo, entre otros, los establecidos en el documento U.S. 2012/0282617.

En ciertas realizaciones, una variedad de moléculas de ácido nucleico modificadas se hace fluir sobre una célula de flujo que comprende una variedad de nanocanales. Como se usa en este documento, el término nanocanal se refiere a un canal estrecho en el que se estira una molécula de ácido nucleico lineal largo. En algunas realizaciones, el número de cadenas es, o no es más que, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900 ó 1000 cadenas largas individuales de ácido nucleico, o un intervalo definido por dos de los valores precedentes, que se extienden a través de cada nanocanal. En algunas realizaciones, los nano-canales individuales están separados por una barrera física que impide que las cadenas largas individuales del ácido nucleico diana interactúen con múltiples nanocanales. En algunas realizaciones, el soporte sólido comprende, o comprende al menos, 10, 50, 100, 200, 500, 1000, 3000, 5000, 10000, 30000, 50000, 80000 ó 100000 nano-canales, o un intervalo definido por cualesquiera dos de los valores precedentes.

En algunas realizaciones, los ácidos nucleicos modificados se dividen una vez que los ácidos nucleicos se han estirado a lo largo del canal. Los fragmentos resultantes se pueden amplificar opcionalmente para formar grupos a lo largo de la superficie del canal. El mapeo de contigüidad se puede realizar, por ejemplo, siguiendo los grupos por la longitud de uno de estos canales. Como ejemplo, puede ser utilizada una celda de flujo con 1000 o más nanocanales con productos de fragmentación inmovilizados asignados en los nano-canales para secuenciar el genoma de un organismo con lecturas cortas 'posicionadas'. En algunas realizaciones, los productos de fragmentación inmovilizados en los nano-canales se pueden utilizar para resolver haplotipos. En algunas realizaciones, los productos de fragmentación inmovilizados asignados en los nano-canales se pueden utilizar para resolver problemas de fase.

En algunas realizaciones, la transposición unilateral se utiliza para insertar ADN artificial en ADNg. En un ejemplo,

se inserta ADN artificial en regiones de repetición de un ADN genómico (u otro ácido nucleico) para preparar regiones de repetición únicas. Las regiones de repetición pueden analizarse, por ejemplo, mediante técnicas de secuenciación, como las que se establecen anteriormente, para contar el número de repeticiones o para orientar otras secuencias en el ADN genómico en relación con las regiones de repetición. En otro ejemplo, el ADN artificial que se inserta por una transposición unilateral hace diferente a la cadena superior y a la inferior de un ácido nucleico bicatenario. Por lo tanto, el producto del método de inserción puede analizarse, por ejemplo, mediante técnicas de secuenciación, como las que se establecen anteriormente, para discriminar una cadena de la otra. Esto puede permitir aún más el ensamblaje independiente de la cadena superior e inferior en una reconstrucción de un ADN genómico (u otro ácido nucleico bicatenario).

10 Ejemplos

Ejemplo 1- Ácidos nucleicos de transposón bloqueados

Se trató ADN diana sin los transposomas (Amplicon), o con transposomas que comprendían transposasa y (1) ácido nucleico transposón bloqueado con un grupo de biotina de 3' (Bio 3'); (2) ácido nucleico de transposón bloqueado con un grupo espaciador 3' (espaciador 3'); o (3) ácido nucleico de transposón no bloqueado (TDE1). La figura 6 muestra los resultados en los que no se produce la transposición con los transposomas compuestos de ácidos nucleicos de transposón bloqueados (Bio 3', y Espaciador 3').

Ejemplo 2-Modelo de inserción y ensamblaje de puntos de referencia

Se insertaron en un ADN diana puntos de referencia que comprendían secuencias aleatorias de 12 bp. El ADN fue secuenciado y las secuencias de fragmentos fueron ensambladas *de novo*. La FIG. 7 muestra gráficos de número de coberturas nominales, y la media de la longitud de lectura sintética para lecturas de 500 bp con una frecuencia de inserción de 100 bp, y para lecturas de 300 bp con una frecuencia de inserción de 50 bp. Se demostró que podían ser ensamblados 6-7 kb *de novo* con una cobertura 50 X.

Ejemplo 3-Transposición unilateral con y sin glicerol

Se incubó ADN diana (pUC19) con transposomas y los productos de ADN se separaron en un 1,2% de gel. Diez muestras fueron puestas en práctica: las primeras 5 muestras no incluían ningún glicerol y las segundas 5 muestras incluían glicerol al 65%. Cada conjunto de 5 muestras se estableció como una titulación de diferentes concentraciones del transposoma. El transposoma consistió en transposasa y ácido nucleico de transposón no-bloqueado (TDE1). Se muestra una fotografía del gel teñido en la FIG. 9. El gel también se cargó con un control para el pUC19 sin cortar (pUC19 solamente), pUC19 linealizado (EcorI) y pUC19 monocatenario mellado (Nb. Bsr DI). Como se muestra en los carriles del gel cargado con las muestras "sin glicerol", el aumento de la concentración de TDE dio como resultado un aumento de los productos de transposición unilateral (es decir, mellados) y de transposición bilateral (es decir, lineales). En comparación, las reacciones que se realizaron en presencia de glicerol al 65% mostraron una mayor cantidad del producto de transposición unilateral como aumento de TDE1, pero hubo poco o ningún aumento del producto de transposición bilateral en presencia de la concentración creciente de TDE1.

35 Ejemplo 4-Las alteraciones en la longitud de los ácidos nucleicos de transposón de la cadena transferida inhiben la transposición

Este ejemplo demuestra que los cambios en la longitud de la cadena transferida de un transposón por la substracción de un nucleótido (n-1) o la adición de un nucleótido (n + 1) reducen la eficacia de la transposición.

Fueron formados transposomas con 3' n-1 y transposón n+1 METS y fueron hibridados con 0, 1%, 5%, 50%, 90%, 99%, o 100% de TDE1 durante la noche a temperatura ambiente. Los transposomas resultantes entonces reaccionaron durante la noche, a temperatura ambiente, con 1 kb de amplicón, seguido por un tratamiento con SDS, y luego la separación en un gel TBE. La figura 10 muestra geles TBE cargados con los productos de reacción junto con una escalera de peso molecular y una muestra de control sin enzima transposasa. Sorprendentemente, incluso con la incubación durante la noche, la mayoría del ADN diana todavía estaba presente en cada muestra, indicando que los transposones n-1 y n+1 tenían un efecto inhibitorio sobre la transposición. Además, el efecto inhibitorio se correlacionó con el aumento del porcentaje de los transposones n-1 y n+1.

La expresión "que comprende", como se usa en este documento, es sinónima de "que incluye," "que contiene", o "caracterizada por" y es inclusiva o abierta y no excluye elementos adicionales, no mencionados o etapas del método.

La descripción anterior describe varios métodos y materiales de la presente invención. Esta invención es susceptible a modificaciones en los métodos y materiales, así como a alteraciones en los métodos y equipos de fabricación. Tales modificaciones se harán evidentes para los expertos en la técnica a partir de una consideración de esta descripción o la práctica de la invención descrita en este documento. En consecuencia, no se pretende que esta invención se limite a las realizaciones específicas descritas en este documento, sino que cubra todas las modificaciones y alternativas que se encuentran dentro del alcance de la invención.

En la medida en que las publicaciones y las patentes o las solicitudes de patente mencionadas en este documento contradigan la descripción contenida en la memoria descriptiva, la memoria descriptiva tiene por objeto sustituir y/o prevalecer sobre cualquier material contradictorio.

REIVINDICACIONES

1. Un método de preparación de una biblioteca de secuenciación a partir de un ácido nucleico diana bicatenario que comprende:

5 (a) proporcionar una variedad de transposomas, comprendiendo cada monómero de transposoma una transposasa y un ácido nucleico de transposón, en el que el transposoma se configura para mellar solamente una cadena del ácido nucleico diana bicatenario; y

10 (b) poner en contacto el ácido nucleico diana con los transposomas de tal manera que el ácido nucleico diana se mella en una variedad de sitios del ácido nucleico diana y los ácidos nucleicos de un solo transposón se unen, al menos, a uno de los ácidos nucleicos diana mellados para generar ácidos nucleicos transpuestos, obteniéndose así una biblioteca de ácidos nucleicos modificados para la secuenciación.

2. El método según la reivindicación 1, en donde dicho método es un método de preparación de una biblioteca de secuenciación con códigos de barras; en donde en la etapa (a) cada transposoma comprende una transposasa y un ácido nucleico de transposón que comprende una secuencia de reconocimiento; y

15 en donde la etapa (b) comprende: (b) insertar los ácidos nucleicos del transposón en hebras del ácido nucleico diana, que comprenden:

(i) poner en contacto el ácido nucleico diana con los transposomas de tal manera que el ácido nucleico diana se mella en una variedad de sitios y los ácidos nucleicos del transposón individual se unen a las cadenas melladas en un lado de los sitios mellados para generar los ácidos nucleicos transpuestos, y

20 (ii) unir los ácidos nucleicos del transposón individual unidos a las cadenas melladas en el otro lado de los sitios mellados, obteniéndose así una biblioteca de ácidos nucleicos modificados para la secuenciación.

3. Un método para capturar información de contigüidad de un ADN diana que comprende:

(a) proporcionar una variedad de transposomas, comprendiendo cada monómero de transposoma una transposasa y un ácido nucleico de transposón, en el que el transposoma se configura para mellar solamente una cadena del ácido nucleico diana bicatenario;

25 (b) poner en contacto el ADN diana con los transposomas de tal manera que el ADN diana es mellado en una pluralidad de sitios del ácido nucleico diana;

(c) añadir o insertar una o más secuencias de reconocimiento a la secuencia de ADN diana para generar un ADN diana tratado;

(d) secuenciar el ADN diana tratado; y

30 (e) capturar información de contigüidad identificando las secuencias de ADN diana o las secuencias de reconocimiento que tienen una propiedad compartida.

4. Un método de captura de la información de contigüidad de un ADN diana que comprende:

35 (a) proporcionar una variedad de transposomas, comprendiendo cada monómero de transposoma una transposasa y un ácido nucleico de transposón que comprende una secuencia de reconocimiento, en donde el transposoma se configura para mellar solamente una cadena del ácido nucleico diana bicatenario;

(b) insertar los ácidos nucleicos del transposón en hebras del ácido nucleico diana, que comprenden:

(i) poner en contacto el ácido nucleico diana con los transposomas, de tal manera que el ácido nucleico diana se mella en una variedad de sitios y los ácidos nucleicos del transposón individual se unen a las cadenas melladas en un lado de los sitios mellados, y

40 (ii) unir los ácidos nucleicos del transposón individual unidos a las cadenas melladas en el otro lado de los sitios mellados, obteniéndose así un ácido nucleico modificado;

(c) amplificar el ácido nucleico modificado, obteniéndose así una pluralidad de ácidos nucleicos que comprenden secuencias de reconocimiento insertadas;

(d) secuenciar el ADN diana tratado; y

45 (e) capturar información de contigüidad identificando las secuencias de ADN diana o las secuencias de reconocimiento que tienen una propiedad compartida.

5. El método de reivindicación 3 ó 4, en el que la secuencia de reconocimiento es un código de barras, y donde el código de barras de, al menos, un ácido nucleico de transposón es diferente; en el que la secuencia de

reconocimiento es un código de barras, y en el que los códigos de barras de los ácidos nucleicos de transposón no son iguales.

6. El método de cualquiera de las reivindicaciones precedentes, comprendiendo además la captura de los ácidos nucleicos modificados en una superficie.

5 **7.** El método de la reivindicación 6, caracterizado porque la superficie comprende una variedad de sondas de captura; opcionalmente en donde

(i) las sondas de captura comprenden ácidos nucleicos y el método comprende la hibridación de los ácidos nucleicos modificados con las sondas de captura; o

10 (ii) los ácidos nucleicos modificados y las sondas de captura comprenden una fracción de afinidad, por ejemplo, biotina, avidina o estreptavidina; o

(iii) opcionalmente, en donde la superficie comprende al menos alrededor de 100.000 ácidos nucleicos capturados por mm².

15 **8.** El método de cualquiera de las reivindicaciones precedentes, en la que los transposomas, que se ponen en contacto con los ácidos nucleicos diana en (b), se unen a una superficie, capturando así los ácidos nucleicos modificados en la superficie.

9. El método de cualquiera de las reivindicaciones 6 a 8, que comprende la secuenciación de los ácidos nucleicos capturados en la superficie.

20 **10.** El método de cualquiera de las reivindicaciones 6 a 9, caracterizado porque la proximidad de la información de la secuencia obtenida a partir de dos ácidos nucleicos capturados en una representación lineal de la secuencia de ácidos nucleicos diana es indicativa de la proximidad de los ácidos nucleicos capturados en la superficie.

11. El método de la reivindicación 10, en el que los ácidos nucleicos capturados con mayor proximidad entre sí en la superficie comprenden secuencias con mayor proximidad en la representación de la secuencia de ácido nucleico diana en comparación con los ácidos nucleicos capturados con menos proximidad.

25 **12.** El método de cualquiera de las reivindicaciones 10 y 11, caracterizado porque la representación de la secuencia de ácidos nucleicos diana comprende una representación de haplotipo.

30 **13.** El método de cualquiera de las reivindicaciones 1-12, en donde el transposoma se compone de una actividad de transposasa unilateral; opcionalmente (i) en el que el transposoma comprende una subunidad monomérica que carece de actividad de transposasa; o (ii) en el que el transposoma comprende subunidades monoméricas unidas covalentemente; o (iii) en donde la transposasa carece de capacidad para formar dímeros; y/o, (iv) en donde la estructura cuaternaria de la transposasa es monomérica.

14. El método de cualquiera de las reivindicaciones 1-13, en donde la transposasa se selecciona del grupo que consiste en Mu, Mu E392Q, Tn5, Tn5 hiperactiva, variantes Tn5, Vibhar, RAG y Tn552.

35 **15.** El método de cualquiera de las reivindicaciones 1-14, en el que uno o más ácidos nucleicos de transposón son no funcionales; opcionalmente (i) en el que se selecciona el extremo 3' del ácido nucleico transposón no funcional del grupo compuesto por un grupo didesoxi, un grupo espaciador, un grupo amina, un grupo alquilo, un grupo arilo, un grupo fosfato, un grupo tiol, un nucleótido inverso, un grupo azido, un grupo sulfato, y un grupo biotina; o (ii) en el que la variedad de los transposomas se preparan poniendo en contacto las transposasas con ácidos nucleicos de transposón funcionales y ácidos nucleicos de transposón no funcionales; o (iii) en el que la proporción de ácidos nucleicos de transposón que comprenden ácidos nucleicos de transposón no funcionales frente a ácidos nucleicos de transposón funcionales es mayor o igual a 1:1.

40 **16.** El método de reivindicación 3, que comprende:

(i) proporcionar una ADN polimerasa al ácido nucleico diana mellado;

(ii) extender el extremo 3' del ácido nucleico diana utilizando la cadena complementaria como plantilla;

(iii) opcionalmente amplificar los ácidos nucleicos extendidos; y

45 opcionalmente, en el que la amplificación de los ácidos nucleicos extendidos es con cebadores de amplificaciones de cola que comprenden una secuencia seleccionada del grupo que consiste en un sitio de anclaje, un sitio de cebador de secuenciación, un sitio de cebador de amplificación, un código de barras y un marcador indicador.

50 **17.** El método de cualquiera de las reivindicaciones 6 a 16, que comprende amplificar los ácidos nucleicos capturados; opcionalmente, en el que la amplificación de los ácidos nucleicos capturados comprende la amplificación del puente.

- 18.** El método de cualquiera de las reivindicaciones 6 a 17, en donde la superficie comprende una variedad de sondas de captura;
- opcionalmente, en donde las sondas de captura comprenden ácidos nucleicos.
- 5 **19.** El método de cualquiera de las reivindicaciones precedentes, en el que el ácido nucleico transposón comprende una secuencia seleccionada del grupo consistente en un sitio de anclaje, un código de barras, un sitio de cebador de secuenciación, un sitio de cebador de amplificación y un marcador indicador.
- 20.** El método de cualquiera de las reivindicaciones 1-19, en donde al menos un transposoma comprende dos ácidos nucleicos de transposón;
- opcionalmente, en donde los dos ácidos nucleicos de transposón tienen diferentes secuencias.
- 10 **21.** El método de cualquiera de las reivindicaciones 1-20, en donde la variedad de los transposomas comprende al menos dos ácidos nucleicos de transposón diferentes.
- 22.** El método de cualquiera de las reivindicaciones 1-21, en donde la variedad de los transposomas comprende al menos una transposasa que tiene la capacidad de formar un transposoma, pero carece de la capacidad de transponer.
- 15 **23.** El método de cualquiera de las reivindicaciones 1-22, en donde el ácido nucleico diana se selecciona del grupo que consiste en ADN genómico, fragmentos de ADN genómico, y ADNc.
- 24.** El método de cualquiera de las reivindicaciones 1-23, que comprende además la codificación de barras combinatoria, en el que el transposón comprende el primer conjunto de códigos de barra;
- 20 el primer conjunto de códigos de barras se introduce en el ácido nucleico diana durante la transposición para generar ácidos nucleicos diana transpuestos que comprenden el primer conjunto de códigos de barras;
- los ácidos nucleicos diana transpuestos se agrupan para generar un primer grupo de ácidos nucleicos diana transpuesto;
- un segundo conjunto de códigos de barras se introduce en el primer grupo de ácidos nucleicos diana transpuesto para generar ácidos nucleicos diana que comprenden los primeros y segundos conjuntos de códigos de barras;
- 25 ácidos nucleicos diana que comprenden los primeros y segundos conjuntos de códigos de barras; se agrupan para generar un segundo grupo de ácidos nucleicos diana transpuesto;
- opcionalmente repitiendo las etapas de introducción de códigos de barras adicionales y la agrupación para generar una biblioteca de ácidos nucleicos diana de código de barras.

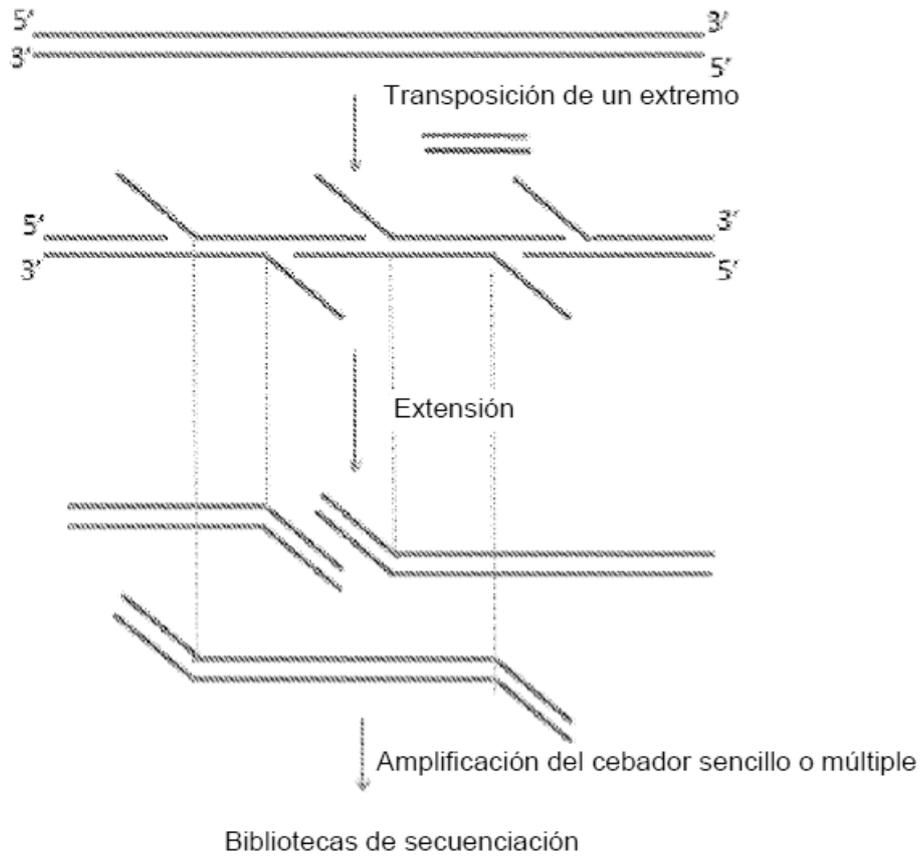


FIG.1

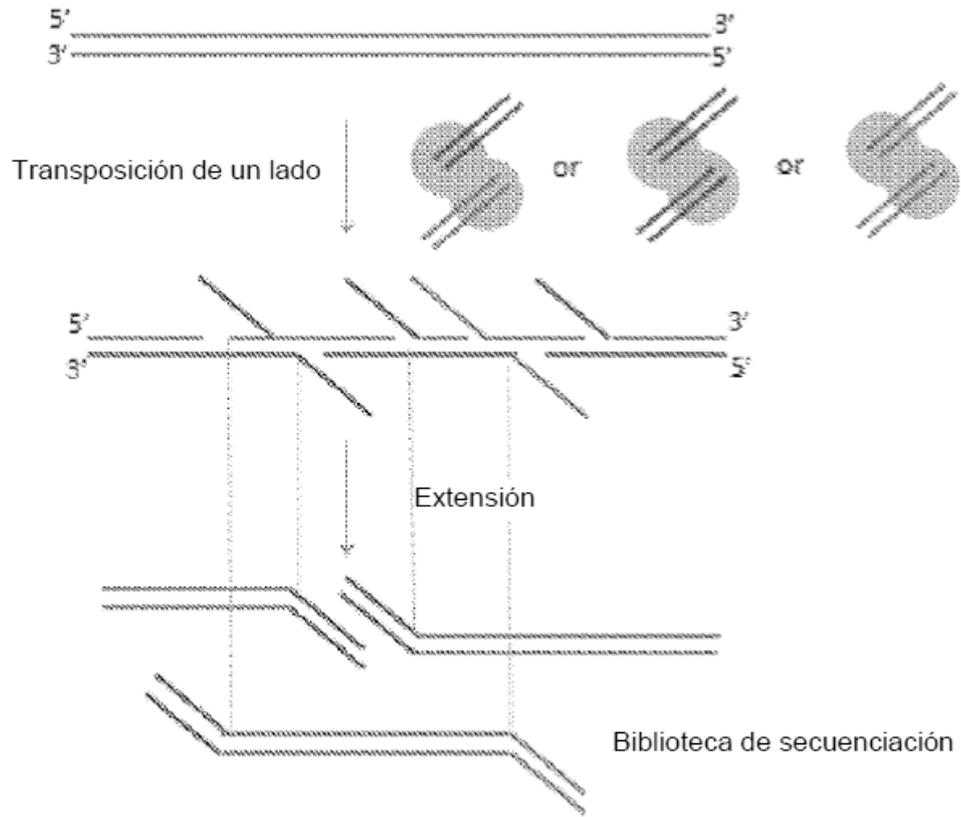


FIG.2

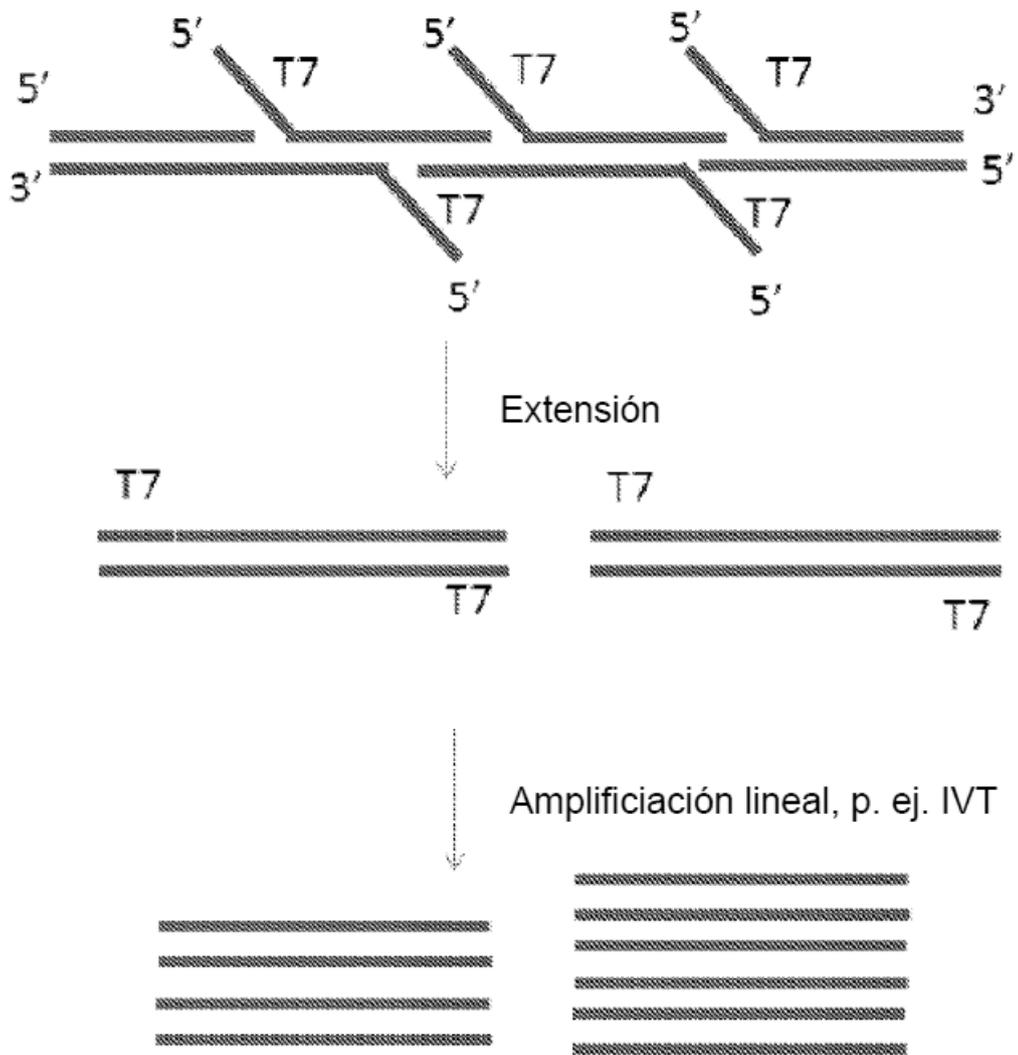


FIG.3

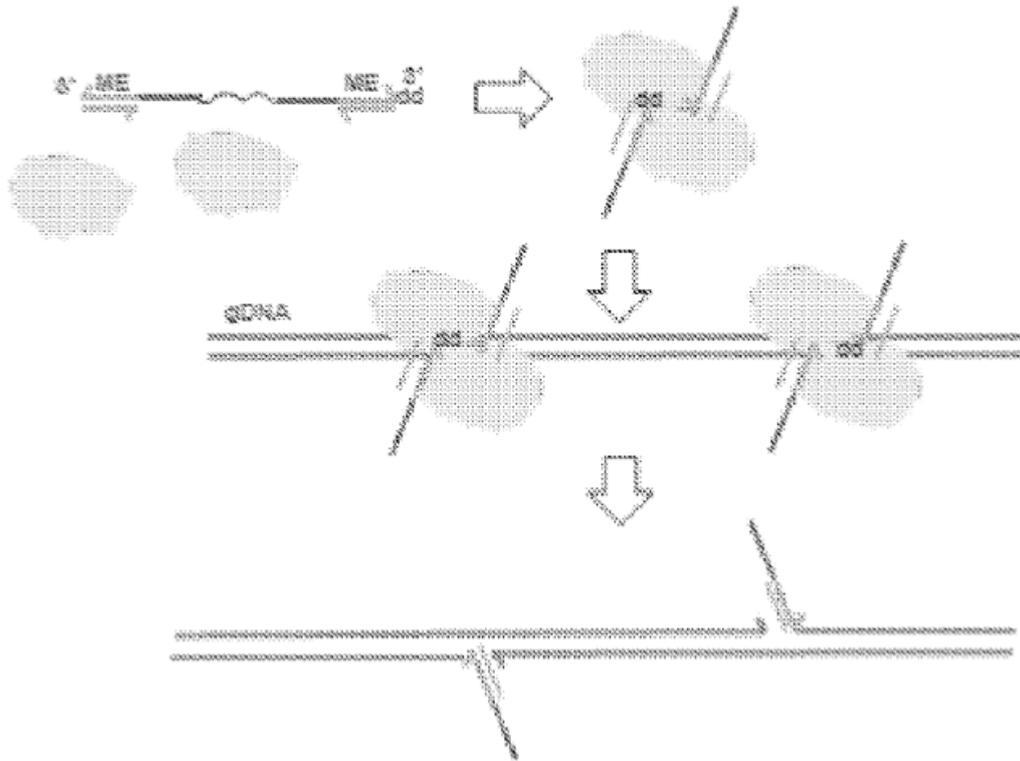


FIG.4

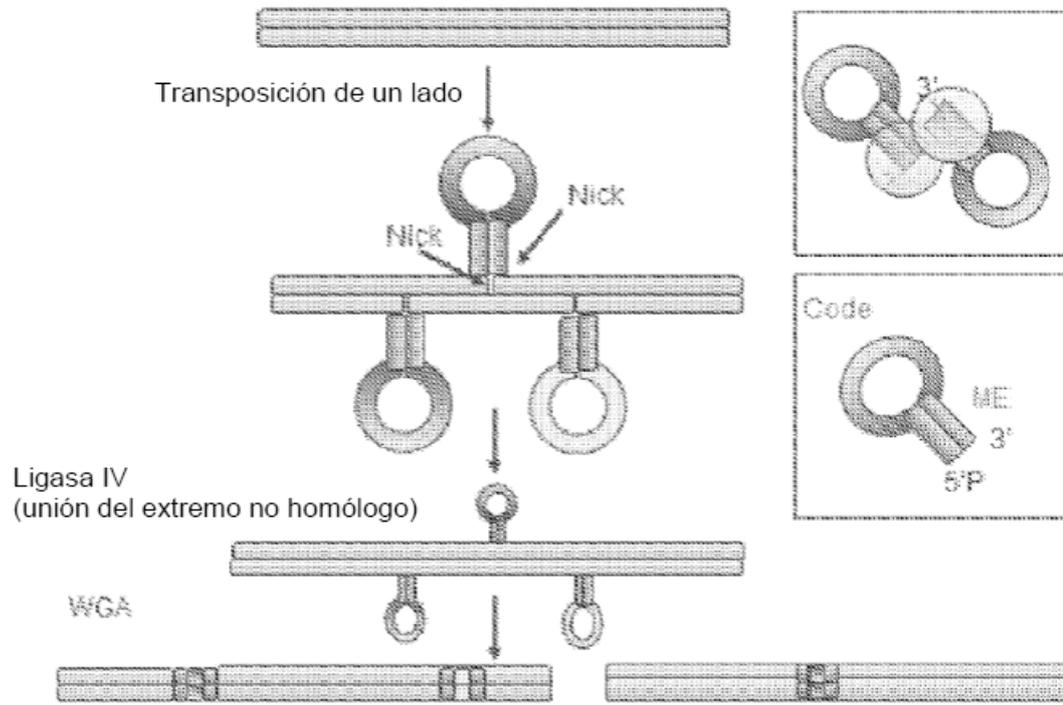


FIG.5

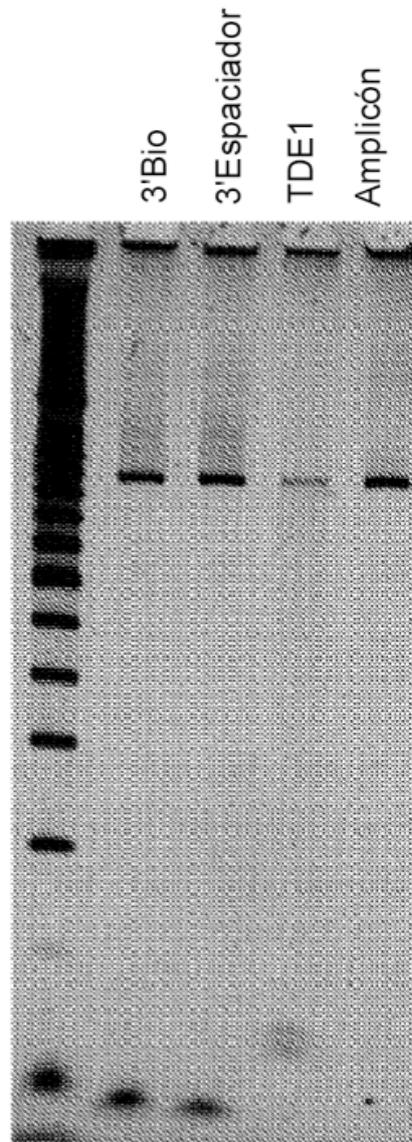


FIG.6

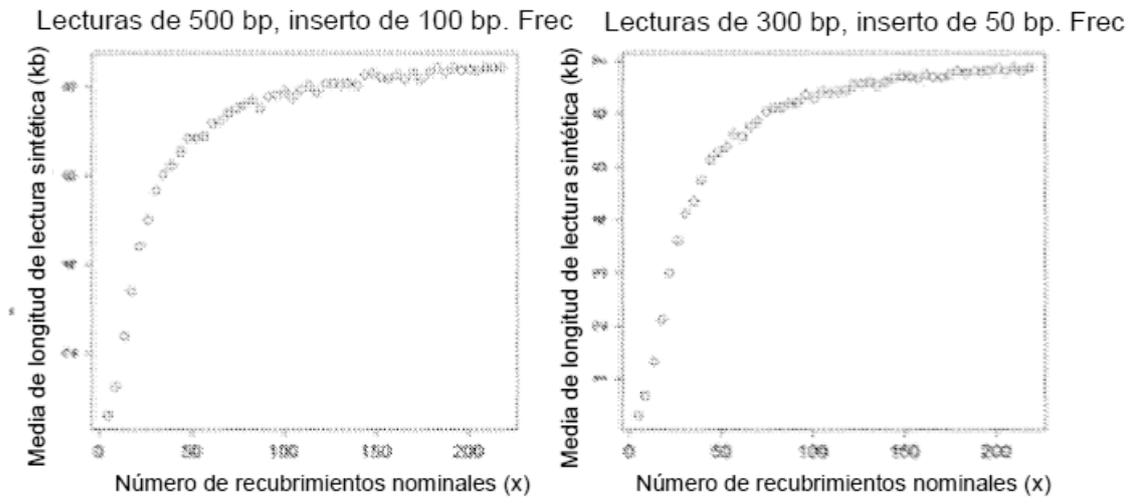


FIG. 7

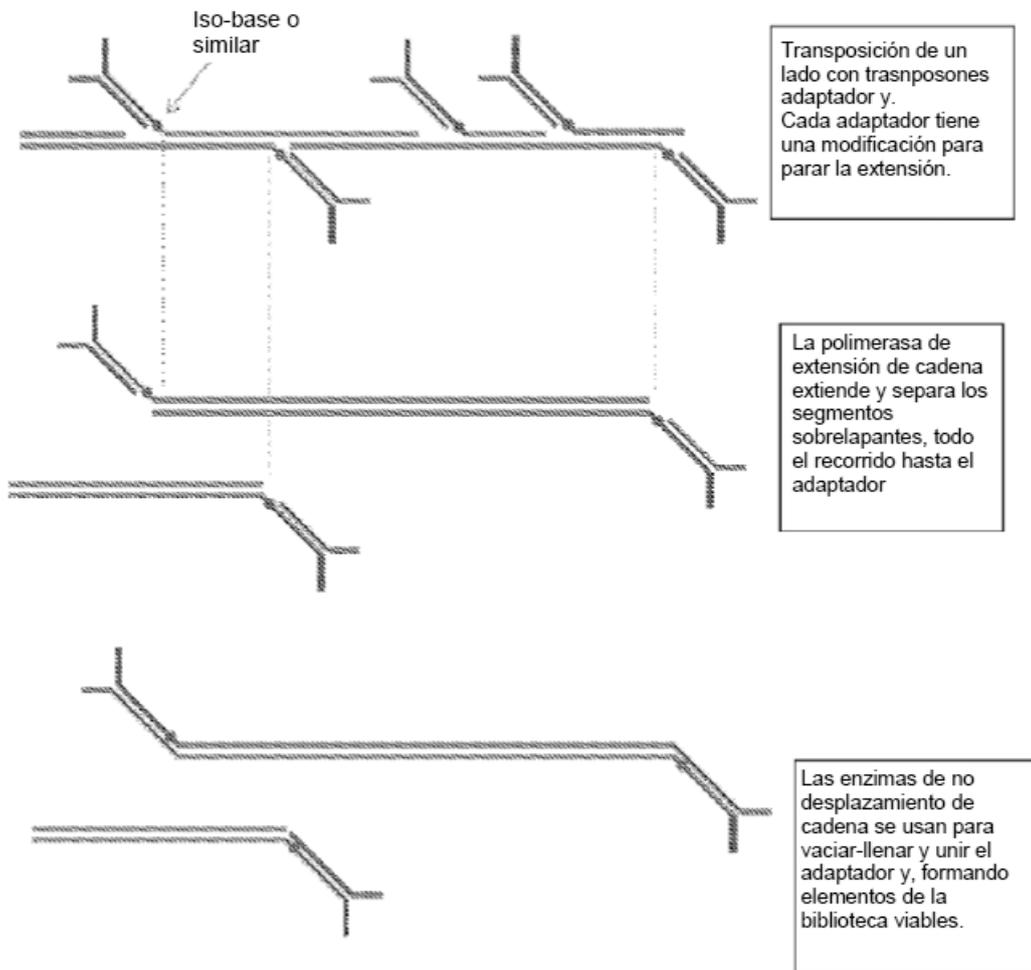


FIG. 8A

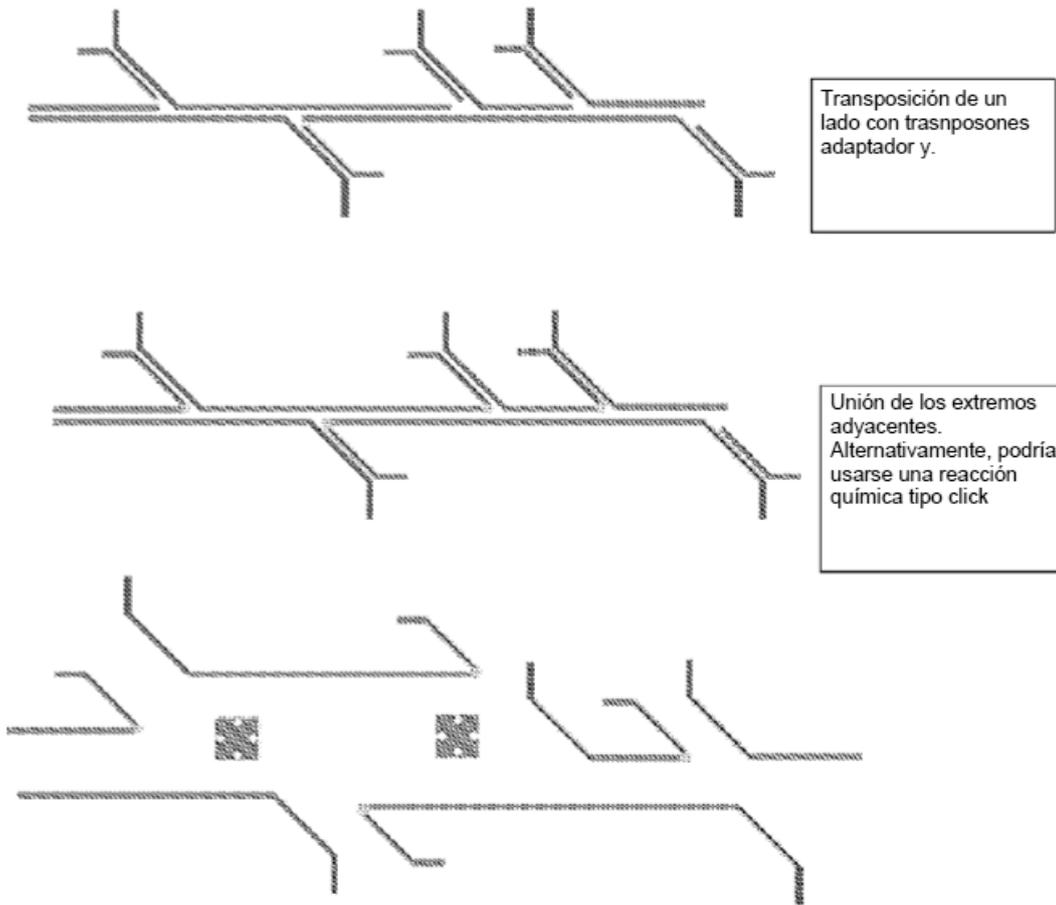


FIG. 8B

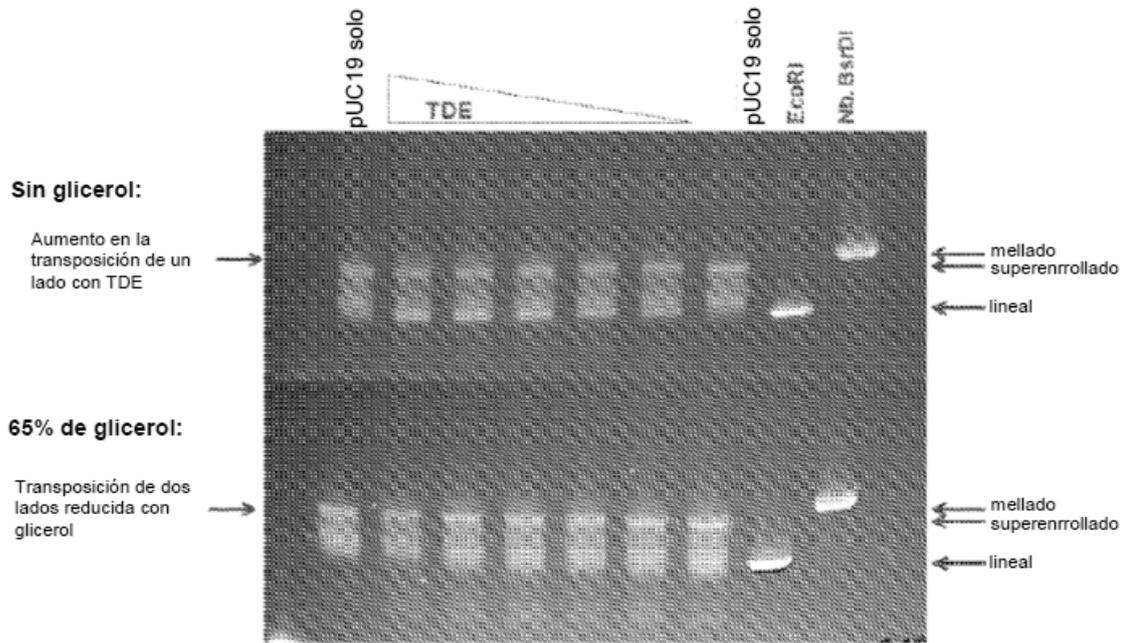


FIG. 9

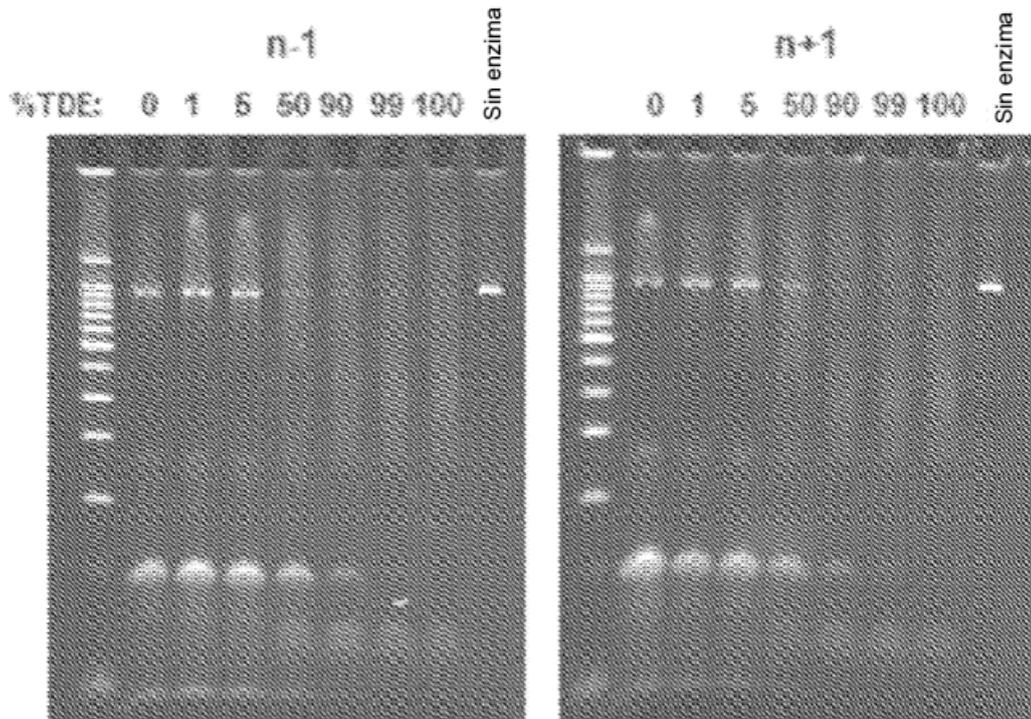


FIG. 10

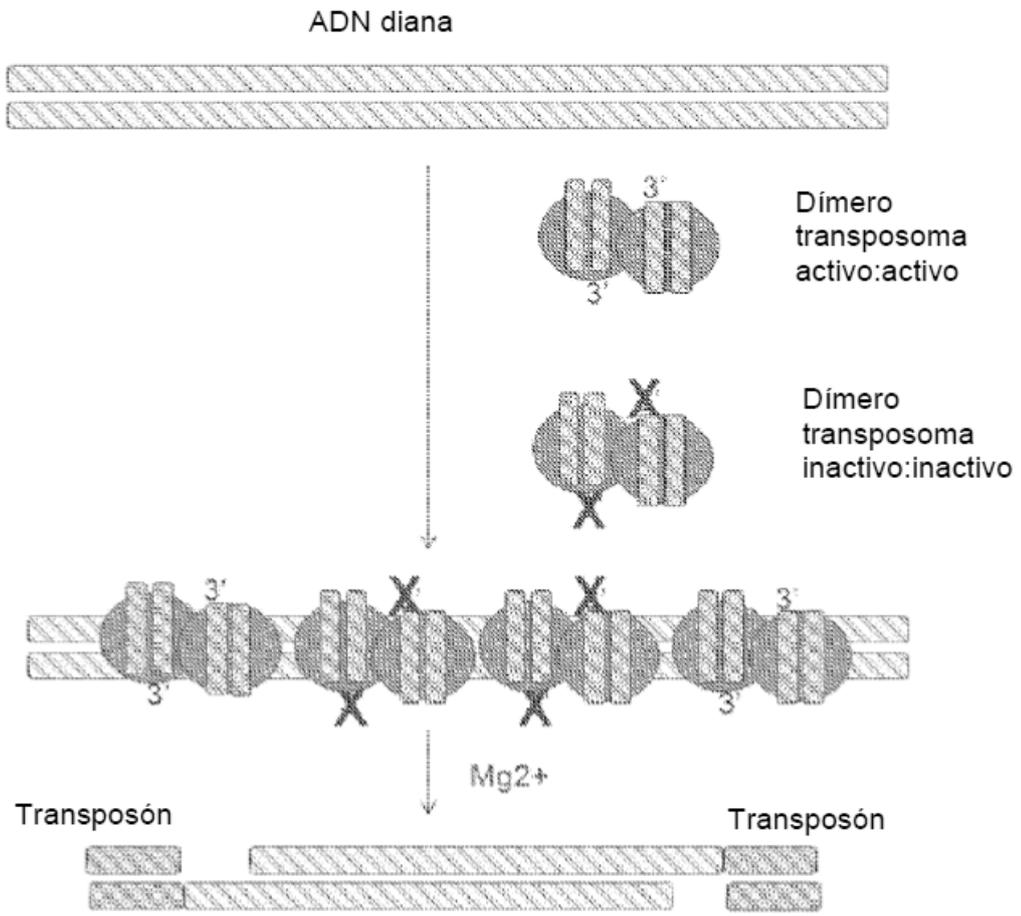


FIG. 11

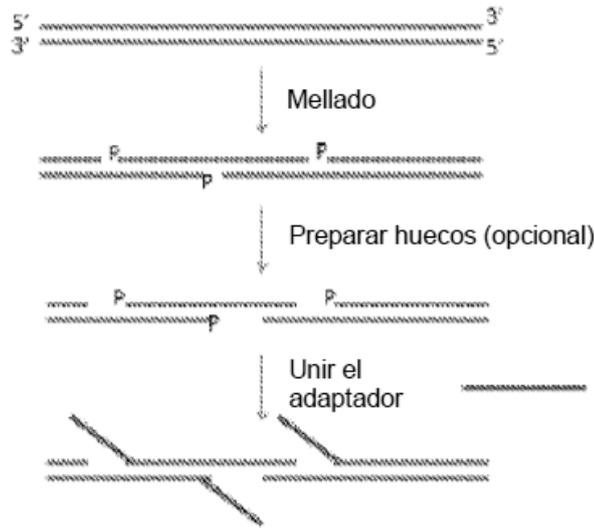


FIG. 12

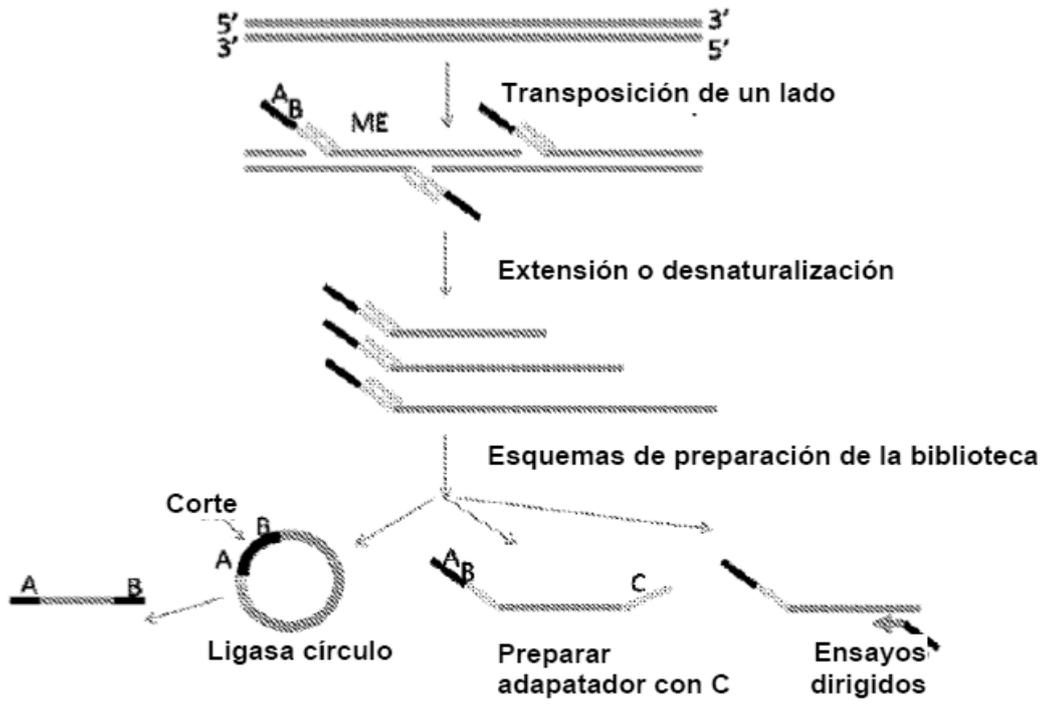


FIG. 13

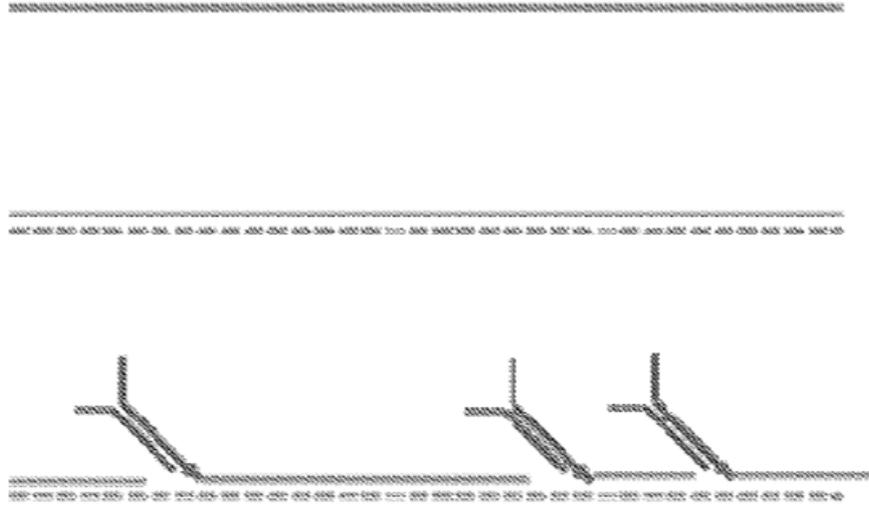


FIG. 14

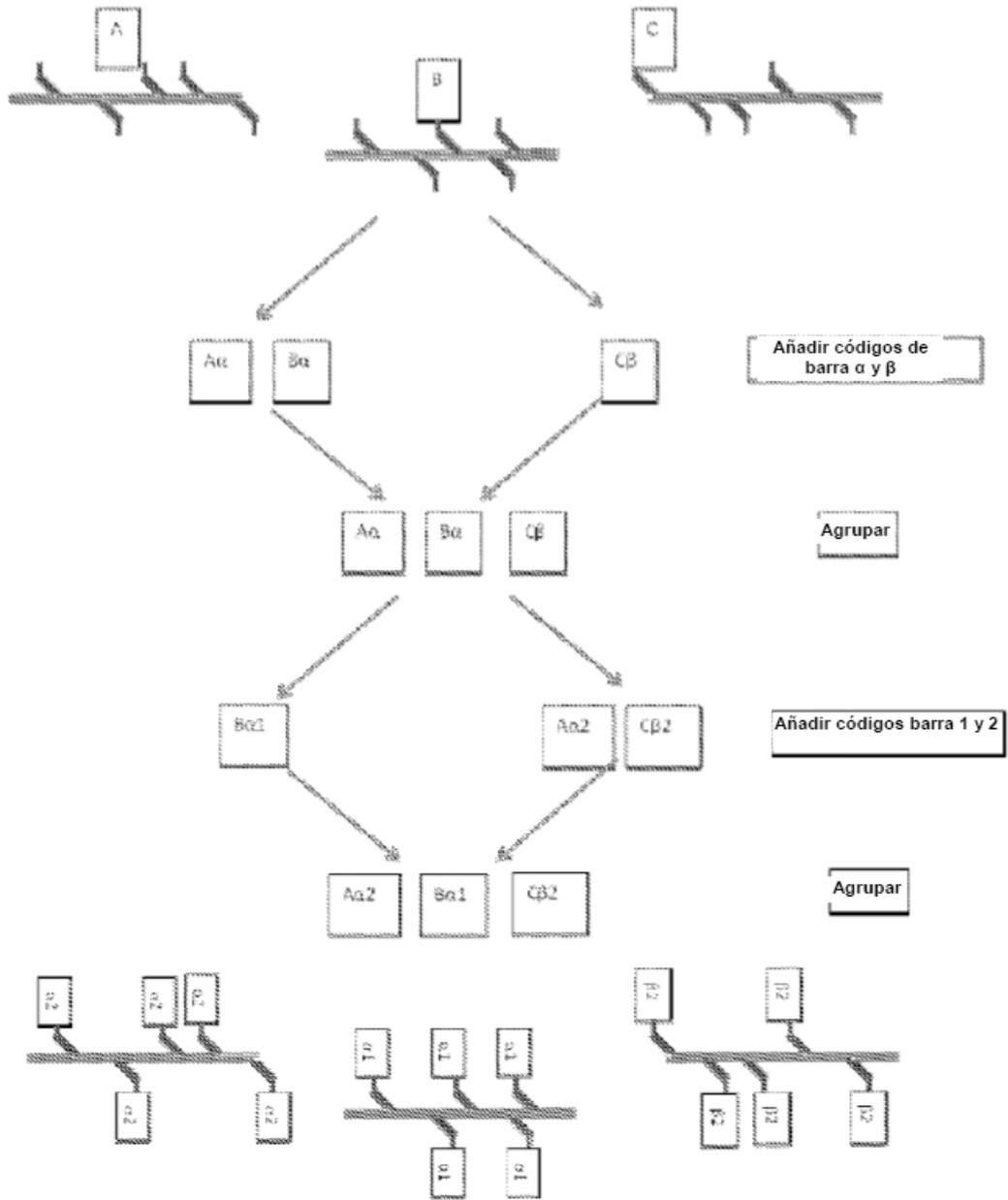


FIG. 15