

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 812 261**

51 Int. Cl.:

C12Q 1/6855 (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **31.07.2017 PCT/EP2017/069308**

87 Fecha y número de publicación internacional: **08.02.2018 WO18024671**

96 Fecha de presentación y número de la solicitud europea: **31.07.2017 E 17752044 (2)**

97 Fecha y número de publicación de la concesión europea: **17.06.2020 EP 3491147**

54 Título: **Procedimientos para la retirada de dímeros de adaptadores de preparaciones para la secuenciación de ácidos nucleicos**

30 Prioridad:

01.08.2016 US 201662369352 P

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

16.03.2021

73 Titular/es:

**F. HOFFMANN-LA ROCHE AG (100.0%)
Grenzacherstrasse 124
4070 Basel, CH**

72 Inventor/es:

**DAVIS, RANDALL W. y
BIBILLO, ARKADIUSZ**

74 Agente/Representante:

LINAGE GONZÁLEZ, Rafael

ES 2 812 261 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

Procedimientos para la retirada de dímeros de adaptadores de preparaciones para la secuenciación de ácidos nucleicos

CAMPO DE LA INVENCION

La invención se refiere a procedimientos y composiciones para la retirada de dímeros de adaptadores no deseados de una mezcla de preparación de muestras para la secuenciación de ácidos nucleicos.

ANTECEDENTES

Las principales tecnologías de secuenciación de ADN se basan en la unión de adaptadores de ADN específicos a los extremos de fragmentos de ADN de muestra para crear sitios de cebador para polimerasa, añadir sitios de captura de muestra, "identificar" muestras y añadir sitios de calibración a muestras, entre muchos usos posibles. Los ácidos nucleicos adaptadores se ligan a cada extremo de una muestra de ADN que se va a secuenciar. La creación de fragmentos de ADN ligados a adaptador a menudo incluye adaptadores que se ligan entre sí, formando "dímeros de adaptadores" no deseados.

Actualmente, los dímeros de adaptadores se retiran por etapas de limpieza de fragmentos pequeños, tales como el uso de microesferas magnéticas para capturar y separar los fragmentos de polinucleótidos grandes de los fragmentos pequeños. Dicho procedimiento no se basa en la secuencia y, por lo tanto, es inespecífico e ineficiente. Si la muestra se circulariza, los adaptadores no circularizados se pueden despolimerizar de sus extremos expuestos por tratamiento con exonucleasas. En particular, la retirada de dímeros de adaptadores circularizados de un conjunto de fragmentos de muestra de ADN de tamaño parecido o similar es especialmente difícil. Las etapas de limpieza no retiran eficazmente los dímeros en base al tamaño o llevan un tiempo considerable. Además, los adaptadores circulares no se pueden despolimerizar por exonucleasas, puesto que no tienen un extremo expuesto.

Se necesitan procedimientos más eficaces y específicos para la retirada de dímeros de adaptadores no deseados.

BREVE SUMARIO DE LA INVENCION

Se proporcionan adaptadores para la secuenciación de polinucleótidos, así como procedimientos y kits para su uso en procedimientos de secuenciación. Los adaptadores y procedimientos descritos en el presente documento se pueden usar para producir muestras de polinucleótido para su secuenciación sin dímeros de adaptadores o con cantidades muy bajas de dímeros de adaptadores, que, de otro modo, interferirían en o reducirían la eficacia de los procedimientos de secuenciación de polinucleótidos.

En un aspecto, se proporciona un procedimiento para preparar dúplex de ADN diana para su secuenciación. En los procedimientos divulgados en el presente documento, los adaptadores incluyen una porción de una secuencia de reconocimiento para endonucleasa dependiente de metilo, que, cuando se une de forma covalente en un dímero de adaptadores, formará la secuencia de reconocimiento completa y estará sujeta a digestión por la endonucleasa. Los dímeros de adaptadores se retiran por digestión por la endonucleasa dependiente de metilo, seguido de digestión con exonucleasa. Los adaptadores enlazados a polinucleótidos diana para su secuenciación no son sensibles a la digestión por las enzimas endonucleasa y exonucleasa, y, por tanto, solo los dímeros de adaptadores no deseados se retiran de la muestra.

En un modo de realización, el procedimiento incluye: (a) enlazar de forma covalente regiones dúplex de polinucleótido bicatenario de una pluralidad de adaptadores de secuenciación a los primer y segundo extremos de una pluralidad de dúplex de ADN diana de extremos romos, produciendo, de este modo, una pluralidad de dúplex de ADN diana enlazados a adaptador con un adaptador de secuenciación enlazado de forma covalente en cada extremo del dúplex de ADN diana, en el que la región dúplex de polinucleótido bicatenario de cada adaptador incluye una porción de una secuencia de reconocimiento para una endonucleasa dependiente de metilo en su extremo, en el que se formará una secuencia de reconocimiento completa para la endonucleasa dependiente de metilo si las regiones dúplex de dos adaptadores se enlazan de forma covalente entre sí para producir un dímero de adaptadores; y (b) retirar los dímeros de adaptadores, si los hay, por digestión con una endonucleasa dependiente de metilo seguido de digestión con una o más exonucleasas.

En otro modo de realización, el procedimiento incluye: (a) proporcionar una pluralidad de adaptadores de secuenciación, en el que cada uno de dichos adaptadores incluye una región dúplex de polinucleótido bicatenario, en el que la región dúplex de polinucleótido bicatenario incluye una porción de una secuencia de reconocimiento para una endonucleasa dependiente de metilo en su extremo, en el que se formará una secuencia de reconocimiento completa para la endonucleasa dependiente de metilo si las regiones dúplex de dos adaptadores se enlazan de forma covalente entre sí para producir un dímero de adaptadores; (b) enlazar de forma covalente las regiones dúplex de dichos adaptadores de secuenciación a los primer y segundo extremos de una pluralidad de dúplex de ADN diana de extremos romos, produciendo, de este modo, una pluralidad de dúplex de ADN diana

enlazados a adaptador con un adaptador de secuenciación enlazado de forma covalente en cada extremo del dúplex de ADN diana; y (c) retirar los dímeros de adaptadores, si los hay, por digestión con una endonucleasa dependiente de metilo seguido de digestión con una o más exonucleasas.

5 Los adaptadores pueden incluir una región en horquilla monocatenaria o pueden ser lineales. En un modo de realización, cada uno de los adaptadores incluye una región en horquilla de polinucleótido monocatenaria y la región dúplex de polinucleótido bicatenario. En otro modo de realización, cada uno de los adaptadores es un polinucleótido lineal que incluye primera y segunda hebras de polinucleótido, en el que cada adaptador incluye la región dúplex de polinucleótido bicatenario y una región con protuberancia hacia 3', en el que la primera hebra incluye la región con protuberancia hacia 3' e incluye un nucleótido modificado que es resistente a la digestión con exonucleasa (por ejemplo, un nucleótido tionado) en o cerca del extremo 3' y la segunda hebra, que es parte del dúplex de polinucleótido, incluye un nucleótido modificado que es resistente a la digestión con exonucleasa (por ejemplo, un nucleótido tionado) en o cerca del extremo 5'.

15 En algunos modos de realización, la región dúplex de polinucleótido bicatenario en cada uno de los adaptadores incluye una primera hebra con un extremo 5' hibridada a una segunda hebra con un extremo 3', en los que la primera hebra incluye la secuencia GG en el extremo 5' hibridada a la secuencia C_{Me}C en el extremo 3' de la segunda hebra. La endonucleasa dependiente de metilo puede ser, por ejemplo, MspI o MspII, en la que los dímeros de adaptadores incluyen la secuencia de reconocimiento CC_{Me}GG.

20 En algunos modos de realización, la región dúplex de polinucleótido bicatenario en cada uno de los adaptadores incluye una primera hebra con un extremo 5' hibridada a una segunda hebra con un extremo 3', en los que la primera hebra incluye la secuencia TC en el extremo 5' hibridada a la secuencia GA_{Me} en el extremo 3'. La endonucleasa dependiente de metilo puede ser, por ejemplo, DpnI o DpnII, en la que los dímeros de adaptadores incluyen la secuencia de reconocimiento GA_{Me}TC.

25 En algunos modos de realización, la(s) exonucleasa(s) usada(s) para la digestión de dímeros de adaptadores escindidos incluye(n) exonucleasa VII, exonucleasa III y/o exonucleasa de T5. En algunos modos de realización, se usan exonucleasa III y exonucleasa VII. En algunos modos de realización, se usan exonucleasa III y exonucleasa de T5.

30 En algunos modos de realización, se usa una enzima ligasa para enlazar de forma covalente las regiones dúplex de polinucleótido bicatenario de los adaptadores de secuenciación a los primer y segundo extremos de los dúplex de ADN diana de extremos romos.

35 En diversos modos de realización, los dúplex de ADN diana pueden incluir polinucleótidos diana amplificados, fragmentos no amplificados de ADN genómico, copias de fragmentos de ADN genómico sintetizado con nucleótidos no metilados y/o ADNc transcrito a partir de fragmentos de ARN. En diversos modos de realización, los dúplex de ADN diana se pueden derivar de una muestra de líquido biológico o tisular y/o de uno o más microorganismos.

40 En otro aspecto, se proporciona un procedimiento para preparar una muestra de polinucleótido para su secuenciación, incluyendo el procedimiento unir de forma covalente adaptadores de secuenciación a dúplex de ADN diana que se van a secuenciar, de acuerdo con los procedimientos descritos en el presente documento, y en el que la muestra de polinucleótido resultante para su secuenciación incluye menos de aproximadamente un 1 % de dímeros de adaptadores.

45 En otro aspecto, se proporciona un procedimiento para preparar una muestra de polinucleótido para su secuenciación, incluyendo el procedimiento unir de forma covalente adaptadores de secuenciación a dúplex de ADN diana que se van a secuenciar, de acuerdo con los procedimientos descritos en el presente documento, incluyendo hibridar un cebador a una secuencia de unión a cebador en los adaptadores enlazados de forma covalente y extender el cebador con una enzima ADN polimerasa, preparando, de este modo, un producto de extensión del cebador para su secuenciación. Por ejemplo, el cebador se puede extender para producir una copia complementaria de una hebra del dúplex de ADN diana, en la que la copia se secuencia a medida que se sintetiza por la polimerasa. En un modo de realización, la secuencia de unión a cebador está en una región en horquilla monocatenaria del adaptador. En otro modo de realización, la secuencia de unión a cebador está en una región con protuberancia monocatenaria hacia 3' del adaptador.

50 En otro aspecto, se proporcionan adaptadores para la secuenciación de polinucleótidos. El adaptador incluye una región dúplex de polinucleótido bicatenario, en el que la región dúplex de polinucleótido bicatenario incluye una porción de una secuencia de reconocimiento para una endonucleasa dependiente de metilo en su extremo, y en el que se formará una secuencia de reconocimiento completa para la endonucleasa dependiente de metilo si las regiones dúplex de dos adaptadores se enlazan de forma covalente entre sí para producir un dímero de adaptadores.

65 En un modo de realización, el adaptador incluye una región en horquilla de polinucleótido monocatenaria y la región dúplex de polinucleótido bicatenario, incluyendo opcionalmente un sitio de unión de cebador en la región en

horquilla.

5 En un modo de realización, el adaptador es un polinucleótido lineal que incluye primera y segunda hebras de polinucleótido, en el que el adaptador incluye la región dúplex de polinucleótido bicatenario y una región con protuberancia hacia 3', en el que la primera hebra incluye la región con protuberancia hacia 3' e incluye un nucleótido modificado (por ejemplo, tionado) que es resistente a la digestión con exonucleasa en o cerca del extremo 3' y la segunda hebra, que es parte del dúplex de polinucleótido, incluye un nucleótido modificado (por ejemplo, tionado) que es resistente a la digestión con exonucleasa en o cerca del extremo 5'.

10 En un modo de realización, el dúplex de polinucleótido bicatenario en el adaptador incluye una primera hebra con un extremo 5' hibridada a una segunda hebra con un extremo 3', en el que la primera hebra incluye la secuencia GG en el extremo 5' hibridada a la secuencia C_{Me}C en el extremo 3' de la segunda hebra.

15 En un modo de realización, el dúplex de polinucleótido de doble hebra en el adaptador incluye una primera hebra con un extremo 5' hibridada a una segunda hebra con un extremo 3', en el que la primera hebra incluye la secuencia TC en el extremo 5' hibridada a la secuencia GA_{Me} en el extremo 3'.

20 En otro aspecto, se proporciona un procedimiento para secuenciar polinucleótidos, incluyendo el procedimiento secuenciar una muestra de polinucleótido que incluye una pluralidad de dúplex de ADN enlazados a adaptador con un adaptador de secuenciación como se describe en el presente documento enlazado de forma covalente en cada extremo del dúplex de ADN diana, en el que la muestra de polinucleótido incluye menos de aproximadamente un 1 % de dímeros de adaptadores.

25 En otro aspecto, se proporciona una muestra de polinucleótido para su secuenciación, incluyendo la muestra una pluralidad de dúplex de ADN enlazados a adaptador con un adaptador de secuenciación como se describe en el presente documento enlazado de forma covalente en cada extremo del dúplex de ADN diana, en el que la muestra de polinucleótido incluye menos de aproximadamente un 1 % de dímeros de adaptadores.

30 En otro aspecto, se proporciona un kit para la secuenciación de polinucleótidos. En algunos modos de realización, el kit incluye: (a) una pluralidad de adaptadores de secuenciación como se describe en el presente documento; y (b) instrucciones para preparar dúplex de ADN diana para su secuenciación como se describe en el presente documento. En algunos modos de realización, el kit incluye además (c) una enzima ligasa; (d) una enzima endonucleasa dependiente de metilo; (e) una o más enzimas exonucleasas; y/o (f) uno o más cebadores de secuenciación.

35 Por tanto, la presente invención proporciona un procedimiento para preparar dúplex de ADN diana para su secuenciación, comprendiendo dicho procedimiento:

40 (a) proporcionar una pluralidad de adaptadores de secuenciación, en el que cada uno de dichos adaptadores comprende una región dúplex de polinucleótido bicatenario, en el que la región dúplex de polinucleótido bicatenario comprende una porción de una secuencia de reconocimiento para una endonucleasa dependiente de metilo en su extremo, en el que se formará una secuencia de reconocimiento completa para la endonucleasa dependiente de metilo si las regiones dúplex de dos adaptadores se enlazan de forma covalente entre sí para producir un dímero de adaptadores;

45 (b) enlazar de forma covalente las regiones dúplex de dichos adaptadores de secuenciación a los primer y segundo extremos de una pluralidad de dúplex de ADN diana de extremos romos, produciendo, de este modo, una pluralidad de dúplex de ADN diana enlazados a adaptador con un adaptador de secuenciación enlazado de forma covalente en cada extremo del dúplex de ADN diana; y

50 (c) retirar los dímeros de adaptadores, si los hay, por digestión con una endonucleasa dependiente de metilo seguido de digestión con una o más exonucleasas.

55 Cada uno de dichos adaptadores puede comprender una región en horquilla de polinucleótido monocatenaria y la región dúplex de polinucleótido bicatenario. Cada uno de dichos adaptadores puede ser un polinucleótido lineal que comprende primera y segunda hebras de polinucleótido, en el que cada adaptador comprende un primer extremo que comprende la región dúplex de polinucleótido bicatenario y un segundo extremo que comprende un extremo 3' de la primera hebra y un extremo 5' de la segunda hebra, en el que la primera hebra comprende una región con protuberancia hacia 3' monocatenaria que comprende un nucleótido tionado en o cerca del extremo 3' y la segunda hebra comprende un nucleótido tionado en o cerca del extremo 5'. La región dúplex de polinucleótido bicatenario en cada uno de dichos adaptadores puede comprender una primera hebra que comprende un extremo 5' hibridada a una segunda hebra que comprende un extremo 3', en la que la primera hebra comprende la secuencia GG en el extremo 5' hibridada a la secuencia CC_{Me} en el extremo 3' de la segunda hebra. En este caso, la endonucleasa dependiente de metilo puede ser Mspl o MspII, en la que los dímeros de adaptadores comprenden la secuencia de reconocimiento CC_{Me}GG.

65

5 La región dúplex de polinucleótido bicatenario en cada uno de dichos adaptadores puede comprender una primera hebra que comprende un extremo 5' hibridada a una segunda hebra que comprende un extremo 3', en la que la primera hebra comprende la secuencia TC en el extremo 5' hibridada a la secuencia G_{Me} en el extremo 3'. Si este es el caso, la endonucleasa dependiente de metilo puede ser DpnI o DpnII, en la que los dímeros de adaptadores comprenden la secuencia de reconocimiento G_{Me}TC.

La(s) exonucleasa(s) puede(n) ser exonucleasa VII, exonucleasa III o exonucleasa de T5. El enlace covalente se puede realizar con una enzima ligasa.

10 El ADN diana puede comprender polinucleótidos diana amplificados o fragmentos no amplificados de ADN genómico. Además, el ADN diana puede comprender copias de fragmentos de ADN genómico sintetizado con nucleótidos no metilados o ADNc transcrito a partir de fragmentos de ARN. El ADN diana se puede derivar de una muestra de líquido biológico o tisular o de uno o más microorganismos.

15 La presente invención también proporciona un procedimiento para preparar una muestra de polinucleótido para su secuenciación, que comprende unir de forma covalente adaptadores de secuenciación a dúplex de ADN diana que se van a secuenciar, en el que la muestra de polinucleótido para su secuenciación se prepara de acuerdo con los procedimientos divulgados anteriormente y comprende menos de aproximadamente un 1 % de dímeros de adaptadores.

20 La presente invención también proporciona un procedimiento para secuenciar polinucleótidos, que comprende preparar dúplex de ADN diana para su secuenciación como se divulga anteriormente, en el que la región en horquilla monocatenaria de cada adaptador comprende una secuencia de unión a cebador, y en el que dicho procedimiento comprende además:

25 (d) hibridar un cebador a la secuencia de unión a cebador y extender dicho cebador con una enzima ADN polimerasa, preparando, de este modo, un producto de extensión del cebador para su secuenciación. Si este es el caso, el cebador se puede extender para producir una copia complementaria de una hebra del dúplex de ADN diana, en la que la copia se secuenciará a medida que se sintetiza por la polimerasa.

30 La presente invención también comprende un procedimiento para secuenciar polinucleótidos, que comprende preparar dúplex de ADN diana para su secuenciación como se divulga anteriormente, en el que la región con protuberancia hacia 3' monocatenaria comprende una secuencia de unión a cebador, y en el que dicho procedimiento comprende además:

35 (d) hibridar un cebador a la secuencia de unión a cebador y extender dicho cebador con una enzima ADN polimerasa, preparando, de este modo, un producto de extensión del cebador para su secuenciación. Si este es el caso, el cebador se puede extender para producir una copia complementaria de una hebra del dúplex de ADN diana, en la que la copia se secuenciará a medida que se sintetiza por la polimerasa.

40 En otro aspecto, la presente invención proporciona un adaptador para la secuenciación de polinucleótidos, que comprende una región dúplex de polinucleótido bicatenario, en el que la región dúplex de polinucleótido bicatenario comprende una porción de una secuencia de reconocimiento para una endonucleasa dependiente de metilo en su extremo, en el que se formará una secuencia de reconocimiento completa para la endonucleasa dependiente de metilo si las regiones dúplex de dos adaptadores se enlazan de forma covalente entre sí para producir un dímero de adaptadores. El adaptador puede comprender una región en horquilla de polinucleótido monocatenaria, que puede comprender una secuencia de unión a cebador, y la región dúplex de polinucleótido bicatenario. Además, el adaptador puede ser un polinucleótido lineal que comprende primera y segunda hebras de polinucleótido, en el que el adaptador comprende un primer extremo que comprende la región dúplex de polinucleótido bicatenario y un segundo extremo que comprende un extremo 3' de la primera hebra y un extremo 5' de la segunda hebra, en el que la primera hebra comprende una región con protuberancia hacia 3' monocatenaria que comprende un nucleótido tionado en o cerca del extremo 3' y la segunda hebra comprende un nucleótido tionado en o cerca del extremo 5'. Si este es el caso, la primera hebra puede comprender una secuencia de unión a cebador en la región con protuberancia hacia 3' monocatenaria.

55 Además, el dúplex de polinucleótido bicatenario puede comprender una primera hebra que comprende un extremo 5' hibridada a una segunda hebra que comprende un extremo 3', en el que la primera hebra comprende la secuencia GG en el extremo 5' hibridada a la secuencia C_{Me} en el extremo 3' de la segunda hebra. Además, el dúplex de polinucleótido bicatenario puede comprender una primera hebra que comprende un extremo 5' hibridada a una segunda hebra que comprende un extremo 3', en el que la primera hebra comprende la secuencia TC en el extremo 5' hibridada a la secuencia G_{Me} en el extremo 3'.

60 Todavía en otro aspecto, la presente invención proporciona un procedimiento para secuenciar polinucleótidos, que comprende secuenciar una muestra de polinucleótido que comprende una pluralidad de dúplex de ADN enlazados a adaptador con un adaptador de secuenciación enlazado de forma covalente en cada extremo del dúplex de ADN diana, en el que dicha muestra de polinucleótido comprende menos de aproximadamente un 1 % de dímeros de

adaptadores.

La presente invención también proporciona una muestra de polinucleótido para su secuenciación, que comprende una pluralidad de dúplex de ADN enlazados a adaptador con un adaptador de secuenciación como se divulga anteriormente enlazado de forma covalente en cada extremo del dúplex de ADN diana, y que comprende menos de aproximadamente un 1 % de dímeros de adaptadores.

La presente invención proporciona además un kit para la secuenciación de polinucleótidos, que comprende:

- (a) una pluralidad de adaptadores e
- (b) instrucciones para preparar dúplex de ADN diana para su secuenciación.

Dicho kit puede comprender además uno o más de:

- (c) una enzima ligasa;
- (d) una enzima endonucleasa dependiente de metilo;
- (e) una o más enzimas exonucleasas; y
- (f) uno o más cebadores de secuenciación.

Finalmente, la presente invención proporciona un procedimiento para preparar dúplex de ADN diana para su secuenciación, comprendiendo dicho procedimiento:

- (a) enlazar de forma covalente regiones dúplex de polinucleótido bicatenario de una pluralidad de adaptadores de secuenciación a los primer y segundo extremos de una pluralidad de dúplex de ADN diana de extremos romos, produciendo, de este modo, una pluralidad de dúplex de ADN diana enlazados a adaptador con un adaptador de secuenciación enlazado de forma covalente en cada extremo del dúplex de ADN diana, en el que la región dúplex de polinucleótido bicatenario de cada adaptador comprende una porción de una secuencia de reconocimiento para una endonucleasa dependiente de metilo en su extremo, en el que se formará una secuencia de reconocimiento completa para la endonucleasa dependiente de metilo si las regiones dúplex de dos adaptadores se enlazan de forma covalente entre sí para producir un dímero de adaptadores; y
- (b) retirar los dímeros de adaptadores, si los hay, por digestión con una endonucleasa dependiente de metilo seguido de digestión con una o más exonucleasas.

Cada uno de dichos adaptadores puede comprender una región en horquilla de polinucleótido monocatenaria y la región dúplex de polinucleótido bicatenario. Además, cada uno de dichos adaptadores puede ser un polinucleótido lineal que comprende primera y segunda hebras de polinucleótido, en el que cada adaptador comprende un primer extremo que comprende la región dúplex de polinucleótido bicatenario y un segundo extremo que comprende un extremo 3' de la primera hebra y un extremo 5' de la segunda hebra, en el que la primera hebra comprende una región con protuberancia hacia 3' monocatenaria que comprende un nucleótido tionado en o cerca del extremo 3' y la segunda hebra comprende un nucleótido tionado en o cerca del extremo 5'.

BREVE DESCRIPCIÓN DE LOS DIBUJOS

La **figura 1** muestra un modo de realización de un adaptador con una región dúplex de polinucleótido bicatenario y una región en horquilla monocatenaria.

La **figura 2** muestra un modo de realización de un adaptador lineal con una región dúplex de polinucleótido bicatenario y una región con protuberancia hacia 3'.

La **figura 3** muestra un modo de realización en el que dos adaptadores se ligan entre sí para formar un dímero de adaptadores que incluye una secuencia de reconocimiento completa para una endonucleasa dependiente de metilo.

DESCRIPCIÓN DETALLADA

Se proporcionan procedimientos para preparar dúplex de ADN enlazados a adaptador para su secuenciación sin cantidades significativas de dúplex de adaptadores no deseados que se forman por ligación de adaptadores entre sí en lugar de a los extremos del ADN diana. En los procedimientos descritos en el presente documento, los adaptadores contienen una secuencia única en los extremos de un dúplex de ADN en los mismos, que contiene una base nucleotídica metilada. Cuando los adaptadores se ligan entre sí para formar un dímero, estas secuencias forman un sitio de restricción para una enzima endonucleasa de restricción dependiente de metilo, proporcionando un sitio de escisión para la retirada específica de dímeros de adaptadores de la mezcla de reacción. Los

adaptadores que se ligan a los extremos de ADN diana, que no contiene las secuencias restantes para el sitio de reconocimiento para la enzima de restricción, no se escinden y permanecen unidos a la diana para su secuenciación.

5 A menos que se defina de otro modo en el presente documento, todos los términos técnicos y científicos usados en el presente documento tienen el mismo significado que se entiende comúnmente por un experto en la técnica a la que pertenece la presente invención. Singleton, *et al.*, Dictionary of Microbiology and Molecular Biology, segunda ed., John Wiley and Sons, New York (1994), y Hale & Markham, The Harper Collins Dictionary of Biology, Harper Perennial, NY (1991) proporcionan a un experto un diccionario general de muchos de los términos usados en la
10 presente invención. Se puede usar cualquier procedimiento y material similar o equivalente a los descritos en el presente documento en la práctica o pruebas de la presente invención.

La práctica de la presente invención empleará, a menos que se indique de otro modo, técnicas convencionales de biología molecular (incluyendo técnicas recombinantes), microbiología, biología celular y bioquímica, que están
15 dentro de la habilidad de la técnica. Dichas técnicas se explican completamente en la literatura, por ejemplo, Molecular Cloning: A Laboratory Manual, segunda edición (Sambrook *et al.*, 1989); Oligonucleotide Synthesis (ed. M. J. Gait, 1984; Current Protocols in Molecular Biology (eds. F. M. Ausubel *et al.*, 1994); PCR: The Polymerase Chain Reaction (eds. Mullis *et al.*, 1994); y Gene Transfer and Expression: A Laboratory Manual (Kriegler, 1990).

20 Los intervalos numéricos proporcionados en el presente documento incluyen los números que definen el intervalo.

A menos que se indique de otro modo, los ácidos nucleicos se escriben de izquierda a derecha en una orientación de 5' a 3'; las secuencias de aminoácidos se escriben de izquierda a derecha en una orientación de amino a carboxi, respectivamente.

25 **Definiciones**

"Uno", "una" y "el/la" incluyen referencias al plural a menos que el contexto lo indique claramente de otro modo.

30 El término "adaptador" en el presente documento se refiere a un ácido nucleico que se une a ambas hebras de una molécula de ADN bicatenario. El adaptador puede estar compuesto por dos moléculas de oligonucleótido distintas que tienen emparejamiento de bases entre sí, es decir, son complementarias. De forma alternativa, el adaptador puede estar compuesto por un único oligonucleótido que comprende una o más regiones de complementariedad, y una o más regiones de no complementariedad.

35 El término "par de bases" o "pb" como se usa en el presente documento se refiere a una asociación (es decir, emparejamiento por enlaces de hidrógeno) de adenina (A) con timina (T), o de citosina (C) con guanina (G) en una molécula de ADN bicatenario. En algunos modos de realización, un par de bases puede comprender A emparejado con uracilo (U), por ejemplo, en un dúplex de ADN/ARN.

40 El término "complementario" en el presente documento se refiere al concepto amplio de complementariedad de secuencias en regiones dúplex de una única hebra de polinucleótido o entre dos hebras de polinucleótido entre pares de nucleótidos a través del emparejamiento de bases. Es conocido que un nucleótido adenina puede formar enlaces de hidrógeno específicos ("emparejamiento de bases") con un nucleótido que sea timina o uracilo. De forma
45 similar, es conocido que un nucleótido citosina puede realizar emparejamiento de bases con un nucleótido guanina. "Esencialmente complementario" en el presente documento se refiere a la complementariedad de secuencias en regiones dúplex de una única hebra de polinucleótido o entre dos hebras de polinucleótido, por ejemplo, hebras de polinucleótido de un adaptador en el que la complementariedad es de menos de un 100 %, pero es mayor de un 90 %, y retiene la estabilidad de la región dúplex, por ejemplo, en condiciones para el enlace covalente del
50 adaptador a un dúplex de ADN diana.

El término "derivado de" engloba los términos "originado de", "obtenido de", "obtenible de", "aislado de" y "creado de" y, en general, indica que un material especificado encuentra su origen en otro material especificado o tiene rasgos característicos que se pueden describir con referencia al otro material especificado.

55 El término "dúplex" en el presente documento se refiere a una región de complementariedad que existe entre dos secuencias de polinucleótido.

60 Los términos "primer extremo" y "segundo extremo", cuando se usan en referencia a una molécula de ácido nucleico, en el presente documento se refieren a extremos de una molécula de ácido nucleico lineal.

Un "gen" se refiere a un segmento de ADN que está implicado en la producción de un polipéptido e incluye regiones que preceden y que siguen a las regiones codificantes, así como secuencias intermedias (intrones) entre segmentos codificantes individuales (exones).

65 El término "introducido", en el contexto de la inserción de una secuencia de ácido nucleico en una célula, incluye

"transfección", "transformación" o "transducción" y se refiere a la incorporación de una secuencia de ácido nucleico en una célula eucariota o procariota en la que la secuencia de ácido nucleico se puede incorporar en el genoma de la célula (por ejemplo, cromosoma, plásmido, plástido o ADN mitocondrial), convertir en un replicón autónomo o expresar de forma transitoria.

5 Los términos "aislado", "purificado", "separado" y "recuperado" como se usa en el presente documento se refieren a un material (por ejemplo, una proteína, ácido nucleico o célula) que se retira de al menos un componente con el que se asocia de forma natural, por ejemplo, a una concentración de al menos un 90 % en peso, o al menos un 95 % en peso, o al menos un 98 % en peso de la muestra en la que está contenido. Por ejemplo, estos términos se pueden referir a un material que está sustancial o esencialmente libre de componentes que normalmente lo acompañan, como se encuentra en su estado natural, tal como, por ejemplo, un sistema biológico intacto. Una molécula de ácido nucleico aislado incluye una molécula de ácido nucleico contenida en células que expresan habitualmente la molécula de ácido nucleico, pero la molécula de ácido nucleico está presente de forma extracromosómica o en una localización cromosómica que es diferente de su localización cromosómica natural.

15 El término "genoteca" en el presente documento se refiere a una colección o pluralidad de moléculas molde, es decir, dúplex de ADN diana, que comparten secuencias comunes en sus extremos 5' y secuencias comunes en sus extremos 3'. El uso del término "genoteca" para referirse a una colección o pluralidad de moléculas molde no se debe tomar como que supone que los moldes que constituyen la genoteca se derivan de una fuente particular, o que la "genoteca" tiene una composición particular. A modo de ejemplo, el uso del término "genoteca" no se debe tomar como que supone que los moldes individuales dentro de la genoteca deban ser de una secuencia de nucleótidos diferente o que los moldes deban estar relacionados en términos de secuencia y/o fuente.

20 Una "endonucleasa dependiente de metilo" es una enzima endonucleasa de restricción que requiere bases metiladas específicas en el sitio de restricción para la escisión de ADN bicatenario.

25 El término "mutación" en el presente documento se refiere a un cambio introducido en una secuencia original, incluyendo, pero sin limitarse a, sustituciones, inserciones, deleciones (incluyendo truncamientos). Las consecuencias de una mutación incluyen, pero no se limitan a, la creación de un nuevo carácter, propiedad, función, fenotipo o rasgo no encontrado en la proteína codificada por la secuencia original.

30 El término "nanoporo" en el presente documento se refiere a un poro, canal o paso formado o de otro modo proporcionado en una membrana. Una membrana puede ser una membrana orgánica, tal como una bicapa lipídica, o una membrana sintética, tal como una membrana formada de un material polimérico. El nanoporo se puede disponer contiguo o cerca de un circuito de detección o un electrodo acoplado a un circuito de detección, tal como, por ejemplo, un circuito con semiconductor complementario de óxido metálico (CMOS) o transistor de efecto campo (FET). En algunos ejemplos, un nanoporo tiene una anchura o diámetro característico del orden de 0,1 nm a aproximadamente 1000 nm. Algunos nanoporos son proteínas. OmpG es un ejemplo de un nanoporo proteico.

35 El término "secuenciación de nueva generación (SNG)" en el presente documento se refiere a procedimientos de secuenciación que permiten la secuenciación masiva en paralelo de moléculas de ácido nucleico clonalmente amplificadas y únicas durante los que una pluralidad, por ejemplo, millones, de fragmentos de ácido nucleico de una única muestra o de múltiples muestras diferentes se secuencian al unísono. Los ejemplos no limitantes de SNG incluyen secuenciación por síntesis, secuenciación por ligación, secuenciación ultrarrápida y secuenciación por nanoporos.

40 El término "nucleótido" en el presente documento se refiere a una unidad monomérica de ADN o ARN que consiste en un resto glucídico (pentosa), un fosfato y una base heterocíclica nitrogenada. La base se enlaza al resto glucídico por medio del carbono glucosídico (carbono 1' de la pentosa) y esa combinación de base y glúcido es un nucleósido. Cuando el nucleósido contiene un grupo fosfato enlazado a la posición 3' o 5' de la pentosa se denomina nucleótido. Una secuencia de nucleótidos poliméricos enlazados de forma funcional típicamente se denomina en el presente documento "secuencia de bases" o "secuencia de nucleótidos", o "hebra" de ácido nucleico o polinucleótido, y se representa en el presente documento por una fórmula en la que su orientación de izquierda a derecha está en el sentido convencional del extremo 5' al extremo 3', en referencia al grupo fosfato 5' terminal y al grupo hidroxilo 3' terminal en los extremos "5'" y "3'" de la secuencia polimérica, respectivamente.

45 El término "análogo de nucleótido" en el presente documento se refiere a análogos de nucleósidos trifosfato, por ejemplo, (S)-glicerol-nucleósidos trifosfato (gNTP) de las nucleobases comunes: adenina, citosina, guanina, uracilo y timidina (Horhota *et al.*, *Organic Letters*, 8:5345-5347 [2006]). También se engloban nucleósidos tetrafosfato, nucleósidos pentafofosfato y nucleósidos hexafofosfato. Los "nucleótidos metilados" son nucleótidos que se han modificado por adición de un grupo metilo (por ejemplo, 3-metilcitosina, 3-metiladenina, N6-metiladenina). Los nucleótidos metilados se indican en el presente documento con las letras "Me" en subíndice después de la designación de una letra para la base nucleotídica (por ejemplo, C_{Me}). Los análogos de nucleótidos incluyen nucleótidos que son resistentes a la digestión con exonucleasa, por ejemplo, nucleótidos tionados. Los nucleótidos "tionados" incluyen un enlace fosforotioato (PS), que sustituye un átomo de azufre por un oxígeno no puente en la cadena principal de fosfato de un oligonucleótido. Esta modificación hace que el enlace internucleotídico sea

resistente a la degradación por la nucleasa. Se pueden introducir enlaces fosforotioato entre los últimos 3-5 nucleótidos en el extremo 5' o 3' de un oligonucleótido para inhibir la degradación por la exonucleasa.

El término "enlazado de forma funcional" se refiere a una yuxtaposición o disposición de elementos especificados que les permite actuar en concierto para producir un efecto. Por ejemplo, un promotor se enlaza de forma funcional a una secuencia codificante si controla la transcripción de la secuencia codificante.

El término "polimerasa" en el presente documento se refiere a una enzima que cataliza la polimerización de nucleótidos (es decir, la actividad polimerasa). El término polimerasa engloba ADN polimerasas, ARN polimerasas y retrotranscriptasas. Una "ADN polimerasa" cataliza la polimerización de desoxirribonucleótidos. Una "ARN polimerasa" cataliza la polimerización de ribonucleótidos. Una "retrotranscriptasa" cataliza la polimerización de desoxirribonucleótidos que son complementarios a un molde de ARN.

Los términos "polinucleótido" y "ácido nucleico" y "oligonucleótido" en el presente documento se usan de manera intercambiable para referirse a una molécula polimérica compuesta por monómeros nucleotídicos enlazados de forma covalente en una cadena. ADN (ácido desoxirribonucleico) y ARN (ácido ribonucleico) son ejemplos de polinucleótidos. Como se usa en el presente documento, el término polinucleótido se refiere a una forma polimérica de nucleótidos de cualquier longitud y cualquier estructura tridimensional y mono o multicatenaria (por ejemplo, monocatenaria, bicatenaria, triple helicoidal, etc.), que contiene desoxirribonucleótidos, ribonucleótidos y/o análogos o formas modificadas de desoxirribonucleótidos o ribonucleótidos, incluyendo nucleótidos o bases modificados o sus análogos. Debido a que el código genético es redundante, se puede usar más de un codón para codificar un aminoácido particular, y la presente invención engloba polinucleótidos que codifican una secuencia de aminoácidos particular. Se puede usar cualquier tipo de nucleótido modificado o análogo de nucleótido, siempre que el polinucleótido retenga la funcionalidad deseada en las condiciones de uso, incluyendo modificaciones que incrementan la resistencia a la nucleasa (por ejemplo, desoxi, 2'-O-Me, fosforotioatos, etc.). También se pueden incorporar marcadores para propósitos de detección o captura, por ejemplo, marcadores o anclajes radioactivos o no radioactivos, por ejemplo, biotina. El término polinucleótido también incluye ácidos peptidonucleicos (PNA). Los polinucleótidos pueden ser naturales o no naturales. Los términos "polinucleótido", "ácido nucleico" y "oligonucleótido" se usan en el presente documento de manera intercambiable. Los polinucleótidos pueden contener ARN, ADN o ambos, y/o formas modificadas y/o análogos de los mismos. Una secuencia de nucleótidos se puede interrumpir por componentes distintos a nucleótidos. Se pueden reemplazar uno o más enlaces fosfodiéster por grupos de enlace alternativos. Estos grupos de enlace alternativos incluyen, pero no se limitan a, modos de realización en los que fosfato se reemplaza por P(O)S ("tioato"), P(S)S ("ditioato"), (O)NR₂ ("amidato"), P(O)R, P(O)OR', CO o CH₂ ("formacetal"), en los que cada R o R' es independientemente H o alquilo (C₁₋₂₀) sustituido o no sustituido que contiene opcionalmente un enlace éter (-O-), arilo, alquenilo, cicloalquilo, cicloalquenilo o araldilo. No todos los enlaces en un polinucleótido necesitan porciones circulares.

Como se usa en el presente documento, "polipéptido" se refiere a una composición compuesta por aminoácidos y reconocida como una proteína por los expertos en la técnica. El código convencional de una letra o de tres letras para residuos de aminoácido se usa en el presente documento. Los términos "polipéptido" y "proteína" se usan de manera intercambiable en el presente documento para referirse a polímeros de aminoácidos de cualquier longitud. El polímero puede ser lineal o ramificado, puede comprender aminoácidos modificados y se puede interrumpir por moléculas distintas a aminoácidos. Los términos también engloban un polímero de aminoácido que se ha modificado de forma natural o por intervención; por ejemplo, formación de enlaces disulfuro, glucosilación, lipidación, acetilación, fosforilación o cualquier otra manipulación o modificación, tal como conjugación con un componente de marcado. También se incluyen dentro de la definición, por ejemplo, polipéptidos que contienen uno o más análogos de un aminoácido (incluyendo, por ejemplo, aminoácidos no naturales, etc.), así como otras modificaciones conocidas en la técnica.

El término "cebador" en el presente documento se refiere a un oligonucleótido, ya sea natural o producido sintéticamente, que puede actuar como un punto de iniciación de la síntesis de ácidos nucleicos cuando se dispone en condiciones en las que se induce la síntesis de un producto de extensión del cebador que es complementario a una hebra de ácido nucleico, por ejemplo, en presencia de cuatro nucleótidos trifosfato diferentes y una enzima polimerasa, por ejemplo, una enzima termoestable, en un tampón apropiado ("tampón" incluye pH, fuerza iónica, cofactores, etc.) y a una temperatura adecuada. El cebador es preferentemente monocatenario para obtener la máxima eficacia en la amplificación, pero, de forma alternativa, puede ser bicatenario. Si es bicatenario, en primer lugar se trata el cebador para separar sus hebras antes de usarse para preparar productos de extensión. Preferentemente, el cebador es un oligodesoxirribonucleótido. El cebador debe ser lo suficientemente largo para cebar la síntesis de productos de extensión en presencia de la polimerasa, por ejemplo, enzima polimerasa termoestable. Las longitudes exactas de un cebador dependerán de muchos factores, incluyendo la temperatura, fuente del cebador y el uso del procedimiento. Por ejemplo, dependiendo de la complejidad de la secuencia diana, el cebador oligonucleotídico típicamente contiene 15-25 nucleótidos, aunque puede contener más o menos nucleótidos. En general, las moléculas cebadoras cortas requieren temperaturas más bajas para formar complejos híbridos lo suficientemente estables con el molde.

Un "promotor" se refiere a una secuencia reguladora que está implicada en la unión de la ARN polimerasa para

iniciar la transcripción de un gen. Un promotor puede ser un promotor inducible o un promotor constitutivo. Un "promotor inducible" es un promotor que es activo en condiciones reguladoras ambientales o del desarrollo.

5 El término "recombinante" se refiere a material genético (es decir, ácidos nucleicos, los polipéptidos que codifican, y vectores y células que comprenden dichos polinucleótidos) que se ha modificado para alterar su secuencia o características de expresión, tal como mutando la secuencia codificante para producir un polipéptido alterado, fusionando la secuencia codificante a la de otro gen, disponiendo un gen bajo el control de un promotor diferente, expresando un gen en un organismo heterólogo, expresando un gen en cantidades disminuidas o elevadas, expresando un gen condicional o constitutivamente de manera diferente de su perfil de expresión natural, y
10 similares. En general, los ácidos nucleicos, polipéptidos y células recombinantes basadas en los mismos se han manipulado por el hombre de modo que no son idénticos a los ácidos nucleicos, polipéptidos y células relacionados encontrados en la naturaleza.

15 El término "marcador selectivo" o "marcador seleccionable" se refiere a un gen que se puede expresar en una célula huésped que permite la facilidad de selección de aquellos huéspedes que contienen un ácido nucleico o vector introducido. Los ejemplos de marcadores seleccionables incluyen, pero no se limitan a, sustancias antimicrobianas (por ejemplo, higromicina, bleomicina o cloranfenicol) y/o genes que confieren una ventaja metabólica, tal como una ventaja nutricional, en la célula huésped.

20 El término "genoteca para secuenciación" en el presente documento se refiere al ADN que se procesa para su secuenciación, por ejemplo, usando procedimientos masivos en paralelo, por ejemplo, la SNG. El ADN se puede amplificar opcionalmente para obtener una población de múltiples copias de ADN procesado, que se puede secuenciar por SNG.

25 Una "secuencia señal" (también denominada "presecuencia", "péptido señal", "secuencia líder" o "péptido líder") se refiere a una secuencia de aminoácidos unidos a la porción N terminal de una proteína que facilita la secreción de la forma madura de la proteína de la célula. La forma madura de la proteína extracelular carece de la secuencia señal que se escinde durante el proceso de secreción.

30 El término "protuberancia monocatenaria" o "protuberancia" se usa en el presente documento para referirse a una hebra de una molécula de ácido nucleico bicatenario (bc.) que se extiende más allá del extremo de la hebra complementaria de la molécula de ácido nucleico bc. El término "protuberancia hacia 5'" o "secuencia protuberante hacia 5'" se usa en el presente documento para referirse a una hebra de una molécula de ácido nucleico bc. que se extiende en un sentido 5' más allá del extremo 3' de la hebra complementaria de la molécula de ácido nucleico bc. El
35 término "protuberancia hacia 3'" o "secuencia protuberante hacia 3'" se usa en el presente documento para referirse a una hebra de una molécula de ácido nucleico bc. que se extiende en un sentido 3' más allá del extremo 5' de la hebra complementaria de la molécula de ácido nucleico bc.

40 Las frases "sustancialmente similar" y "sustancialmente idéntico" en el contexto de al menos dos ácidos nucleicos o polipéptidos típicamente significan que un polinucleótido o polipéptido comprende una secuencia que tiene al menos aproximadamente un 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o incluso un 99,5 % de identidad de secuencia, en comparación con un polinucleótido o polipéptido de referencia (por ejemplo, natural).

45 La identidad de secuencia se puede determinar usando programas conocidos, tales como BLAST, ALIGN y CLUSTAL usando parámetros estándar. (Véanse, por ejemplo, Altshul *et al.* (1990) J. Mol. Biol. 215:403-410; Henikoff *et al.* (1989) Proc. Natl. Acad. Sci. 89:10915; Karin *et al.* (1993) Proc. Natl. Acad. Sci. 90:5873; y Higgins *et al.* (1988) Gene 73:237). El programa informático para realizar los análisis por BLAST está disponible públicamente a través del Centro Nacional para la Información Biotecnológica (National Center for Biotechnology Information).
50 Además, se pueden hacer búsquedas en bases de datos usando FASTA (Person *et al.* (1988) Proc. Natl. Acad. Sci. 85:2444-2448). En algunos modos de realización, los polipéptidos sustancialmente idénticos difieren solo en una o más sustituciones aminoacídicas conservadoras. En algunos modos de realización, los polipéptidos sustancialmente idénticos son inmunológicamente reactivos de forma cruzada. En algunos modos de realización, las moléculas de ácido nucleico sustancialmente idénticas se hibridan entre sí en condiciones rigurosas (por ejemplo, dentro de un
55 intervalo de rigurosidad de media a alta).

La "síntesis" de ácidos nucleicos en el presente documento se refiere a cualquier procedimiento *in vitro* para preparar una nueva hebra de polinucleótido o alargar un polinucleótido existente (es decir, ADN o ARN) de una manera dependiente de molde. La síntesis, de acuerdo con la invención, puede incluir la amplificación, que
60 incrementa el número de copias de una secuencia molde de polinucleótido con el uso de una polimerasa. La síntesis de polinucleótidos (por ejemplo, amplificación) da como resultado la incorporación de nucleótidos en un polinucleótido (por ejemplo, extensión a partir de un cebador), formando, de este modo, una nueva molécula de polinucleótido complementaria al molde de polinucleótido. La molécula de polinucleótido formada y su molde se pueden usar como moldes para sintetizar moléculas de polinucleótido adicionales. La "síntesis de ADN", como se usa en el presente documento, incluye, pero no se limita a, reacción en cadena de la polimerasa (PCR), y puede
65 incluir el uso de nucleótidos marcados, por ejemplo, para sondas y cebadores oligonucleotídicos, o para la

secuenciación de polinucleótidos.

5 El término "marca" se refiere a un resto detectable que puede ser uno o más átomos o moléculas, o una colección de átomos y moléculas. Una marca puede proporcionar una firma óptica, electroquímica, magnética o electrostática (por ejemplo, inductiva, capacitiva). Una marca puede bloquear el flujo de corriente a través de un nanoporo.

10 El término "nucleótido marcado" en el presente documento se refiere a un nucleótido que incluye una marca (o especie de marca) que se acopla a cualquier localización del nucleótido, incluyendo, pero sin limitarse a, un fosfato (por ejemplo, fosfato terminal), resto glucídico o de base nitrogenada del nucleótido. Las marcas pueden ser uno o más átomos o moléculas, o una colección de átomos y moléculas. Una marca puede proporcionar una firma óptica, electroquímica, magnética o electrostática (por ejemplo, inductiva, capacitiva), pudiéndose detectar dicha firma con la ayuda de un nanoporo (documento US2014/013616). También se puede unir una marca a un polifosfato como se muestra en la figura 13 del documento US2014/013616.

15 El término "dúplex de ADN diana" en el presente documento se refiere a una molécula de ADN bicatenario que se deriva de un polinucleótido de muestra que es ADN, por ejemplo, ADN genómico o libre de células y/o ARN.

20 El término "molécula de ADN molde" en el presente documento se refiere a una hebra de un ácido nucleico a partir de la que se sintetiza una hebra de ácido nucleico complementaria por una ADN polimerasa, por ejemplo, en una reacción de extensión del cebador.

25 El término "manera dependiente de molde" se refiere a un procedimiento que implica la extensión dependiente de molde de una molécula de cebador (por ejemplo, la síntesis de ADN por ADN polimerasa). El término "manera dependiente de molde" típicamente se refiere a la síntesis de polinucleótidos de ARN o ADN en la que la secuencia de la hebra de polinucleótido recién sintetizada está dictada por las reglas bien conocidas del emparejamiento de bases complementarias (véase, por ejemplo, Watson, J. D. *et al.*, en: *Molecular Biology of the Gene*, 4.^a ed., W. A. Benjamin, Inc., Menlo Park, Calif. (1987)).

30 Como se usa en el presente documento, un "vector" se refiere a una secuencia de polinucleótido diseñada para introducir ácidos nucleicos en uno o más tipos de células. Los vectores incluyen vectores de clonación, vectores de expresión, vectores transportadores, plásmidos, partículas de fago, casetes y similares.

35 Las proteínas relacionadas (y derivadas) engloban proteínas "variantes". Las proteínas variantes difieren de otra proteína (es decir, original) y/o entre sí en un pequeño número de residuos de aminoácido. Una variante puede incluir una o más mutaciones aminoacídicas (por ejemplo, delección, inserción o sustitución aminoacídica) en comparación con la proteína original de la que se deriva. En algunos modos de realización, el número de residuos de aminoácido diferentes es cualquiera de aproximadamente 1, 2, 3, 4, 5, 10, 20, 25, 30, 35, 40, 45 o 50. En algunos modos de realización, las variantes difieren en de aproximadamente 1 a aproximadamente 10 aminoácidos. De forma alternativa o adicionalmente, las variantes pueden tener un grado especificado de identidad de secuencia con una proteína o ácido nucleico de referencia, por ejemplo, como se determina usando una herramienta de alineación de secuencias, tal como BLAST, ALIGN y CLUSTAL (véase más abajo). Por ejemplo, las proteínas o ácido nucleico variantes pueden tener al menos aproximadamente un 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o incluso un 99,5 % de identidad de secuencia de aminoácidos con una secuencia de referencia.

45 Como se usa en el presente documento, las proteínas "naturales" son las que se encuentran en la naturaleza. El término "secuencia natural" se refiere a una secuencia de aminoácidos o de ácido nucleico que se encuentra en la naturaleza o que es natural. En algunos modos de realización, una secuencia natural es el punto de partida de un proyecto de genomaniplulación de proteínas, por ejemplo, la producción de proteínas variantes.

50 **Adaptadores**

55 Se proporcionan adaptadores para la secuenciación de polinucleótidos en el presente documento. Los adaptadores contienen una región dúplex de polinucleótido bicatenario (por ejemplo, dúplex de ADN) que tiene una porción de una secuencia de reconocimiento para una endonucleasa de restricción dependiente de metilo en su extremo. Dos adaptadores que se enlazan de forma covalente entre sí (por ejemplo, se ligan) en los extremos de sus respectivas regiones dúplex de polinucleótido, es decir, formando un dímero de adaptadores, formarán una secuencia completa para la endonucleasa dependiente de metilo y estarán sujetos a escisión en presencia de la endonucleasa. Los adaptadores que se enlazan de forma covalente a los extremos de los dúplex de polinucleótido diana que se van a secuenciar no estarán sujetos a escisión por la endonucleasa, excepto en el caso infrecuente donde el extremo del polinucleótido diana contiene las secuencias restantes para el sitio de reconocimiento para endonucleasa.

65 En algunos modos de realización, como se muestra esquemáticamente en la **fig. 1**, un adaptador incluye una región en horquilla monocatenaria **2** y la región dúplex de polinucleótido bicatenario **1** con una porción de la secuencia de endonucleasa dependiente de metilo en el extremo de la región dúplex. Opcionalmente, la región en horquilla monocatenaria incluye una secuencia de unión a cebador **3**.

En algunos modos de realización, como se muestra esquemáticamente en la **fig. 2**, un adaptador es un polinucleótido lineal (por ejemplo, ADN) que contiene primera y segunda hebras de polinucleótido (por ejemplo, ADN) y primer y segundo extremos. El primer extremo incluye el dúplex de polinucleótido bicatenario **20** con una porción de la secuencia de endonucleasa dependiente de metilo. El segundo extremo incluye una región con protuberancia hacia 3' monocatenaria **21** con uno o más nucleótidos modificados **22** que son resistentes a la digestión con exonucleasa (por ejemplo, nucleótido(s) tionado(s)) en o cerca del extremo 3' de la primera hebra y uno o más nucleótidos modificados **23** que son resistentes a la digestión con exonucleasa (por ejemplo, nucleótido(s) tionado(s)) en o cerca del extremo 5' de la segunda hebra, que es parte de la región dúplex de polinucleótido bicatenario. Opcionalmente, la protuberancia hacia 3' monocatenaria incluye una o más secuencias de unión a cebador **24**. En algunos modos de realización, la región con protuberancia hacia 3' monocatenaria **21** tiene de 8 a aproximadamente 100 nucleótidos de longitud.

En algunos modos de realización, la región dúplex de polinucleótido bicatenario del adaptador incluye la secuencia GG en el extremo 5' y CC_{Me} en el extremo 3', formando la secuencia CC_{Me}GG cuando dos adaptadores se enlazan entre sí. Esta secuencia es una secuencia de reconocimiento para las endonucleasas dependientes de metilo MspI o MspII.

En algunos modos de realización, la región dúplex de polinucleótido bicatenario del adaptador incluye la secuencia TC en el extremo 5' y GA_{Me} en el extremo 3', formando la secuencia GA_{Me}TC cuando dos adaptadores se enlazan entre sí. Esta secuencia es una secuencia de reconocimiento para las endonucleasas dependientes de metilo DpnI o DpnII.

Polinucleótidos diana

Una muestra de ácido nucleico a la que se pueden aplicar los procedimientos descritos en el presente documento se puede derivar de una muestra biológica, tal como una muestra tisular, una muestra de líquido biológico o una muestra celular, y fracciones procesadas de la misma. Una muestra de líquido biológico incluye, como ejemplos no limitantes, muestras de sangre, plasma, suero, sudor, lágrimas, esputo, orina, secreción del oído, linfa, líquido intersticial, saliva, líquido cefalorraquídeo, lavados, suspensión de médula ósea, secreción vaginal, lavado transcervicouterino, líquido cerebral, ascitis, leche, secreciones de las vías respiratorias, tubo intestinal y aparato genitourinario, líquido amniótico y de leucoforesis. En algunos modos de realización, la muestra fuente es una muestra que se puede obtener fácilmente por procedimientos no invasivos, por ejemplo, sangre, plasma, suero, sudor, lágrimas, esputo, orina, secreción del oído y saliva. En algunos modos de realización, la muestra biológica es una muestra de sangre periférica, o las fracciones de plasma y suero. En otros modos de realización, la muestra biológica es un hisopo o frotis, una pieza de biopsia o un cultivo celular. En otro modo de realización, la muestra es una mezcla de dos o más muestras biológicas, por ejemplo, una muestra biológica que comprende dos o más de una muestra de líquido biológico, una muestra tisular y una muestra de cultivo celular. Como se usa en el presente documento, los términos "sangre", "plasma" y "suero" engloban fracciones o porciones procesadas de los mismos. De forma similar, cuando se toma una muestra de una biopsia, hisopo, frotis, etc., la "muestra" puede englobar una fracción o porción procesada derivada de la biopsia, hisopo, frotis, etc.

En algunos modos de realización, las muestras se pueden obtener de fuentes, incluyendo, pero sin limitarse a, muestras de diferentes individuos, diferentes fases del desarrollo de los mismos o diferentes individuos, diferentes individuos enfermos (por ejemplo, individuos con cáncer o que se sospecha que tienen un trastorno genético), individuos normales (por ejemplo, individuos que carecen de una afección de interés), muestras obtenidas en diferentes fases de una enfermedad en un individuo, muestras obtenidas de un individuo sometido a diferentes tratamientos para una enfermedad, muestras de individuos sometidos a diferentes factores ambientales, individuos con predisposición a una patología, individuos con exposición a un patógeno, tal como un agente de enfermedades infecciosas (por ejemplo, VIH) e individuos que son receptores de células, tejidos y/u órganos de donantes. En algunos modos de realización, la muestra es una muestra que comprende una mezcla de muestras fuente diferentes derivadas de los mismos o diferentes sujetos. Por ejemplo, una muestra puede incluir una mezcla de células derivadas de dos o más individuos, como a menudo se encuentra en las escenas de un crimen. En un modo de realización, la muestra es una muestra materna que se obtiene de una hembra gestante, por ejemplo, una mujer embarazada. En este caso, la muestra se puede analizar usando los procedimientos descritos en el presente documento para proporcionar un diagnóstico prenatal de posibles trastornos del feto. A menos que se especifique de otro modo, una muestra materna incluye una mezcla de ADN fetal y materno, por ejemplo, ADNic. En algunos modos de realización, la muestra materna es una muestra de líquido biológico, por ejemplo, una muestra de sangre. En otros modos de realización, la muestra materna es una muestra de ADNic purificado.

Una muestra puede ser una muestra biológica no procesada, por ejemplo, una muestra de sangre completa. Una muestra fuente puede ser una muestra biológica parcialmente procesada, por ejemplo, una muestra de sangre que se ha fraccionado para proporcionar una fracción de plasma sustancialmente libre de células. Una muestra fuente puede ser una muestra biológica que contiene ácidos nucleicos purificados, por ejemplo, una muestra de ADNic purificado derivado de una muestra de plasma esencialmente libre de células. El procesamiento de las muestras puede incluir, por ejemplo, congelación (por ejemplo, muestras de biopsia de tejido), fijación (por ejemplo, fijación

con formol) e inclusión (por ejemplo, inclusión en parafina). El procesamiento parcial de las muestras puede incluir, por ejemplo, fraccionamiento de las muestras (por ejemplo, obtención de fracciones de plasma de muestras de sangre) y otras etapas de procesamiento requeridas para los análisis de las muestras obtenidas durante el trabajo clínico habitual, tal como en el contexto de los ensayos clínicos y/o investigación científica. Las etapas de procesamiento adicional pueden incluir, por ejemplo, etapas para aislar y purificar ácidos nucleicos de muestra. El procesamiento adicional de muestras purificadas puede incluir, por ejemplo, etapas para la modificación requerida de ácidos nucleicos de muestra en preparación para su secuenciación. En algunos modos de realización, la muestra es una muestra no procesada o una parcialmente procesada.

Las muestras también se pueden obtener de tejidos, células u otras fuentes que contienen polinucleótidos cultivados *in vitro*. Las muestras cultivadas se pueden tomar de fuentes que incluyen, pero sin limitarse a, cultivos (por ejemplo, tejidos o células) mantenidos en diversos medios y condiciones (por ejemplo, pH, presión, temperatura), mantenidos durante diferentes periodos de tiempo o tratados con diferentes factores o reactivos (por ejemplo, un candidato terapéutico o un modulador).

Se pueden obtener muestras biológicas de una variedad de sujetos, incluyendo, pero sin limitarse a, seres humanos y otros organismos, incluyendo mamíferos, plantas o células de los sujetos, microorganismos (por ejemplo, bacterias, hongos) o virus.

Los polinucleótidos de muestra que se pueden analizar como se describe en el presente documento incluyen ADN celular genómico, ADN libre de células (ADNlc), ADN mitocondrial, ARN y ADNc. La preparación de genotecas para secuenciación para algunas plataformas de secuenciación SNG requiere que los polinucleótidos sean de un intervalo de tamaños de fragmentos específico, y requiere que los polinucleótidos grandes, por ejemplo, ADN genómico celular, estén fragmentados. Por lo tanto, se puede requerir la fragmentación de polinucleótidos, por ejemplo, ADN genómico celular. La fragmentación de moléculas de polinucleótido por medios mecánicos escinde la cadena principal del ADN en C-O, P-O y C-C, dando como resultado una mezcla heterogénea de extremos romos y protuberantes hacia 3' y 5' con enlaces C-O, P-O y C-C rotos, que se necesitan reparar para las reacciones enzimáticas posteriores, por ejemplo, la ligación de adaptadores de secuenciación, que se requieren para preparar el ADN para su secuenciación. De forma alternativa, puede que no sea necesaria la fragmentación del ADNlc, que existe como fragmentos de <300 bases, para generar una genoteca para secuenciación usando muestras de ADNlc. Una vez que el ADN o ADNc de partida se ha fragmentado, los fragmentos se convierten en romos, es decir, se reparan en sus extremos.

En algunos modos de realización, un ácido nucleico que se va a secuenciar se extiende con bases nucleotídicas no metiladas estándar para producir un producto de extensión de ADNc. Si la diana de ácido nucleico contiene nucleótidos metilados, el dúplex que se produce con el ADNc no contendrá una secuencia de reconocimiento interna para una endonucleasa dependiente de metilo, puesto que se requieren nucleótidos metilados en ambas hebras, y el ácido nucleico bicatenario así producido no se escindirán internamente por la endonucleasa. Si existe una base metilada en o cerca del extremo del ácido nucleico diana, y la ligación del extremo del dúplex diana con un adaptador, como se describe en el presente documento, da como resultado la producción de una secuencia de reconocimiento para la endonucleasa dependiente de metilo, el adaptador se puede escindir de la diana durante la escisión de los dímeros de adaptadores, pero se espera que esto sea un acontecimiento infrecuente.

En algunos modos de realización, un ácido nucleico que se va a secuenciar está amplificado, por ejemplo, se amplifica por un procedimiento de amplificación, tal como, por ejemplo, reacción en cadena de la polimerasa (PCR). La amplificación de una muestra de ácido nucleico que contiene bases metiladas (por ejemplo, ADN genómico) con bases nucleotídicas no metiladas estándar dará como resultado productos de ADN que no contengan bases metiladas. Dicho ADN diana amplificado no contendrá ningún sitio de restricción para una endonucleasa dependiente de metilo, y no se escindirán cuando la muestra se trate con dicha endonucleasa para la retirada de los dímeros de adaptadores, como se describe en el presente documento.

Preparación de polinucleótidos diana para su secuenciación

Se proporcionan procedimientos en el presente documento para preparar dúplex de polinucleótido diana para su secuenciación. Cada extremo de un dúplex de polinucleótido diana se enlaza de forma covalente a la región dúplex de polinucleótido bicatenario de un adaptador de secuenciación, como se describe en el presente documento. Se producen dúplex de polinucleótido diana enlazados a adaptador, con un adaptador de secuenciación en cada extremo. También se pueden producir dímeros de adaptadores, como se muestra esquemáticamente en la **fig. 3**. Los dímeros de adaptadores incluyen dos adaptadores **31** y **32** que se ligan entre sí en los extremos de las regiones dúplex de polinucleótido bicatenario de cada adaptador, mostrado esquemáticamente como la línea discontinua **34** en la **fig. 3**. La ligación de los dos adaptadores produce el sitio de restricción **33** para una endonucleasa dependiente de metilo. Después se trata la mezcla de muestras con una endonucleasa dependiente de metilo, y los dímeros de adaptadores que incluyen la secuencia de restricción para la endonucleasa, como se describe anteriormente, se escinden. Después se trata la mezcla de muestras con una o más exonucleasas que despolimerizan los adaptadores escindidos de sus extremos 5' y/o 3' libres, retirando así los dímeros de adaptadores y cualquier adaptador libre no ligado de la mezcla de polinucleótidos que se van a secuenciar. Los

adaptadores que se ligan a los extremos de los dúplex de polinucleótido diana no tienen un extremo libre que sea susceptible a la digestión con exonucleasa. Como se describe anteriormente, los adaptadores tienen un bucle en horquilla o bien, en el caso de un adaptador lineal, nucleótidos modificados, por ejemplo, nucleótidos tionados, en sus extremos no ligados, de modo que no se digieran por una exonucleasa.

En algunos modos de realización, se preparan polinucleótidos diana para su secuenciación de acuerdo con los procedimientos divulgados en el presente documento, y después del tratamiento con endonucleasa dependiente de metilo y exonucleasa, la mezcla de muestras que contiene dúplex de polinucleótido diana enlazados a adaptador contiene menos de aproximadamente un 1 %, 0,5 %, 0,2 %, 0,1 %, 0,05 % o 0,01 % de dímeros de adaptadores. En algunos modos de realización, la mezcla de muestras no contiene ningún o sustancialmente ningún dímero de adaptadores.

Endonucleasas dependientes de metilo

Son conocidas una serie de enzimas endonucleasas dependientes de metilo. En un modo de realización, la endonucleasa dependiente de metilo es MspI o MspII, que escinde ADN bicatenario en la secuencia de restricción palindrómica CC_{Me}GG. En otro modo de realización, la endonucleasa dependiente de metilo es DpnI o DpnII, que escinde ADN bicatenario en la secuencia de restricción palindrómica GA_{Me}TC. Se apreciará que cualquier endonucleasa dependiente de metilo puede ser útil en los procedimientos descritos en el presente documento si escinde en una secuencia de reconocimiento que incluye al menos un nucleótido metilado y que se puede producir por la ligación de secuencias de nucleótidos en los extremos de regiones dúplex de adaptadores como se describe en el presente documento para proporcionar la secuencia de reconocimiento completa para la enzima.

Procedimientos para secuenciación

Se proporcionan procedimientos para secuenciar polinucleótidos. Se preparan dúplex de polinucleótido para su secuenciación enlazando de forma covalente adaptadores como se describe en el presente documento en cada extremo. Los dímeros de adaptadores se retiran como se describe en el presente documento, dando como resultado una mezcla de secuenciación que contiene menos de aproximadamente un 1 %, 0,5 %, 0,2 %, 0,1 %, 0,05 % o 0,01 % de dímeros de adaptadores. En algunos modos de realización, la mezcla de muestras no contiene ningún o sustancialmente ningún dímero de adaptadores.

En algunos modos de realización, el procedimiento incluye secuenciar una muestra de polinucleótido que incluye una pluralidad de dúplex de ADN enlazados a adaptador con un adaptador de secuenciación como se describe en el presente documento enlazado en cada extremo del dúplex diana, en el que la muestra de polinucleótido contiene menos de aproximadamente un 1 %, 0,5 %, 0,2 %, 0,1 %, 0,05 % o 0,01 % de dímeros de adaptadores, o ningún o sustancialmente ningún dímero de adaptadores.

En algunos modos de realización, un cebador se hibrida a una secuencia de unión a cebador en el adaptador y se extiende con una enzima ADN polimerasa, preparando, de este modo, un producto de extensión del cebador para su secuenciación. En algunos modos de realización, el cebador se extiende para producir una copia complementaria de una hebra del dúplex de polinucleótido diana y la copia se secuencia a medida que se sintetiza por la polimerasa.

Son conocidos en la técnica una serie de procedimientos de secuenciación de polinucleótidos. Los procedimientos de secuenciación que se pueden usar incluyen tecnologías de secuenciación de nueva generación (SNG), que permiten secuenciar individualmente múltiples muestras (es decir, secuenciación simple) o como muestras agrupadas como moléculas de ADN diana indexado (es decir, secuenciación múltiple) en una única tanda de secuenciación, y generar hasta varios cientos de millones de lecturas de secuencias de ADN. Las secuencias de ácidos nucleicos diana y de ácidos nucleicos diana indexados se pueden determinar usando procedimientos de SNG en los que los moldes de ADN clonalmente amplificados o moléculas de ADN únicas, respectivamente, se secuencian de una manera masiva en paralelo (por ejemplo, como se describe en Voelkerding *et al.* (2008) *Clin Chem* 55:641-658; Metzker (2010) *Nature Rev* 11:31-46. Las tecnologías de SNG a veces se subclasifican como secuenciación de primera, segunda y tercera generación (Pareek y Smoczynski (2011) *J Appl Genetics* 52:413-435). Además de la información de secuencia de alto rendimiento, SNG proporciona información cuantitativa, en tanto que cada lectura de secuencia puede ser una "marca de secuencia" contable que representa un molde de ADN clónico individual o una molécula de ADN única. Las tecnologías de secuenciación de SNG incluyen, sin limitación, pirosecuenciación, secuenciación por síntesis con terminadores con tinte reversibles, secuenciación por ligación con sondas oligonucleotídicas, secuenciación por semiconductores iónicos y secuenciación por nanoporos.

Las principales etapas implicadas en la SNG que son aplicables, en general, a todas las tecnologías actuales son la elección/construcción de genotecas, la preparación de genotecas para su secuenciación y la secuenciación masiva en paralelo.

En un modo de realización, el presente procedimiento se puede aplicar a la secuenciación de 454 (<http://www.454.com/>) (por ejemplo, como se describe en Margulies, M. *et al.* (2005) *Nature* 437:376-380). El enfoque global para 454 se basa en pirosecuenciación. La preparación para la secuenciación comienza con

longitudes de ADN (por ejemplo, amplicones o ADN genómico/metagenómico nebulizado) que tienen adaptadores en cada extremo. Estos se fijan a pequeñas microesferas (idealmente, una microesfera tendrá un fragmento de ADN) que se suspenden en una emulsión de agua en aceite. A continuación, se realiza una etapa de PCR en emulsión para hacer múltiples copias de cada fragmento de ADN, lo que da como resultado un conjunto de microesferas en las que cada una contiene muchas copias clonadas del mismo fragmento de ADN. Un chip de fibra óptica lleno de un campo de micropocillos, conocido como PicoTiterPlate, se lava, a continuación, con la emulsión, permitiendo que una única microesfera caiga en cada pocillo. Los pocillos también están llenos de un conjunto de enzimas para el procedimiento de secuenciación (por ejemplo, ADN polimerasa, ATP sulfúrica y luciferasa). En este punto, puede comenzar la secuenciación por síntesis, con la adición de bases que desencadenan la liberación de pirofosfato, lo que produce destellos de luz que se registran para inferir la secuencia de los fragmentos de ADN en cada pocillo a medida que se añade cada tipo de base (A, C, G, T).

En otro modo de realización, el presente procedimiento se puede aplicar a secuenciadores de Illumina. La secuenciación de Illumina es un procedimiento de secuenciación por síntesis, que difiere del de 454 en dos maneras principales: (1) usa una cubeta de lectura con un campo de oligonucleótidos unidos, en lugar de un chip que contiene micropocillos individuales con microesferas y (2) no implica pirosecuenciación, sino, más bien, terminadores con tinte reversibles. El enfoque de terminación con tinte se asemeja a la secuenciación de Sanger "tradicional". Sin embargo, es diferente del de Sanger en que los terminadores con tinte son reversibles, por lo que se retiran después de cada ciclo de obtención de imágenes para dar paso al siguiente nucleótido terminado con tinte reversible. La preparación para la secuenciación comienza lavando longitudes de ADN que tienen adaptadores específicos en cada extremo sobre una cubeta de lectura llena de oligonucleótidos específicos que se hibridan a los extremos de los fragmentos. A continuación, cada fragmento se replica para formar una agrupación de fragmentos idénticos. A continuación, los nucleótidos con terminador con tinte reversible se lavan sobre la cubeta de lectura y se les da tiempo para unirse; los nucleótidos en exceso se eliminan por lavado, se obtienen imágenes de la cubeta de lectura y los terminadores se revierten de modo que el procedimiento se pueda repetir y los nucleótidos puedan continuar añadiéndose en ciclos posteriores.

En otro modo de realización, al presente procedimiento se le puede aplicar el procedimiento SOLiD de Applied Biosystems (<http://solid.appliedbiosystems.com>). El procedimiento SOLiD comienza con una etapa de PCR en emulsión semejante al usado por 454, pero la secuenciación por sí misma es completamente diferente de los sistemas descritos previamente. La secuenciación implica un sistema de incorporación escalonada de dibases en múltiples rondas. Se usa ADN ligasa para la incorporación, lo que la convierte en un enfoque de "secuenciación por ligación", a diferencia de los enfoques de "secuenciación por síntesis" mencionados previamente. Mardis, E.R. (2008) *Annu Rev Genomics Hum Genet* 9:387–402, proporciona una visión general completa de los complejos procedimientos de secuenciación y descodificación implicados con el uso de este sistema.

En otro modo de realización, al presente procedimiento se le puede aplicar el sistema Ion Torrent (<http://www.iontorrent.com/>). El sistema Ion Torrent comienza de manera similar a 454, con una placa de micropocillos que contienen microesferas a las que se unen fragmentos de ADN. Sin embargo, difiere de todos los demás sistemas en la manera en que se detecta la incorporación de bases. Cuando se añade una base a una hebra de ADN creciente, se libera un protón, que altera ligeramente el pH circundante. Los microdetectores sensibles al pH se asocian con los pocillos en la placa, que por sí misma es un chip semiconductor, y registran cuándo se producen estos cambios. A medida que las diferentes bases (A, C, G, T) se lavan secuencialmente de forma exhaustiva, se registran las adiciones, lo que permite inferir la secuencia de cada pocillo.

En otro modo de realización, al presente procedimiento se le puede aplicar el enfoque de secuenciación ultrarrápida de molécula única PacBio (<http://www.pacificbiosciences.com/>). El sistema de secuenciación PacBio no implica ninguna etapa de amplificación, lo que le distingue de otros sistemas de SNG principales. La secuenciación se realiza en un chip que contiene muchos detectores con guía de onda de modo cero (ZMW). Las ADN polimerasas se unen a los detectores con ZMW y se obtienen imágenes de la incorporación de nucleótidos marcados con tinte enlazados por grupos fosfato en tiempo real a medida que se sintetizan las hebras de ADN. RS II C2 y XL de PacBio actualmente ofrece tanto las mayores longitudes de lectura (promediando alrededor de 4.600 bases) como el mayor número de lecturas por tanda (aproximadamente 47.000). El enfoque de "extremos emparejados" típico no se usa con PacBio, puesto que las lecturas típicamente son bastante largas como para que los fragmentos, a través de CCS, se puedan cubrir múltiples veces sin tener que secuenciar de cada extremo independientemente. La multiplexación con PacBio no implica una lectura independiente, sino, más bien, sigue el modelo de identificación "en línea" estándar.

En otro modo de realización, el presente procedimiento se puede aplicar a secuenciación por nanoporos (por ejemplo, como se describe en Soni, G.V. y Meller, A. (2007) *Clin Chem* 53:1996-2001). Las técnicas de análisis de ADN por secuenciación por nanoporos se han desarrollado por una serie de empresas, incluyendo Oxford Nanopore Technologies (Oxford, Reino Unido), Roche e Illumina. En un modo de realización, se usa secuenciación por síntesis, con lo que los nucleótidos que se incorporan en la copia creciente de una molécula de ADN se marcan con una marca de polímero que es única para cada tipo de nucleótido: A, G, C y T. Durante la incorporación de los nucleótidos marcados durante la extensión enzimática de la nueva hebra, las marcas de polímero quedan atrapadas en el poro y el bloqueo de la corriente iónica por la marca designa la base que se está incorporando. Las sucesivas

incorporaciones de nucleótidos durante la síntesis de la hebra dan lugar a sucesivas capturas de marcas de polímero, lo que permite detectar la secuencia de la nueva hebra.

Kits

- 5 Se proporcionan kits para su uso en los procedimientos descritos en el presente documento. Los kits incluyen adaptadores de secuenciación como se describe en el presente documento. Opcionalmente, se proporcionan instrucciones de uso, por ejemplo, para preparar dúplex de polinucleótido diana para su secuenciación. Las instrucciones se pueden proporcionar en forma impresa o en forma de un medio electrónico tal como un CD, DVD o
- 10 USB, o en forma de una dirección de página web donde se puedan obtener dichas instrucciones. Opcionalmente, se pueden incluir otros componentes para preparar dúplex de polinucleótido diana para su secuenciación y/o reactivos de secuenciación. Por ejemplo, un kit puede incluir uno o más de: una enzima ligasa; una enzima endonucleasa dependiente de metilo; una o más enzimas exonucleasas; y uno o más cebadores de secuenciación.
- 15 Se proporciona un envase adecuado. Como se usa en el presente documento, "envase" se refiere a una matriz o material sólido usado habitualmente en un sistema y que puede contener una composición dentro de límites fijados. Dichos materiales incluyen frascos de vidrio y plástico (por ejemplo, polietileno, polipropileno y policarbonato), viales, papel, plástico y sobres laminados con plástico-papel metalizado y similares.

REIVINDICACIONES

1. Un procedimiento para preparar dúplex de ADN diana para su secuenciación, comprendiendo dicho procedimiento:
- 5 (a) proporcionar una pluralidad de adaptadores de secuenciación, en el que cada uno de dichos adaptadores comprende una región dúplex de polinucleótido bicatenario, en el que la región dúplex de polinucleótido bicatenario comprende una porción de una secuencia de reconocimiento para una endonucleasa dependiente de metilo en su extremo, en el que se formará una secuencia de reconocimiento completa para la endonucleasa dependiente de metilo si las regiones dúplex de dos adaptadores se enlazan de forma covalente entre sí para producir un dímero de adaptadores;
- 10 (b) enlazar de forma covalente las regiones dúplex de dichos adaptadores de secuenciación a los primer y segundo extremos de una pluralidad de dúplex de ADN diana de extremos romos, produciendo, de este modo, una pluralidad de dúplex de ADN diana enlazados a adaptador con un adaptador de secuenciación enlazado de forma covalente en cada extremo del dúplex de ADN diana; y
- 15 (c) retirar los dímeros de adaptadores, si los hay, por digestión con una endonucleasa dependiente de metilo seguido de digestión con una o más exonucleasas.
- 20 2. Un procedimiento de acuerdo con la reivindicación 1, en el que cada uno de dichos adaptadores comprende una región en horquilla de polinucleótido monocatenaria y la región dúplex de polinucleótido bicatenario.
3. Un procedimiento de acuerdo con la reivindicación 1, en el que cada uno de dichos adaptadores es un polinucleótido lineal que comprende primera y segunda hebras de polinucleótido, en el que cada adaptador comprende un primer extremo que comprende la región dúplex de polinucleótido bicatenario y un segundo extremo que comprende un extremo 3' de la primera hebra y un extremo 5' de la segunda hebra, en el que la primera hebra comprende una región con protuberancia hacia 3' monocatenaria que comprende un nucleótido tionado en o cerca del extremo 3' y la segunda hebra comprende un nucleótido tionado en o cerca del extremo 5'.
- 25 30 4. Un procedimiento de acuerdo con la reivindicación 1, en el que la región dúplex de polinucleótido bicatenario en cada uno de dichos adaptadores comprende una primera hebra que comprende un extremo 5' hibridada a una segunda hebra que comprende un extremo 3', en el que la primera hebra comprende la secuencia GG en el extremo 5' hibridada a la secuencia CC_{Me} en el extremo 3' de la segunda hebra.
- 35 5. Un procedimiento de acuerdo con la reivindicación 1, en el que la región dúplex de polinucleótido bicatenario en cada uno de dichos adaptadores comprende una primera hebra que comprende un extremo 5' hibridada a una segunda hebra que comprende un extremo 3', en el que la primera hebra comprende la secuencia TC en el extremo 5' hibridada a la secuencia GA_{Me} en el extremo 3'.
- 40 6. Un procedimiento de acuerdo con la reivindicación 1, en el que la etapa (b) comprende el enlace covalente con una enzima ligasa.
7. Un procedimiento de acuerdo con la reivindicación 1, en el que el ADN diana comprende copias de fragmentos de ADN genómico sintetizado con nucleótidos no metilados.
- 45 8. Un procedimiento para preparar una muestra de polinucleótido para su secuenciación, que comprende unir de forma covalente adaptadores de secuenciación a dúplex de ADN diana que se van a secuenciar, en el que la muestra de polinucleótido para su secuenciación se prepara de acuerdo con el procedimiento de la reivindicación 1 y comprende menos de aproximadamente un 1 % de dímeros de adaptadores.
- 50 9. Un procedimiento para secuenciar polinucleótidos, que comprende preparar dúplex de ADN diana para su secuenciación de acuerdo con la reivindicación 2,
- 55 en el que la región en horquilla monocatenaria de cada adaptador comprende una secuencia de unión a cebador, y en el que dicho procedimiento comprende además:
- (d) hibridar un cebador a la secuencia de unión a cebador y extender dicho cebador con una enzima ADN polimerasa, preparando, de este modo, un producto de extensión del cebador para su secuenciación.
- 60 10. Un procedimiento para secuenciar polinucleótidos, que comprende preparar dúplex de ADN diana para su secuenciación de acuerdo con la reivindicación 3,
- 65 en el que la región con protuberancia hacia 3' monocatenaria comprende una secuencia de unión a cebador, y en el que dicho procedimiento comprende además:

- (d) hibridar un cebador a la secuencia de unión a cebador y extender dicho cebador con una enzima ADN polimerasa, preparando, de este modo, un producto de extensión del cebador para su secuenciación.
- 5 11. Un adaptador para la secuenciación de polinucleótidos, que comprende una región dúplex de polinucleótido bicatenario, en el que la región dúplex de polinucleótido bicatenario comprende una porción de una secuencia de reconocimiento para una endonucleasa dependiente de metilo en su extremo, en el que se formará una secuencia de reconocimiento completa para la endonucleasa dependiente de metilo si las regiones dúplex de dos adaptadores se enlazan de forma covalente entre sí para producir un dímero de adaptadores.
- 10 12. Un procedimiento para secuenciar polinucleótidos, que comprende secuenciar una muestra de polinucleótido que comprende una pluralidad de dúplex de ADN enlazados a adaptador con un adaptador de secuenciación de acuerdo con la reivindicación 11 enlazado de forma covalente en cada extremo del dúplex de ADN diana, en el que dicha muestra de polinucleótido comprende menos de aproximadamente un 1 % de dímeros de adaptadores.
- 15 13. Una muestra de polinucleótido para su secuenciación, que comprende una pluralidad de dúplex de ADN enlazados a adaptador con un adaptador de secuenciación de acuerdo con la reivindicación 11 enlazado de forma covalente en cada extremo del dúplex de ADN diana, y que comprende menos de aproximadamente un 1 % de dímeros de adaptadores.
- 20 14. Un kit para la secuenciación de polinucleótidos, que comprende:
- (a) una pluralidad de adaptadores de acuerdo con la reivindicación 11; e
- 25 (b) instrucciones para preparar dúplex de ADN diana para su secuenciación.
15. Un procedimiento para preparar dúplex de ADN diana para su secuenciación, comprendiendo dicho procedimiento:
- 30 (a) enlazar de forma covalente regiones dúplex de polinucleótido bicatenario de una pluralidad de adaptadores de secuenciación a los primer y segundo extremos de una pluralidad de dúplex de ADN diana de extremos romos, produciendo, de este modo, una pluralidad de dúplex de ADN diana enlazados a adaptador con un adaptador de secuenciación enlazado de forma covalente en cada extremo del dúplex de ADN diana,
- 35 en el que la región dúplex de polinucleótido bicatenario de cada adaptador comprende una porción de una secuencia de reconocimiento para una endonucleasa dependiente de metilo en su extremo, en el que se formará una secuencia de reconocimiento completa para la endonucleasa dependiente de metilo si las regiones dúplex de dos adaptadores se enlazan de forma covalente entre sí para producir un dímero de adaptadores; y
- 40 (b) retirar los dímeros de adaptadores, si los hay, por digestión con una endonucleasa dependiente de metilo seguido de digestión con una o más exonucleasas.

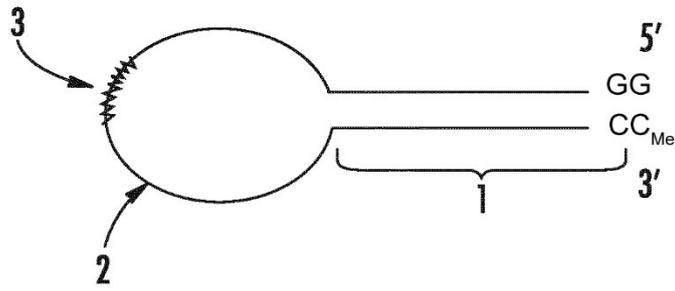


FIG. 1

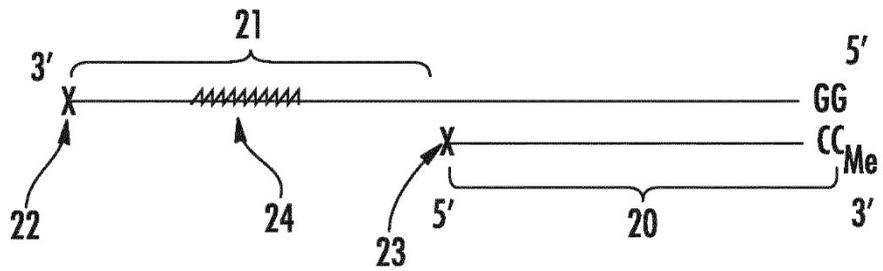


FIG. 2

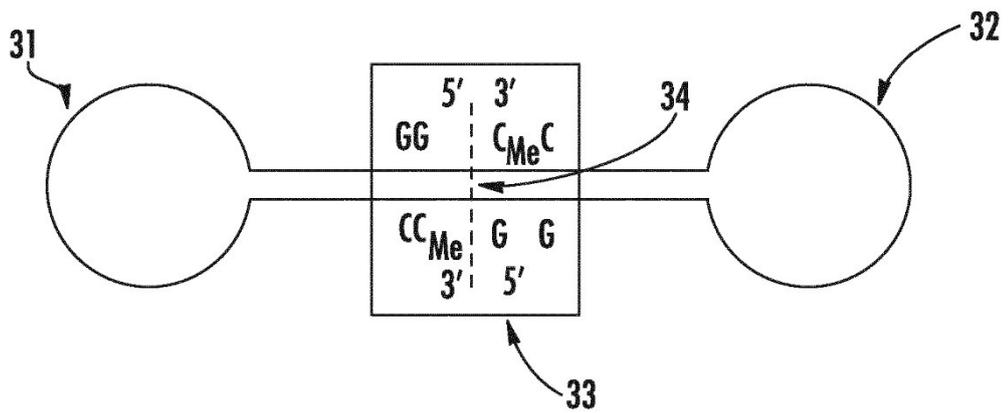


FIG. 3