

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 719 579**

51 Int. Cl.:

C12Q 1/6874 (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **25.06.2014** **E 17164818 (1)**

97 Fecha y número de publicación de la concesión europea: **20.02.2019** **EP 3241913**

54 Título: **Sistema para secuenciación por síntesis ortogonal**

30 Prioridad:

03.07.2013 US 201361842501 P

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

11.07.2019

73 Titular/es:

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

72 Inventor/es:

**FABANI, MARTIN MARIA;
MOON, JOHN A. y
ROBERT BACIGALUPO, MARIA CANDELARIA**

74 Agente/Representante:

ELZABURU, S.L.P

ES 2 719 579 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

Sistema para secuenciación por síntesis ortogonal

Referencia cruzada a solicitudes relacionadas

5 La presente solicitud reivindica el beneficio de prioridad de la solicitud provisional de patente n.º 61/842.501, presentada el 3 de julio de 2013.

Antecedentes

Esta descripción se refiere, en general, al análisis de ácidos nucleicos y, más específicamente, a la secuenciación de ácidos nucleicos.

10 Las plataformas disponibles en el mercado en la actualidad para secuenciar ADN son relativamente caras. Estas plataformas usan una estrategia de "secuenciación por síntesis", llamada así porque los polímeros de ADN se sintetizan a medida que detectan la adición de cada monómero (concretamente, nucleótido) a la estructura polimérica en crecimiento. Debido a que una hebra de ADN molde dirige estrictamente la síntesis de un nuevo polímero de ADN, se puede inferir la secuencia del ADN molde a partir de la serie de monómeros de nucleótidos que se añaden a la hebra en crecimiento durante la síntesis. La capacidad de detectar adiciones de monómeros se ve
15 facilitada por variantes especialmente modificados de componentes bioquímicos que normalmente realizan la síntesis de ADN en sistemas biológicos. Estos componentes modificados son relativamente caros de fabricar y son consumidos en cantidades relativamente grandes durante la secuenciación por síntesis. Además, el control de la reacción emplea un hardware relativamente caro, tal como láseres, ópticas de detección y sistemas de transporte de fluidos complejos. Las plataformas con más éxito del mercado hasta la fecha también requieren reactivos y hardware
20 caros para amplificar los moldes de ADN antes incluso de que empiece la secuenciación por síntesis. La complejidad y el coste de estas plataformas han impedido su uso en algunos contextos clínicos y de investigación en los que hay una necesidad evidente de esta tecnología.

Por tanto, son necesarias mejoras en las plataformas de secuenciación por síntesis para que sean más baratas, rápidas y cómodas. La presente descripción soluciona estas necesidades y también proporciona otras ventajas.

Breve resumen

25 La presente invención se refiere a un sistema para secuenciar moldes de ácidos nucleicos, según se define en las reivindicaciones adjuntas.

La presente descripción proporciona un método para secuenciar moldes de ácidos nucleicos. El método puede incluir las etapas de (a) proporcionar una matriz de sitios, en el que cada sitio incluye un primer molde de ácido
30 nucleico y un segundo molde de ácido nucleico, en el que el primer molde de ácido nucleico tiene una secuencia que es diferente de la secuencia del segundo molde de ácido nucleico; (b) extender un primer cebador unido al primer molde usando una primera especie de polimerasa y un primer conjunto de análogos de nucleótidos, produciendo con ello un primer producto de extensión de cebador que presenta un primer análogo de nucleótido en cada uno de los sitios; (c) extender un segundo cebador unido al segundo molde usando una segunda especie de polimerasa y un
35 segundo conjunto de análogos de nucleótidos, produciendo con ello un segundo producto de extensión de cebador que presenta un segundo análogo de nucleótido en cada uno de los sitios, en el que la primera especie de polimerasa es diferente de la segunda especie de polimerasa, y en el que el primer conjunto de análogos de nucleótidos es diferente del segundo conjunto de análogos de nucleótidos; (d) detectar el primer producto de extensión de cebador y el segundo producto de extensión de cebador en cada uno de los sitios; y (e) repetir las
40 etapas (b) a (d), determinando con ello las diferentes secuencias del primer molde y el segundo molde en cada uno de los sitios.

En la presente también se proporciona un método para secuenciar moldes de ácido nucleicos que incluye las etapas de (a) proporcionar un primer molde de ácido nucleico y un segundo molde de ácido nucleico, en el que el primer
45 molde de ácido nucleico tiene una secuencia que es diferente de la secuencia del segundo molde de ácido nucleico; (b) extender un primer cebador unido al primer molde usando una primera especie de polimerasa y un primer conjunto de análogos de nucleótidos, produciendo con ello un primer producto de extensión de cebador que presenta un primer análogo de nucleótido; (c) extender un segundo cebador unido al segundo molde usando una segunda especie de polimerasa y un segundo conjunto de análogos de nucleótidos, produciendo con ello un
50 segundo producto de extensión de cebador que presenta un segundo análogo de nucleótido, en el que la primera especie de polimerasa es diferente de la segunda especie de polimerasa, y en el que el primer conjunto de análogos de nucleótidos es diferente del segundo conjunto de análogos de nucleótidos; (d) detectar el primer producto de extensión de cebador y el segundo producto de extensión usando un detector que tiene una resolución que es menor que la separación espacial entre el primer producto de extensión de cebador y el segundo producto de extensión de cebador; y (e) repetir las etapas (b) a (d), determinando con ello las diferentes secuencias del primer
55 molde y el segundo molde.

Breve descripción de los dibujos

La figura 1 muestra la representación en forma de diagrama de una reacción de secuenciación por síntesis ortogonal que incluye una etapa de detección (figura 1A) y una etapa de extensión con polimerasa (figura 1B).

5 La figura 2A muestra ejemplos de conjuntos de desoxirribonucleótidos terminadores reversibles (d-FFN) y ribonucleótidos terminadores reversibles (r-FFN) útiles para una reacción de secuenciación por síntesis ortogonal de 1 tinte.

La figura 2B muestra la representación en forma de diagrama de un ciclo de reacción usado para una reacción de secuenciación por síntesis ortogonal de 1 tinte que incluye datos simulados para señales detectadas en 2 canales de emisión diferentes.

10 La figura 3A muestra ejemplos de conjuntos de desoxirribonucleótidos terminadores reversibles (d-FFN) y ribonucleótidos terminadores reversibles (r-FFN) útiles para una reacción de secuenciación por síntesis ortogonal de 2 tintes.

15 La figura 3B muestra la representación en forma de diagrama de un ciclo de reacción usado para una reacción de secuenciación por síntesis ortogonal de 2 tintes que incluye datos simulados para señales detectadas en 4 canales de emisión diferentes.

La figura 4 muestra la representación en forma de diagrama de un método para preparar moldes para una reacción de secuenciación por síntesis ortogonal.

La figura 5 muestra una construcción para la secuenciación por síntesis ortogonal en un formato de extremos apareados.

20 Descripción detallada

Esta descripción proporciona un método para la detección de alta densidad de ácidos nucleicos. Las realizaciones concretas de los métodos de la presente descripción aprovechan técnicas conocidas para manipular y detectar ácidos nucleicos. Sin embargo, las mejoras indicadas a continuación proporcionan un procesamiento ortogonal, de modo que aumenta la densidad de información obtenida por el uso de estas técnicas.

25 El ejemplo de una técnica de detección basada en la extensión de un cebador es ilustrativa del aumento en la densidad de información que puede obtenerse. De modo específico, una secuencia diana de un ácido nucleico puede hibridarse con un cebador, y el cebador puede extenderse por medio de una ADN polimerasa para añadir un nucleótido marcado. Puede usarse un formato de matriz con múltiples sitios, y cada sitio contiene una única secuencia diana que es diferente de las secuencias diana presentes en otros sitios. Opcionalmente se emplean también varias especies de nucleótidos diferentes, cada una con un marcador distinguible. La extensión del cebador da como resultado el reclutamiento del nucleótido marcado hacia el ácido nucleico que contiene la secuencia diana. En un formato de matriz en la que se emplean diferentes nucleótidos marcados, se puede distinguir el marcador que es reclutado hacia cada sitio y usar esta información para identificar el ácido nucleico diana en ese sitio. La densidad de información obtenida a partir de este formato de matriz es una secuencia diana identificada por sitio.

35 En el formato ortogonal de la presente descripción, cada sitio de la matriz puede contener dos o más secuencias diana diferentes que se detectan de modo simultáneo y pueden distinguirse entre sí. En este caso, la información derivada de la matriz puede al menos doblarse. Por ejemplo, dos cebadores diferentes pueden hibridarse con las dos secuencias diana diferentes en cada sitio individual. El primer cebador puede ser un cebador de ADN que es capaz de ser extendido mediante la adición de un desoxirribonucleótido marcado usando una ADN polimerasa, y el segundo cebador puede ser un cebador de ARN que es capaz de ser extendido mediante la adición de un ribonucleótido marcado usando una ADN polimerasa. Los reactivos usados para extender el cebador de ADN son ortogonales con respecto a los reactivos usados para extender el cebador de ARN, en el sentido de que los reactivos de extensión del cebador de ADN no presentan reacción cruzada con el cebador de ARN, y los reactivos de extensión del cebador de ARN no presentan reacción cruzada con el cebador de ADN. Los desoxirribonucleótidos pueden portar marcadores que son distinguibles de los marcadores sobre los ribonucleótidos. La ortogonalidad resultante en la reactividad bioquímica y la gestión de los marcadores permite que el acontecimiento de extensión del cebador de ADN pueda distinguirse del acontecimiento de extensión del cebador de ARN en cada sitio. Así, las dos secuencias diana pueden detectarse de modo distinguible.

40 La ortogonalidad puede surgir de cualquiera de una diversidad de componentes bioquímicos o condiciones de reacción que confieren selectividad a dos acontecimientos de detección diferentes. Tal como se ejemplifica mediante las reacciones que usan ADN polimerasa y ARN polimerasa, la ortogonalidad puede derivarse de la especificidad de dos polimerasas diferentes por diferentes especies de cebadores y por diferentes clases de nucleótidos. En algunas realizaciones, la ortogonalidad puede derivarse de la selectividad de diferentes polimerasas para una especie concreta de molde (por ejemplo, ADN frente a ARN), tanto si las polimerasas son selectivas para una especie concreta de cebador o clase de nucleótido como si no lo son. Así, es posible usar diferentes polimerasas que extienden los cebadores de ADN con desoxirribonucleótidos, pero que son diferencialmente selectivas para un

molde de ADN y un molde de ARN, respectivamente. Por ejemplo, las ADN polimerasas son selectivas, en general, para moldes de ADN, y las transcriptasas inversas son selectivas, en general, para moldes de ARN; sin embargo, ambas enzimas pueden usar un cebador de ADN y desoxirribonucleótidos. Se contemplan combinaciones de polimerasas nativas y/o modificadas para su uso en los sistemas de reacción ortogonales.

5 Los conceptos de ortogonalidad ejemplificados anteriormente para una técnica de detección basada en la extensión de cebadores pueden aplicarse con facilidad a una técnica de secuenciación por síntesis ("sequencing-by-synthesis", SBS). Tal como se presenta en forma de diagrama en las figuras 1A y 1B y tal como se indica con más detalle a continuación, cada ciclo de SBS puede realizarse usando cebadores, polimerasas y nucleótidos ortogonales para proporcionar una mayor adquisición de información a partir de una célula de flujo u otro sustrato empleado en la técnica de SBS. Para fines ilustrativos, la ortogonalidad se ejemplificará para una estrategia de secuenciación denominada secuenciación por síntesis ortogonal ("sequencing-by-orthogonal-synthesis", SBOS); sin embargo, otros métodos también pueden aprovechar la manipulación y detección ortogonal, tal como se indica con más detalle a continuación. Sin embargo, las composiciones, los aparatos y los métodos indicados en la presente no se limitan a aplicaciones de secuenciación.

15 La ortogonalidad puede aprovecharse para aumentar la densidad de adquisición de información en 2 o más veces. Por ejemplo, puede obtenerse un aumento mayor que 2 veces en la densidad de información usando más de dos conjuntos de reactivos ortogonales. Como ejemplo, pueden usarse conjuntos de 3 reactivos que incluyen: (1) reactivos de extensión basados en ADN polimerasa, (2) reactivos de extensión basados en ARN polimerasa, y (3) una polimerasa modificada acoplada a HNA (ácidos nucleicos de 1,5-anhidrohexitol).

20 Tal como se demostró anteriormente y como se indica con más detalle a continuación, la presente descripción proporciona la ventaja de una formación de imágenes de superresolución de una matriz, por lo cual el número de secuencias diana de resolución simultánea en un sitio concreto es mayor que uno. La formación de imágenes de superresolución puede proporcionar el beneficio de distinguir simultáneamente una serie de diferentes ácidos nucleicos diana cuyo número es mayor que el número de sitios en la matriz. De modo similar, se proporciona una superresolución en el sentido de que pueden distinguirse dos secuencias diana diferentes sobre un sustrato en fase sólida usando un detector que tiene una resolución inferior a la resolución espacial que se requeriría, en otros casos, para distinguir las dos secuencias diana sobre el sustrato.

En realizaciones concretas, esta descripción proporciona configuraciones de reactivos y hardware para una detección eficaz de ácidos nucleicos. Un ejemplo de configuración emplea menos marcadores que el número de especies de nucleótidos que se van a distinguir en una etapa de extensión de cebadores. Por ejemplo, pueden distinguirse cuatro especies de desoxirribonucleótidos basándose en la detección de una sola especie de marcador. Tal como se indica con más detalle a continuación, esto puede lograrse usando un primer conjunto de nucleótidos que incluye las siguientes cuatro especies: (1) una especie que porta un primer marcador, (2) una especie que porta un ligando, (3) una especie que presenta un enlace rompible con el primer marcador, y (4) una especie que carece de cualquier marcador o ligando usado en una etapa posterior. Un conjunto ortogonal de nucleótidos (por ejemplo, ribonucleótidos que son ortogonales con respecto a desoxirribonucleótidos) puede incluir cuatro especies: (5) una especie que porta un segundo marcador, (6) una especie que porta una mezcla del primer y segundo marcador, (7) una especie que presenta un enlace rompible con el segundo marcador, y (8) una especie que carece de cualquier marcador o ligando usado en una etapa posterior. La especie dentro de cada conjunto puede distinguirse de las otras basándose en el recuento apropiado de cuáles son los marcadores que aparecen o desaparecen después de ciertas etapas fluidicas específicas, y los dos conjuntos ortogonales de nucleótidos pueden distinguirse basándose en los dos marcadores diferentes. Volviendo al anterior ejemplo de las 8 especies, las especies (1) y (5) pueden distinguirse entre sí basándose en diferentes marcadores y pueden distinguirse de todas las demás especies debido a su aparición después de una etapa de marcaje inicial y a su resistencia a un agente de ruptura; la especie (2) puede distinguirse basándose en la aparición del marcador después de una incubación con un receptor marcado; las especies (3) y (7) pueden distinguirse entre sí basándose en diferentes marcadores y se distinguen de todas las demás especies basándose en la aparición inicial del marcador, seguida de su desaparición después de un tratamiento con un reactivo de ruptura; la especie (6) puede distinguirse de todas las demás especies basándose en la presencia de ambos marcadores a una intensidad que es la mitad de la intensidad para las especies totalmente marcadas; y las especies (4) y (8) pueden distinguirse basándose en la inferencia que surge de la falta de detección de cualquier otra especie en los respectivos conjuntos. Son posibles muchas otras configuraciones para alterar el número de marcadores, el número de manipulaciones fluidicas durante una fase de detección y/o la complejidad del dispositivo de detección para distinguir un cierto número de marcadores. Así, la configuración puede adaptarse para que se ajuste a una estrategia o una aplicación concretas.

55 Se considerará que los términos y expresiones usados en la presente tienen su significado habitual, a menos que se especifique lo contrario. A continuación, se ofrecen ejemplos de varios términos y expresiones empleados en la presente y sus definiciones.

Tal como se emplea en la presente, el término "matriz" se refiere a una población de sitios que pueden diferenciarse entre sí según su localización relativa. Diferentes moléculas que se encuentran en diferentes sitios de una matriz pueden diferenciarse entre sí según las localizaciones de los sitios en la matriz. Un sitio individual de una matriz puede incluir una o más moléculas de un tipo concreto. Por ejemplo, un sitio puede incluir una única molécula de

ácido nucleico diana que tiene una secuencia concreta, o un sitio puede incluir varias moléculas de ácidos nucleicos que tienen la misma secuencia (y/o su secuencia complementaria). Los sitios de una matriz pueden ser diferentes elementos localizados sobre el mismo sustrato. Los ejemplos de elementos incluyen, sin limitación, pocillos en un sustrato, esferas (u otras partículas) en un sustrato o sobre un sustrato, proyecciones desde un sustrato, rebordes sobre un sustrato o canales en un sustrato. Los sitios de una matriz pueden ser sustratos distintos que portan cada uno una molécula diferente. Pueden identificarse diferentes moléculas unidas a sustratos distintos según las localizaciones de los sustratos sobre una superficie con la cual los sustratos están asociados, o según las localizaciones de los sustratos en un líquido o un gel. Los ejemplos de matrices en las que los distintos sustratos están localizados sobre una superficie incluyen, sin limitación, los que presentan esferas en pocillos.

Tal como se emplea en la presente, el término “agrupamiento”, cuando se emplea en referencia a ácidos nucleicos, se refiere a una población de ácidos nucleicos que está unida a una fase sólida para formar un elemento o un sitio. En general, los ácidos nucleicos son de una sola especie, formando con ello un agrupamiento homogéneo. Sin embargo, en algunas realizaciones, los ácidos nucleicos pueden ser heterogéneos, de modo que están presentes moléculas individuales con diferentes secuencias en el sitio o elemento. En general, los ácidos nucleicos están unidos covalentemente, por ejemplo, a través de sus extremos 5', pero en algunos casos son posibles otros medios de unión. Los ácidos nucleicos en un agrupamiento pueden ser monocatenarios o bicatenarios. En algunas, pero no todas las realizaciones, los agrupamientos se forman mediante un método de amplificación en fase sólida, conocido como amplificación de puente. Se indican ejemplos de configuraciones para agrupamientos y métodos para su producción, por ejemplo, en la patente de EE. UU. n.º 5.641.658; la patente de EE. UU. n.º de publicación 2002/0055100; la patente de EE. UU. n.º 7.115.400; la patente de EE. UU. n.º de publicación 2004/0096853; la patente de EE. UU. n.º de publicación 2004/0002090; la patente de EE. UU. n.º de publicación 2007/0128624; y la patente de EE. UU. n.º de publicación 2008/0009420.

Tal como se emplea en la presente, el término “diferente”, cuando se emplea en referencia a ácidos nucleicos, significa que los ácidos nucleicos contienen secuencias de nucleótidos que no son iguales entre sí. Dos o más ácidos nucleicos pueden presentar secuencias de nucleótidos que son diferentes a lo largo de su longitud completa. Como alternativa, dos o más ácidos nucleicos pueden presentar secuencias de nucleótidos que son diferentes a lo largo de una porción sustancial de su longitud. Por ejemplo, dos o más ácidos nucleicos pueden presentar porciones de secuencias de nucleótidos diana que son diferentes entre sí, mientras que también presentan una región de secuencia universal que es la misma en todos. En general, cuando en la presente se indica que dos especies son “diferentes”, una de las especies tendrá una propiedad estructural que no es igual que las propiedades estructurales de la segunda especie. Por ejemplo, dos especies poliméricas diferentes (tales como dos proteínas) pueden tener diferentes secuencias de subunidades monoméricas (tal como diferentes secuencias de aminoácidos para dos proteínas diferentes).

Tal como se emplea en la presente, la expresión “cada uno”, cuando se emplea en referencia a una colección de artículos, pretende identificar un artículo individual en la colección, pero no se refiere necesariamente a cada artículo de la colección. Se producen excepciones si el contexto o una descripción explícita lo indican claramente.

Tal como se emplea en la presente, el término “sitio” significa una localización en una matriz en la que está presente una especie concreta de molécula. Un sitio puede contener una única molécula o puede contener una población de varias moléculas de la misma especie. Los sitios de una matriz generalmente son discretos. Los sitios discretos pueden estar contiguos o puede haber espacio entre ellos.

Tal como se emplea en la presente, la expresión “ácido nucleico” pretende ser coherente con su uso en la técnica e incluye ácidos nucleicos naturales o sus análogos funcionales. Los análogos funcionales particularmente útiles son capaces de hibridarse con un ácido nucleico de una manera específica de secuencia, o son capaces de ser utilizados como moldes para la replicación de una secuencia de nucleótidos concreta. Los ácidos nucleicos naturales en general presentan un esqueleto que contiene enlaces fosfodiéster. Una estructura análoga puede presentar un enlace del esqueleto alternativo, que incluye cualquiera de una diversidad de los conocidos en la técnica. Los ácidos nucleicos naturales en general presentan un azúcar desoxirribosa (por ejemplo, tal como se encuentra en el ácido desoxirribonucleico (ADN)) o un azúcar ribosa (por ejemplo, tal como se encuentra en el ácido ribonucleico (ARN)). Un ácido nucleico puede contener cualquiera de una diversidad de análogos de estos restos azúcar que son conocidos en la técnica. Un ácido nucleico puede incluir bases nativas o no nativas. A este respecto, un ácido desoxirribonucleico nativo puede presentar una o más bases seleccionadas del grupo que consiste en adenina, timina, citosina o guanina, y un ácido ribonucleico puede presentar una o más bases seleccionadas del grupo que consiste en uracilo, adenina, citosina o guanina. Las bases no nativas útiles que pueden incluirse en un ácido nucleico son conocidas en la técnica. El término “diana”, cuando se emplea en referencia a un ácido nucleico, es un identificador semántico para el ácido nucleico en el contexto de un método o una composición indicada en la presente, y no limita necesariamente la estructura o la función del ácido nucleico más allá de los que se indica de forma explícita.

Tal como se emplea en la presente, la expresión “molde de ácido nucleico” se refiere a un ácido nucleico, o a una de sus porciones, que es capaz de emplearse como guía para la replicación catalizada por polimerasa. Una molécula de ácido nucleico puede incluir múltiples moldes a lo largo de su longitud o, como alternativa, puede usarse un único molde en una realización concreta en la presente. Un molde de ácido nucleico también puede actuar como guía para

la extensión de cebadores catalizada por ligasa.

Tal como se emplean en la presente, el término “nucleótido” o la expresión “análogo de nucleótido” incluye nucleótidos naturales, nucleótidos no naturales, ribonucleótidos, desoxirribonucleótidos, didesoxirribonucleótidos y otras moléculas conocidas como nucleótidos. Por ejemplo, el término y la expresión usados en la presente se refieren en general a un resto nucleósido (ribosa, desoxirribosa o uno de sus análogos) que incluye un resto básico y uno o más restos fosfatos unidos opcionalmente. El término y la expresión pueden usarse para referirse a una unidad monomérica que está presente en un polímero, por ejemplo, para identificar una subunidad presente en una hebra de ADN o ARN. El término y la expresión también pueden usarse para referirse a una molécula monomérica que no está necesariamente presente en un polímero, por ejemplo, una molécula que es capaz de ser incorporada en un polinucleótido por una polimerasa de una manera dependiente del molde.

Los ejemplos de nucleótidos incluyen, pero no se limitan a ribonucleótido monofosfato (a veces denominado ribonucleósido monofosfato), ribonucleótido difosfato (a veces denominado ribonucleósido difosfato), ribonucleótido trifosfato (a veces denominado ribonucleósido trifosfato), desoxinucleótido monofosfato (a veces denominado desoxirribonucleósido monofosfato), desoxinucleótido difosfato (a veces denominado desoxirribonucleósido difosfato) y desoxinucleótido trifosfato (a veces denominado desoxirribonucleósido trifosfato). Para ayudar a la aclaración en el caso de que se desee distinguir los componentes de ARN de los componentes de ADN, el término “ribonucleótido” puede usarse para especificar nucleótidos de ARN, tales como ribouridina trifosfato, riboguanidina trifosfato, ribocitidina trifosfato o riboadenosina trifosfato; y el término “desoxinucleótido” puede usarse para especificar nucleótidos de ADN, tales como desoxitimidina trifosfato, desoxiguanidina trifosfato, desoxicitidina trifosfato y desoxiadenosina trifosfato. En realizaciones concretas, los nucleótidos son “extensibles”, por ejemplo, carecen de un resto de bloqueo de la extensión en el 3’ hidroxilo o en cualquier otra posición del nucleótido. En otras realizaciones, los nucleótidos están “bloqueados”, y tienen un resto que evita que la posición 3’ participe en la extensión por medio de la adición de otro nucleótido u oligonucleótido.

Tal como se emplea en la presente, el término “cebador” significa un ácido nucleico que presenta una secuencia que se une a un sitio de unión de cebador en una secuencia molde o cerca de esta. En general, el cebador se une en una configuración que permite la replicación del molde, por ejemplo, a través de la extensión del cebador con polimerasa. El cebador puede ser una primera porción de una molécula de ácido nucleico que se une a una segunda porción de la molécula de ácido nucleico, siendo la primera porción una secuencia de cebador y siendo la segunda porción una secuencia de unión de cebador (por ejemplo, un cebador de horquilla). Como alternativa, el cebador puede ser una primera molécula de ácido nucleico que se une a una segunda molécula de ácido nucleico que contiene la secuencia de molde. Un cebador puede consistir en ADN, ARN o sus análogos.

Tal como se emplea en la presente, la expresión “producto de extensión de cebador” significa un cebador que se ha modificado mediante la adición de al menos un análogo de nucleótido. Por ejemplo, un cebador puede modificarse mediante la adición de uno o más análogos de nucleótidos a su extremo 3’ (por ejemplo, a través de una catálisis con polimerasa), formando con ello un producto de extensión de cebador. Como alternativa, un producto de extensión de cebador puede producirse mediante el acoplamiento de un oligonucleótido al extremo 3’ o 5’ de un cebador. En este caso, el producto de extensión de cebador se extiende en una longitud equivalente a la longitud del oligonucleótido. Un producto de extensión de cebador puede ser al menos 1, 2, 5, 10, 500, 1000 o más nucleótidos más largos que el cebador. Como alternativa o además, un producto de extensión de cebador puede ser no más de 1, 2, 5, 10, 500, o 1000 nucleótidos más largo que el cebador. Por ejemplo, el uso de un nucleótido bloqueado proporciona un producto de extensión que es al menos 1 nucleótido más largo que el cebador y también no más de 1 nucleótido más largo que el cebador.

Tal como se emplea en la presente, la referencia a manipular “selectivamente” (o una manipulación “selectiva”) un primer artículo comparado con un segundo artículo significa que la manipulación tiene un mayor efecto sobre el primer artículo, comparado con el efecto sobre el segundo artículo. No es necesario que la manipulación tenga efecto sobre el segundo artículo. La manipulación puede tener un efecto sobre el primer artículo que es al menos 1%, 5%, 10%, 25%, 50%, 75%, 90%, 95%, o 99% mayor que el efecto sobre el segundo artículo. La manipulación puede tener un efecto sobre el primer artículo que es al menos 2 veces, 3 veces, 4 veces, 5 veces, 10 veces, 100 veces, 1×10^3 veces, 1×10^4 veces o 1×10^6 veces mayor que el efecto sobre el segundo artículo. La manipulación puede incluir, por ejemplo, modificar, ponerse en contacto, tratar, cambiar, romper (por ejemplo, un enlace químico), romper fotoquímicamente (por ejemplo, un enlace químico), formar (por ejemplo, un enlace químico), formar fotoquímicamente (por ejemplo, un enlace químico), modificar covalentemente, modificar no covalentemente, destruir, fotoablacionar, retirar, sintetizar, polimerizar, fotopolimerizar, amplificar (por ejemplo, un ácido nucleico), copiar (por ejemplo, un ácido nucleico), extender (por ejemplo, un ácido nucleico), acoplar (por ejemplo, un ácido nucleico) u otras manipulaciones indicadas en la presente o conocidas de otro modo en la técnica.

Tal como se emplea en la presente, el término “secuencia”, cuando se emplea en referencia a un ácido nucleico, se refiere al orden de nucleótidos (o bases) en los ácidos nucleicos. En los casos en que estén presentes diferentes especies de nucleótidos en el ácido nucleico, la secuencia incluye una identificación de la especie de nucleótido (o base) en las posiciones respectivas en el ácido nucleico. Una secuencia es una propiedad de toda o parte de una molécula de ácido nucleico. El término puede usarse de modo similar para describir el orden y la identidad posicional de unidades monoméricas en otros polímeros, tales como unidades monoméricas de aminoácidos de polímeros de

proteínas.

Tal como se emplea en la presente, el término “especie” se emplea para identificar moléculas que comparten la misma estructura química. Por ejemplo, una mezcla de nucleótidos puede incluir varias moléculas de dCTP. Se entenderá que las moléculas de dCTP son la misma especie entre sí, pero son una especie diferente comparadas con dATP, dGTP, dTTP, etc. De modo similar, las moléculas de ADN individuales que tienen la misma secuencia de nucleótidos son de la misma especie, mientras que las moléculas de ADN con secuencias diferentes son especies diferentes. Como otro ejemplo, una ADN polimerasa es una especie de polimerasa diferente comparada con una ARN polimerasa (incluso si las dos polimerasas surgen del mismo organismo).

Las realizaciones indicadas a continuación y descritas en las reivindicaciones pueden entenderse a la vista de las anteriores definiciones.

La presente descripción proporciona un método para secuenciar moldes de ácidos nucleicos. El método puede incluir las etapas de (a) proporcionar una matriz de sitios, en el que cada sitio incluye un primer molde de ácido nucleico y un segundo molde de ácido nucleico, en el que el primer molde de ácido nucleico tiene una secuencia que es diferente de la secuencia del segundo molde de ácido nucleico; (b) extender un primer cebador unido al primer molde usando una primera especie de polimerasa y un primer conjunto de análogos de nucleótidos, produciendo con ello un primer producto de extensión de cebador que presenta un primer análogo de nucleótido en cada uno de los sitios; (c) extender un segundo cebador unido al segundo molde usando una segunda especie de polimerasa y un segundo conjunto de análogos de nucleótidos, produciendo con ello un segundo producto de extensión de cebador que presenta un segundo análogo de nucleótido en cada uno de los sitios, en el que la primera especie de polimerasa es diferente de la segunda especie de polimerasa, y en el que el primer conjunto de análogos de nucleótidos es diferente del segundo conjunto de análogos de nucleótidos; (d) detectar el primer producto de extensión de cebador y el segundo producto de extensión de cebador en cada uno de los sitios; y (e) repetir las etapas (b) a (d), determinando con ello las diferentes secuencias del primer molde y el segundo molde en cada uno de los sitios.

En la presente también se proporciona un método para secuencia moldes de ácido nucleicos que incluye las etapas de (a) proporcionar un primer molde de ácido nucleico y un segundo molde de ácido nucleico, en el que el primer molde de ácido nucleico tiene una secuencia que es diferente de la secuencia del segundo molde de ácido nucleico; (b) extender un primer cebador unido al primer molde usando una primera especie de polimerasa y un primer conjunto de análogos de nucleótidos, produciendo con ello un primer producto de extensión de cebador que presenta un primer análogo de nucleótido; (c) extender un segundo cebador unido al segundo molde usando una segunda especie de polimerasa y un segundo conjunto de análogos de nucleótidos, produciendo con ello un segundo producto de extensión de cebador que presenta un segundo análogo de nucleótido, en el que la primera especie de polimerasa es diferente de la segunda especie de polimerasa, y en el que el primer conjunto de análogos de nucleótidos es diferente del segundo conjunto de análogos de nucleótidos; (d) detectar el primer producto de extensión de cebador y el segundo producto de extensión usando un detector que tiene una resolución que es menor que la separación espacial entre el primer producto de extensión de cebador y el segundo producto de extensión de cebador; y (e) repetir las etapas (b) a (d), determinando con ello las diferentes secuencias del primer molde y el segundo molde.

Tal como se indicó anteriormente, un método de la presente descripción puede incluir una etapa de proporcionar un primer y un segundo molde de ácido nucleico, en los que las secuencias de los dos moldes son diferentes. Las dos secuencias de molde pueden ser porciones de una única molécula de ácido nucleico o, como alternativa, las dos secuencias de molde pueden estar localizadas sobre moléculas distintas. Tal como se indica con más detalle en otro punto de la presente, las dos secuencias de molde pueden estar tan cerca que no pueden resolverse espacialmente con el sistema de detección usado. No obstante, los métodos de detección ortogonal de la presente invención permiten distinguir estas secuencias de molde. El esquema de detección ortogonal se ejemplifica para dos secuencias de molde, pero puede emplearse con dos o más secuencias de molde. Por consiguiente, un sistema o método indicado en la presente pueden incluir al menos 2, 3, 4, 5, 10 o más secuencias de molde que están muy cercanas, por ejemplo, sobre una única molécula de ácido nucleico, un único sitio de una matriz o pueden estar tan cerca que no pueden resolverse espacialmente con el sistema de detección usado.

Los ácidos nucleicos diana usados en la presente pueden estar compuestos de ADN, ARN o sus análogos. La fuente de los ácidos nucleicos diana puede ser ADN genómico, ARN mensajero u otros ácidos nucleicos procedentes de fuentes nativas. En algunos casos, los ácidos nucleicos diana que se derivan de dichas fuentes pueden amplificarse antes de su uso en un método una composición de la presente.

Los ejemplos de muestras biológicas a partir de las cuales pueden obtenerse ácidos nucleicos diana incluyen, por ejemplo, las procedentes de un mamífero, tal como un roedor, ratón, rata, conejo, cobaya, ungulado, caballo, oveja, cerdo, cabra, vaca, gato, perro, primate, humano o primate no humano; una planta, tal como *Arabidopsis thaliana*, maíz, sorgo, avena, trigo, arroz, canola, o soja; un alga, tal como *Chlamydomonas reinhardtii*; un nemátodo, tal como *Caenorhabditis elegans*; un insecto, tal como *Drosophila melanogaster*, mosquito, mosca de la fruta, abeja melífera o araña; un pez, tal como el pez cebra; un reptil; un anfibio, tal como una rana o *Xenopus laevis*; *Dictyostelium discoideum*; un hongo, tal como *Pneumocystis carinii*, *Takifugu rubripes*, levadura, *Saccharomyces cerevisiae* o

Schizosaccharomyces pombe; o *Plasmodium falciparum*. Los ácidos nucleicos diana también pueden obtenerse a partir de un procariota, tal como una bacteria, *Escherichia coli*, estafilococos o *Mycoplasma pneumoniae*; una arqueobacteria; un virus, tal como el virus de la hepatitis C o el virus de la inmunodeficiencia humana; o un viroide. Los ácidos nucleicos diana pueden obtenerse a partir de un cultivo homogéneo o población de los anteriores organismos o, como alternativa, a partir de una colección de varios organismos diferentes, por ejemplo, en una comunidad o ecosistema. Los ácidos nucleicos pueden aislarse usando métodos conocidos en la técnica que incluyen, por ejemplo, los descritos en Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, 3ª edición, Cold Spring Harbor Laboratory, Nueva York (2001), o en Ausubel *et al.*, *Current Protocols in Molecular Biology*, John Wiley and Sons, Baltimore, Md. (1998).

En realizaciones concretas, una muestra de ácido nucleico puede modificarse o prepararse para su uso en uno o más de los métodos indicados en la presente. En algunos casos se desea añadir uno o más sitios de unión a cebadores a un ácido nucleico. Pueden utilizarse técnicas de biología molecular conocidas para introducir sitios de unión a cebadores cadena arriba de las respectivas secuencias de molde, por ejemplo, a través de la inserción de un adaptador que contenga el sitio de unión al cebador, una mutación para crear el sitio de unión al cebador, el acoplamiento de un adaptador que contenga el sitio de unión al cebador, etc. Se describen métodos útiles en Sambrook *et al.*, *supra*, y Ausubel *et al.*, *supra*. El ejemplo I proporciona una ilustración de una técnica basada en la tagmentación. La tagmentación es particularmente útil para introducir uno o más sitios de unión a cebadores y puede llevarse a cabo, por ejemplo, usando las técnicas indicadas en las patentes de EE. UU. n.ºs 6.294.385 y 8.383.345, y la publicación PCT n.º WO 2012/1 06546. Se entenderá que, en algunos casos, las regiones de la secuencia natural que residen cadena arriba de las respectivas secuencias de molde pueden aprovecharse como sitios de unión a cebadores en un método indicado en la presente. Pueden emplearse métodos similares a los ejemplificados anteriormente para los sitios de unión a cebadores para introducir otros elementos de secuencia deseados, tales como elementos de promotor para la extensión basada en ARN polimerasa o secuencias de marcadores. Un ejemplo de método para crear fragmentos de ácidos nucleicos que tengan, cada uno, una primera secuencia de molde cerca de un sitio de cebado de ADN y una segunda secuencia de molde cerca de un sitio de cebado de ARN y un promotor de ARN polimerasa se muestra en la figura 4 y se describe en el ejemplo I.

Los sitios de cebado universal son particularmente útiles para aplicaciones de múltiple de los métodos indicados en la presente. El término "universal", cuando se emplea en referencia a ácidos nucleicos, significa una región de la secuencia que es común para dos o más moléculas de ácidos nucleicos, en las que las moléculas también contienen regiones de secuencia diferente. Una secuencia universal presente en diferentes miembros de una colección de moléculas puede permitir la replicación, la amplificación o la detección de múltiples secuencias diferentes usando una única especie de cebador universal que es complementario con la secuencia universal. Así, un cebador universal incluye una secuencia que puede hibridarse específicamente con una secuencia universal. Pueden encontrarse ejemplos de métodos para unir secuencias universales a una colección de ácidos nucleicos diana en la solicitud de patente de EE. UU. n.º de publicación 2007/0128624 A1.

Puede utilizarse cualquiera de una diversidad de promotores que sean apropiados para la ARN polimerasa concreta que se está utilizando. Por ejemplo, puede usarse un promotor bacteriano con una ARN polimerasa bacteriana, o puede usarse un promotor eucariota con una ARN polimerasa eucariota. En general, un promotor estará localizado cerca del molde que se va a detectar, cadena arriba del sitio de unión al cebador de ARN y sobre la misma hebra que el molde. Pueden usarse técnicas de biología molecular y/o de síntesis de ácidos nucleicos convencionales para crear una construcción de promotor funcional en un ácido nucleico diana. Un promotor particularmente útil es un promotor bidireccional, tales como los presentes en parejas de genes bidireccionales de mamífero. Un promotor bidireccional puede ser útil para aplicaciones de secuenciación de extremos apareados, tales como las indicadas con más detalle a continuación.

En algunas realizaciones, los ácidos nucleicos diana pueden obtenerse como fragmentos de uno o más ácidos nucleicos más grandes. La fragmentación puede realizarse usando cualquiera de una diversidad de técnicas conocidas en la técnica que incluyen, por ejemplo, nebulización, sonicación, ruptura química, ruptura enzimática o cizallamiento físico. La fragmentación también puede surgir del uso de una técnica de amplificación concreta que produce amplicones copiando solo una porción de un ácido nucleico más grande. Por ejemplo, la amplificación con PCR produce fragmentos que tienen un tamaño definido por la longitud del fragmento entre los cebadores flanqueantes usados para la amplificación.

Una población de ácidos nucleicos diana, o sus amplicones, puede tener una longitud promedio de hebra que es la deseada o la apropiada para una aplicación concreta de los métodos o las composiciones indicados en la presente. Por ejemplo, la longitud promedio de hebra puede ser menor que aproximadamente 100.000 nucleótidos, 50.000 nucleótidos, 10.000 nucleótidos, 5.000 nucleótidos, 1.000 nucleótidos, 500 nucleótidos, 100 nucleótidos, o 50 nucleótidos. Como alternativa o además, la longitud promedio de hebra puede ser mayor que aproximadamente 10 nucleótidos, 50 nucleótidos, 100 nucleótidos, 500 nucleótidos, 1.000 nucleótidos, 5.000 nucleótidos, 10.000 nucleótidos, 50.000 nucleótidos, o 100.000 nucleótidos. La longitud promedio de hebra para una población de ácidos nucleicos diana, o sus amplicones, puede estar en un intervalo entre un valor máximo y mínimo indicado anteriormente. Se entenderá que los amplicones generados en un sitio de amplificación (o fabricados o usados de otra manera en la presente) pueden tener una longitud promedio de hebra que esté en un intervalo entre un límite superior e inferior seleccionados de los ejemplificados anteriormente.

En algunos casos, puede producirse o configurarse de otro modo una población de ácidos nucleicos diana para que sus miembros tengan una longitud máxima. Por ejemplo, la longitud máxima para los miembros que se producen o se usan como se indica en la presente puede ser menor que aproximadamente 100.000 nucleótidos, 50.000 nucleótidos, 10.000 nucleótidos, 5.000 nucleótidos, 1.000 nucleótidos, 500 nucleótidos, 100 nucleótidos, o 50 nucleótidos. Como alternativa o además, puede producirse o configurarse de otro modo una población de ácidos nucleicos diana, o sus amplicones, bajo condiciones para que sus miembros tengan una longitud mínima. Por ejemplo, la longitud mínima para los miembros que se usan en una o más etapas de un método indicado en la presente o que están presentes en una composición concreta puede ser mayor que aproximadamente 10 nucleótidos, 50 nucleótidos, 100 nucleótidos, 500 nucleótidos, 1.000 nucleótidos, 5.000 nucleótidos, 10.000 nucleótidos, 50.000 nucleótidos, o 100.000 nucleótidos. La longitud de hebra máxima y mínima para los ácidos nucleicos diana en una población puede estar en un intervalo entre un valor máximo y mínimo indicado anteriormente. Se entenderá que los amplicones generados en un sitio de amplificación (o fabricados o usados de otra manera en la presente) pueden tener una longitud de hebra máxima y/o mínima que esté en un intervalo entre los límites superiores e inferiores ejemplificados anteriormente.

Puede usarse cualquiera de una diversidad de técnicas de amplificación conocidas para aumentar la cantidad de secuencias de molde presentes para su uso en un método indicado en la presente. Los ejemplos de técnicas incluyen, pero no se limitan a la reacción en cadena con polimerasa (PCR), la amplificación de círculo rodante ("rolling circle amplification", RCA), la amplificación de desplazamiento múltiple ("multiple displacement amplification", MDA), o la amplificación de cebadores aleatorios ("random prime amplification", RPA) de moléculas de ácidos nucleicos que contengan secuencias de molde. Se entenderá que la amplificación de ácidos nucleicos diana antes de su uso en un método o una composición indicados en la presente es opcional. Como tal, los ácidos nucleicos diana no se amplificarán antes del uso en algunas realizaciones de los métodos y las composiciones indicados en la presente. Los ácidos nucleicos diana pueden obtenerse opcionalmente a partir de bancos sintéticos. Los ácidos nucleicos sintéticos pueden tener composiciones de ADN o ARN nativas o pueden ser sus análogos. También pueden usarse métodos de amplificación en fase sólida que incluyen, por ejemplo, la amplificación de agrupamiento, la amplificación de puente u otros métodos indicados a continuación en el contexto de los métodos basados en matrices.

Un ácido nucleico usado en un método indicado en la presente puede estar en fase de disolución o en fase sólida. El ácido nucleico, cuando está en fase de disolución, en general es soluble, pero también puede estar en una forma suspendida que es capaz de precipitar, como es el caso de algunas especies de ácidos nucleicos grandes, tales como cromosomas o nanobolas de ácidos nucleicos (véase, por ejemplo, la patente de EE. UU. n.º de publicación 2007/0099208 A1). Un ácido nucleico que está en fase sólida puede proporcionarse en un soporte en fase sólida o sobre este. Los ejemplos de soportes en fase sólida incluyen los fabricados de vidrio, nitrocelulosa, sílice, metal, plástico y otros materiales indicados en otro punto en la presente, por ejemplo, con respecto a los formatos de ensayo y células de flujo. De forma similar, un ácido nucleico puede proporcionarse en un soporte semisólido o sobre este, tal como un gel. Los ejemplos de geles que son útiles incluyen, pero no se limitan a los geles que tienen una estructura coloidal, tal como agarosa; una estructura de malla de polímero, tal como gelatina; o una estructura de polímero reticulado, tal como poliacrilamida. Los hidrogeles son particularmente útiles, tales como los indicados en la patente de EE. UU. n.º de publicación 2011/0059865 A1, y el documento US 2014/0079923.

La unión de un ácido nucleico a un soporte, tanto rígido como semirrígido, puede producirse a través de uno o más enlaces covalentes o no covalentes. Se indican ejemplos de enlaces en las patentes de EE. UU. n.ºs 6.737.236; 7.259.258; 7.375.234 y 7.427.678; y la patente de EE. UU. n.º de publicación 2011/0059865 A1. En algunas realizaciones, un ácido nucleico u otro componente de reacción puede unirse a un gel u otro soporte semisólido que, a su vez, está unido o adherido a un soporte en fase sólida. En estas realizaciones, se entiende que el ácido nucleico u otro componente de reacción está en fase sólida.

Los sistemas de detección ortogonal pueden basarse en el uso de dos o más sistemas de reactivos para la extensión de los cebadores, en los que los componentes de los dos sistemas de reactivos no presentan una reactividad cruzada sustancial. Por ejemplo, un primer sistema de reactivos puede incluir una ADN polimerasa, desoxirribonucleótidos y un cebador de ADN; y un segundo sistema de reactivos puede incluir una ARN polimerasa, ribonucleótidos y un cebador de ARN. Ambos sistemas son capaces de actuar sobre un molde de ADN, por ejemplo, conteniendo uno de ellos un primer sitio de cebado que es complementario con el cebador de ADN, y un segundo sitio de cebado que es complementario con el cebador de ARN. Sin embargo, la ADN polimerasa es específica para el cebador de ADN y los desoxirribonucleótidos, de tal forma que extiende selectivamente el cebador de ADN con los desoxirribonucleótidos, en lugar de los ribonucleótidos. A la inversa, la ARN polimerasa es específica para el cebador de ARN y los ribonucleótidos, de tal forma que extiende selectivamente el cebador de ARN con los ribonucleótidos en lugar de los desoxirribonucleótidos. De modo similar, puede lograrse la ortogonalidad usando otros sistemas de reactivos específicos, tales como una polimerasa modificada que incorpora selectivamente HNA (ácidos nucleicos de 1,5-anhidrohexitol) a un cebador fabricado a partir de monómeros de HNA. La extensión de cebadores basada en HNA es ortogonal con respecto a los sistemas de extensión de ADN polimerasa y ARN polimerasa. Se describen ejemplos de condiciones y reactivos que pueden usarse para la extensión de cebadores basada en HNA en Pinheiro *et al.*, *Science*, 336(6079):341-344 (2012), y Cozens *et al.*, *Proc. Nat'l. Acad. Sci. USA*, 109(21):8067-8072 (2012).

Según los ejemplos de realizaciones indicados anteriormente, los desoxirribonucleótidos pueden considerarse una clase ortogonal de nucleótidos con respecto a los ribonucleótidos y HNA. De modo similar, en el contexto de realizaciones concretas, las clases de ADN polimerasas y ARN polimerasas son ortogonales entre sí, y las clases de cebadores de ADN y cebadores de ARN son ortogonales entre sí. En general, la ortogonalidad puede aprovecharse en un método indicado en la presente cuando una primera polimerasa es selectiva para una primera clase de análogos de nucleótidos, comparado con una segunda clase de análogos de nucleótidos, y cuando una segunda polimerasa es selectiva para la segunda clase de análogos de nucleótidos, comparado con la primera clase de análogos de nucleótidos. De forma similar, puede existir ortogonalidad cuando la primera polimerasa es selectiva para una primera clase de cebadores, comparado con una segunda clase de cebadores, y cuando la segunda polimerasa es selectiva para la segunda clase de cebadores, comparado con la primera clase de cebadores.

Puede usarse cualquiera de una diversidad de polimerasas en un método o una composición indicados en la presente, incluyendo, por ejemplo, enzimas basadas en proteínas aisladas a partir de sistemas biológicos y sus variantes funcionales. Se entenderá que la referencia a una polimerasa concreta, tales como las ejemplificadas a continuación, incluye sus variantes funcionales, a menos que se indique lo contrario. Una función particularmente útil de una polimerasa es catalizar la polimerización de una hebra de ácido nucleico usando un ácido nucleico existente como molde. Otras funciones que son útiles se describen en otro punto en la presente. Los ejemplos de polimerasas útiles incluyen ADN polimerasas, transcriptasas inversas y ARN polimerasas.

Una polimerasa que tiene una actividad exonucleasa correctora de errores 3' a 5' intrínseca puede ser útil en algunas realizaciones. Las polimerasas que carecen sustancialmente de actividad exonucleasa correctora de errores 3' a 5' también son útiles en algunas realizaciones, por ejemplo, en la mayoría de las realizaciones de secuenciación. La ausencia de actividad exonucleasa puede ser una característica de tipo salvaje o una característica impartida por una estructura de polimerasa variante o modificada. Por ejemplo, el fragmento de Klenow menos exo es una versión mutada del fragmento de Klenow que carece de actividad exonucleasa correctora de errores 3' a 5'.

Dependiendo de la realización que se vaya a usar, una polimerasa puede ser termófila o inactivable por calor. Las polimerasas termófilas son generalmente útiles para condiciones de alta temperatura o en condiciones de termociclación, tales como las empleadas para las técnicas de la reacción en cadena con polimerasa (PCR). Los ejemplos de polimerasas termófilas incluyen, pero no se limitan a 9° N ADN polimerasa, Taq ADN polimerasa, Phusion® ADN polimerasa, Pfu ADN polimerasa, RB69 ADN polimerasa, KOD ADN polimerasa y VentR® ADN polimerasa. La mayoría de las polimerasas aisladas a partir de organismos no termófilos son inactivables por calor. Los ejemplos son ADN polimerasas de fagos. Se entenderá que las polimerasas procedentes de cualquiera de una diversidad de fuentes pueden modificarse para aumentar o disminuir su tolerancia a condiciones de alta temperatura. Se describen polimerasas particularmente útiles para incorporar nucleótidos que contienen marcadores y/o restos terminadores en el documento US 2006/0281109 A1.

Otro sistema de reactivos ortogonales de extensión de cebadores es un sistema basado en ligasa que es selectivo para la incorporación de oligonucleótidos, en lugar de nucleótidos monoméricos, que son incorporados por los sistemas de extensión basados en polimerasa descritos anteriormente. Un sistema de reactivos de ADN ligasa es totalmente ortogonal con un sistema de reactivos basado en ARN polimerasa cuando se usa bajo condiciones en las que el cebador de ADN es extendido por la ADN ligasa, pero no por la ARN polimerasa, y en las que un cebador de ARN es extendido por la ARN polimerasa, pero no por la ADN ligasa. La extensión mediante acoplamiento puede realizarse en una aplicación de secuenciación usando una población de sondas oligonucleotídicas parcialmente aleatorias que presentan un esquema de codificación de una o dos bases. Las técnicas extensión basadas en el acoplamiento que pueden usarse para la detección de una reacción de extensión, tal como en un contexto de secuenciación, se describen en McKernan *et al.*, *Genome Research*, 19(9): 1527-41 (2009); Shendure *et al.*, *Science*, 309:1728-1732 (2005); y las patentes de EE. UU. n.ºs 5.599.675 y 5.750.341.

La manipulación y detección ortogonal según la presente descripción no requiere que dos secuencias de molde se diferencien en cada posición a lo largo de su longitud. Por el contrario, el mismo resto básico puede estar presente en posiciones que son detectadas sobre un primer molde y un segundo molde, respectivamente. Las dos posiciones pueden distinguirse basándose en las características distinguibles de los marcadores presentes en los sistemas de reactivos ortogonales y en la especificidad de los sistemas de reactivos para extender el cebador apropiado. Esta información puede emplearse, a su vez, para detectar de modo distinguible las dos secuencias de molde diferentes, incluso si las dos posiciones se detectan simultáneamente usando un detector que tenga una resolución que es demasiado baja para resolver puntos a una distancia equivalente al espaciamiento de las dos secuencias de molde.

Puede usarse cualquiera de una diversidad de marcadores. Un resto de marcador que es particularmente útil cuando se emplea para la detección de un análogo de nucleótido puede ser cualquier parte del análogo de nucleótido que proporciona una característica distinguible cuando se compara con otras moléculas presentes en su entorno. La característica distinguible puede ser, por ejemplo, una señal óptica, tal como la absorbancia de radiación, emisión de fluorescencia, emisión de luminiscencia, vida de la fluorescencia, polarización de la fluorescencia, o similares; la afinidad de unión por un ligando o receptor; propiedades magnéticas; propiedades eléctricas; carga; masa; radiactividad o similares. Los ejemplos de restos de marcadores incluyen, sin limitación, un fluoróforo, un luminóforo, un cromóforo, un isótopo radiactivo, un marcador de masa, un marcador de carga, un marcador de espín, un receptor, un ligando o similares. El resto marcador puede ser parte de un nucleótido que es

una unidad monomérica presente en un polímero de ácido nucleico, o el resto de marcador puede ser parte de un análogo de nucleótido libre (por ejemplo, un nucleótido trifosfato).

Los fluoróforos son particularmente útiles e incluyen, por ejemplo, nanocristales fluorescentes; puntos cuánticos, fluoresceína, rodamina, tetrametilrodamina, eosina, eritrosina, cumarina, metilcumarinas, pireno, verde de malaquita, Cy3, Cy5, estilbeno, amarillo de Lucifer, azul de cascada, rojo de Texas, tintes Alexa, tintes SETA, tintes Atto, ficoeritrina, BODIPY, y sus análogos. Se describen sondas ópticas útiles en Lakowicz, *Principles of Fluorescence Spectroscopy*, 3ª ed., Springer (2006); Haugland, *Handbook of Fluorescent Probes and Research Products* 9ª ed., Molecular Probes, Inc., (2002); Shapiro, *Practical Flow Cytometry*, 4ª ed., John Wiley & Sons (2003); documento WO 98/59066; documento WO 91/06678 o solicitud de patente de EE. UU. n.º de publicación 2010/0092957 A1.

Pueden usarse otros marcadores, algunos de los cuales son marcadores no ópticos, en diversas realizaciones de los métodos y las composiciones indicados en la presente. Los ejemplos incluyen, sin limitación, un marcador isotópico, tal como un isótopo pesado o radiactivo no abundante en la naturaleza; una sustancia magnética; un material rico en electrones, tal como un metal; un marcador electroquimioluminiscente, tal como Ru(bpy)³²⁺; o un resto que pueda detectarse basándose en una característica nuclear magnética, paramagnética, eléctrica, de carga a masa o térmica. Los marcadores también incluyen partículas magnéticas o nanopartículas codificadas ópticamente. Estos marcadores pueden detectarse usando métodos apropiados conocidos por los expertos en la técnica. Por ejemplo, un marcador cargado puede detectarse usando un detector eléctrico, tales como los que se emplean en sistemas de secuenciación disponibles en el mercado en Ion Torrent (Guilford, Conn., una filial de Life Technologies) o los sistemas de detección descritos en las solicitudes de patente de EE. UU. n.ºs de publicación 2009/0026082 A1; 2009/0127589 A1; 2010/0137143 A1; y 2010/0282617 A1. Se entenderá que, para algunas realizaciones, no es necesario que un análogo de nucleótido contenga uno o más de los marcadores indicados en la presente.

Un resto de marcador puede unirse a un nucleótido mediante una diversidad de formas. Se indican ejemplos de uniones y composiciones de marcadores que son útiles para los nucleótidos indicados en la presente en Bentley *et al.*, *Nature*, 456:53-59 (2008); documento WO 04/018497; documento US 7.057.026; documento WO 91/06678; documento WO 07/123744; documento US 7.329.492; documento US 7.211.414; documento US 7.315.019; documento US 7.405.281; y documento US 2008/01 08082.

En realizaciones concretas, por ejemplo, en las que se utiliza la extensión de cebadores cíclica en una estrategia de secuenciación por síntesis ortogonal, los nucleótidos pueden incluir restos terminadores reversibles. Los restos terminadores reversibles proporcionan una manera conveniente para controlar una reacción de extensión mediante la adición de un único nucleótido a un cebador hasta que se realiza una posterior etapa de desbloqueo. Esto puede entenderse en el contexto de una estrategia de secuenciación como sigue. Para iniciar un primer ciclo de secuenciación, pueden administrarse uno o más nucleótidos marcados, ADN polimerasas, etc., a una matriz de moldes de ácidos nucleicos unidos a cebadores. Opcionalmente, los nucleótidos pueden incluir un resto terminador reversible, de modo que no pueda producirse una posterior extensión hasta que se administre un agente desbloqueante para retirar el resto. Pueden detectarse dos o más marcadores añadidos a los sitios mediante las reacciones de extensión de cebadores, por ejemplo, usando métodos o aparatos indicados en la presente. Puede ponerse en contacto un agente desbloqueante con la matriz (antes o después de que se produzca la detección) para retirar el terminador reversible. Pueden realizarse lavados entre las diversas etapas de administración. El ciclo después puede repetirse n veces para extender los cebadores en n nucleótidos, detectando con ello secuencias de longitud n. Se describen ejemplos de técnicas de secuenciación y reactivos útiles, por ejemplo, en Bentley *et al.*, *Nature*, 456:53-59 (2008); documento WO 04/018497; documento US 7.057.026; documento WO 91/06678; documento WO 07/123744; documento US 7.329.492; documento US 7.211.414; documento US 7.315.019; documento US 7.405.281; y documento US 2008/01 08082.

Un método de secuenciación ortogonal indicado en la presente puede utilizarse en una estrategia de secuenciación de extremos apareados. En general, la secuenciación de extremos apareados implica determinar las secuencias en dos extremos de una región de secuencia de molde, en la que se conoce la longitud de la región de secuencia de molde. Se conocen métodos para fragmentar una muestra de ácido nucleico diana (por ejemplo, una muestra de ADN genómico), de unir los cebadores para adaptarse a las lecturas de los extremos apareados y de leer la secuencia desde los extremos de los fragmentos y pueden realizarse como se describe, por ejemplo, en las patentes de EE. UU. n.ºs 7.754.429; 8.017.335; y 8.192.930.

En el caso de una realización de secuenciación por síntesis ortogonal, pueden construirse fragmentos de ácidos nucleicos para que incluyan dos secuencias de molde y pueden obtenerse las lecturas apareadas de cada uno de los dos moldes para obtener 4 lecturas a partir de un único fragmento. Las lecturas de los extremos apareados pueden facilitarse mediante el uso de un promotor bidireccional flanqueado por sitios de unión de ARN polimerasa. En la figura 5 se muestra un ejemplo de construcción. En este ejemplo, la construcción incluye los sitios de cebado de la lectura 1 y la lectura 1' para una primera lectura ortogonal. En la primera lectura ortogonal, un cebador de ADN puede hibridarse con el sitio de cebado de la lectura 1 para permitir una lectura catalizada por ADN polimerasa de la secuencia en un primer extremo del molde 1 (indicado por la flecha negra en el diagrama superior). Además, en la primera lectura ortogonal, el sitio de cebado de la lectura 1' puede hibridarse con un cebador de ARN y, debido a la proximidad del promotor bidireccional, la ARN polimerasa pueden leer un primer extremo del molde 2 (indicado por

la flecha blanca en el diagrama superior). Puede obtenerse una segunda lectura ortogonal hibridando un cebador de ADN con el sitio de cebado de la lectura 2 y leyendo el segundo extremo del molde 2 e hibridando un cebador de ARN al sitio de cebado de la lectura 2' y leyendo el segundo extremo del molde 1 (indicado por la flecha negra en el diagrama inferior). La proximidad del promotor bidireccional al sitio de cebado de la lectura 2' permite que se produzca la extensión por la ARN polimerasa (indicado por la flecha blanca en el diagrama inferior). La construcción ejemplificada en la figura 5 puede realizarse, por ejemplo, usando los métodos descritos en la figura 4.

No es necesario un promotor bidireccional para las lecturas de extremos apareados usando una ARN polimerasa en una realización de secuenciación por síntesis ortogonal. Por el contrario, los sitios de cebado de ARN y sus promotores pueden localizarse en los extremos de una construcción de 2 moldes, y el adaptador que une los dos moldes puede contener sitios de cebado de ADN. Tomando la construcción de la figura 5 como ejemplo, las posiciones de los sitios de cebado de la lectura 1 y la lectura 1' pueden intercambiarse, las posiciones de los sitios de cebado de la lectura 2 y la lectura 2' pueden intercambiarse, el promotor bidireccional puede retirarse y los promotores de ARN distintos pueden localizarse cadena arriba de los sitios de cebado de la lectura 1' y la lectura 2', respectivamente.

Una reacción de extensión de ácidos nucleicos, u otra reacción cíclica, que se realiza usando los métodos indicados en la presente puede desarrollarse durante uno o más ciclos. En realizaciones concretas, una reacción de múltiples ciclos puede incluir al menos 2 ciclos, 5 ciclos, 10 ciclos, 50 ciclos, 100 ciclos, 500 ciclos, 1.000 ciclos, 5.000 ciclos, 10.000 ciclos o más. Como alternativa o además, una reacción puede tener un límite superior, según el cual no se realizan más de 1 ciclo, 2 ciclos, 5 ciclos, 10 ciclos, 50 ciclos, 100 ciclos, 500 ciclos, 1.000 ciclos, 5.000 ciclos, o 10.000 ciclos. En algunas realizaciones, cada ciclo provoca la incorporación de un único análogo de nucleótido a un cebador extendido. En este caso, puede entenderse que el número mínimo o máximo de ciclos ejemplificados anteriormente ejemplifica el número mínimo o máximo de nucleótidos incorporados en un producto de extensión en una reacción catalizada por polimerasa.

Algunas realizaciones pueden emplear reacciones de extensión no cíclicas, tales como reacciones de extensión de una sola base ("single base extension", SBE) o de extensión de cebadores específica de alelo ("allele specific primer extension", ASPE). Pueden usarse restos terminadores reversibles para la extensión no cíclica. Puesto que no es necesaria una etapa de desbloqueo para estas reacciones no cíclicas, los nucleótidos pueden terminarse de una manera no reversible. Por ejemplo, pueden emplearse didesoxinucleótidos. Se describen ejemplos de reactivos y técnicas relacionadas para SBE, ASPE y otras técnicas de extensión no cíclicas útiles, por ejemplo, en la patente de EE. UU. n.º 7.670.810 y las solicitudes de patente de EE. UU. n.ºs de publicaciones 2003/0108867; 2003/0108900; 2003/0170684; 2003/0207295; o 2005/0181394. Un ejemplo de un producto disponible en el mercado que emplea una técnica de extensión no cíclica y que puede modificarse para aumentar el contenido de información por medio de los métodos de detección ortogonal indicados en la presente es el producto de genotipificación Infinium®, disponible en Illumina, Inc. (San Diego, Calif.).

Tanto las reacciones cíclicas como las reacciones no cíclicas pueden incluir etapas en las que los componentes de reacción están separados entre sí o se retiran del entorno de reacción. Uno o más componentes de reacción pueden separarse, por ejemplo, por separación de los componentes en fase sólida de los componentes en fase líquida. Pueden incluirse opcionalmente etapas de lavada para retirar más completamente uno o más componentes en fase líquida no deseados de uno o más componentes en fase sólida. Un recipiente de reacción particularmente útil para dichas separaciones es una célula de flujo, tales como las que se emplean habitualmente en procedimientos de secuenciación cíclicos. Se describen ejemplos de células de flujo, métodos para su fabricación y métodos para su uso en las solicitudes de patente de EE. UU. n.ºs de publicación 2010/0111768 A1 y 2012/0270305 A1; y el documento WO 05/065814. Tanto si se emplean o no métodos de separación en fase sólida, los componentes de reacción pueden retirarse mediante cualquiera de una diversidad de otras técnicas conocidas en la técnica que incluye extracción de líquido-líquido, extracción en fase sólida, cromatografía, filtración, centrifugación o similares.

La detección puede realizarse en un método de la presente descripción usando un aparato adecuado para el marcador concreto en uso. Por ejemplo, puede emplearse un detector óptico, tal como un detector de fluorescencia, un detector de absorbancia, un detector de luminiscencia o similares, para detectar los marcadores ópticos apropiados. Los sistemas diseñados para una detección basada en matrices son particularmente útiles. Por ejemplo, pueden construirse sistemas ópticos para su uso con los métodos indicados en la presente para que incluyan diversos componentes y ensamblajes tal como se describe en las patentes de EE. UU. n.ºs 8.241.573; 7.329.860 y 8.039.817; y las solicitudes de patente de EE. UU. n.ºs de publicación 2009/0272914 A1 y 2012/0270305 A1.

Tal como se indicó anteriormente, un método de la presente descripción puede incluir dos etapas de extensión de cebadores ortogonales. Por ejemplo, anteriormente se indica un método que incluye, entre otras, las etapas de (b) extender un primer cebador unido a un primer ácido nucleico usando una primera especie de polimerasa y un primer conjunto de análogos de nucleótidos, produciendo con ello un primer producto de extensión de cebador que presenta un primer análogo de nucleótido en cada uno de los sitios; (c) extender un segundo cebador unido a un segundo ácido nucleico usando una segunda especie de polimerasa y un segundo conjunto de análogos de nucleótidos, produciendo con ello un segundo producto de extensión de cebador que presenta un segundo análogo de nucleótido en cada uno de los sitios, en el que la primera especie de polimerasa es diferente de la segunda especie de polimerasa, y en el que el primer conjunto de análogos de nucleótidos es diferente del segundo conjunto

de análogos de nucleótidos. En algunas realizaciones, las etapas (b) y (c) se realizan simultáneamente. Como alternativa, las etapas (b) y (c) pueden realizarse de modo secuencial en cualquier orden. En cualquiera de los casos, la ortogonalidad de las reacciones de extensión de cebadores permite distinguir los dos productos de extensión. Así, ambos productos de extensión pueden estar presentes simultáneamente durante una etapa de
5 detección y no es necesario que sean espacialmente resueltos por el detector usado.

Una reacción de múltiplex puede utilizar un soporte en fase sólida. Un soporte en fase sólida puede ser útil para separar reacciones individuales de modo que cada una pueda ser interrogada por separado o de modo individual. Por ejemplo, varios ácidos nucleicos diferentes en una mezcla pueden unirse al soporte en fase sólida. Los ácidos nucleicos pueden unirse al soporte en fase sólida en un formato de matriz.

10 En algunas realizaciones, se proporciona una matriz de sitios, en la que cada sitio incluye un primer molde de ácido nucleico y un segundo molde de ácido nucleico, y en la que el primer molde de ácido nucleico tiene una secuencia que es diferente de la secuencia del segundo molde de ácido nucleico. Los ejemplos de matrices que pueden ser útiles incluyen, sin limitación, una matriz BeadChip disponible en Illumina®, Inc. (San Diego, Calif.) o matrices tales como las descritas en las patentes de EE. UU. n.ºs 6.266.459; 6.355.431; 6.770.441; 6.859.570; o 7.622.294; o
15 publicación PCT n.º WO 00/63437. Otros ejemplos de matrices disponibles en el mercado que pueden usarse incluyen, por ejemplo, una matriz Affymetrix® GeneChip® u otra matriz sintetizada según técnicas que a veces se denominan tecnologías VLSIPS™ ("Very Large Scale Immobilized Polymer Synthesis", síntesis de polímeros inmovilizados a escala muy grande). También puede usarse una matriz de manchas según algunas realizaciones. Un ejemplo de matriz de manchas es una matriz CodeLink™ disponible en Amersham Biosciences. Otra matriz que
20 es útil es una matriz fabricada usando métodos de impresión de inyección de tinta, tal como la tecnología SurePrint™ disponible en Agilent Technologies.

Otras matrices útiles incluyen las que se emplean en aplicaciones de secuenciación de ácidos nucleicos. Por ejemplo, las matrices que contienen amplicones de fragmentos genómicos (a menudo denominados agrupamientos) son particularmente útiles, tales como las descritas en Bentley *et al.*, *Nature*, 456:53-59 (2008); el documento WO
25 04/018497; documento US 7.057.026; documento WO 91/06678; documento WO 07/123744; documento US 7.329.492; documento US 7.211.414; documento US 7.315.019; documento US 7.405.281; o documento US 2008/0108082.

Pueden crearse agrupamientos de ácidos nucleicos mediante métodos de amplificación en fase sólida. Por ejemplo, un ácido nucleico que contenga una o más secuencias de molde que se van a detectar puede unirse a una
30 superficie y amplificarse usando la amplificación de puente. Se describen métodos de amplificación de puente útiles, por ejemplo, en la patente de EE. UU. n.º 5.641.658; la patente de EE. UU. n.º de publicación 2002/0055100; la patente de EE. UU. n.º 7.115.400; la patente de EE. UU. n.º de publicación 2004/0096853; la patente de EE. UU. n.º de publicación 2004/0002090; la patente de EE. UU. n.º de publicación 2007/0128624; y la patente de EE. UU. n.º de publicación 2008/0009420. Otro método útil para amplificar ácidos nucleicos sobre una superficie es la
35 amplificación de círculo rodante ("rolling circle amplification", RCA), por ejemplo, como se describe en Lizardi *et al.*, *Nat. Genet.*, 19:225-232 (1998), y la solicitud de patente de EE. UU. n.º de publicación 2007/0099208 A1. Otro tipo de matriz que es útil es una matriz de partículas producida a partir de una técnica de amplificación de PCR en emulsión. Se describen ejemplos en Dressman *et al.*, *Proc. Natl. Acad. Sci. USA*, 100:8817-8822 (2003), y los documentos WO 05/010145, US 2005/0130173 o US 2005/0064460. Aunque las anteriores matrices se han descrito
40 en el contexto de aplicaciones de secuenciación, se entenderá que las matrices pueden usarse en otras realizaciones que incluyen, por ejemplo, las que emplean una técnica de extensión de cebadores no cíclica.

La detección puede realizarse a un nivel de molécula individual o de conjunto de moléculas en una matriz. La detección a nivel de conjunto es la detección que se produce de tal forma que se detectan varias copias de una
45 única secuencia de molde en cada sitio individual, y las copias individuales en el sitio no se distinguen entre sí. Así, la detección de conjunto proporciona una señal promedio de una secuencia molde concreta en el sitio. Por ejemplo, el sitio puede contener al menos 10, 100, 1000 o más copias de una secuencia de molde concreta. Por supuesto, un sitio puede contener múltiples secuencias de molde diferentes, cada una de las cuales está presente como un conjunto. Como alternativa, la detección al nivel de una sola molécula incluye la detección que se produce de tal forma que las secuencias de molde individuales se resuelven individualmente sobre la matriz, cada una en un sitio
50 diferente. Así, la detección de una única molécula proporciona una señal procedente de una molécula individual que se distingue de una o más señales que pueden surgir de una población de moléculas dentro de la cual está presente la molécula individual. Por supuesto, incluso en una matriz de moléculas individuales, un sitio puede contener varias secuencias de molde diferentes (por ejemplo, dos o más regiones de secuencia de molde localizadas a lo largo de una única molécula de ácido nucleico).

55 Una matriz de sitios puede aparecer como una cuadrícula de manchas o parches. Los sitios pueden localizarse en un patrón repetitivo o en un patrón no repetitivo irregular. Los patrones particularmente útiles son patrones hexagonales, patrones rectilíneos, patrones en cuadrícula, patrones que tienen una simetría especular, patrones que tienen una simetría rotacional o similares. Los patrones asimétricos también pueden ser útiles.

60 El tamaño de los sitios y/o el espaciado entre los sitios en una matriz puede variar para lograr una alta densidad, una densidad intermedia o una densidad más baja. Las matrices de alta densidad se caracterizan porque presentan

5 sitios separados por menos de aproximadamente 15 μm . Las matrices de densidad intermedia presentan sitios separados por aproximadamente 15 a 30 μm , mientras que las matrices de baja densidad presentan sitios separados por más de 30 μm . Una matriz útil en algunas realizaciones puede presentar sitios que están separados por menos de 100 μm , 50 μm , 10 μm , 5 μm , 1 μm , o 0,5 μm . Una realización de los métodos indicados en la presente puede usarse para formar una imagen de una matriz con una resolución suficiente para distinguir sitios a las anteriores densidades o intervalos de densidad. Sin embargo, la etapa de detección generalmente empleará un detector que tenga una resolución espacial que es demasiado baja para resolver puntos a una distancia equivalente al espaciamiento entre el primer producto de extensión de cebador y el segundo producto de extensión de cebador en cada uno de los sitios. En realizaciones concretas, los sitios de una matriz pueden tener cada uno área que es mayor que aproximadamente 100 nm^2 , 250 nm^2 , 500 nm^2 , 1 μm^2 , 2,5 μm^2 , 5 μm^2 , 10 μm^2 , 100 μm^2 , o 500 μm^2 . Como alternativa o además, los sitios de una matriz pueden tener cada uno área que es menor que aproximadamente 1 mm^2 , 500 μm^2 , 100 μm^2 , 25 μm^2 , 10 μm^2 , 5 μm^2 , 1 μm^2 , 500 nm^2 , o 100 nm^2 . En efecto, un sitio puede tener un tamaño que esté en el intervalo entre un límite superior e inferior seleccionado a partir de los ejemplificados anteriormente.

15 Los métodos indicados en la presente pueden emplear matrices que presentan sitios con una cualquiera de una diversidad de densidades que incluyen, por ejemplo, al menos aproximadamente 10 sitios/ cm^2 , 100 sitios/ cm^2 , 500 sitios/ cm^2 , 1.000 sitios/ cm^2 , 5.000 sitios/ cm^2 , 10.000 sitios/ cm^2 , 50.000 sitios/ cm^2 , 100.000 sitios/ cm^2 , 1.000.000 sitios/ cm^2 , 5.000.000 sitios/ cm^2 , o más.

20 Un sistema de detección ortogonal, tal como un sistema usado para la secuenciación por síntesis ortogonal, puede emplear diferentes marcadores para distinguir diferentes nucleótidos que se añaden a cada cebador. En una realización, cada especie de nucleótido contendrá un marcador óptico exclusivo que produce una señal exclusiva para distinguir esa especie de nucleótido. Un ejemplo es la estrategia de SBOS de 4 tintes descrita en el siguiente ejemplo I y mostrada en la figura 1A y la figura 1B. En este ejemplo, se emplea un primer conjunto de 4 tintes fluorescentes diferentes para distinguir los 4 análogos de dNTP diferentes entre sí, y un segundo conjunto de 4 tintes fluorescentes diferentes para distinguir los 4 análogos de rNTP diferentes entre sí. Los dos conjuntos de tintes son exclusivos, de modo que los 8 tintes producen 8 señales distinguibles, respectivamente.

25 En realizaciones en las que todos los nucleótidos están marcados de modo distinguible, tales como la estrategia de SBOS de 4 tintes, una pareja de secuencias de molde puede ponerse en contacto con todos los nucleótidos y después la detección puede realizarse posteriormente. En este caso, la capacidad para distinguir todos los nucleótidos debido a los marcadores ópticos exclusivos proporciona el beneficio de una manipulación fluidica relativamente sencilla, por la cual todos los nucleótidos pueden administrarse a las secuencias de molde de modo que estén simultáneamente presentes. En una realización de SBOS preferida y relativamente sencilla, los 8 nucleótidos se administran a la vez; sin embargo, pueden administrarse uno o más subconjuntos secuencialmente si se desea. La detección puede producirse durante o después de la administración de los nucleótidos. Este proceso fluidoico relativamente sencillo está alojado en un dispositivo de detección relativamente complejo que tiene la capacidad de distinguir todas las señales. Por ejemplo, puede utilizarse un sistema de detección de fluorescencia capaz de distinguir 8 señales fluorescentes diferentes para una estrategia de SBOS que utiliza 8 nucleótidos marcados de modo fluorescente distintos. Los expertos en la técnica sabrán o serán capaces de determinar el aparato de detección de fluorescencia apropiado para lograr este tipo de diferenciación de señal. Por ejemplo, las propiedades de excitación y emisión de los marcadores fluorescentes pueden hacerse corresponder de modo apropiado con una combinación de longitudes de onda de excitación producidas y unas longitudes de onda de emisión detectadas por un fluorómetro. Se proporcionan ejemplos de guías para la óptica y los marcadores útiles para la detección de fluorescencia de múltiples longitudes de onda en Lakowicz, *Principles of Fluorescence Spectroscopy*, 3ª ed., Springer (2006); Haugland, *Handbook of Fluorescent Probes and Research Products*, 9ª ed., Molecular Probes, Inc. (2002); y Shapiro, *Practical Flow Cytometry*, 4ª ed., John Wiley & Sons (2003).

30 Los principios ejemplificados anteriormente para un sistema en el que todos los nucleótidos están marcados de modo distinguible pueden extenderse con facilidad a un formato de matriz. Se espera que una matriz que tenga un número y una diversidad suficiente de secuencias de molde diferentes incorporará todos los nucleótidos marcados cuando se trata con sistemas de reacción de extensión de cebadores. De modo más específico, en una estrategia de SBOS basada en una matriz que presenta una amplia diversidad de ácidos nucleicos a través de los sitios de la matriz y que contiene dos moldes diferentes por sitio, se espera que se produzcan todas las combinaciones de 2 tintes en la matriz, tras un ciclo de extensión de cebador en el que los 8 nucleótidos se administran a la matriz. Los sitios pueden distinguirse espacialmente usando dispositivos ópticos conocidos en la técnica, por ejemplo, los descritos en las patentes de EE. UU. n.ºs 8.241.573; 7.329.860 y 8.039.817; y las solicitudes de patente de EE. UU. n.ºs de publicación 2009/0272914 A1 y 2012/0270305 A1. Estos sistemas de detección pueden modificarse con facilidad para adaptarse a la detección fluorescente de 8 colores, tal como se indicó anteriormente. Un sistema de detección que se modifica de esta forma será capaz de una detección ortogonal de múltiple, de modo que se distinguen dos moldes diferentes (por ejemplo, mediante secuenciación) en múltiples sitios que tienen, cada uno, una composición de secuencia diferente.

60 En algunas realizaciones, el número de señales diferentes que se distinguen en un método concreto es menor que el número de especies de nucleótidos diferentes empleadas en dicho método. Por ejemplo, múltiples especies de nucleótidos diferentes pueden contener el mismo marcador y/o un subconjunto de especies de nucleótidos puede

estar no marcado. Un ejemplo de una configuración que emplea el mismo marcador para múltiples especies de nucleótidos diferentes es el caso de un método de extensión de cebadores ortogonal (o SBOS), en el que 4 desoxirribonucleótidos diferentes contienen un primer marcador en común y 4 ribonucleótidos diferentes contienen un segundo marcador en común. En esta configuración, los 4 desoxirribonucleótidos diferentes pueden distinguirse entre sí por medio de ciclos secuenciales de administración de uno de los desoxirribonucleótidos y la detección de los desoxirribonucleótidos antes de administrar el posterior desoxirribonucleótido. Con la condición de que el primer marcador y el segundo marcador en este ejemplo sean distinguibles, los desoxirribonucleótidos y los ribonucleótidos pueden administrarse en parejas (1 de cada de una única especie de desoxirribonucleótido y una única especie de ribonucleótido) en 4 ciclos de administración y detección. Por tanto, los miembros de un primer conjunto de análogos de nucleótidos usados en una reacción de extensión de cebadores (por ejemplo, dNTP) puede incluir solo un tipo de marcador óptico que es detectado, y un segundo conjunto de análogos de nucleótidos, que es ortogonal con respecto al primer conjunto (por ejemplo, rNTP) también puede incluir solo un tipo de marcador óptico que es detectado, y el marcador usado en el primer conjunto puede distinguirse ópticamente del marcador usado en el segundo conjunto.

Una escala de grises permite el uso de múltiples especies de nucleótidos diferentes que contienen el mismo marcador. En este caso pueden distinguirse diferentes especies de nucleótidos basándose en la intensidad de la señal del marcador detectada. Por ejemplo, cada especie de nucleótido puede administrarse como una mezcla con una proporción exclusiva de esta especie de nucleótido en forma marcada y no marcada. La variación en la proporción de nucleótido marcado: no marcado para cada especie produce una salida de señal en escala de grises exclusiva para cada mezcla. Como ejemplo más específico, un primer nucleótido puede estar totalmente marcado (no hay mezcla de un primer nucleótido marcado y no marcado), un segundo nucleótido puede estar 75% marcado (una mezcla de un segundo nucleótido 75% marcado y el mismo nucleótido 25% marcado), un tercer nucleótido puede estar 50% marcado (una mezcla de un tercer nucleótido 50% marcado y el mismo nucleótido 50% no marcado), y un cuarto nucleótido puede estar 25% marcado (una mezcla de un cuarto nucleótido 25% marcado y el mismo nucleótido 75% no marcado). Estas 4 especies de nucleótidos pueden distinguirse basándose en las diferencias resultantes en la intensidad de la señal, por lo cual una población de cebadores (por ejemplo, en un sitio de una matriz) producirá una señal completa debido a la incorporación del primer nucleótido, 75% de la señal debido a la incorporación del segundo nucleótido, 50% de la señal debido a la incorporación del tercer nucleótido, y 25% de la señal debido a la incorporación del cuarto nucleótido.

En realizaciones concretas, al menos una de las especies de nucleótidos puede estar totalmente no marcada. Así, en el caso en que estén presentes marcadores ópticos sobre los otros nucleótidos en un conjunto de nucleótidos, también puede existir un nucleótido "oscuro". La extensión de un cebador para que incorpore un nucleótido oscuro, o no marcado, puede determinarse mediante una inferencia basada en la ausencia de un marcador cuya presencia se esperaría si los otros nucleótidos en el conjunto se hubieran incorporado a través de la reacción de extensión. Por tanto, en algunas realizaciones solo es necesario que un subconjunto de los nucleótidos usados en una reacción de extensión de cebador contenga un marcador.

El uso de una especie de nucleótido totalmente no marcado puede combinarse con la escala de grises. Por ejemplo, tres de cuatro especies de nucleótidos diferentes en un conjunto puede contener cantidades distinguibles distintas de cero de un marcador concreto (por ejemplo, proporciones de nucleótidos marcados y no marcados en una mezcla), y la cuarta especie de nucleótido puede carecer de ese marcador. Como alternativa o además, la escala de grises puede combinarse con el uso de varios marcadores ópticamente distinguibles. Por ejemplo, algunas especies de nucleótidos pueden estar representadas en una reacción de extensión en forma de una mezcla de nucleótidos del mismo tipo pero con diferentes marcadores. Esta configuración se ejemplifica en el siguiente ejemplo I, en el que una especie de nucleótido se proporciona como una mezcla de 50% de rCTP-F_{rojo}/50% de rCTP-F_{azul}. Otros ejemplos de escalas de grises y marcadores mixtos que pueden modificarse para su uso en un método ortogonal de la presente descripción se indican en el documento US 2013/0079232 A1.

Como alternativa o además del uso de múltiples marcadores diferentes, escala de grises y/o especies no marcadas, una realización indicada en la presente puede emplear un nucleótido que contiene un ligando, un conector rompible u otro resto que proporciona la ganancia o la pérdida de un marcador debido a un tratamiento definido. En el siguiente ejemplo I se ilustran sistemas de reactivos de este tipo, en los que algunas especies de nucleótidos contienen un ligando de modo que pueden distinguirse de otros nucleótidos basándose en la ausencia inicial de una señal detectable, seguido de la aparición de una señal después de un tratamiento con un receptor marcado de forma apropiada. El ejemplo I también ilustra el uso de un nucleótido que puede distinguirse basándose en una señal detectable inicial que posteriormente se pierde o al menos se reduce debido a un tratamiento con un reactivo que modifica el marcador (por ejemplo, a través de la ruptura química de un conector entre el marcador y el nucleótido). En este caso, la otra especie de nucleótido en el conjunto no es susceptible a la modificación (por ejemplo, carece del conector rompible) y ambas se distinguen basándose en la persistencia de la generación de señal después del tratamiento.

Tal como se ejemplifica anteriormente y en el ejemplo I, en algunas realizaciones puede unirse un marcador a un análogo de nucleótido a través de un conector rompible. En realizaciones concretas pueden usarse conectores fotorrompibles en lugar del conector que puede romperse de modo químico ejemplificado anteriormente. En algunas realizaciones, el conector se selecciona de conectores lábiles frente a ácidos (que incluyen conectores de

dialcoxibencilo, conectores de Sieber, conectores de indol, conectores de Sieber de t-butilo), conectores que pueden romperse electrofílicamente, conectores que pueden romperse nucleofílicamente, conectores fotorrumpibles, conectores que pueden romperse bajo condiciones reductoras u oxidativas, conectores de cierre de seguridad ("safety-catch"), y conectores que pueden romperse por medio de mecanismos de eliminación. En algunas de estas realizaciones, el conector se selecciona de un conector de disulfuro (—S—S—), éster, nitrobenzoceno, imina, péptidos y polinucleótidos, tales como ADN, que pueden romperse de modo enzimático o químico.

En algunas realizaciones, los miembros de un primer conjunto de análogos de nucleótidos usados en una reacción de extensión de cebadores (por ejemplo, dNTP) incluirán solo un tipo de marcador óptico que es detectado, y un segundo conjunto de análogos de nucleótidos, que es ortogonal con respecto al primer conjunto por ejemplo, (por ejemplo, rNTP) también incluirá solo un tipo de marcador óptico que es detectado, y el marcador usado en el primer conjunto puede distinguirse ópticamente del marcador usado en el segundo conjunto. En esta realización, dicho un tipo de marcador óptico puede unirse sustancialmente a todos los análogos de nucleótidos de una primera especie en el primer conjunto, dicho un tipo de marcador óptico puede unirse a un subconjunto de los análogos de nucleótidos de una segunda especie en el primer conjunto, sustancialmente todos los análogos de nucleótidos de una tercera especie en el primer conjunto pueden unirse a un ligando, y sustancialmente todos los análogos de nucleótidos de una cuarta especie en el primer conjunto no se unen a dicho un tipo de marcador óptico o al ligando.

En otra realización, los miembros de un primer conjunto de análogos de nucleótidos usados en una reacción de extensión de cebadores (por ejemplo, dNTP) incluirán solo dos tipos de marcadores ópticos que son detectados, y un segundo conjunto de análogos de nucleótidos, que es ortogonal con respecto al primer conjunto (por ejemplo, rNTP) también incluirá solo dos tipos de marcadores ópticos que son detectados. En esta realización, un primero de los dos tipos de marcadores ópticos puede unirse sustancialmente a todos los análogos de nucleótidos de una primera especie en el primer conjunto, un segundo de los dos tipos de marcadores ópticos puede unirse sustancialmente a todos los análogos de nucleótidos de una segunda especie en el primer conjunto, el primero de los dos tipos de marcadores ópticos y el segundo de los dos tipos de marcadores ópticos puede unirse a los análogos de nucleótidos de una tercera especie en el primer conjunto, y sustancialmente todos los análogos de nucleótidos de una cuarta especie en el primer conjunto no se unen a uno de los dos tipos de marcadores ópticos o al segundo de los dos tipos de marcadores ópticos.

A partir de los anteriores ejemplos, se entenderá que la reducción en el número de marcadores diferentes en un sistema de detección ortogonal tendrá la ventaja de reducir la complejidad del dispositivo de detección necesario para distinguir la adición de diferentes nucleótidos a un cebador unido a un molde. Sin embargo, en muchas realizaciones esto se logra aumentando la complejidad de las etapas fluídicas de modo que el número de manipulaciones fluídicas usadas durante las etapas de detección aumenta, comparado con las etapas fluídicas usadas cuando cada una de las especies de nucleótidos contiene un marcador exclusivo. Una ventaja general de los presentes métodos es que los expertos en la técnica pueden seleccionar una combinación apropiada de marcadores, etapas fluídicas y dispositivos de detección para que se ajusten a una aplicación o circunstancia concretas.

La presente descripción proporciona mezclas de reacción (también denominadas en la presente sistemas de reactivos) que incluyen diversas combinaciones de componentes. En varios casos se describen componentes de reacción y varias combinaciones de componentes en el contexto de ejemplos de métodos. Se entenderá que las mezclas de reacción y sus componentes no se limitan al uso en los métodos ejemplificados en la presente. También se contemplan otros usos. Por consiguiente, los componentes pueden ensamblarse en una diversidad de combinaciones útiles, por ejemplo, para crear kits. Los kits pueden ser útiles para el almacenaje, el transporte o la transacción comercial de los componentes indicados en la presente. Los kits pueden incluir opcionalmente instrucciones para realizar uno o más de los métodos indicados en la presente.

Los siguientes ejemplos pretenden ilustrar la presente invención, pero no limitarla.

Ejemplo I: Secuenciación por síntesis ortogonal con dos cebadores

Este ejemplo describe una nueva plataforma de secuenciación que permite doblar los resultados de secuenciación cuando se compara con la secuenciación por síntesis (SBS) tradicional. Esta plataforma es mejor que SBS porque la información de secuenciación se deriva de un alargamiento en etapas del cebador de secuenciación (figura 1A). Sin embargo, en la nueva plataforma, el acontecimiento de alargamiento en el primer sitio se produce en paralelo a un segundo acontecimiento de alargamiento de secuenciación, que se produce en un segundo sitio cadena abajo del primer sitio (figura 1B).

La ortogonalidad entre los sitios 1 y 2 de SBS es proporcionada por el uso de dos combinaciones diferentes de polimerasas y sustrato. Tal como como se ejemplifica a continuación, el sistema puede usar una ADN polimerasa de SBS y nucleótidos totalmente funcionales ("fully functional nucleotides", FFN), tales como los disponibles en Illumina, Inc. (San Diego, Calif.), en combinación con una ARN polimerasa (por ejemplo, T7 ARN polimerasa) y los correspondientes rNTP marcados (figura 1C). Aunque la discriminación en la especificidad de sustrato entre las ARN y ADN polimerasas no es absoluta, se ha demostrado que la diferencia en la preferencia de sustrato entre los rNTP y los dNTP es tan alta como de 10000 veces en ciertas condiciones (Joyce, *Proc. Nat'l. Acad. Sci. USA*, 94:1619-

1622 (1997); y Gao *et al.*, *Proc. Nat'l. Acad. Sci. USA*, 94:407-411 (1997)). Como alternativa, puede lograrse un sistema ortogonal empleando análogos de ADN/ARN sintéticos (AXN) y una correspondiente polimerasa modificada. Dicho sistema es posible empleando una síntesis con moldes de ADN de HNA largos (ácidos nucleicos de 1,5-anhidrohexitol) con una ADN polimerasa evolucionada. Se describen ejemplos de condiciones y reactivos que pueden usarse para la síntesis con moldes de ADN de HNA largos en Pinheiro *et al.*, *Science*, 336 (6079):341-344 (2012); y Cozens *et al.*, *Proc. Nat'l. Acad. Sci. USA*, 109(21):8067-8072 (2012).

Discriminación de la señal

Pueden usarse los siguientes ejemplos de configuraciones para la discriminación de la señal en una plataforma de secuenciación por síntesis ortogonal (SBOS).

10 (1) Estrategia basada en la química de SBOS de 4 tintes

En esta estrategia se emplea un conjunto de cuatro desoxirribonucleótidos terminadores reversibles (rtdNTP), en la que cada una de las especies de rtdNTP contiene un fluoróforo que es ópticamente distinguible de los fluoróforos usados en las otras tres especies de rtdNTP. Esto es parecido a la combinación de nucleótidos usados en las plataformas de SBS de 4 tintes disponibles en el mercado en Illumina, Inc. (San Diego, Calif.). También se emplea un conjunto de cuatro ribonucleótidos terminadores reversibles (rtrNTP), y cada una de las especies de rtrNTP contiene un fluoróforo que es ópticamente distinguible de los fluoróforos usados en las otras tres especies de rtrNTP. Además, los fluoróforos usados para los rtdNTP son ópticamente distinguibles de los fluoróforos para los rtrNTP. Así, se emplean 8 fluoróforos diferentes a través de los dos conjuntos de nucleótidos. La tabla 1 muestra un ejemplo de conjunto, en el que el subíndice se refiere a la longitud de onda de emisión de cada fluoróforo (F).

20 Tabla 1 - Nucleótidos para la configuración de SBS de 4 tintes

Conjunto de rtdNTP	Conjunto de rtrNTP
rtdTTP-F _{rojo lejano}	rtrUTP-F _{verde}
rtdCTP-F _{rojo cercano}	rtrCTP-F _{azul}
rtdGTP-F _{naranja}	rtrGTP-F _{índigo}
rtdATP-F _{amarillo}	rtrATP-F _{violeta}

En la estrategia de 4 tintes, los componentes ópticos usados para distinguir los 8 fluoróforos diferentes pueden ser relativamente complejos, por ejemplo, empleando hasta ocho canales de emisión diferentes y/o hasta ocho líneas de excitación diferentes.

25 (2) Estrategia basada en la química de SBOS de 1 tinte

Para reducir la complejidad del dispositivo óptico descrito para la estrategia de 4 tintes, puede usarse otra estrategia que emplea menos fluoróforos que especies de nucleótidos. Una estrategia de 1 tinte puede emplear un conjunto de rtdNTP que tengan un único fluoróforo de un primer tipo (por ejemplo, un fluoróforo de emisión en azul). Las diferentes especies de rtdNTP pueden distinguirse entre sí debido a la presencia del fluoróforo sobre una primera especie en el conjunto de los rtdNTP (por ejemplo, rtdTTP-fluoróforo de emisión en azul), la presencia de un ligando de unión sobre una segunda especie en el conjunto de los rtdNTP (por ejemplo, rtdCTP-biotina), la ausencia del fluoróforo y ligando de unión sobre una tercera especie en el conjunto de los rtdNTP (por ejemplo, rtdGTP no marcado), y la unión del fluoróforo a través de un conector rompible a una cuarta especie en el conjunto de los rtdNTP (por ejemplo, rtdATP-enlace disulfuro-fluoróforo de emisión en azul). Se indican con más detalle ejemplos de combinaciones de especies de nucleótidos marcados con fluoróforos, marcados con ligandos y no marcados que pueden usarse para crear un conjunto de nucleótidos para la detección de 1 tinte en el documento US 2013/0079232 A1.

La estrategia de 1 tinte también puede emplear un conjunto de rtrNTP que contienen un segundo tipo de fluoróforo. El segundo tipo de fluoróforo (por ejemplo, un fluoróforo de emisión en rojo) es ópticamente distinguible del fluoróforo usado para los miembros del conjunto de rtdNTP. Además, las especies individuales en el conjunto de rtrNTP pueden distinguirse entre sí mediante una combinación de especies marcadas y no marcadas similar a la combinación ejemplificada anteriormente para el conjunto de rtdNTP, con la siguiente modificación. Para evitar la introducción de un nuevo sistema de biotina-estreptavidina, uno de los rtrNTP puede ser una mezcla de los dos fluoróforos en uso. En la tabla 2 se muestra un ejemplo de conjunto de nucleótidos (se muestra un conjunto similar en la figura 2A).

Tabla 2 - Nucleótidos para la configuración de SBS de 1 tinte

Conjunto de rtdNTP	Conjunto de rtrNTP
rtdTTP-F _{azul}	rtrUTP-F _{rojo}
rtdCTP-biotina	50% de rtrCTP-F _{rojo} /50% de rtrCTP-F _{azul}
rtdGTP (no marcado)	rtrGTP (no marcado)
rtdATP-S-S-F _{azul}	rtrATP-S-S-F _{rojo}

Para una estrategia de SBOS que emplea la configuración de la tabla 2, solo es necesario que los componentes ópticos incluyan 2 canales de emisión diferentes y solo 1 o 2 líneas de excitación diferentes. En esta estrategia, tal como se muestra en la figura 2A, se adquiere un total de 4 imágenes por ciclo (2 por color). Dos imágenes se registran antes y dos después del tratamiento con estreptavidina--F_{azul}. En la figura 2A y la figura 2B, “oscuro” indica la ausencia de marcador de fluoróforo, “NR550C4” es el fluoróforo de emisión en azul, “THP” es tris(3-hidroxipropil)fosfina que rompe el enlace disulfuro (“SS”), “Estrep” es estreptavidina, “d-FFN” es rtdNTP, “r-FFN” es rtrNTP y “rojo” es el fluoróforo de emisión en rojo.

Aunque la mezcla de especies marcadas con rtrCTP se ejemplifica anteriormente como una proporción molar 1:1 (es decir, 50% de F_{rojo} y 50% de F_{azul}), se entenderá que pueden usarse otras proporciones. Por ejemplo, puede resultar deseable ajustar la proporción para adaptarse a las diferentes propiedades ópticas para los dos fluoróforos (por ejemplo, uno de los fluoróforos puede estar presente en un ligero exceso molar para ajustarse a una intensidad de emisión relativamente menor que el otro fluoróforo a la longitud de onda de detección que se está usando). La proporción también puede desviarse de 1:1 para adaptarse a las diferentes propiedades bioquímicas de los nucleótidos marcados (por ejemplo, diferente afinidad por la polimerasa).

(3) Estrategia basada en la química de SBOS de 2 tintes

Esta estrategia proporciona otra combinación de nucleótidos con menos fluoróforos que las especies de nucleótidos. Puede usarse un conjunto de rtrNTP que contienen fluoróforos con dos emisiones diferentes (por ejemplo, emisión en el rojo lejano y azul). De modo similar, puede usarse un conjunto de rtdNTP que contienen fluoróforos con dos emisiones diferentes. Las cuatro emisiones diferentes son ópticamente distinguibles a través de los dos conjuntos de rNTP. En la tabla 3 se muestra un ejemplo de conjunto de nucleótidos.

Tabla 3 - Nucleótidos para la configuración de SBS de 2 tintes

Conjunto de rtdNTP	Conjunto de rtrNTP
50% de rtdTTP-F _{verde} /50% de rtdTTP-F _{rojo}	rtrUTP-F _{rojo lejano}
rtdCTP-F _{rojo}	50% de rtrCTP-F _{rojo lejano} /50% de rtrCTP-F _{azul}
rtdGTP (no marcado)	rtrGTP (no marcado)
rtdATP-F _{verde}	rtrATP-F _{azul}

La estrategia de 2 tintes ejemplificada anteriormente proporciona una ventaja frente a la estrategia de 1 tinte previamente ejemplificada en que no se emplean etapas de procesamiento químico, tal como la unión de estreptavidina y la ruptura de THP, durante la fase de detección de la estrategia de 2 tintes. Esto se demuestra por medio del protocolo de ciclo de extensión mostrado para la estrategia de 2 tintes en la figura 3. Sin embargo, en el caso de esta estrategia de 2 tintes concreta, se emplea un dispositivo óptico capaz de distinguir cuatro colores diferentes. Tal como se ilustra mediante esta comparación, pueden emplearse diferentes estrategias para ajustar la complejidad óptica o la complejidad fluidica para adaptarse a una plataforma de secuenciación concreta. Por ejemplo, la adición de etapas fluidicas para modificar los nucleótidos durante un ciclo de detección de SBS permite usar menos componentes ópticos y/o componentes ópticos más sencillos. A la inversa, en los casos en que están disponibles componentes ópticos más complejos, pueden usarse menos manipulaciones fluidicas y/o manipulaciones fluidicas más sencillas.

Preparación de la muestra

Puede prepararse una muestra de ADN genómico o de otro ácido nucleico para una SBS de 2 cebadores como se ejemplifica en la figura 4. El método permite la creación de un molde de ADN que tiene dos sitios de cebado diferentes. El método puede realizarse bajo condiciones que proporcionan fragmentes en un intervalo de tamaño

deseado. Tal como se muestra en la figura 4, puede emplearse una transposasa para fragmentar una muestra de ADN genómico con dos secuencias de marcador diferentes. Las condiciones pueden seleccionarse para producir fragmentos que tengan una longitud promedio de aproximadamente 300 nucleótidos y que contengan un primer marcador en un extremo y un segundo marcador en el otro extremo. Los marcadores formarán prolongaciones monocatenarias con las que pueden hibridarse los adaptadores. El acoplamiento de adaptadores hibridados, tal como se muestra en la figura 4, producirá concatámeros de 2 fragmentos con el siguiente orden de regiones de secuencia: una secuencia de p5, un sitio de cebado de la lectura 1, un fragmento de ADN 1, un promotor de ARN polimerasa, un sitio de cebado de la lectura 1', un fragmento de ADN 2, un índice y una secuencia de p7. Las secuencias de p5 y p7 en el extremo de los concatámeros permiten la captura y amplificación de puente, por ejemplo, usando células de flujo de secuenciación y kits disponibles en Illumina, Inc. (San Diego, Calif.). El sitio de cebado de la lectura 1 es complementario con un cebador de ADN que va a ser extendido por una ADN polimerasa, y el sitio de cebado de la lectura 1' es complementario con un cebador de ARN que va a ser extendido por una ARN polimerasa, respectivamente, en la reacción de SBS ortogonal. El promotor de ARN polimerasa se encuentra cadena abajo del sitio de cebado de la ARN polimerasa para activar la actividad ARN polimerasa. El fragmento de ADN 1 se encuentra cadena abajo del sitio de cebado de la ADN polimerasa y, así, está colocado para la detección por una extensión con ADN polimerasa en la reacción de SBS, y el fragmento de ADN 2 se encuentra cadena abajo del sitio de cebado de la ARN polimerasa y, así, está colocado para la detección por una extensión con ARN polimerasa en la reacción de SBS. El índice está opcionalmente disponible para objetivos de seguimiento de la muestra.

De modo global, se espera que la plataforma proporcionada por este ejemplo aumenta el resultado de la secuenciación doblando la información de secuenciación por ciclo, es decir, un ensayo de 150 ciclos de SBS con 2 cebadores sería equivalente a un ensayo de 2×150 ciclos de SBS de extremos apareados disponible en el mercado (Illumina, Inc., San Diego Calif.), pero con un mayor ahorro de tiempo y utilización de reactivos. Si se aplica en un formato de extremos apareados, un ensayo de 2×75 ciclos de SBS de extremos apareados con 2 cebadores sería equivalente a un ensayo de 2×150 extremos apareados tradicional.

En la presente, el término "comprende" pretende ser abierto e incluir no solo los elementos mencionados, sino también cualquier elemento adicional.

Aunque la invención se ha descrito haciendo referencia a los ejemplos proporcionados anteriormente, debe entenderse que pueden realizarse diversas modificaciones sin apartarse de la invención, que se define mediante las reivindicaciones adjuntas.

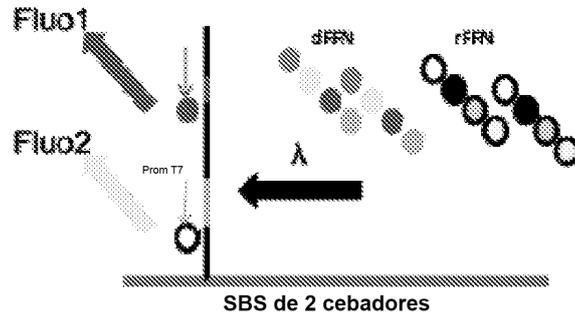
REIVINDICACIONES

- 1.- Un sistema para secuenciar moldes de ácidos nucleicos, que comprende:
- (a) una matriz de sitios, en el que cada sitio incluye un primer molde de ácido nucleico y un segundo molde de ácido nucleico,
 - 5 en el que el primer molde de ácido nucleico tiene una secuencia que es diferente de la secuencia del segundo molde de ácido nucleico;
 - (b) un primer cebador unido al primer molde;
 - (c) una primera especie de polimerasa adecuada para extender el primer cebador unido al primer molde;
 - 10 (d) un primer conjunto de análogos de nucleótidos adecuados para producir un primer producto de extensión de cebador que se extiende desde el primer cebador, en el que el primer producto de extensión de cebador comprende un primer análogo de nucleótido en cada uno de los sitios;
 - (e) un segundo cebador unido al segundo molde;
 - (f) una segunda especie de polimerasa adecuada para extender el segundo cebador unido al segundo molde;
 - 15 (g) un segundo conjunto de análogos de nucleótidos adecuados para producir un segundo producto de extensión de cebador que se extiende desde el segundo cebador, en el que el segundo producto de extensión de cebador comprende un segundo análogo de nucleótido en cada uno de los sitios;
- en el que la primera especie de polimerasa es diferente de la segunda especie de polimerasa, y en el que el primer conjunto de análogos de nucleótidos es diferente del segundo conjunto de análogos de nucleótidos; y
- 20 (h) un detector para detectar el primer producto de extensión de cebador y el segundo producto de extensión de cebador en cada uno de los sitios.
- 2.- El sistema de la reivindicación 1, en el que la primera polimerasa es selectiva para el primer conjunto de análogos de nucleótidos, comparado con el segundo conjunto de análogos de nucleótidos, y en el que la segunda polimerasa es selectiva para el segundo conjunto de análogos de nucleótidos, comparado con el primer conjunto de análogos de nucleótidos.
- 25 3.- El sistema de la reivindicación 1, en el que la primera polimerasa es selectiva para el primer cebador, comparado con el segundo cebador, y en el que la segunda polimerasa es selectiva para el segundo cebador, comparado con el primer cebador.
- 4.- El sistema de la reivindicación 1, en el que el detector tiene una resolución espacial que es demasiado baja para resolver puntos a una distancia equivalente al espaciamiento entre el primer producto de extensión de cebador y el
- 30 segundo producto de extensión de cebador en cada uno de los sitios.
- 5.- El sistema de la reivindicación 4, en el que el detector es un detector óptico.
- 6.- El sistema de la reivindicación 5, en el que los análogos de nucleótidos comprenden marcadores ópticos.
- 7.- El sistema de la reivindicación 6, en el que los marcadores ópticos del primer conjunto de análogos de nucleótidos son diferentes de los marcadores ópticos del segundo conjunto de análogos de nucleótidos.
- 35 8.- El sistema de la reivindicación 5, en el que un subconjunto de los análogos de nucleótidos en el primer conjunto de análogos de nucleótidos comprende marcadores ópticos.
- 9.- El sistema de la reivindicación 8, en el que un subconjunto de los análogos de nucleótidos en el segundo conjunto de análogos de nucleótidos comprende marcadores ópticos.
- 40 10.- El sistema de la reivindicación 9, en el que el primer conjunto de análogos de nucleótidos comprende marcadores ópticos que son diferentes de los marcadores ópticos del segundo conjunto de análogos de nucleótidos.
- 11.- El sistema de la reivindicación 4, en el que un píxel del detector adquiere señales del primer producto de extensión de cebador y del segundo producto de extensión de cebador.
- 12.- El sistema de la reivindicación 1, en el que la primera polimerasa es una ARN polimerasa y la segunda polimerasa es una ADN polimerasa y, opcionalmente, en el que el primer cebador es un cebador de ARN y el segundo cebador es un cebador de ADN.
- 45 13.- El sistema de la reivindicación 1, en el que los sitios tienen un área que no es mayor que $100 \mu\text{m}^2$.

14.- El sistema de la reivindicación 1, en el que los sitios comprenden múltiples copias del primer molde de ácido nucleico y del segundo molde de ácido nucleico.

15.- El sistema de la reivindicación 1, en el que la primera polimerasa es selectiva para el primer molde, comparado con el segundo molde, y en el que la segunda polimerasa es selectiva para el segundo molde, comparado con el primer molde.

5



ADN pol + cebador de ADN + dNTP marcados + ARN pol (concretamente, T7) + cebador de ARN + rNTP marcados

Figura 1A

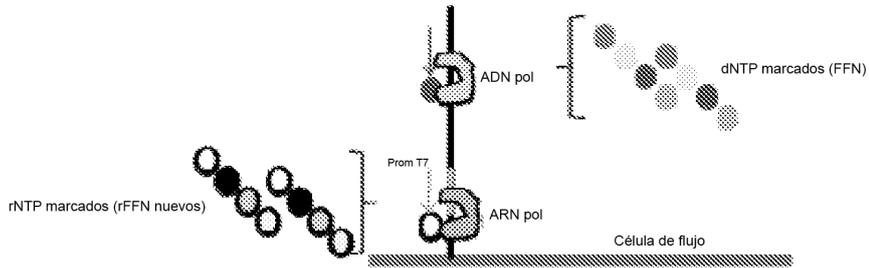


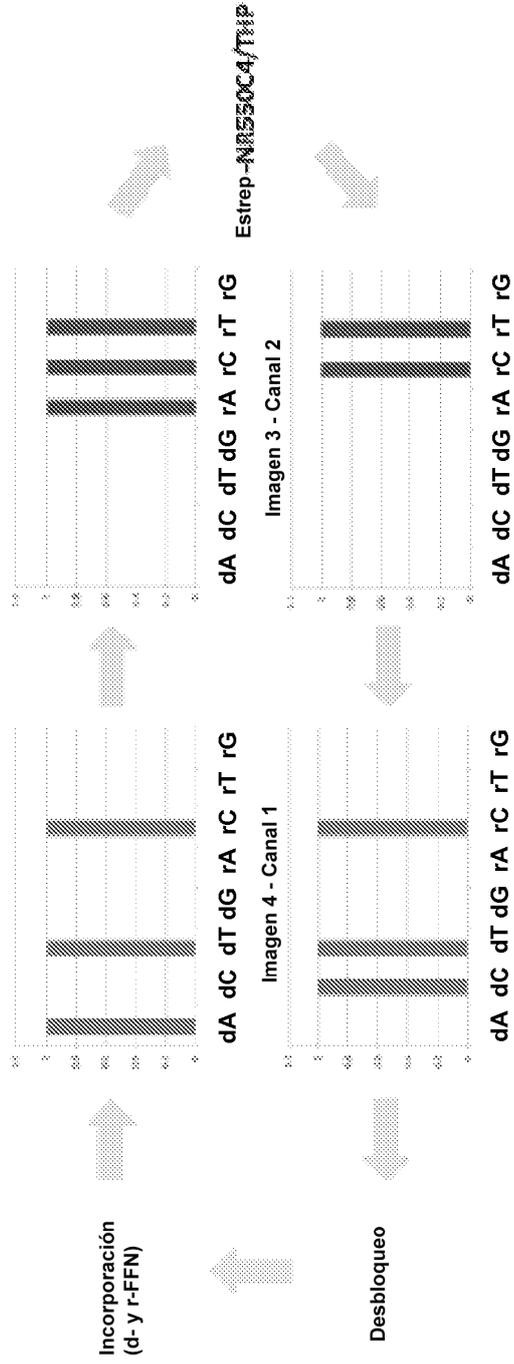
Figura 1B

A

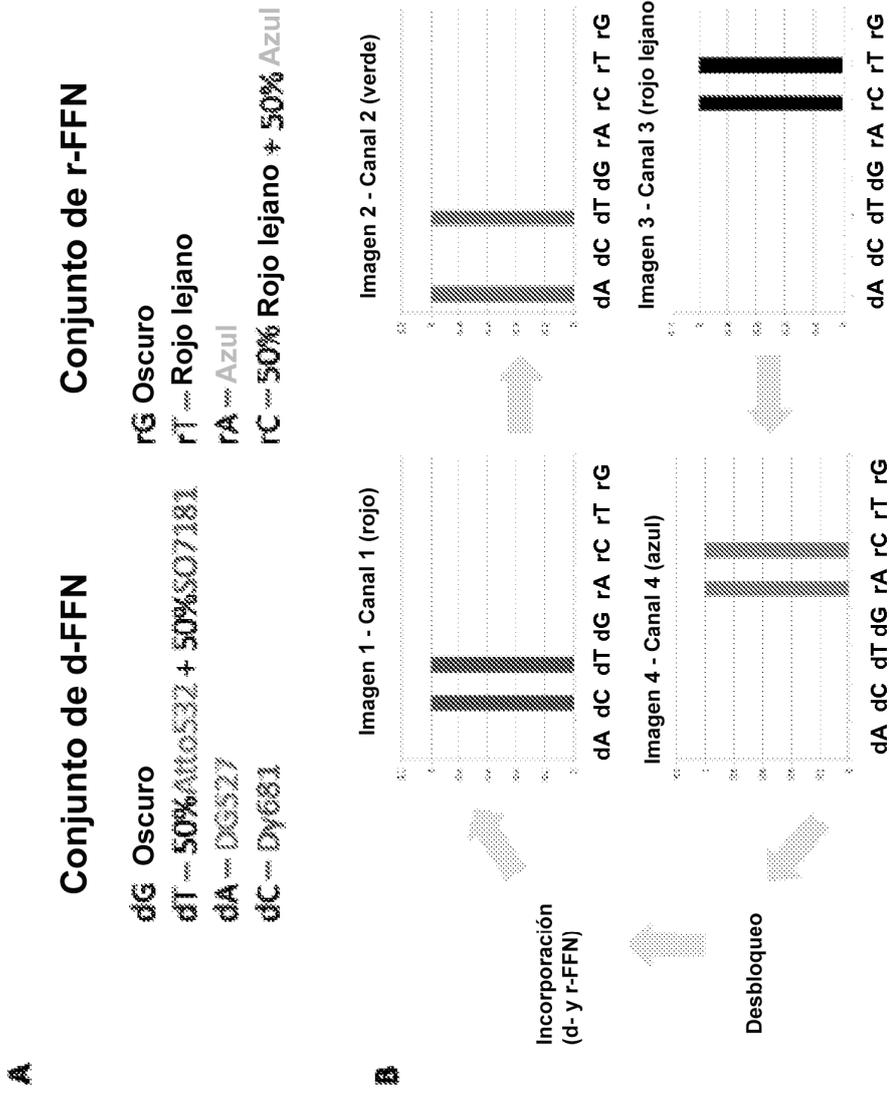
Conjunto de d-FFN Conjunto de r-FFN

dG Oscuro **rG Oscuro**
dT ~ NR550C4 **rT ~ Rojo**
dA ~ SS ~ NR550C4 **rA ~ SS ~ Rojo**
dC ~ biotina **rC ~ 50% NR550C4 + 50% Rojo**

B



Figuras 2A y 2B



Figuras 3A y 3B

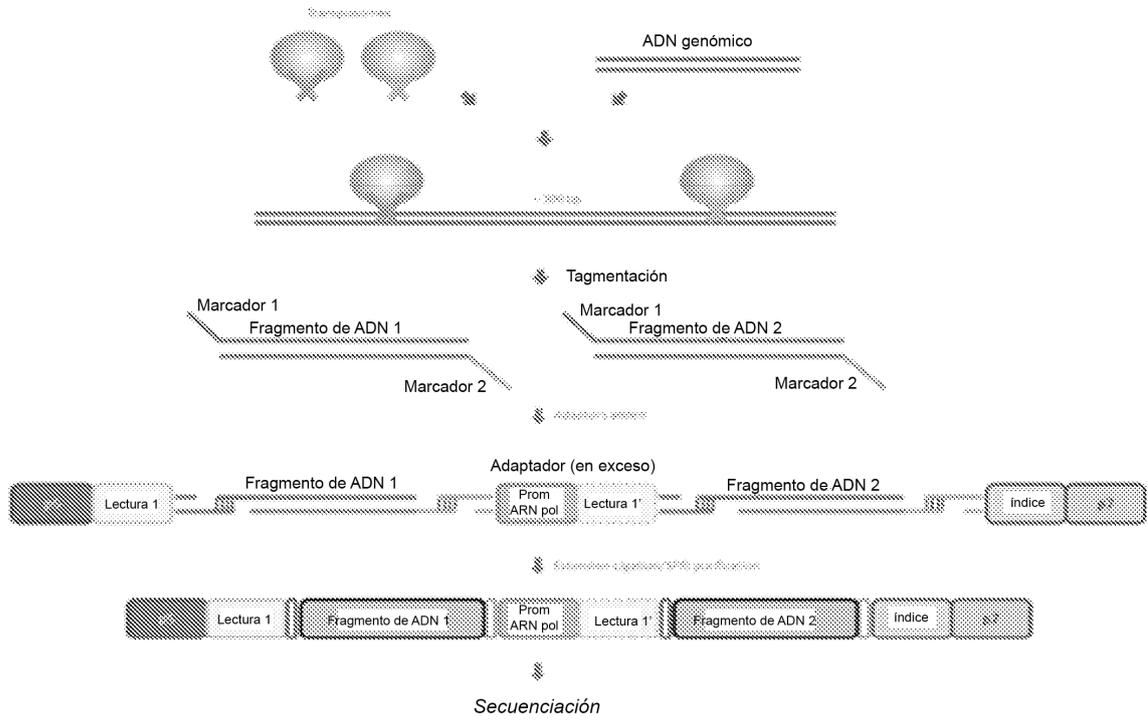
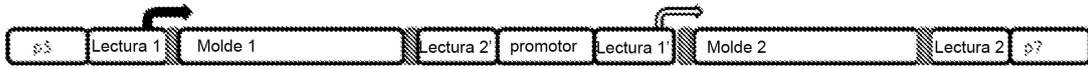


Figura 4

Lectura 1 de PA ortogonal



Lectura 2 de PA ortogonal



Figura 5