

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 683 979**

51 Int. Cl.:

C12Q 1/68 (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **02.05.2013 PCT/US2013/039297**

87 Fecha y número de publicación internacional: **07.11.2013 WO13166304**

96 Fecha de presentación y número de la solicitud europea: **02.05.2013 E 13785146 (5)**

97 Fecha y número de publicación de la concesión europea: **18.07.2018 EP 2844775**

54 Título: **Secuenciación de ADN**

30 Prioridad:
02.05.2012 US 201261641718 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
01.10.2018

73 Titular/es:
**IBIS BIOSCIENCES, INC. (100.0%)
Suite 150 2251 Faraday Avenue
Carlsbad, CA 92008, US**

72 Inventor/es:
ESHOO, MARK W.

74 Agente/Representante:
IZQUIERDO BLANCO, María Alicia

ES 2 683 979 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

Secuenciación de ADN

5 **Referencia cruzada a solicitudes relacionadas**

Campo de la invención

10 En el presente documento se proporciona tecnología relacionada con la secuenciación de ácidos nucleicos y particularmente, pero no exclusivamente, con procedimientos, composiciones y sistemas para secuenciar un ácido nucleico usando uno o más marcadores y amplitud de señal para distinguir bases.

Antecedentes

15 La secuenciación del ADN está impulsando la investigación y el descubrimiento de la genómica. La finalización del Proyecto del Genoma Humano fue un logro monumental con una increíble cantidad de esfuerzos combinados entre los centros del genoma y los científicos de todo el mundo. Este proyecto de una década de duración se completó utilizando el procedimiento de secuenciación de Sanger, que sigue siendo la metodología de secuenciación del genoma básica en los centros de secuenciación del genoma de alto rendimiento. La razón principal detrás del éxito prolongado de este procedimiento es su procedimiento básico y eficiente, pero elegante, de terminación de la cadena didesoxi. Con mejoras cada vez mayores en la secuenciación de Sanger, incluido el uso de excitación fluorescente inducida por láser de colorantes de transferencia de energía, ADN polimerasas modificadas, electroforesis capilar, preparación de muestras, informática y software de análisis de secuencias, la plataforma de secuenciación de Sanger ha podido mantener su estado. Los secuenciadores de ADN basados en Sanger según la técnica actual pueden producir más de 700 bases de secuencia claramente legible en una sola ronda a partir de moldes de hasta 30 kb de longitud. Sin embargo, como ocurre con la mayoría de las invenciones tecnológicas, las continuas mejoras en esta plataforma de secuenciación han llegado a un punto muerto, con el coste estimado actual para producir un borrador de una secuencia de genoma microbiano de alta calidad a aproximadamente 10.000 \$ por par de megabases. Los secuenciadores de ADN actuales basados en el procedimiento de Sanger permiten analizar hasta 384 muestras en paralelo.

Es evidente que la explotación de la secuencia completa del genoma humano para la medicina clínica y la atención sanitaria requiere procedimientos precisos de secuenciación de ADN de bajo coste y alto rendimiento. De hecho, el sector de ciencia genómica tanto público (National Human Genome Research Institute, NHGRI) como privado (The J. Craig Venter Science Foundation and Archon X prize for genomics) han lanzado una llamada para el desarrollo de la tecnología de secuenciación de próxima generación que reducirá el coste de secuenciar hasta una décima parte de su coste actual en los próximos diez años. De acuerdo con ello, para superar las limitaciones de las actuales tecnologías de secuenciación convencionales, se han investigado diversos nuevos procedimientos de secuenciación de ADN, que incluyen procedimientos de secuenciación por síntesis (SBS) tales como pirosecuenciación (Ronaghi y col. (1998) *Science* 281: 363–365), secuenciación de moléculas de ADN sencillas (Braslaysky y col. (2003) *Proc. Natl. Acad. Sci. USA* 100: 3960–3964), y colonias de polimerasa (secuenciación "polonia") (Mitra y col. (2003) *Anal. Biochem.* 320: 55–65).

45 El concepto de secuenciación por síntesis de ADN (SBS) se reveló en 1988 con un intento de secuenciar el ADN mediante la detección del grupo pirofosfato que se genera cuando se incorpora un nucleótido mediante una reacción de ADN polimerasa (Hyman (1999) *Anal. Biochem.* 174: 423–436). Las posteriores tecnologías de SBS se basaron en formas adicionales para detectar la incorporación de un nucleótido a una cadena de ADN en crecimiento. En general, el SBS convencional usa un cebador oligonucleotídico diseñado para hibridar con una posición predeterminada de la molécula molde de muestra que se va a secuenciar. Se presenta al complejo cebador-molde un nucleótido en presencia de una enzima polimerasa. Si el nucleótido es complementario a la posición en la molécula molde de muestra que está directamente en el extremo 3' del cebador oligonucleotídico, la ADN polimerasa extenderá el cebador con el nucleótido. La incorporación del nucleótido y la identidad del nucleótido insertado se puede detectar, por ejemplo, mediante la emisión de luz, un cambio en la fluorescencia, un cambio en el pH (véase, por ejemplo, la patente de Estados Unidos n.º 7.932.034), un cambio en la conformación de la enzima o algún otro cambio físico o químico en la reacción (véanse, por ejemplo, los documentos WO 1993/023564 y WO 1989/009283; Seo y col. (2005) "Four-color DNA sequencing by synthesis on a chip using photocleavable fluorescent nucleotides," *PNAS* 102: 5926–59). Tras cada incorporación exitosa de un nucleótido, se detecta una señal que refleja la aparición, identidad y número de incorporaciones de nucleótidos. A continuación, pueden eliminarse los nucleótidos no incorporados (por ejemplo, mediante degradación química o mediante lavado) y la siguiente posición en el cebador-molde puede consultarse con otra especie de nucleótido.

Sumario

65 En la secuenciación por síntesis de ADN convencional usando monómeros de nucleótidos marcados, se unen cuatro restos diferentes (por ejemplo, un colorante o un marcador fluorescente) a las cuatro bases nucleotídicas para permitir que el detector distinga las bases entre sí. Por ejemplo, algunos procedimientos marcan cada una de las

bases A, C, G y T con un resto fluorescente que emite luz a una longitud de onda que se distingue de la luz emitida por los otros tres restos fluorescentes, por ejemplo, para producir luz de cuatro colores diferentes asociados con cada una de las cuatro bases.

5 Por el contrario, la presente tecnología se basa en la detección de diferencias en la amplitud de la señal en lugar de diferencias en la longitud de onda de la señal (por ejemplo, el color) para identificar cada base incorporada durante una reacción de secuenciación. En este esquema, cada base individual se marca con el mismo resto (por ejemplo, un colorante, un marcador fluorescente, etc.) en un porcentaje conocido diferente (por ejemplo, una "fracción del marcador" o "extensión del marcaje"). Como una realización de ejemplo, el 25 % de las moléculas de ATP están marcadas, el 50 % de las moléculas de TTP están marcadas, el 75 % de las moléculas de GTP están marcadas y el 100 % de las moléculas de CTP están marcadas. A continuación, de acuerdo con algunas realizaciones, se realiza un enfoque de secuenciación en conjunto (por ejemplo, una colonia o una colonia clonal) y la secuencia se determina detectando una intensidad de señal después de cada incorporación de base y asociando las intensidades con las bases.

15 En algunas realizaciones, un elemento de la tecnología que permite separar y asignar las intensidades de señal en "contenedores" específicos de base apropiados es el uso de una secuencia de calibración de 4 bases al comienzo de una ronda de secuenciación. Esta secuencia de calibración contiene cada una de las 4 bases en un orden conocido para proporcionar una referencia de calibración, por ejemplo, para calibrar un instrumento de secuenciación para reconocer las intensidades de señal apropiadas para cada una de las bases.

20 Como consecuencia, las realizaciones de la tecnología reducen el número de colorantes fluorescentes necesarios para identificar las cuatro bases (por ejemplo, permitiendo el uso de solo el colorante o colorantes más óptimos para adquirir una secuencia), reducen el número de láseres usados para excitar el resto o restos fluorescente, reducen o eliminan la óptica utilizada para dividir la señal óptica por longitud de onda y reducen la cantidad de detectores para registrar eventos de incorporación.

25 De acuerdo con lo anterior, en el presente documento se proporciona tecnología relacionada con un procedimiento para secuenciar un ácido nucleico diana, comprendiendo el procedimiento detectar una amplitud de una señal producida a partir de una pluralidad de una base de nucleótidos; y asociar la amplitud con la base de nucleótidos para identificar la base de nucleótidos. En algunas realizaciones, la amplitud de la señal producida por la pluralidad de la base de nucleótidos es diferente de forma detectable de una segunda amplitud de una segunda señal producida por una segunda pluralidad de una segunda base de nucleótidos. Por ejemplo, en algunas realizaciones, una fracción de la pluralidad de la base de nucleótidos produce una señal, por ejemplo, en algunas realizaciones, la pluralidad de la base de nucleótidos está marcada de forma detectable.

30 En un aspecto, la tecnología se refiere a la identificación de las bases de nucleótidos en una secuencia de nucleótidos mediante una reacción de secuenciación. Por tanto, la tecnología proporciona, en algunas realizaciones, un procedimiento para secuenciar un ácido nucleico en el que una fracción de la pluralidad de la base de nucleótidos está marcada detectablemente y dicha fracción es diferente de una segunda fracción de una segunda pluralidad de una segunda base de nucleótidos que está marcada de forma detectable. Algunas realizaciones de los procedimientos comprenden proporcionar una primera pluralidad de una primera base de nucleótidos y una segunda pluralidad de una segunda base de nucleótidos, en la que una primera fracción de la primera pluralidad de la primera base está marcada con un marcador y una segunda fracción de la segunda pluralidad de la segunda base de nucleótidos está marcada con dicho marcador.

35 En algunas realizaciones, los procedimientos comprenden proporcionar una primera pluralidad de una primera base de nucleótidos, una segunda pluralidad de una segunda base de nucleótidos, una tercera pluralidad de una tercera base de nucleótidos y una cuarta pluralidad de una cuarta base de nucleótidos, en los que una primera fracción de la primera pluralidad de la primera base está marcada con un marcador, una segunda fracción de la segunda pluralidad de la segunda base de nucleótidos está marcada con dicho marcador, una tercera fracción de la tercera pluralidad de la tercera base de nucleótidos está marcada con dicho marcador, y una cuarta fracción de la cuarta pluralidad de la cuarta base de nucleótidos está marcada con dicho marcador. Por ejemplo, en algunas realizaciones, la primera base de nucleótidos es A, la segunda base de nucleótidos es C, la tercera base de nucleótidos es G y la cuarta base de nucleótidos es T. Además, en algunas realizaciones, el marcador es un resto fluorescente.

40 Se proporcionan esquemas alternativos para identificar bases de nucleótidos en el presente documento. Por ejemplo, la tecnología incluye un procedimiento que comprende proporcionar una primera pluralidad de una primera base de nucleótidos, una segunda pluralidad de una segunda base de nucleótidos, una tercera pluralidad de una tercera base de nucleótidos y una cuarta pluralidad de una cuarta base de nucleótidos, en el que una primera fracción de la primera pluralidad de la primera base está marcada con un primer marcador, una segunda fracción de la segunda pluralidad de la segunda base de nucleótidos está marcada con el primer marcador, una tercera fracción de la tercera pluralidad de la tercera base de nucleótidos está marcada con un segundo marcador, y una cuarta fracción de la cuarta pluralidad de la cuarta base de nucleótidos está marcada con el segundo marcador.

Los procedimientos proporcionados en el presente documento se refieren en algunos aspectos a la secuenciación de un ácido nucleico usando un procedimiento de secuenciación por síntesis. Por ejemplo, algunas realizaciones proporcionan un procedimiento que comprende incorporar por polimerización la pluralidad de la base de nucleótidos en una pluralidad de un ácido nucleico que es complementario al ácido nucleico diana.

5 En algunas realizaciones, la señal es una onda electromagnética, por ejemplo, una señal que tiene una longitud de onda en el rango visible. . Por ejemplo, en algunas realizaciones, la señal es fluorescencia. Y, además, en algunas realizaciones, los procedimientos proporcionan la detección de la señal con un dispositivo óptico.

10 Los nucleótidos de una secuencia de ácido nucleico se identifican mediante el control de la amplitud de la señal producida durante la secuenciación. Las asociaciones entre una amplitud de señal y la identidad o tipo de una base de nucleótidos se definen en algunas realizaciones calibrando un aparato de secuenciación con una secuencia de calibración que comprende un orden conocido de bases de nucleótidos. Por consiguiente, la tecnología se refiere a un procedimiento que comprende proporcionar un oligonucleótido de calibración que comprende una secuencia conocida. Además, los procedimientos proporcionados se refieren a la determinación de la secuencia de nucleótidos de un ácido nucleico; así, algunas realizaciones de procedimientos comprenden además analizar un conjunto de datos de amplitudes ordenadas para producir una secuencia de nucleótidos del ácido nucleico diana.

20 Otro aspecto de la tecnología proporciona una composición que comprende una pluralidad de una base de nucleótidos en la que una fracción de la pluralidad está marcada de forma detectable con un marcador. Dicha composición encuentra uso, por ejemplo, en procedimientos para secuenciar un ácido nucleico. En algunas realizaciones, las composiciones proporcionadas en el presente documento comprenden además una segunda pluralidad de una segunda base de nucleótidos en la que una segunda fracción de la segunda pluralidad está marcada de forma detectable con dicho marcador. Además, en otras realizaciones adicionales, se proporcionan composiciones que comprenden una tercera pluralidad de una tercera base de nucleótidos y una cuarta pluralidad de una cuarta base de nucleótidos, en la que una tercera fracción de la tercera pluralidad de la tercera base de nucleótidos está marcada con dicho marcador y una la cuarta fracción de la cuarta pluralidad de la cuarta base de nucleótidos está marcada con dicho marcador. En algunas realizaciones, la primera base de nucleótidos es A, la segunda base de nucleótidos es C, la tercera base de nucleótidos es G y la cuarta base de nucleótidos es T (o U, según sea adecuado), y, además, algunas realizaciones proporcionan que el marcador sea un resto fluorescente. En algunas realizaciones, la primera, la segunda, la tercera y/o la cuarta base es una base modificada o un análogo de base tal como una inosina, isoguanina, isocitosina, una diaminopirimidina, una xantina, un nitroazol, una base de tamaño expandido, etc.

35 En esquemas de marcaje alternativos, las realizaciones proporcionan composiciones que comprenden además una segunda pluralidad de una segunda base de nucleótidos en la que una segunda fracción de la segunda pluralidad está marcada detectablemente con un segundo marcador o composiciones que comprenden además una tercera pluralidad de una tercera base de nucleótidos y una cuarta pluralidad de una cuarta base de nucleótidos, en las que una tercera fracción de la tercera pluralidad de la tercera base de nucleótidos está marcada con dicho marcador y una cuarta fracción de la cuarta pluralidad de la cuarta base de nucleótidos está marcada con dicho segundo marcador.

45 Aunque la tecnología se refiere a composiciones de nucleótidos marcados, debe entenderse que la tecnología también incluye composiciones que comprenden además un ácido nucleico diana, un cebador de secuenciación y una polimerasa y/o a composiciones que comprenden además un ácido nucleico que comprende la base de nucleótidos.

50 La tecnología está incorporada en sistemas que incorporan características de los procedimientos y composiciones proporcionados. Por ejemplo, las realizaciones particulares proporcionan un sistema para secuenciar un ácido nucleico, en el que el sistema comprende una composición que comprende una pluralidad de una base de nucleótidos en la que una fracción de la pluralidad se marca detectablemente con un marcador; y un oligonucleótido de calibración. Algunas realizaciones comprenden además un aparato de secuenciación, algunas realizaciones comprenden además un procesador configurado para asociar una amplitud de una señal con una base de nucleótidos, y algunas realizaciones comprenden además una funcionalidad de salida para proporcionar una secuencia de nucleótidos del ácido nucleico.

60 En realizaciones particulares, se proporcionan sistemas que comprenden además una segunda pluralidad de una segunda base de nucleótidos, una tercera pluralidad de una tercera base de nucleótidos y una cuarta pluralidad de una cuarta base de nucleótidos, en los que una segunda fracción de la segunda pluralidad del segundo nucleótido la base está marcada con dicho marcador, una tercera fracción de la tercera pluralidad de la tercera base de nucleótidos está marcada con dicho marcador y una cuarta fracción de la cuarta pluralidad de la cuarta base de nucleótidos está marcada con dicho marcador. Además, algunas realizaciones comprenden una funcionalidad para detectar el marcador y/o una funcionalidad para diferenciar la base de nucleótidos, la segunda base de nucleótidos, la tercera base de nucleótidos y la cuarta base de nucleótidos entre sí.

65

Se proporcionan realizaciones de kits para la secuenciación de ácidos nucleicos. Por ejemplo, algunas realizaciones de la tecnología son kits para secuenciar un ácido nucleico, en los que los kits comprenden una composición que comprende una pluralidad de una base de nucleótidos en la que una fracción de la pluralidad se marca detectablemente con un marcador; y un oligonucleótido de calibración. Realizaciones adicionales comprenden además una segunda pluralidad de una segunda base de nucleótidos, una tercera pluralidad de una tercera base de nucleótidos y una cuarta pluralidad de una cuarta base de nucleótidos, en los que una segunda fracción de la segunda pluralidad del segundo nucleótido la base está marcada con dicho marcador, una tercera fracción de la tercera pluralidad de la tercera base de nucleótidos está marcada con dicho marcador y una cuarta fracción de la cuarta pluralidad de la cuarta base de nucleótidos está marcada con dicho marcador.

Las realizaciones adicionales serán evidentes para los expertos en la técnica relevante según las enseñanzas contenidas en el presente documento.

Descripción detallada

En el presente documento se proporciona tecnología relacionada con la secuenciación de ácidos nucleicos y particularmente, pero no exclusivamente, con procedimientos, composiciones, sistemas y kits para secuenciar un ácido nucleico usando uno o más marcadores y amplitud de señal para distinguir bases.

Los encabezados de sección usados en el presente documento son únicamente con motivos organizativos y no deben interpretarse de ningún modo como limitantes de la materia objeto descrita.

En esta descripción detallada de las diversas realizaciones, para fines de explicación, se exponen numerosos detalles específicos para proporcionar una comprensión completa de las realizaciones desveladas. Un experto en la materia apreciará, sin embargo, que estas diversas realizaciones se pueden poner en práctica con o sin estos detalles específicos. En otros casos, las estructuras y los dispositivos se muestran en forma de diagrama de bloques. Además, un experto en la técnica puede apreciar fácilmente que las secuencias específicas en las que se presentan y realizan los procedimientos son ilustrativas y se contempla que las secuencias pueden variarse y que sigan dentro del espíritu y alcance de las diversas realizaciones desveladas en el presente documento. A menos que se defina otra cosa, todos los términos técnicos y científicos usados en el presente documento tienen el mismo significado que un experto en la técnica a la que pertenecen las diversas realizaciones descritas.

Se apreciará que hay un "aproximadamente" implícito antes de las temperaturas, concentraciones, tiempos, etc. tratados en las presentes enseñanzas, de modo que las desviaciones leves e insustanciales están dentro del alcance de las presentes enseñanzas. En la presente solicitud, el uso del singular incluye el plural a menos que específicamente se indique lo contrario. Asimismo, el uso de "comprenden", "comprende", "que comprende", "contienen", "contiene", "que contiene", "incluyen", "incluye" y "que incluye" no pretende ser limitante. Debe entenderse que tanto la descripción general anterior como la descripción detallada siguiente son ejemplos y explicaciones únicamente y no son restrictivas de las presentes enseñanzas.

Adicionalmente, a menos que el contexto requiera lo contrario, los términos en singular incluyen pluralidades y los términos en plural incluirán el singular. Generalmente, las nomenclaturas utilizadas en relación y las técnicas de cultivo celular y tisular, biología molecular y química e hibridación de proteínas y oligonucleótidos o polinucleótidos descritas en el presente documento son las bien conocidas y usadas habitualmente en la técnica. A menos que se indique lo contrario, se usan técnicas estándar, por ejemplo, para la purificación y preparación de ácidos nucleicos, análisis químico, ácido nucleico recombinante y síntesis de oligonucleótidos. Se realizan reacciones enzimáticas y técnicas de purificación de acuerdo con las especificaciones del fabricante, como normalmente se efectúan en la técnica o tal como se describe en el presente documento. Las técnicas y procedimientos descritos en el presente documento, en general, se realizan de acuerdo con procedimientos convencionales bien conocidos en la técnica y tal como se describen en varias referencias generales y más específicas que se citan y tratan a lo largo de la presente especificación. Véase, por ejemplo, Sambrook y col., *Molecular Cloning: A Laboratory Manual* (Tercera ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (2000)). Las nomenclaturas utilizadas en relación con, y los procedimientos y técnicas de laboratorio descritos en el presente documento, son los bien conocidos y de uso habitual en la técnica.

Definiciones

Para facilitar la comprensión de la presente tecnología, a continuación se definen una serie de términos y frases. A lo largo de la descripción detallada se exponen definiciones adicionales.

A lo largo de la memoria descriptiva y las reivindicaciones, los siguientes términos toman los significados asociados explícitamente en el presente documento, a menos que el contexto indique claramente lo contrario. La frase "en una realización" como se usa en el presente documento no necesariamente se refiere a la misma realización, aunque puede. Adicionalmente, la frase "en otra realización" como se usa en el presente documento no necesariamente se refiere a una realización diferente, aunque puede. Por tanto, como se describe a continuación, varias realizaciones de la invención pueden combinarse fácilmente, sin apartarse del alcance o espíritu de la invención.

Además, como se utiliza en el presente documento, el término "o" es un operador de inclusión "o" y es equivalente a la expresión "y/o", a menos que el contexto indique claramente lo contrario. El término "basándose en" no es exclusivo y permite el fundamento en factores adicionales no descritos, a menos que el contexto indique claramente lo contrario. Además, a lo largo de la memoria descriptiva, el significado de "un", "una" y "el/la" incluyen referencias en plural. El significado de "en" incluye "en" y "sobre".

Un "sistema" denota un conjunto de componentes, reales o abstractos, que comprenden un todo donde cada componente interacciona o está relacionado con al menos otro componente dentro del todo.

Como se usa en el presente documento, la frase "dNTP" significa desoxinucleótido-trifosfato, en el que el nucleótido comprende una base de nucleótidos, tal como A, T, C, G o U.

El término "monómero" tal como se usa en el presente documento significa cualquier compuesto que pueda incorporarse a una cadena molecular en crecimiento por una polimerasa dada. Tales monómeros incluyen, sin limitaciones, nucleótidos de origen natural (por ejemplo, ATP, GTP, TTP, UTP, CTP, dATP, dGTP, dTTP, dUTP, dCTP, análogos sintéticos), precursores para cada nucleótido, nucleótidos no naturales y sus precursores o cualquier otra molécula que pueda incorporarse en una cadena polimérica en crecimiento por una polimerasa dada.

Como se usa en el presente documento, un "ácido nucleico" significará cualquier molécula de ácido nucleico, que incluye, sin limitación, ADN, ARN e híbridos de los mismos. Las bases de ácido nucleico que forman moléculas de ácido nucleico pueden ser las bases A, C, G, T y U, así como sus derivados. Los derivados de estas bases son bien conocidos en la técnica. El término debe entenderse que incluye, como equivalentes, análogos de ADN o ARN preparados a partir de análogos de nucleótidos. El término, tal como se usa en el presente documento, también abarca ADNc, que es ADN complementario o copia, producido a partir de un molde de ARN, por ejemplo mediante la acción de una transcriptasa inversa. Es bien sabido que el ADN (ácido desoxirribonucleico) es una cadena de nucleótidos que consiste en 4 tipos de nucleótidos: A (adenina), T (timina), C (citosina) y G (guanina), y ese ARN (ácido ribonucleico) es una cadena de nucleótidos que consiste en 4 tipos de nucleótidos: A, U (uracilo), G y C. También se sabe que todos estos 5 tipos de nucleótidos se unen específicamente entre sí en combinaciones llamadas apareamiento de bases complementarias. Es decir, la adenina (A) se aparea con timina (T) (en el caso del ARN, sin embargo, la adenina (A) se aparea con uracilo (U)) y la citosina (C) se aparea con guanina (G), de modo que cada uno de estos pares de bases forman una doble cadena. Como se usa en el presente documento, "datos de secuenciación de ácido nucleico", "información de secuenciación de ácido nucleico", "secuencia de ácido nucleico", "secuencia genómica", "secuencia genética", "secuencia de fragmento", o "lectura de secuencia de ácido nucleico" denota cualquier información o datos que son indicativos del orden de las bases de nucleótidos (por ejemplo, adenina, guanina, citosina y timina/uracilo) en una molécula (por ejemplo, un genoma completo, un transcriptoma completo, un exoma, oligonucleótido, polinucleótido, fragmento, etc.) de ADN o ARN.

La referencia a una base, un nucleótido o a otra molécula puede ser en singular o en plural. Es decir, "una base" puede hacer referencia a una sola molécula de dicha base o a una pluralidad de la base, por ejemplo, en una solución.

Como se usa en el presente documento, la expresión "una pluralidad clonal de ácidos nucleicos" o "una población clonal de ácidos nucleicos" o "un grupo" o "una colonia" se refiere a un conjunto de productos de ácido nucleico que son sustancial o completamente o esencialmente idénticos a entre sí, y son copias complementarias de la cadena de ácido nucleico molde a partir de la cual se sintetizan.

Tal como se usa en el presente documento, un "polinucleótido", también llamado ácido nucleico, es una serie de nucleótidos unidos covalentemente en la que la posición 3' de la pentosa de un nucleótido está unida por un grupo fosfodiéster a la posición 5' de la siguiente. El ADN (ácido desoxirribonucleico) y el ARN (ácido ribonucleico) son polinucleótidos que se producen biológicamente en los que los restos de nucleótidos están unidos en una secuencia específica mediante enlaces fosfodiéster. Como se usa en el presente documento, los términos "polinucleótido" u "oligonucleótido" abarcan cualquier compuesto polimérico que tenga una cadena principal lineal de nucleótidos. Los oligonucleótidos, también denominados oligómeros, son generalmente polinucleótidos de cadena más corta.

Como se usa en el presente documento, "complementario" generalmente se refiere a dúplex de nucleótidos específicos para formar pares de bases de Watson-Crick canónicos, como entienden los expertos en la técnica. Sin embargo, complementario también incluye apareamiento de bases de análogos de nucleótidos que son capaces de aparearse bases universales con los nucleótidos A, T, G o C y ácidos nucleicos bloqueados que mejoran la estabilidad térmica de los dúplex. Un experto en la técnica reconocerá que la rigurosidad de la hibridación es un factor determinante en el grado de coincidencia o falta de coincidencia en el dúplex formado por hibridación.

Como se usa en el presente documento, "resto" se refiere a una de dos o más partes en las que algo puede dividirse, tal como, por ejemplo, las diversas partes de una correa, una molécula o una sonda.

Una "polimerasa" es una enzima generalmente para unir nucleótidos 3'-OH 5'-trifosfato, oligómeros y análogos. Las polimerasas incluyen, pero no están limitadas a, ADN polimerasas dependientes de ADN, ARN polimerasas

dependientes de ADN, ADN polimerasas dependientes de ARN, ARN polimerasas dependientes de ARN, ADN polimerasa T7, ADN polimerasa de T3, ADN polimerasa de T4, ARN polimerasa de T7, ARN polimerasa de T3, ARN polimerasa de SP6, ADN polimerasa 1, fragmento Klenow, ADN polimerasa de *Thermophilus aquaticus*, ADN polimerasa de *Tth*, ADN polimerasa de Vent (New England Biolabs), ADN polimerasa de Deep Vent (New England Biolabs), fragmento grande de ADN polimerasa de *Bst*, fragmento Stoeffel, ADN polimerasa de 9° N, ADN polimerasa de *Pfu*, ADN polimerasa de *Tfl*, RepliPHI Phi29 Polimerasa, ADN polimerasa de *Tli*, ADN polimerasa beta eucariótica, telomerasa, polimerasa Terminator (New England Biolabs), KOD HiFi. ADN polimerasa (Novagen), ADN polimerasa KOD1, Q-beta replicasa, transferasa terminal, transcriptasa inversa de AMV, transcriptasa inversa de M-MLV, transcriptasa inversa de Phi6, transcriptasa inversa de Hiv-1, nuevas polimerasas descubiertas mediante bioprospección y las polimerasas citadas en la publicación de solicitud de patente de Estados Unidos n.º 2007/0048748 y en la patente de Estados Unidos n.º 6.329.178; 6.602.695; y 6.395.524. Estas polimerasas incluyen isoformas mutantes silvestres y variantes modificadas genéticamente, tales como exo- polimerasas y otros mutantes, por ejemplo que toleran los nucleótidos marcados y los incorporan en una cadena de ácido nucleico.

La expresión "cebador" se refiere a un oligonucleótido, natural como en un digesto de restricción purificado o producido sintéticamente, que es capaz de actuar como punto de inicio de la síntesis cuando se introduce en las condiciones en las que la síntesis de un producto de extensión del cebador, que es complementario a una hebra de ácido nucleico, se induce (por ejemplo, en presencia de nucleótidos y un agente inductor tal como una ADN polimerasa y a una temperatura y pH adecuados). Preferentemente, el cebador es monocatenario para una máxima eficiencia en la amplificación, pero, como alternativa, puede ser bicatenario. Si es bicatenario, el cebador se trata primero para separar las hebras antes de usar para preparar productos de extensión. Preferentemente, el cebador es un oligodesoxirribonucleótido. El cebador debe ser lo bastante largo como para cebar la síntesis de los productos de extensión en presencia del agente inductor. Las longitudes exactas de los cebadores dependerán de muchos factores, incluidos la temperatura, la fuente del cebador y el uso del procedimiento.

Realizaciones de la tecnología

La tecnología se refiere generalmente a procedimientos, composiciones, sistemas y kits para la secuenciación de ADN utilizando un enfoque de secuenciación por síntesis. Aunque la divulgación en el presente documento se refiere a ciertas realizaciones ilustradas, debe entenderse que estas realizaciones se presentan a modo de ejemplo y no a modo de limitación.

Procedimientos

Algunas realizaciones de la tecnología proporcionan procedimientos de secuenciación por síntesis de ADN en los que las diferencias en la amplitud de la señal, en lugar de diferencias en la longitud de onda de la señal, identifican bases incorporadas durante, por ejemplo, una reacción de secuenciación por síntesis. En algunas realizaciones, se usa un enfoque de secuenciación basado en un conjunto (por ejemplo, una colonia de polimerasa ("polonia") o una colonia clonal). Estos enfoques secuencian múltiples copias idénticas o sustancialmente idénticas de una molécula de ADN que forman un racimo de moléculas molde. Se proporcionan procedimientos para formar racimos, por ejemplo, en la patente de Estados Unidos n.º 7.115.400. En algunas realizaciones, los grupos se inmovilizan sobre un soporte sólido tal como una esfera. Estos racimos típicamente son el resultado de la amplificación de una única molécula de ADN original; por lo tanto, cada racimo representa la molécula sencilla que inició la amplificación. Por ejemplo, en el proceso de "amplificación puente" utilizado en la secuenciación de Solexa, aproximadamente 1 millón de copias del fragmento de la molécula de ADN original están presentes en dicho racimo. A continuación, dependiendo de la química de la secuenciación y la metodología de las realizaciones particulares, se añaden bases a la colección de racimos (o, de manera equivalente, colonias, polonias). En un procedimiento de conjunto de acuerdo con la presente tecnología, el grado de marcaje está directamente asociado con la intensidad de la señal producida. Por ejemplo, una base que tiene una fracción marcada de 0,25 producirá una señal de aproximadamente el 25 % de la señal para una base que tiene una fracción marcada de 1,00.

En general, se utilizan dos enfoques para la adición de bases en la secuenciación por síntesis basada en conjunto: en el primero, las bases se proporcionan de a una por vez; en el segundo, las bases se modifican con restos identificadores para que el tipo de base del nucleótido incorporado se identifique a medida que avanza la síntesis.

En algunas realizaciones, la síntesis se controla de forma sincronizada añadiendo una base cada vez (véase, por ejemplo, Margulies, M. y col. "Genome sequencing in microfabricated high-density picolitre reactors", Nature 437: 376-380 (2005); Harris, T.D. y col. "Single-molecule DNA sequencing of a viral genome", Science 320: 106-109 (2008)) o mediante el uso de nucleótidos que están bloqueados de forma reversible. En realizaciones particulares, la extensión se bloquea momentáneamente después de cada adición de base usando nucleótidos modificados (por ejemplo, terminadores reversibles de nucleótidos como se describe en, por ejemplo, el documento WO2004/018497; la publicación de solicitud de patente de Estados Unidos n.º 2007/0166705; Bentley, D.R. y col. "Accurate whole human genome sequencing using reversible terminator chemistry", Nature 456: 53-59 (2008); Turcatti, G. y col. "A new class of cleavable fluorescent nucleotides: synthesis and optimization as reversible terminators for DNA sequencing by synthesis", Nucleic Acids Res. 36: e25 (2008); Guo, J. y col. "Four-color DNA sequencing with 3'-O-modified nucleotide reversible terminators and chemically cleavable fluorescent dideoxynucleotides", Proc. Natl. Acad. Sci. USA 105: 9145-9150 (2008); Ju, J. y col. "Four-color DNA sequencing by synthesis using cleavable

fluorescent nucleotide reversible terminators", Proc. Natl. Acad. Sci. USA 103: 19635–19640 (2006); Seo, T.S. y col. "Four-color DNA sequencing by synthesis using cleavable fluorescent nucleotide reversible terminators", Proc. Natl. Acad. Sci. USA 102: 5926–5931 (2005); Wu, W. y col. "Termination of DNA synthesis by N6-alkylated, not 3'-O-alkylated, photocleavable 2'-deoxyadenosine triphosphates", Nucleic Acids Res. 35: 6339–6349 (2007)) u omitiendo componentes de la reacción, tales como iones metálicos divalentes (véase, por ejemplo, el documento WO 2005/123957; la publicación de solicitud de patente de Estados Unidos n.º 20060051807).

Normalmente, a cada adición de base le sigue una etapa de lavado para eliminar el exceso de reactivos. A continuación, mientras se detiene la síntesis, se crean imágenes de los racimos para determinar qué base se ha añadido. En realizaciones en las que se añade una base por ciclo de reacción, la incorporación exitosa de una base indica la base (y por lo tanto la secuencia) en esa posición. Estas adiciones de bases se detectan normalmente mediante fluorescencia (véase, por ejemplo, Harris, citado anteriormente) o por cascadas enzimáticas que identifican la liberación de pirofosfato mediante la producción de luz (véase, por ejemplo, Margulies, citado anteriormente; Bentley, citado anteriormente). De acuerdo con la tecnología proporcionada en el presente documento, la identidad de la base se asocia con la intensidad de la señal generada, que, a su vez, está asociada con el grado de marcaje de las bases (y, en algunas realizaciones, con el grado de marcaje y el color).

Cuando todas las bases se añaden simultáneamente, las bases se discriminan convencionalmente mediante marcas diferentes (por ejemplo, restos fluorescentes) unidas a cada base (véase, por ejemplo, Korlach, J. y col. "Selective aluminum passivation for targeted immobilization of single DNA polymerase molecules in zero-mode waveguide nanostructures", Proc. Natl. Acad. Sci. USA 105: 1176–1181 (2008); la publicación de solicitud de patente de Estados Unidos n.º US 20030194740 la publicación de solicitud de patente de Estados Unidos n.º US 20030064366; Turcatti, G., y col. "A new class of cleavable fluorescent nucleotides: synthesis and optimization as reversible terminators for DNA sequencing by synthesis", Nucleic Acids Res. 36: e25 (2008); Guo, J. y col. "Four-color DNA sequencing with 3'-O-modified nucleotide reversible terminators and chemically cleavable fluorescent dideoxynucleotides", Proc. Natl. Acad. Sci. USA 105: 9145–9150 (2008); Ju, J. y col. "Four-color DNA sequencing by synthesis using cleavable fluorescent nucleotide reversible terminators", Proc. Natl. Acad. Sci. USA 103: 19635–19640 (2006); Seo, T.S. y col. "Four-color DNA sequencing by synthesis using cleavable fluorescent nucleotide reversible terminators", Proc. Natl. Acad. Sci. USA 102: 5926–5931 (2005); Wu, W. y col. "Termination of DNA synthesis by N6-alkylated, not 3'-O-alkylated, photocleavable 2'-deoxyadenosine triphosphates", Nucleic Acids Res. 35: 6339–6349 (2007); el documento WO 2006/084132). De acuerdo con la tecnología proporcionada en el presente documento, la identidad de la base se asocia con la intensidad (por ejemplo, la amplitud) de la señal generada, que, a su vez, está asociada con el grado de marcaje de las bases (y, en algunas realizaciones, con el grado de marcaje y el color).

Por ejemplo, en algunas realizaciones, los cuatro nucleótidos se añaden simultáneamente a la reacción que comprende ADN polimerasa y los racimos de complejos molde-cebador. En algunas realizaciones, los nucleótidos llevan un marcador fluorescente y el grupo hidroxilo en 3' está químicamente bloqueado (por ejemplo, con un terminador reversible marcado) de modo que la síntesis se detiene después de que se incorpora una base en la cadena de ADN en crecimiento (sintetizada). Un paso de formación de imágenes sigue cada etapa de incorporación de bases, durante el cual se crean imágenes de los racimos. Para obtener imágenes de los racimos, en algunas realizaciones los marcadores fluorescentes son excitados por un láser y luego se registra la fluorescencia emitida desde los racimos. En algunas realizaciones, la formación de imágenes registra el color y/o la intensidad de la fluorescencia. De acuerdo con realizaciones de la presente tecnología, al menos dos bases están marcadas en diferentes grados y, por lo tanto, las diferencias en intensidad distinguen las bases entre sí. Después, antes de iniciar el siguiente ciclo sintético, los grupos bloqueantes en el extremo 3' se eliminan para proporcionar un sustrato para la incorporación de la siguiente base. Los ciclos se repiten de esta manera para determinar la secuencia de los moldes una base cada vez.

En algunas realizaciones, cada nucleótido se añade de uno en uno a una mezcla de reacción que contiene el complejo de ácido nucleico diana-cebador y la polimerasa, controlando la reacción para detectar una señal y eliminando la base de la reacción. Por ejemplo, una realización ilustrativa del procedimiento comprende:

1. proporcionar un cebador de secuenciación, un molde, una polimerasa y soluciones de las cuatro bases A, C, G y T
2. hibridar el cebador con el molde en condiciones químicas y físicas apropiadas
3. añadir una alícuota de una solución que comprende la base A a la reacción
4. controlar la reacción para la producción de una señal
5. eliminar la base A de la solución
6. añadir una alícuota de una solución que comprende la base C a la reacción
7. controlar la reacción para la producción de una señal
8. eliminar la base C de la solución
9. añadir una alícuota de una solución que comprende la base G a la reacción
10. controlar la reacción para la producción de una señal
11. eliminar la base G de la solución
12. añadir una alícuota de una solución que comprende la base T a la reacción

13. controlar la reacción para la producción de una señal
14. eliminar la base T de la solución
15. repetir las etapas 3-14 hasta que el molde esté secuenciado

5 Durante cada etapa de monitorización, la detección de una señal de salida apropiada para la base añadida en la etapa anterior indica una incorporación exitosa de esa base y, por lo tanto, identifica la base incorporada en dicha etapa.

10 La detección puede ser por modos convencionales. Por ejemplo, si el marcador es un resto fluorescente, la detección de una base incorporada se puede llevar a cabo utilizando un microscopio de barrido confocal para escanear la colección de racimos (por ejemplo, unidos a una superficie) con un láser para formar imágenes de un fluoróforo unido directamente a la base incorporada. Como alternativa, se puede usar un detector 2-D sensible, tal como un detector acoplado a carga (CCD), para visualizar las señales individuales generadas. Sin embargo, hay disponibles otras técnicas, tales como microscopía óptica de barrido campo cercano (SNOM), y se pueden usar al

15 generar imágenes de matrices densas. Por ejemplo, usando SNOM, se pueden distinguir polinucleótidos individuales cuando están separados por una distancia de menos de 100 nm, por ejemplo de 10 nm a 10 fm. Para obtener una descripción de la microscopía óptica de barrido de campo cercano, consulte Moyer y col., *Laser Focus World* (1993) 29:10. Se conocen aparatos adecuados usados para formar imágenes de matrices de polinucleótidos y la configuración técnica será evidente para el experto. La detección se usa, preferentemente, en combinación con un

20 sistema de análisis para determinar el número y la naturaleza de los nucleótidos incorporados para cada etapa de la síntesis. Este análisis, que puede llevarse a cabo inmediatamente después de cada etapa de síntesis, o más tarde utilizando datos registrados, permite determinar la secuencia del molde de ácido nucleico dentro de una colonia dada.

25 Si bien esta realización de ejemplo indica la adición de las bases en el orden A, C, G y T, la tecnología no está limitada a este orden. De hecho, en algunas realizaciones, las bases se añaden en cualquier orden permutado del conjunto {A C G T} o {A C G U}, por ejemplo, A, G, C, T; A, T, C, G; T, C, G, A, etc. Además, algunas realizaciones proporcionan que se añadan análogos de las bases, bases modificadas y otras moléculas en lugar de A, C, G y T. Debe entenderse que los nucleótidos que comprenden uridina ("U") puede usarse en lugar de T y viceversa. Si la

30 secuencia que se está determinando es desconocida, los nucleótidos añadidos se aplican generalmente en un orden elegido que luego se repite a lo largo del análisis, por ejemplo como se ha tratado anteriormente. Sin embargo, si la secuencia que se está determinando es conocida y se vuelve a secuenciar, por ejemplo, para determinar si hay pequeñas diferencias presentes en la secuencia relativa a la secuencia conocida, el proceso de determinación de secuenciación puede hacerse más rápido añadiendo los nucleótidos en cada etapa en el orden apropiado, elegido de acuerdo con la secuencia conocida. Por lo tanto, las diferencias de la secuencia dada se detectan por la falta de

35 incorporación de ciertos nucleótidos en etapas particulares de la extensión del cebador.

Como un procedimiento mejorado para detectar la adición de bases en SBS, la tecnología es generalmente aplicable a los procedimientos de SBS en los que las bases están marcadas diferencialmente para identificarlas. Sin embargo,

40 si bien las tecnologías convencionales diferencian las bases solo por color, la tecnología proporcionada en el presente documento diferencia las bases por diferencias de intensidad. En algunas realizaciones, las diferencias en intensidad y color diferencian las bases. Por ejemplo, en algunas realizaciones, las cuatro bases están marcadas en una extensión diferente. En algunas realizaciones, se usan dos colores y dos intensidades en combinación para diferenciar las bases, por ejemplo, dos bases se marcan con un primer marcador fluorescente en dos extensiones

45 diferentes y las otras dos bases se marcan con un segundo marcador fluorescente en dos extensiones diferentes. Las discusiones sobre los esquemas de identificación y marcaje se tratan más detalladamente a continuación.

Con respecto a los procedimientos y esquemas de secuenciación por síntesis que encuentran uso, por ejemplo, según se adapte apropiadamente a los procedimientos provistos en el presente documento, Morozova y Marra ofrecen una revisión de algunas de esas tecnologías en *Genomics* 92: 255 (2008); discusiones adicionales se encuentran en *Mardis, Annu. Rev. Genomics Hum. Genet.* (2008) 9:387-402 y en *Fuller, y col.* (2009) *Nat. Biotechnol.* 27: 1013.

50

Más específicamente, algunas realizaciones proporcionan el uso de bases marcadas en diferentes extensiones en una técnica de secuenciación por síntesis de conjunto, tal como la siguiente: secuenciación paralela de amplicones divididos (publicación PCT N.º: WO2006084132); extensión de oligonucleótidos paralelos (véase, por ejemplo, la patente de Estados Unidos n.º 5.750.341; la patente de Estados Unidos n.º 6.306.597); secuenciación de colonias (Mitra y col. (2003) *Analytical Biochemistry* 320: 55-65; Shendure y col. (2005) *Science* 309: 1728-1732; la patente de Estados Unidos n.º 6.432.360, la patente de Estados Unidos n.º 6.485.944, la patente de Estados Unidos n.º 6.511.803;); la tecnología de adición de bases individuales Solexa (véase, por ejemplo, Bennett y col. (2005), *Pharmacogenomics* 6: 373-382; la patente de Estados Unidos n.º 6.787.308; la patente de Estados Unidos n.º 6.833.246; la tecnología de secuenciación masiva de firmas paralelas de Lynx (Brenner y col. (2000). *Nat. Biotechnol.* 18: 630-634; la patente de Estados Unidos n.º 5.695.934; la patente de Estados Unidos n.º 5.714.330) y la tecnología de colonias por PCR de Adessi (Adessi y col. (2000). *Nucleic Acid Res.* 28: E87; documento WO 00018957).

65

En una realización a modo de ejemplo, se usa la secuenciación de Solexa. En la plataforma Solexa/Illumina (Voelkerding y col., *Clinical Chem.*, 55: 641–658, 2009; MacLean y col., *Nature Rev. Microbiol.* 7: 287–296; patente de Estados Unidos n.º 6.833.246; patente de Estados Unidos n.º 7.115.400; patente de Estados Unidos n.º 6.969.488; y patente de Estados Unidos n.º 6.787.308, los datos de secuenciación se producen en forma de lecturas de longitud más corta. En este procedimiento, el ADN fragmentado monocatenario se repara en los extremos para generar extremos romos fosforilados en 3', seguido de la adición mediada por el fragmento Klenow de una única base A al extremo 3' de los fragmentos. La adición de A facilita la adición de oligonucleótidos adaptadores sobresalientes de T, que se utilizan posteriormente para capturar las moléculas adaptadoras del molde sobre la superficie de una celda de flujo que está tachonada con anclajes de oligonucleótidos. El anclaje se utiliza como cebador de PCR, pero debido a la longitud del molde y su proximidad a otros oligonucleótidos de anclaje cercanos, la extensión por PCR da como resultado el "arqueamiento" de la molécula para hibridar con un oligonucleótido de anclaje adyacente para formar una estructura de puente (y después de varias rondas de amplificación, un racimo) sobre la superficie de la celda de flujo. Estos bucles de ADN son desnaturalizados y escindidos. A continuación, las hebras delanteras se secuencian con terminadores colorantes reversibles. La secuencia de los nucleótidos incorporados se determina mediante detección de fluorescencia posterior a la incorporación (por ejemplo, mediante diferencias en intensidad), con cada fluorescencia y bloque eliminados antes del siguiente ciclo de adición de dNTP. La longitud de la lectura de la secuencia varía desde 36 nucleótidos a más de 50 nucleótidos, con una producción total superior a mil millones de pares de nucleótidos por análisis.

En algunas realizaciones, se usa una secuencia de calibración para diferenciar las intensidades de señal asociadas con las bases. Por ejemplo, dicha secuencia de calibración comprende, en algunas realizaciones, cada una de las cuatro bases en un orden conocido, de modo que un instrumento de secuenciación se calibra para reconocer las intensidades de señal (debido a la fracción del marcador) esperadas para cada una de las bases complementarias de la secuencia de calibración. En algunas realizaciones, la secuencia de calibración está unida al comienzo de cada ácido nucleico diana que se va a secuenciar. En algunas realizaciones, la secuencia de calibración no está unida a la secuencia diana, sino que se usa para calibrar el instrumento de secuenciación antes de adquirir la secuencia del ácido nucleico diana. En algunas realizaciones, la calibración se usa para más de una ronda de secuenciación. Para las realizaciones en las que una combinación de colores y extensiones de marcaje (y las consecuentes intensidades de señal) diferencia bases entre sí, la secuencia de calibración se ajusta apropiadamente. La secuencia de calibración es de cualquier longitud que proporcione la calibración adecuada. En algunas realizaciones, la secuencia de calibración tiene cuatro bases de longitud; en algunas realizaciones, la secuencia de calibración es de 5, 6, 7, 8, 9, 10, 16, 20, 24, 28, 32, 64 o más bases de longitud.

Algunas realizaciones proporcionan procedimientos para la detección de moléculas o marcaje diferencial de muestras usando reactivos de detección marcados en diferentes grados. Las diferencias de intensidad identifican las moléculas y diferencian las moléculas entre sí. Por ejemplo, algunos procedimientos comprenden poner en contacto una muestra (por ejemplo, una célula, tejido, fluido, etc.) con dos o más anticuerpos en los que cada anticuerpo está marcado en una extensión diferente; algunos procedimientos comprenden poner en contacto una muestra (por ejemplo, una célula, tejido, fluido, etc.) con dos o más sondas oligonucleotídicas marcadas en las que cada sonda está marcada en un grado diferente. En algunas realizaciones, el mismo marcador (por ejemplo, un marcador o resto fluorescente) se une a los reactivos de detección y, en algunas realizaciones, se usan combinaciones de color y fracción marcada para identificar y diferenciar los reactivos de detección. Los procedimientos comprenden diferenciar dos o más moléculas, muestras, tejidos, células, etc. entre sí mediante la asociación de una diferencia en intensidad (y, en algunas realizaciones, también diferencias en el color) con un reactivo de detección.

Composiciones

La tecnología proporciona composiciones, por ejemplo, composiciones de bases de nucleótidos solos o en combinación, en los que el grado de marcaje difiere para al menos dos de las bases. Como se ha indicado anteriormente, la intensidad de señal producida y detectada durante la reacción de SBS varía proporcionalmente con la fracción marcada de cada base. Por ejemplo, la reducción del grado de marcaje reduce la intensidad de la señal y el aumento del grado de marcaje aumenta la intensidad de la señal.

En algunas realizaciones, el grado de marcaje (o "fracción marcada") difiere entre las cuatro bases, permitiendo diferenciar cada base de las otras tres, por ejemplo, ya que cada base se incorpora en una reacción de SBS de conjunto y se produce una señal. Como se usa en el presente documento, el "grado de marcaje" o "fracción marcada" se refiere a la fracción o porción de moléculas de base de un tipo que está marcado. La extensión del etiquetado o la fracción marcada puede ser una fracción de 0,00 a 1,00 (como alternativa, un porcentaje de 0 % a 100 %). Por ejemplo, si el número de moléculas de base A individuales (por ejemplo, en una solución) es 100 y el número de moléculas de base A individuales que están marcadas es 25, el grado de marcaje para A es 0,25 o 25 %. En esta realización de ejemplo, el grado de marcaje para las otras tres bases C, G y T es 0,50 (50 %), 0,75 (75 %) y 1,00 (100 %), respectivamente. Varias realizaciones proporcionan grados de marcaje distintos de estos valores de ejemplo. De hecho, la tecnología contempla cualquier combinación de fracciones marcadas siempre que las cuatro bases se puedan distinguir entre sí según las diferencias en los grados de marcaje y las señales posteriores producidas en una reacción de SBS. En diversas realizaciones, cualquiera de las cuatro bases está marcada en un grado de marcaje que es 10 %, 20 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 % y/o 100 %, y valores intermedios

entre estos valores enumerados, siempre que la fracción marcada para las bases sea suficiente para diferenciar las bases entre sí.

En algunas realizaciones, se elige un primer valor que representa la fracción marcada para una base y un segundo valor que representa la fracción marcada para una segunda base, de modo que los múltiplos bajos de los dos valores rara vez son los mismos (por ejemplo, generalmente, cuando se multiplican por un número menor de 20). Tal esquema encuentra uso en diferenciar rondas homopoliméricas de bases individuales entre sí y a partir de bases individuales, por ejemplo, en algunos esquemas de secuenciación en las que la presencia de la misma base en posiciones consecutivas produce una señal más fuerte que una señal producida por una base.

En algunas realizaciones, las bases se diferencian por color y fracción marcada. Por ejemplo, en algunas realizaciones, dos bases están marcadas con el mismo marcador fluorescente en dos fracciones diferentes y las otras dos bases están marcadas con un marcador fluorescente diferente en dos fracciones diferentes, que pueden ser las mismas o diferentes fracciones que las fracciones utilizadas para la primeras dos bases. Este enfoque se extiende para incluir realizaciones en las que tres bases se marcan con el mismo marcador fluorescente en tres fracciones diferentes y la cuarta base se marca con un marcador fluorescente diferente y para incluir realizaciones en las que tres bases se marcan usando tres marcadores fluorescentes diferentes y la cuarta base se marca usando uno de los mismos marcadores fluorescentes como se usa para las primeras tres bases, pero en un grado de marcaje diferente. Algunas realizaciones contemplan el uso de cuatro marcadores diferentes (por ejemplo, restos fluorescentes) y cuatro fracciones de marcador diferentes, por ejemplo, para proporcionar un esquema de identificación redundante.

Las composiciones proporcionadas por la presente tecnología incluyen soluciones de cuatro bases en las que al menos dos bases están marcadas en un grado diferente. Las realizaciones de las composiciones generalmente comprenden un tampón conocido en la técnica y, opcionalmente, comprenden otras sales y conservantes conocidos por los expertos en la técnica, por ejemplo, para mantener la estabilidad de la composición. Varias realizaciones incluyen composiciones que comprenden una base o mezclas de 2, 3, 4 o más bases. Las bases en estas composiciones se marcan en diferentes grados y/o con diferentes marcadores fluorescentes usando esquemas de identificación como se ha tratado anteriormente.

Algunas realizaciones proporcionan una composición que comprende un oligonucleótido de calibración que comprende o que consiste en una secuencia conocida de bases. En algunas realizaciones, el oligonucleótido de calibración comprende o consiste en 4, 5, 6, 7, 8 o más bases cuya secuencia es conocida. El oligonucleótido, en algunas realizaciones, se sintetiza químicamente.

De acuerdo con la tecnología, las bases se marcan con un resto que da como resultado la producción de una señal detectable tras la incorporación de la base en la cadena de ADN que se está sintetizando. En algunas realizaciones, el resto produce una señal (por ejemplo, fluorescencia) antes de la incorporación y/o después de la incorporación. En algunas realizaciones, el resto está unido de tal manera que es apropiado para eliminar el resto después de la incorporación o después de la formación de imágenes. El resto marcador es, en algunas realizaciones, un colorante orgánico fluorescente derivatizado para unirse a la base directamente o a través de un enlazador. La guía práctica está disponible en la literatura que proporciona una lista de moléculas fluorescentes y cromogénicas y sus propiedades ópticas relevantes (véase, por ejemplo, Berlman, Handbook of Fluorescence Spectra of Aromatic Molecules, 2ª Edición (Academic Press, New York, 1971); Griffiths, Colour and Constitution of Organic Molecules (Academic Press, New York, 1976); Bishop, Ed., Indicators (Pergamon Press, Oxford, 1972); Haugland, Handbook of Fluorescent Probes and Research Chemicals (Molecular Probes, Eugene, 1992) Pringsheim, Fluorescence and Phosphorescence (Interscience Publishers, New York, 1949); y similares.

Además, existe una guía en la bibliografía para derivar moléculas de fluoróforo para unión covalente a través de grupos reactivos comunes que pueden añadirse a un nucleótido, como se ilustra a modo de ejemplo en Haugland (citado anteriormente); Ullman et al, patente de Estados Unidos n.º 3.996.345; Khanna et al, patente de Estados Unidos n.º 4.351.760. Existen muchos restos de unión y metodologías para unir restos fluorescentes o inactivadores a nucleótidos, como se ilustra como ejemplo en las siguientes referencias: Eckstein, editor, Oligonucleotides and Analogues: A Practical Approach (IRL Press, Oxford, 1991); Zuckerman y col. (1987), Nucleic Acids Research 15: 5305–5321; Sharma y col. (1991), Nucleic Acids Research 19: 3019; Giusti y col., PCR Methods and Applications 2: 223–227 (1993); Fung y col., patente de Estados Unidos n.º 4.757.141; Stabinsky, patente de Estados Unidos n.º 4.739.044; Agrawal y col. (1990), Tetrahedron Letters 31: 1543–1546; Sproat y col. (1987), Nucleic Acids Research 15: 4837; Nelson y col. (1989), Nucleic Acids Research 17: 7187–7194; y similares. Se describen varias metodologías de secuenciación basadas en fluorescencia de ADN, por ejemplo, en Birren y col., Genome Analysis: Analyzing DNA, (Cold Spring Harbor, N.Y.).

Las realizaciones de la tecnología comprenden composiciones que comprenden un molde de ácido nucleico diana. En algunas realizaciones, la composición comprende un cebador, por ejemplo, en algunas realizaciones, que está unido al molde de ácido nucleico diana.

El ácido nucleico objetivo no es crucial y puede provenir de diversas fuentes estándar. Puede ser ARNm, ARN ribosómico, ADN genómico o ADNc. Cuando la diana procede de una fuente biológica, se conocen procedimientos para extraer ácido nucleico y, opcionalmente, amplificarlo a una concentración conveniente para el genotipado o el trabajo de secuencia. El ácido nucleico se puede obtener de cualquier célula viva de una persona, animal o planta (y en muchos casos de células muertas o cualquier otra materia de origen biológico). Los seres humanos, los microbios patogénicos y los virus son fuentes particularmente interesantes. También se conocen procedimientos de amplificación de ácidos nucleicos. Preferentemente, la amplificación se lleva a cabo mediante reacción en cadena de la polimerasa (PCR) (patentes de Estados Unidos n.º 4.683.202; 4.683.195; y 4.889.818; Gyllenstein et al (1988) Proc. Natl. Acad. Sci. USA 85: 7652–7656; Ochman y col. (1988) Genetics 120: 621–623; Loh et al (1989) Science 243: 217–220; Innis et al (1990) PCR Protocols (Academic Press, San Diego, CA). Se pueden usar otros procedimientos de amplificación conocidos en la técnica, que incluyen, pero no se limitan a los mismos, la reacción en cadena de la ligasa (véase, por ejemplo, el documento EP 320308), el uso de Q-beta replicasa, o los procedimientos enumerados en Kricka y col., 1995, Molecular Probing, Blotting, and Sequencing, Cap. 1 y la Tabla IX (Academic Press, New York).

La tecnología proporcionada en el presente documento se refiere al uso de una polimerasa en una reacción de secuenciación. En general, las polimerasas que encuentran uso en la tecnología toleran los marcadores en diversas posiciones, por ejemplo, en la base nucleotídica, en el fosfato gamma, en el hidroxilo en 3'. Por ejemplo, las polimerasas que encuentran uso en la tecnología son la ADN polimerasa de T7, la ADN polimerasa de T5, la transcriptasa inversa del VIH, la ADN pol I de *E. coli*, la ADN polimerasa de T4, la ARN polimerasa de T7, la ADN polimerasa de Taq y la ARN polimerasa de *E. coli*. En algunas realizaciones, se usa una polimerasa defectuosa en exonucleasa. En algunas realizaciones (por ejemplo, una tecnología de terminación reversible), se usa una polimerasa que tiene una actividad de exonucleasa para algunas o todas las etapas.

Los cebadores (por ejemplo, como se usan para la síntesis mediante ADN polimerasa) o promotores (por ejemplo, como se usan para la síntesis por ARN polimerasa) se sintetizan usando tecnología de síntesis de ácido nucleico convencional, por ejemplo, usando un sintetizador de ADN automático y químicas estándar, tales como química de fosforamidita, por ejemplo, como se desvela en las siguientes referencias: Beaucage y Iyer, Tetrahedron 48: 2223–211 (1992); patente de Estados Unidos n.º 4.980.460; patente de Estados Unidos n.º 4.725.677; patentes de Estados Unidos n.º 4.415.732; 4.458.066; y 4.973.679; y similares. También pueden emplearse químicas alternativas, por ejemplo, que dan como resultado grupos estructurales no naturales, tales como fosforotioato, fosforamidato y similares, con la condición de que los oligonucleótidos resultantes sean compatibles con la polimerasa. Se pueden solicitar comercialmente a diversas empresas que se especializan en oligonucleótidos adaptados, tales como Operon, IDT, Dharmacon, etc.

Se usan cebadores en combinación con polimerasas para secuenciar el ADN diana. La longitud del cebador se selecciona para proporcionar la hibridación con ADN molde complementario. Los cebadores generalmente tendrán al menos 10 nt de longitud, normalmente al menos entre 15 y 30 nt de longitud. Los cebadores están diseñados para hibridar con sitios internos conocidos en el ADN diana del sujeto. Como alternativa, los cebadores pueden unirse a adaptadores de oligonucleótidos sintéticos unidos a los extremos del ADN diana mediante una ligasa. De forma similar, cuando se usan promotores, pueden ser internos al ADN diana o ligados como adaptadores a los extremos.

La mezcla de reacción para la secuenciación comprende un medio tampón acuoso que está optimizado para la polimerasa particular elegida. En general, el tampón incluye una fuente de iones monovalentes, una fuente de cationes divalentes y un agente tampón. Se puede emplear cualquier fuente conveniente de iones monovalentes, tales como cloruro de potasio, acetato de potasio, acetato de potasio, glutamato de potasio, cloruro de amonio, sulfato de amonio y similares.

El catión divalente puede ser magnesio, manganeso, cinc y similares, en el que el catión será típicamente magnesio. Se puede utilizar cualquier fuente conveniente de catión magnesio, incluido $MgCl_2$, acetato de magnesio y similares. La cantidad de ion Mg presente en el tampón puede variar de 0,5 a 20 mM, pero, preferentemente, variará de aproximadamente 1 a 12 mM, más preferentemente de 2 a 10 mM, e idealmente será de aproximadamente 5 mM.

Los agentes tampón o sales representativas que pueden estar presentes en composiciones según la tecnología descrita (por ejemplo, en una composición que comprende un nucleótido marcado o en una reacción de SBS) incluyen Tris, Tricina, HEPES, MOPS y similares, en las que la cantidad de agente tampón típicamente variará de aproximadamente 5 a 150 mM, normalmente de aproximadamente 10 a 100 mM y más habitualmente de aproximadamente 20 a 50 mM, cuando en ciertas realizaciones preferentes, el agente tampón estará presente en una cantidad suficiente para proporcionar un pH que varía desde aproximadamente 6,0 a 9,5. Otros agentes que pueden estar presentes en el medio tampón incluyen agentes quelantes, tales como EDTA, EGTA y similares.

Algunas realizaciones proporcionan composiciones para la detección de moléculas o para el marcaje diferencial de muestras usando reactivos de detección marcados en diferentes grados. Las diferencias de intensidad identifican las moléculas y diferencian las moléculas entre sí. Por ejemplo, algunas composiciones comprenden dos o más anticuerpos en los que cada anticuerpo está marcado en un grado diferente; algunas composiciones comprenden dos o más sondas oligonucleotídicas marcadas en las que cada sonda está marcada en un grado diferente. En

algunas realizaciones, el mismo marcador (por ejemplo, un marcador o resto fluorescente) se une a los reactivos de detección y, en algunas realizaciones, se usan combinaciones de color y fracción marcada para identificar y diferenciar los reactivos de detección. En algunas realizaciones, uno o más de los reactivos de detección marcados es/son uno de los siguientes: un aptámero, una proteína, un ácido nucleico, un ácido nucleico bloqueado, un ARN, un ADN, un fragmento de anticuerpo, una molécula pequeña, una proteína, un dominio de unión a ADN de una proteína, un dominio de unión a proteína de una proteína y otros reactivos de detección conocidos en la técnica.

Análisis de datos

Algunas realizaciones comprenden un sistema informático sobre el que pueden implementarse las realizaciones de las presentes enseñanzas. En diversas realizaciones, un sistema informático incluye un bus u otro mecanismo de comunicación para comunicar información, y un procesador acoplado con el bus para procesar información. En diversas realizaciones, el sistema informático incluye una memoria, que puede ser una memoria de acceso aleatorio (RAM) u otro dispositivo de almacenamiento dinámico, acoplado al bus para identificar bases (por ejemplo, hacer "llamadas de bases") e instrucciones para ser ejecutadas por el procesador. La memoria también se puede usar para almacenar variables temporales u otra información intermedia durante la ejecución de las instrucciones que debe ejecutar el procesador. En diversas realizaciones, el sistema informático puede incluir además una memoria de solo lectura (ROM) u otro dispositivo de almacenamiento estático acoplado al bus para almacenar información estática e instrucciones para el procesador. Se puede proporcionar un dispositivo de almacenamiento, tal como un disco magnético o disco óptico, y acoplarlo al bus para almacenar información e instrucciones.

En diversas realizaciones, el sistema informático está acoplado a través del bus a una pantalla, tal como un tubo de rayos catódicos (CRT) o una pantalla de cristal líquido (LCD), para mostrar información a un usuario de ordenador. Un dispositivo de entrada, que incluye teclas alfanuméricas y otras teclas, se puede acoplar al bus para comunicar información y selecciones de comandos al procesador. Otro tipo de dispositivo de entrada del usuario es un control del cursor, tal como un ratón, una rueda de desplazamiento o teclas de dirección del cursor para comunicar información de dirección y selecciones de comandos al procesador y para controlar el movimiento del cursor en la pantalla. Este dispositivo de entrada típicamente tiene dos grados de libertad en dos ejes, un primer eje (por ejemplo, x) y un segundo eje (por ejemplo, y), que permite al dispositivo especificar posiciones en un plano.

Un sistema informático puede efectuar realizaciones de la tecnología actual. De acuerdo con ciertas implementaciones de las presentes enseñanzas, el sistema informático puede proporcionar los resultados en respuesta al procesador que ejecuta una o más secuencias de una o más instrucciones contenidas en la memoria. Dichas instrucciones se pueden leer en la memoria desde otro medio legible por ordenador, tal como un dispositivo de almacenamiento. La ejecución de las secuencias de instrucciones contenidas en la memoria puede hacer que el procesador realice los procedimientos descritos en el presente documento. Como alternativa, puede utilizarse una circuitería programada en lugar de, o en combinación con, instrucciones de software para implementar las presentes enseñanzas. Por lo tanto, las implementaciones de las presentes enseñanzas no están limitadas a ninguna combinación específica de circuitería de hardware y software.

La expresión "medio legible por ordenador", tal como se utiliza en el presente documento, se refiere a cualquier medio o que participa en proporcionar instrucciones al procesador para su ejecución. Tal medio puede adoptar muchas formas, incluyendo, pero sin limitaciones, medios no volátiles, medios volátiles y medios de transmisión. Los ejemplos de medios no volátiles pueden incluir, pero no están limitados a los mismos, discos ópticos o magnéticos, tales como un dispositivo de almacenamiento. Los ejemplos de medios volátiles pueden incluir, pero no están limitados a la misma, una memoria dinámica. Los ejemplos de medios de transmisión pueden incluir, pero sin limitaciones a los mismos, cables coaxiales, hilos de cobre y fibras ópticas, incluidos los cables que componen el bus.

Las formas frecuentes de medios legibles por ordenador incluyen, por ejemplo, un disquete, un disco flexible, disco duro, cinta magnética o cualquier otro medio magnético, un CD-ROM, cualquier otro medio óptico, tarjetas perforadas, cinta de papel, cualquier otro medio físico con patrones de agujeros, una RAM, PROM y EPROM, una FLASH-EPROM, cualquier otro chip o cartucho de memoria, o cualquier otro medio tangible desde el cual una pueda leer un ordenador.

Varias formas de medios legibles por ordenador pueden estar implicadas en llevar una o más secuencias de una o más instrucciones al procesador para su ejecución. Por ejemplo, las instrucciones pueden llevarse inicialmente en el disco magnético de un ordenador remoto. El ordenador remoto puede cargar las instrucciones en su memoria dinámica y enviar las instrucciones a través de una conexión de red (por ejemplo, una LAN, una WAN, Internet, una línea telefónica). Un sistema informático local puede recibir los datos y transmitirlos al bus. El bus puede llevar los datos a la memoria, desde la cual el procesador recupera y ejecuta las instrucciones. Las instrucciones recibidas por la memoria pueden almacenarse opcionalmente en un dispositivo de almacenamiento antes o después de la ejecución por parte del procesador.

De acuerdo con diversas realizaciones, las instrucciones configuradas para ser ejecutadas por un procesador para realizar un procedimiento se almacenan en un medio legible por ordenador. El medio legible por ordenador puede

ser un dispositivo que almacena información digital. Por ejemplo, un medio legible por ordenador incluye una memoria de solo lectura de disco compacto (CD-ROM) como se conoce en la técnica para almacenar software. Al medio legible por ordenador se accede a través de un procesador adecuado para ejecutar instrucciones configuradas para ejecutarse.

5 De acuerdo con dicho sistema informático, algunas realizaciones de la tecnología proporcionada en la presente memoria comprenden además funcionalidades para recoger, almacenar y/o analizar datos (por ejemplo, datos de secuencias de nucleótidos). Por ejemplo, algunas realizaciones contemplan un sistema que comprende un procesador, una memoria y/o una base de datos para, por ejemplo, almacenar y ejecutar instrucciones, analizar
10 datos de imágenes de una reacción de SBS, realizar cálculos usando los datos, transformar los datos y almacenar los datos. En algunas realizaciones, un algoritmo de llamada de bases asigna una secuencia de bases a los datos y asocia puntuaciones de calidad a las llamadas de bases según un modelo estadístico. En algunas realizaciones, el sistema está configurado para ensamblar una secuencia a partir de múltiples subsecuencias, en algunos casos teniendo en cuenta la superposición y calculando una secuencia consenso. En algunas realizaciones, una secuencia
15 determinada a partir de una reacción de SBS se alinea con una secuencia de referencia o con un almacén.

Muchos diagnósticos implican determinar la presencia de, o una secuencia de nucleótidos de, uno o más ácidos nucleicos. Por lo tanto, en algunas realizaciones, una ecuación que comprende variables que representan la presencia o las propiedades de secuencia de múltiples ácidos nucleicos produce un valor que encuentra uso al
20 hacer un diagnóstico o evaluar la presencia o las cualidades de un ácido nucleico. Como tal, en algunas realizaciones, este valor es presentado por un dispositivo, por ejemplo, mediante un indicador relacionado con el resultado (por ejemplo, un LED, un icono en una LCD, un sonido o similar). En algunas realizaciones, un dispositivo almacena el valor, transmite el valor o usa el valor para cálculos adicionales.

25 Además, en algunas realizaciones, un procesador está configurado para controlar las reacciones de secuenciación y recopilar los datos (por ejemplo, imágenes). En algunas realizaciones, el procesador se usa para iniciar y/o finalizar cada ronda de secuenciación y recopilación de datos relacionada con una reacción de secuenciación. Algunas realizaciones comprenden un procesador configurado para analizar el conjunto de datos de intensidades y/o colores adquiridos durante la reacción de SBS y discernir la secuencia del ácido nucleico diana y/o de su complementario.

30 En algunas realizaciones, el procesador utiliza un dispositivo que comprende una interfaz de usuario (por ejemplo, un teclado, botones, diales, conmutadores y similares) para recibir la entrada del usuario para dirigir una medición. En algunas realizaciones, el dispositivo comprende además una salida de datos para transmitir (por ejemplo, mediante una conexión por cable o inalámbrica) datos a un destino externo, por ejemplo, un ordenador, una pantalla,
35 una red y/o un medio de almacenamiento externo.

En algunas realizaciones, la tecnología encuentra uso en el ensayo de la presencia de uno o más ácidos nucleicos y/o proporciona la secuencia de uno o más ácidos nucleicos. En consecuencia, la tecnología proporcionada en el presente documento encuentra uso en los campos médico, clínico y de medicina de urgencias. En algunas
40 realizaciones, se usa un dispositivo para analizar muestras biológicas. En tal ensayo, la muestra biológica comprende un ácido nucleico y la secuenciación del ácido nucleico es indicativa de un estado o una propiedad de la muestra y, en algunas realizaciones, el sujeto del que se tomó la muestra. Algunas muestras relevantes incluyen, pero sin limitaciones a las mismas, sangre entera, linfa, plasma, suero, saliva, orina, heces, sudoración, moco, lágrimas, líquido cefalorraquídeo, secreción nasal, secreción cervical o vaginal, semen, líquido pleural, líquido
45 amniótico, fluido peritoneal, fluido del oído medio, fluido articular, aspirado gástrico, un homogeneizado tisular, un homogeneizado celular o similares.

La secuencia de las señales de salida proporciona la secuencia del ADN sintetizado y, por las reglas de complementariedad de las bases, también proporciona de este modo la secuencia de la cadena molde.

50 Aparatos

Un aspecto adicional de la invención proporciona un aparato para llevar a cabo los procedimientos o para preparar las composiciones de la tecnología. Tal aparato podría comprender, por ejemplo, una pluralidad de moldes de ácido nucleico y cebadores unidos, preferentemente de forma covalente, a un soporte sólido, junto con una ácido nucleico polimerasa, una pluralidad de nucleótidos tales como los descritos anteriormente, una proporción de los cuales están marcados (la fracción marcada), y una funcionalidad para controlar la temperatura y/o las adiciones de nucleótidos. Preferentemente, el aparato también comprende una funcionalidad de detección para detectar y distinguir señales de racimos de ácidos nucleicos individuales. Dicha funcionalidad de detección podría comprender un dispositivo de
60 acoplamiento de carga conectado operativamente a un dispositivo de aumento, tal como un microscopio. Preferentemente, cualquier aparato de la invención se proporciona de forma automatizada, por ejemplo, bajo el control de un programa de etapas y decisiones, por ejemplo, como se implementa en un software de ordenador.

65 Algunas realizaciones de un aparato de este tipo incluyen una unidad de suministro y control de fluidos, una unidad de procesamiento de muestras, una unidad de detección de señales y una unidad de adquisición, análisis y control de datos. Diversas realizaciones del aparato pueden proporcionar una secuenciación automática que puede usarse

para recopilar información de secuencia de una pluralidad de secuencias en paralelo, por ejemplo, sustancialmente de forma simultánea.

5 En diversas realizaciones, la unidad de suministro y control de fluidos incluye un sistema de suministro de reactivos. El sistema de suministro de reactivos puede incluir un depósito de reactivo para el almacenamiento de diversos reactivos (por ejemplo, composiciones de nucleótidos de acuerdo con la tecnología). Los reactivos pueden incluir cebadores basados en ARN, cebadores de ADN directos/inversos, mezclas de oligonucleótidos para la secuenciación por ligación, mezclas de nucleótidos para la secuenciación por síntesis, tampones, reactivos de lavado, reactivo de bloqueo, reactivos de extracción y similares. Adicionalmente, el sistema de suministro de reactivos puede incluir un sistema de pipeteo o un sistema de flujo continuo que conecta la unidad de procesamiento de muestras con el depósito de reactivos.

15 En diversas realizaciones, la unidad de procesamiento de muestras puede incluir una cámara de muestra, tal como una celda de flujo, un sustrato, una micromatriz, una bandeja de múltiples pocillos o similar. La unidad de procesamiento de muestras puede incluir múltiples carriles, múltiples canales, múltiples pocillos u otros modos de procesamiento de conjuntos de muestras múltiples de manera sustancialmente simultánea. Adicionalmente, la unidad de procesamiento de muestras puede incluir múltiples cámaras de muestras para permitir el procesamiento de múltiples rondas simultáneamente. En realizaciones particulares, el sistema puede realizar la detección de señal en una cámara de muestras mientras que procesa de manera sustancialmente simultánea otra cámara de muestras. Adicionalmente, la unidad de procesamiento de muestras puede incluir un sistema de automatización para mover o manipular la cámara de muestras.

25 En diversas realizaciones, la unidad de detección de señal puede incluir un sensor de formación de imágenes o detección. La unidad de detección de señal puede incluir un sistema de excitación para provocar que una sonda, tal como un colorante fluorescente, emita una señal. El sistema de excitación puede incluir una fuente de iluminación, tal como una lámpara de arco, un láser, un diodo emisor de luz (LED) o similar. En realizaciones particulares, la unidad de detección de señal puede incluir ópticas para la transmisión de luz desde una fuente de iluminación a la muestra o desde la muestra al sensor de formación o detección de imágenes. Como alternativa, la unidad de detección de señal puede no incluir una fuente de iluminación, tal como, por ejemplo, cuando se produce una señal espontáneamente como resultado de una reacción de secuenciación. Por ejemplo, puede producirse una señal mediante la interacción de un resto liberado, tal como un ion liberado que interacciona con una capa sensible a iones, o un pirofosfato que reacciona con una enzima u otro catalizador para producir una señal quimioluminiscente.

35 En diversas realizaciones, la unidad de control y adquisición de datos puede controlar diversos parámetros del sistema. Los parámetros del sistema pueden incluir la temperatura de varias partes del instrumento, tales como la unidad de procesamiento de muestras o los depósitos de reactivos, los volúmenes de varios reactivos, el estado de varios subcomponentes del sistema, tal como un manipulador, un motor de pasos, una bomba o similares, o cualquier combinación de los mismos.

40 Un experto en la técnica apreciará que pueden usarse diversas realizaciones de dicho instrumento para practicar diversos procedimientos de secuenciación, que incluyen procedimientos basados en ligamiento, secuenciación por síntesis, procedimientos de molécula única y otras técnicas de secuenciación. La secuenciación de ligamiento puede incluir técnicas de ligamiento único, o cambiar las técnicas de ligación en las que se realizan ligamientos múltiples en secuencia en un solo primario. La secuenciación por síntesis puede incluir la incorporación de nucleótidos marcados con colorante, terminación de cadena o similares. Las técnicas de una sola molécula pueden incluir secuenciación escalonada, en las que las reacciones de secuenciación se detienen para determinar la identidad del nucleótido incorporado.

50 En diversas realizaciones, el instrumento de secuenciación puede determinar la secuencia de un ácido nucleico, tal como un polinucleótido o un oligonucleótido. El ácido nucleico puede incluir ADN o ARN, y puede ser monocatenario, tal como ADNss y ARN, o bicatenario, tal como ADNds o un par de ARN/ADNc. En diversas realizaciones, el ácido nucleico puede incluir o derivar de una biblioteca de fragmentos, una biblioteca de pares de parejas, un fragmento CHIP o similar. En realizaciones particulares, el instrumento de secuenciación puede obtener la información de secuencia de una única molécula de ácido nucleico o de un grupo de moléculas de ácido nucleico sustancialmente idénticas.

60 En diversas realizaciones, el instrumento de secuenciación puede dar salida a datos de lectura de la secuenciación de ácido nucleico en diversos tipos/formatos de archivos de datos de salida diferentes, incluyendo, pero sin limitaciones a los mismos: *.fasta, *.csfasta, *.seq.txt, *.qseq.txt, *.fastq, *.sff, *.prb.txt, *.sms, *.srs y/o *.qv.

65 Algunas realizaciones comprenden un sistema para reconstruir una secuencia de ácido nucleico de acuerdo con las diversas realizaciones proporcionadas en el presente documento. El sistema puede incluir un secuenciador de ácido nucleico, un almacenamiento de datos de la secuencia de muestra, un almacenamiento de datos de la secuencia de referencia y un dispositivo/servidor/nodo informático de análisis. En diversas realizaciones, el dispositivo/servidor/nodo de computación analítico puede ser una estación de trabajo, un ordenador central, un ordenador personal, un dispositivo móvil, etc.

El secuenciador de ácido nucleico puede configurarse para analizar (por ejemplo, interrogar) un fragmento de ácido nucleico (por ejemplo, fragmento único, fragmento de pares de parejas, fragmento de extremo apareado, etc.) utilizando todas las variedades de técnicas disponibles, plataformas o tecnologías para obtener información de la secuencia de ácido nucleico, por ejemplo, usando una secuenciación de conjunto por síntesis. En diversas realizaciones, el secuenciador de ácido nucleico puede estar en comunicación con el almacenamiento de datos de la secuencia de muestra directamente a través de un cable de datos (por ejemplo, un cable en serie, una conexión de cable directo, etc.) o enlace de bus o, como alternativa, a través de una conexión de red (por ejemplo, Internet, LAN, WAN, VPN, etc.). En diversas realizaciones, la conexión de red puede ser una conexión física "cableada". Por ejemplo, el secuenciador de ácido nucleico puede conectarse de forma comunicativa (a través de Categoría 5 (CAT5), fibra óptica o un cableado equivalente) a un servidor de datos que puede conectarse de forma comunicativa (a través de CAT5, fibra óptica o un cableado equivalente) a través de Internet y al almacenamiento de datos de la secuencia de muestra. En diversas realizaciones, la conexión de red puede ser una conexión de red inalámbrica (por ejemplo, Wi-Fi, WLAN, etc.), por ejemplo, utilizando un formato de transmisión 802.11b/g o equivalente. En la práctica, la conexión de red utilizada depende de los requisitos particulares del sistema. En diversas realizaciones, el almacenamiento de datos de la secuencia de muestra puede ser una parte integrada del secuenciador de ácido nucleico.

En diversas realizaciones, el almacenamiento de datos de la secuencia de muestra puede ser cualquier dispositivo, sistema o implementación de almacenamiento de base de datos (por ejemplo, partición de almacenamiento de datos, etc.) que está configurado para organizar y almacenar datos de lectura de la secuencia de ácido nucleico generados por secuenciador de ácido nucleico de modo que los datos se pueden buscar y recuperar manualmente (por ejemplo, por un administrador de base de datos/operador de cliente) o automáticamente a través de un programa de ordenador/aplicación/script de software. En diversas realizaciones, el almacenamiento de datos de referencia puede ser cualquier dispositivo, sistema de almacenamiento o implementación de base de datos (por ejemplo, partición de almacenamiento de datos, etc.) que está configurado para organizar y almacenar secuencias de referencia (por ejemplo, genoma completo/parcial, exoma completo/parcial etc.), de forma que los datos se pueden buscar y recuperar manualmente (por ejemplo, por un administrador de base de datos/operador de cliente) o automáticamente a través de un programa de ordenador/aplicación/script de software. En diversas realizaciones, los datos de lectura de la secuencia de ácido nucleico de muestra pueden almacenarse en el almacenamiento de datos de secuencia de muestra y/o el almacenamiento de datos de referencia en diversos tipos/formatos de archivo de datos diferentes, incluyendo, pero sin limitaciones a los mismos: *.fasta, *.csfasta, *seq.txt, *qseq.txt, *.fastq, *.sff, *.prb.txt, *.sms, *srs y/o *.qv.

En diversas realizaciones, el almacenamiento de datos de la secuencia de muestra y el almacenamiento de datos de referencia son dispositivos/sistemas autónomos independientes o implementados en diferentes dispositivos. En diversas realizaciones, el almacenamiento de datos de la secuencia de muestra y el almacenamiento de datos de referencia se implementan en el mismo dispositivo/sistema. En diversas realizaciones, el almacenamiento de datos de la secuencia de muestra y/o el almacenamiento de datos de referencia se pueden implementar en el mismo dispositivo/sistema de computación analítica.

El dispositivo/servidor/nodo de computación analítica puede estar en comunicación con el almacenamiento de datos de la secuencia de muestra y el almacenamiento de datos de referencia directamente a través de un cable de datos (por ejemplo, un cable en serie, una conexión de cable directo, etc.) o enlace de bus o, como alternativa, a través de una conexión de red (por ejemplo, Internet, LAN, WAN, VPN, etc.). En diversas realizaciones, el dispositivo/servidor/nodo de computación analítica puede alojar un motor de mapeo de referencia, un módulo de mapeo de novo y/o un motor de análisis terciario. En diversas realizaciones, el motor de mapeo de referencia puede configurarse para obtener lecturas de la secuencia de ácido nucleico del almacenamiento de datos de muestra y mapearlas contra una o más secuencias de referencia obtenidas del almacenamiento de datos de referencia para ensamblar las lecturas en una secuencia similar pero no necesariamente idéntica a la secuencia de referencia utilizando todas las variedades de técnicas y procedimientos de mapeo/alineación de referencia. La secuencia reensamblada puede analizarse posteriormente mediante uno o más motores de análisis terciarios opcionales para identificar diferencias en la composición genética (genotipo), la expresión génica o el estado epigenético de los individuos que pueden dar lugar a grandes diferencias en las características físicas (fenotipo). Por ejemplo, en diversas realizaciones, el motor de análisis terciario se puede configurar para identificar diversas variantes genómicas (en la secuencia ensamblada) debido a mutaciones, recombinación/reticulación, o deriva genética. Ejemplos de tipos de variantes genómicas incluyen, pero sin limitaciones: polimorfismos de un solo nucleótido (SNP), variaciones en el número de copias (CNV), inserciones/deleciones (Indels), inversiones, etc.

El módulo de mapeo de novo opcional puede configurarse para ensamblar lecturas de secuencias de ácidos nucleicos de muestra a partir del almacenamiento de datos de muestras en secuencias nuevas y previamente desconocidas.

Sin embargo, debe entenderse que los diversos motores y módulos alojados en el dispositivo/servidor/nodo de computación analítica pueden combinarse o contraerse en un único motor o módulo, dependiendo de los requisitos de la aplicación particular o de la arquitectura del sistema. Además, en diversas realizaciones, el

dispositivo/servidor/nodo de computación analítica puede alojar motores o módulos adicionales según lo necesite la aplicación particular o la arquitectura del sistema.

5 En diversas realizaciones, los motores de mapeo y/o análisis terciario están configurados para procesar las lecturas de ácido nucleico y/o secuencia de referencia en el espacio de amplitud de señal. En diversas realizaciones, los motores de mapeo y/o análisis terciario están configurados para procesar las lecturas de ácido nucleico y/o secuencia de referencia en el espacio de color. Sin embargo, debe entenderse que los motores de mapeo y/o análisis terciario desvelados en el presente documento pueden procesar o analizar datos de secuencia de ácido nucleico en cualquier esquema o formato siempre que el esquema o formato pueda transmitir la identidad de la base y la posición de la secuencia de ácido nucleico.

10 En diversas realizaciones, los datos de lectura de secuenciación de ácido nucleico lectura y de la secuencia de referencia se pueden suministrar al dispositivo/servidor/nodo de computación analítica en diversos tipos/formatos de archivos de datos de entrada diferentes, que incluyen, entre otros: *.fasta, *.csfasta, *seq.txt, *qseq.txt, *.fastq, *.sff, *.prb.txt, *.sms, *srs y/o *.qv.

Usos

20 La tecnología proporciona el uso de los procedimientos de la tecnología, o las composiciones de la tecnología, para secuenciar y/o volver a secuenciar moléculas de ácido nucleico para el control de la expresión génica, perfiles de diversidad genética, diagnóstico, cribado, secuenciación del genoma completo, polimorfismo del genoma completo descubrimiento y puntuación, o cualquier otra aplicación que implique el análisis de ácidos nucleicos cuando la información de secuencia o de secuencia parcial sea relevante.

25 Kits

30 Un aspecto adicional más de la invención proporciona un kit para su uso en secuenciación, resecuenciación, monitorización de la expresión génica, perfiles de diversidad genética, diagnóstico, cribado, secuenciación del genoma completo, descubrimiento y puntuación de polimorfismos del genoma completo o cualquier otra aplicación que implique la secuenciación de ácidos nucleicos. En algunas realizaciones, los kits comprenden al menos un nucleótido marcado en un grado conocido (por ejemplo, que tiene una fracción marcada conocida) y, opcionalmente, un oligonucleótido de calibración que comprende una secuencia conocida. En algunas realizaciones, se proporciona un kit para la detección de moléculas usando reactivos de detección marcados en diferentes grados. Las diferencias de intensidad identifican las moléculas y diferencian las moléculas entre sí. Por ejemplo, algunos kits comprenden dos o más anticuerpos en los que cada anticuerpo está marcado en un grado diferente; algunos kits comprenden dos o más sondas oligonucleotídicas marcadas en las que cada sonda está marcada en un grado diferente. En algunas realizaciones, el mismo marcador (por ejemplo, un marcador o resto fluorescente) se une a los reactivos de detección y, en algunas realizaciones, se usan combinaciones de color y fracción marcada para identificar y diferenciar los reactivos de detección.

40 Además, los procesos y sistemas de secuenciación que pueden adaptarse para su uso con la tecnología se describen, por ejemplo, en las patentes de Estados Unidos n.º 7.405.281, titulada "Análogos nucleotídicos fluorescentes y usos de los mismos", expedida el 29 de Julio de 2008 para Xu y col.; 7.315.019, titulada "Matrices de confinamientos ópticos y usos de los mismos", expedida el January 1, 2008 a Turner y col.; 7.313.308, titulada "Optical análisis of molecules", expedida el 25 de diciembre de 2007 a Turner y col.; 7.302.146, titulada "Aparato y procedimiento de análisis de moléculas", expedida el 27 de noviembre de 2007 a Turner y col.; y 7.170.050, titulada "Aparato y procedimientos par el análisis óptico de moléculas", expedida el 30 de enero de 2007 a Turner y col.; y las publicaciones de patente de Estados Unidos n.º 20080212960, tituladas "Procedimientos y sistemas para control simultáneo en tiempo real de señales ópticas de múltiples fuentes", presentada el 26 de octubre de 2007 por Lundquist y col.; 20080206764, titulada "Sistema de celda de flujo para detección de moléculas sencillas", presentada el 26 de octubre de 2007 por Williams y col.; 20080199932, titulada "Polimerasas acopladas a superficie activa", presentada el 26 de octubre de 2007 por Hanzel y col.; 20080199874, titulada "ESCISIÓN CONTROLABLE DE HEBRAS DE ADN CIRCULAR MINI", presentada el 11 de febrero de 2008 por Otto y col.; 20080176769, titulada "Artículos que tienen moléculas localizadas dispuestas sobre los mismos y procedimientos de producir las mismas", presentada el 26 de octubre de 2007 por Rank y col.; 20080176316, titulada "Mitigación del fotodaño en reacciones analíticas", presentada el 31 de octubre de 2007 por Eid y col.; 20080176241, titulada "Mitigación del fotodaño en reacciones analíticas", presentado el 31 de octubre de 2007 por Eid y col.; 20080165346, titulada "Procedimientos y sistemas para el control simultáneo en tiempo real de señales ópticas de múltiples fuentes", presentado el 26 de octubre de 2007 por Lundquist y col.; 20080160531, titulada "Superficies uniformes para sustratos materiales híbridos y procedimientos para fabricar y usar las mismas", presentada el 31 de octubre de 2007 por Korlach; 20080157005, titulada "Procedimientos y sistemas para el control simultáneo en tiempo real de múltiples fuentes", presentada el 26 de octubre de 2007 por Lundquist y col.; 20080153100, titulada "Artículos que tienen moléculas localizadas dispuestas sobre los mismos y procedimientos de producir los mismos", presentada el 31 de octubre de 2007 por Rank y col.; 20080153095, titulada "NUCLEÓTIDOS DE CAMBIO DE CARGA", presentada el 26 de octubre de 2007 por Williams y col.; 20080152281, titulada "Sustratos, sistemas y procedimientos para analizar materiales", presentada el 31 de octubre de 2007 por Lundquist y col.; 20080152280, titulada "Sustratos, sistemas y

procedimientos para analizar materiales", presentada el 31 de octubre de 2007 por Lundquist y col.; 20080145278, titulada "Superficies uniformes para sustratos materiales híbridos y procedimientos para fabricarlas y usar las mismas", presentada el 31 de octubre de 2007 por Korlach; 20080128627, titulada "SUSTRATOS, SISTEMAS Y PROCEDIMIENTOS PARA ANALIZAR MATERIALES", presentada el 31 de agosto de 2007 por Lundquist y col.;
5 20080108082, titulada "Enzimas polimerasas y reactivos para secuenciación mejorada de ácidos nucleicos", presentada el 22 de octubre de 2007 por Rank y col.; 20080095488, titulada "SUSTRATOS PARA REALIZAR REACCIONES ANALÍTICAS", presentada el 11 de junio de 2007 por Foquet y col.; 20080080059, titulada "COMPONENTES ÓPTICOS MODULARES Y SISTEMAS QUE INCORPORAN LOS MISMOS", presentada el 27 de septiembre de 2007 por Dixon y col.; 20080050747, titulada "Artículos que tienen moléculas localizadas dispuestas
10 sobre los mismos y procedimientos de producir y usar los mismos", presentada el 14 de agosto de 2007 por Korlach y col.; 20080032301, titulada "Artículos que tienen moléculas localizadas dispuestas sobre los mismos y procedimientos de producir los mismos", presentada el 29 de marzo de 2007 por Rank y col.; 20080030628, titulada "Procedimientos y sistemas para el control simultáneo en tiempo real de señales ópticas de múltiples fuentes", presentada el 9 de febrero de 2007 por Lundquist y col.; 20080009007, titulada "INICIACIÓN CONTROLADA DE EXTENSIÓN DE CEBADORES", presentada el 15 de junio de 2007 por Lyle y col.; 20070238679, titulada "Artículos que tienen moléculas localizadas dispuestas sobre los mismos y procedimientos de producir los mismos", presentada el 30 de marzo de 2006 por Rank y col.; 20070231804, titulada "Procedimientos, sistemas y composiciones para controlar la actividad enzimática y aplicaciones de los mismos", presentada el 31 de marzo de 2006 por Korlach y col.; 20070206187, titulada "Procedimientos y sistemas para el control simultáneo en tiempo real de señales ópticas de múltiples fuentes", presentada el 9 de febrero de 2007 por Lundquist y col.; 20070196846, titulada "Polimerasas para incorporación de análogos nucleotídicos", presentada el 21 de diciembre 2006 por Hanzel y col.; 20070188750, titulada "Procedimientos y sistemas para el control simultáneo en tiempo real de señales ópticas de múltiples fuentes", presentada el 7 de julio de 2006 por Lundquist y col.; 20070161017, titulada "MITIGACIÓN DEL FOTODAÑO EN REACCIONES ANALÍTICAS", presentada el 1 de diciembre de 2006 por Eid y col.;
25 20070141598, titulada "Composiciones nucleotídicas y usos de los mismos", presentada el 3 de noviembre de 2006 por Turner y col.; 20070134128, titulada "Superficies uniformes para sustratos materiales híbridos y procedimientos para fabricar y usar las mismas", presentada el 27 de noviembre de 2006 por Korlach; 20070128133, titulada "Mitigación del fotodaño en reacciones analíticas", presentada el 2 de diciembre de 2005 por Eid y col.; 20070077564, titulada "Superficies reactivas, sustratos y procedimientos de producir las mismas", presentada el 30 de septiembre de 2005 por Roitman y col.; 20070072196, titulada "Análogos nucleotídicos fluorescentes y usos de los mismos", presentada el 29 de septiembre de 2005 por Xu et al; y 20070036511, titulada "Procedimientos y sistemas para controlar múltiples señales ópticas de una sola fuente ", presentada el 11 de agosto de 2005 por Lundquist y col.; y Korlach y col. "Selective aluminum passivation for targeted immobilization of single DNA polymerase molecules in zero-mode waveguide nanostructures", Proc. Natl. Acad. Sci. USA 105: 1176–81.

35 Serán evidentes para los expertos en la técnica diversas modificaciones y variaciones de las composiciones, procedimientos y usos de la tecnología descritos, sin desviarse del alcance y el espíritu de la tecnología tal como se ha descrito. Aunque la tecnología se ha descrito en relación con realizaciones específicas a modo de ejemplo, debe entenderse que la invención reivindicada no debe limitarse indebidamente a dichas realizaciones específicas. De
40 hecho, se pretende que diversas modificaciones de los modos descritos para llevar a cabo la invención que serán evidentes para los expertos en los campos relevantes estén dentro del alcance de las reivindicaciones siguientes.

45

50

55

60

65

REIVINDICACIONES

1. Un procedimiento para secuenciar un ácido nucleico diana usando un procedimiento de secuenciación por síntesis (SBS), comprendiendo el procedimiento:

5 a) proporcionar una primera pluralidad de una primera base de nucleótidos, una segunda pluralidad de una segunda base de nucleótidos, una tercera pluralidad de una tercera base de nucleótidos, y una cuarta pluralidad de una cuarta base de nucleótidos;

10 en el que una primera fracción de la primera pluralidad de la primera base de nucleótidos está marcada con un marcador, una segunda fracción de la segunda pluralidad de la segunda base de nucleótidos está marcada con dicho marcador, una tercera fracción de la tercera pluralidad de la tercera base de nucleótidos está marcada con un marcador y una cuarta fracción de la cuarta pluralidad de la cuarta base de nucleótidos está marcada con un marcador, y

15 en el que el grado de marcaje difiere para la primera pluralidad de la primera base de nucleótidos y la segunda pluralidad de la segunda base de nucleótidos de manera que dicha primera fracción es diferente de dicha segunda fracción;

20 b) incorporar por polimerización la primera pluralidad de la primera base de nucleótidos, la segunda pluralidad de la segunda base de nucleótidos, la tercera pluralidad de la tercera base de nucleótidos o la cuarta pluralidad de la cuarta base de nucleótidos en una pluralidad de copias de un ácido nucleico que es complementario al ácido nucleico diana;

25 c) detectar una amplitud de una señal producida a partir de la primera pluralidad de la primera base de nucleótidos, o una amplitud de una señal producida a partir de la segunda pluralidad de la segunda base de nucleótidos, o una amplitud de una señal producida a partir de la tercera pluralidad de la tercera base de nucleótidos, o una amplitud de una señal producida a partir de la cuarta pluralidad de la cuarta base de nucleótidos en la pluralidad de copias del ácido nucleico que es complementaria al ácido nucleico diana,

30 en el que la amplitud producida por la primera pluralidad de la primera base de nucleótidos es detectablemente diferente de una segunda amplitud producida por la segunda pluralidad de la segunda base de nucleótidos y en el que el grado de marcaje de la primera base de nucleótidos está directamente asociado con la intensidad de la señal producida para la primera base de nucleótidos y el grado de marcaje de la segunda base de nucleótidos está directamente asociado con la intensidad de la señal producida para la segunda base de nucleótidos; y

35 d) asociar la amplitud de la señal con la primera base de nucleótidos o la segunda base de nucleótidos, o la tercera base de nucleótidos, o la cuarta base de nucleótidos para identificar la base de nucleótidos en la pluralidad de copias de un ácido nucleico que es complementario al ácido nucleico diana,

40 y en el que el procedimiento de SBS es un enfoque de secuenciación basado en conjunto y la secuencia del ácido nucleico diana se determina detectando una intensidad de señal después de la incorporación de cada base y asociando las intensidades con las bases,

en el que:

45 (i) cada uno de los nucleótidos comprende un grupo hidroxilo en 3' que está bloqueado químicamente y los cuatro nucleótidos se añaden simultáneamente a una reacción que comprende ADN polimerasa y racimos de complejos molde-cebador; o

50 (ii) cada nucleótido se añade de uno en uno a una mezcla de reacción que contiene un complejo de ácido nucleico diana-cebador y una polimerasa.

2. El procedimiento de la reivindicación 1, en el que una primera fracción de la primera pluralidad de la primera base de nucleótidos está marcada con un marcador, una segunda fracción de la segunda pluralidad de la segunda base de nucleótidos está marcada con dicho marcador, una tercera fracción de la tercera pluralidad de la tercera base de nucleótidos está marcada con dicho marcador y una cuarta fracción de la cuarta pluralidad de la cuarta base de nucleótidos está marcada con dicho marcador.

3. El procedimiento de la reivindicación 2, en el que el marcador es un resto fluorescente.

60 4. El procedimiento de la reivindicación 3, en el que cada base individual está marcada con el mismo marcador fluorescente en un porcentaje conocido diferente y en el que, opcionalmente, el 25 % de las moléculas de ATP están marcadas, el 50 % de las moléculas de TTP están marcadas, el 75 % de las moléculas de GTP están marcadas y el 100 % de las moléculas de CTP están marcadas.

65 5. El procedimiento de la reivindicación 1, en el que una primera fracción de la primera pluralidad de la primera base está marcada con un primer marcador, una segunda fracción de la segunda pluralidad de la segunda base de nucleótidos está marcada con el primer marcador, una tercera fracción de la tercera pluralidad de la tercera base de

nucleótidos está marcada con un segundo marcador y una cuarta fracción de la cuarta pluralidad de la cuarta base de nucleótidos está marcada con el segundo marcador.

5 6. El procedimiento de la reivindicación 5, en el que el primer marcados es una primera marca fluorescente y en el que el segundo marcador es una marca fluorescente diferente.

7. El procedimiento de la reivindicación 2 o la reivindicación 5, en el que la primera base de nucleótidos es A, la segunda base de nucleótidos es C, la tercera base de nucleótidos es G y la cuarta base de nucleótidos es T.

10 8. El procedimiento de la reivindicación 1, en el que la señal es una onda electromagnética.

9. El procedimiento de la reivindicación 1 que comprende además detectar la señal con un dispositivo óptico.

15 10. El procedimiento de la reivindicación 1 que comprende además proporcionar un oligonucleótido de calibración que comprende una secuencia conocida.

20 11. El procedimiento de la reivindicación 10, en el que la secuencia de calibración contiene cada una de las cuatro bases en un orden conocido, y en el que el procedimiento comprende además opcionalmente el uso de la secuencia de calibración para calibrar un instrumento de secuenciación para reconocer las intensidades de señal apropiadas para cada una de las bases.

12. El procedimiento de la reivindicación 10 u 11, en el que:

- 25 (i) la secuencia de calibración está unida al comienzo de cada ácido nucleico diana que se va a secuenciar; o
(ii) la secuencia de calibración no está unida a la secuencia diana, sino que se usa para calibrar el instrumento de secuenciación antes de adquirir la secuencia del ácido nucleico diana.

30 13. El procedimiento de cualquiera de las reivindicaciones precedentes, en el que el ácido nucleico diana es un ARNm, ARN ribosómico, ADN genómico o ADNc.

35

40

45

50

55

60

65