

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 910 099**

51 Int. Cl.:

C12Q 1/6869 (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **19.11.2015** **E 20151469 (2)**

97 Fecha y número de publicación de la concesión europea: **05.01.2022** **EP 3696280**

54 Título: **Secuenciación sin enzimas y sin amplificación**

30 Prioridad:

21.11.2014 US 201462082883 P

45 Fecha de publicación y mención en BOPI de la
traducción de la patente:
11.05.2022

73 Titular/es:

NANOSTRING TECHNOLOGIES, INC (100.0%)
530 Fairview Ave N, Ste 2000
Seattle, WA 98109, US

72 Inventor/es:

BEECHEM, JOSEPH M. y
KHAFIZOV, RUSTEM

74 Agente/Representante:

LEHMANN NOVO, María Isabel

ES 2 910 099 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 sin enzimas y sin amplificación

Remisión a solicitudes relacionadas

5 Esta solicitud reivindica el beneficio de la solicitud de provisional de Estados Unidos n.º 62/082.883, presentada el 21 de noviembre de 2014.

Lista de secuencias

La presente solicitud contiene una lista de secuencias que se ha presentado en formato ASCII mediante EFS-Web y se incorpora por la presente por referencia en su totalidad. Dicha copia ASCII, creada el 19 de noviembre de 2015, se denomina NATE-025_ST25.txt y es de 20 860 *bytes* de tamaño.

10 Antecedentes de la invención

Actualmente hay una diversidad de métodos para secuenciación de ácidos nucleicos, es decir, el proceso de determinar el orden preciso de nucleótidos dentro de una molécula de ácido nucleico. Los métodos actuales requieren amplificar un ácido nucleico enzimáticamente, por ejemplo, PCR, y/o por clonación. Se requieren polimerizaciones enzimáticas adicionales para producir una señal detectable por un medio de detección con luz. Dichas etapas de amplificación y polimerización son costosas y/o requieren mucho tiempo. La publicación internacional WO2013055995 se refiere a secuenciación mediante ensamblaje de estructuras. Por tanto, hay una necesidad en la técnica de un método de secuenciación de ácidos nucleicos que sea sin amplificación y sin enzimas. La presente invención aborda estas necesidades.

Compendio de la invención

20 La presente invención se define en las reivindicaciones adjuntas. La presente divulgación se refiere a sondas de secuenciación, métodos, kits y aparatos que proporcionan secuenciación de ácidos nucleicos sin enzimas, sin amplificación y sin colecciones que tiene longitudes de lectura largas y con baja tasa de error. Además, los métodos, kits y aparatos tienen una capacidad de muestreo a respuesta rápida. Estos rasgos característicos son particularmente útiles para secuenciación en un entorno clínico.

25 En la presente memoria se divulgan sondas de secuenciación que comprenden un dominio de unión a la diana y un dominio de código de barras. El dominio de unión a la diana y el dominio de código de barras pueden estar unidos de forma funcional, por ejemplo, unidos covalentemente. Una sonda de secuenciación opcionalmente comprende un espaciador entre el dominio de unión a la diana y el dominio de código de barras. El espaciador puede ser cualquier polímero con propiedades mecánicas apropiadas, por ejemplo, un espaciador de ADN mono- o bicatenario (de 1 a 30 100 nucleótidos, por ejemplo, de 2 a 50 nucleótidos). Ejemplos no limitantes de espaciadores de ADN bicatenarios incluyen las secuencias cubiertas por SEQ ID NO: 25 a SEQ ID NO: 29.

El dominio de unión a la diana comprende al menos cuatro nucleótidos (por ejemplo, 4, 5, 6, 7, 8, 9, 10, 11, 12 o más) y puede unirse a un ácido nucleico diana (por ejemplo, ADN, ARN y APN). El dominio de código de barras comprende una cadena principal sintética, teniendo el dominio de código de barras al menos una primera posición que comprende una o más regiones de fijación. El dominio de código de barras puede tener una, dos, tres, cuatro, cinco, seis, siete, ocho, nueve, diez, once, doce o más posiciones; teniendo cada posición una o más (por ejemplo, de una a cincuenta) regiones de fijación; cada región de fijación comprende al menos una (es decir, de una a cincuenta, por ejemplo, de diez a treinta copias de una o más secuencias de ácido nucleico) que puede unirse de forma reversible a una molécula de ácido nucleico complementaria (ARN o ADN). Determinadas posiciones en un dominio de código de barras pueden tener más regiones de fijación que otras posiciones; como alternativa, cada posición en un dominio de código de barras tiene el mismo número de regiones de fijación. La secuencia de ácido nucleico de una primera región de fijación determina la posición e identidad de un primer nucleótido en el ácido nucleico diana que se une mediante un primer nucleótido del dominio de unión a la diana, mientras que la secuencia de ácido nucleico de una segunda región de fijación determina la posición e identidad de un segundo nucleótido en el ácido nucleico diana que se une mediante un segundo nucleótido del dominio de unión a la diana. Asimismo, la secuencia de ácido nucleico de una sexta región de fijación determina la posición e identidad de un sexto nucleótido en el ácido nucleico diana que se une mediante un sexto nucleótido del dominio de unión a la diana. En realizaciones, la cadena principal sintética comprende un polisacárido, un polinucleótido (por ejemplo, ADN o ARN mono- o bicatenario), un péptido, un ácido peptidonucleico o un polipéptido. El número de nucleótidos en un dominio de unión a la diana es igual a o es mayor de (por ejemplo, 1, 2, 3, 4 o más) el número de posiciones en el dominio de código de barras. Cada región de fijación en una posición específica del dominio de código de barras puede incluir una copia de la misma secuencia de ácido nucleico y/o múltiples copias de la misma secuencia de ácido nucleico. Sin embargo, una región de fijación incluirá una secuencia de ácido nucleico diferente de una región de fijación en una posición diferente del dominio de código de barras, incluso cuando ambas regiones de fijación identifican el mismo tipo de nucleótido, por ejemplo, adenina, timina, citosina, guanina, uracilo y análogos de los mismos. Una región de fijación puede ligarse a un monómero modificado, por ejemplo, un nucleótido modificado, en la cadena principal sintética, creando de ese modo una ramificación con respecto a la cadena principal. Una región de fijación puede formar parte de una secuencia polinucleotídica de la

cadena principal. Una o más regiones de fijación pueden estar adyacentes a al menos un polinucleótido monocatenario flanqueante, es decir, una región de fijación puede unirse de forma funcional a un polinucleótido monocatenario flanqueante 5' y/o a un polinucleótido monocatenario flanqueante 3'. Una región de fijación con o sin uno o dos polinucleótidos monocatenarios flanqueantes puede hibridarse con una molécula de ácido nucleico de hibridación que carece de un marcador detectable. Una molécula de ácido nucleico de hibridación que carece de un marcador detectable puede ser entre aproximadamente 4 y aproximadamente 20 nucleótidos de longitud, por ejemplo, de 12 nucleótidos o más larga.

Una región de fijación puede unirse mediante un ácido nucleico complementario que comprende un marcador detectable. Cada ácido nucleico complementario puede comprender un marcador detectable.

Como alternativa, una región de fijación puede unirse mediante un ácido nucleico complementario que forma parte de un complejo indicador (que comprende marcadores detectables). Un ácido nucleico complementario (que comprende un marcador detectable o de un complejo indicador) puede ser entre aproximadamente 4 y aproximadamente 20 nucleótidos de longitud, por ejemplo, aproximadamente 8, 10, 12 y 14 nucleótidos o más. En un complejo indicador, un ácido nucleico complementario se liga (directamente o indirectamente) a una molécula de ácido nucleico primaria. Un ácido nucleico complementario puede ligarse indirectamente a una molécula de ácido nucleico primaria mediante un conector de ácido nucleico mono- o bicatenario (por ejemplo, un polinucleótido que comprende de 1 a 100 nucleótidos). Un ácido nucleico primario se hibrida con uno o más (por ejemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más) ácidos nucleicos secundarios. Cada ácido nucleico secundario se hibrida con uno o más (por ejemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más) ácidos nucleicos terciarios; los ácidos nucleicos terciarios comprenden uno o más marcadores detectables. Uno o cada ácido nucleico secundario puede comprender una región que no hibrida con una molécula de ácido nucleico primaria y no hibrida con una molécula de ácido nucleico terciaria (un "mango adicional"); esta región puede ser de cuatro o más (por ejemplo, de aproximadamente 6 a aproximadamente 40, por ejemplo, aproximadamente 8, 10, 12 y 14) nucleótidos de longitud. La región que no hibrida con una molécula de ácido nucleico primaria y no hibrida con una molécula de ácido nucleico terciaria puede comprender la secuencia de nucleótidos de la molécula de ácido nucleico complementaria que se liga a la molécula de ácido nucleico primaria. Esta región puede estar localizada cerca del extremo del ácido nucleico secundario distal a su extremo que hibrida con el ácido nucleico primario. Al tener "mangos adicionales" que comprenden la secuencia de nucleótidos del ácido nucleico complementario, se aumenta enormemente la probabilidad y velocidad a la que un complejo indicador se une a una sonda de secuenciación. En cualquier realización o aspecto de la presente divulgación, cuando un complejo indicador comprende "mangos adicionales", el complejo indicador puede hibridar con una sonda de secuenciación mediante el ácido nucleico complementario del complejo indicador o mediante el "mango adicional". Por tanto, por ejemplo, la expresión "unir a la primera región de fijación... una primera molécula de ácido nucleico complementaria de un primer complejo indicador" se entendería de acuerdo con su significado sencillo y también se entendería que significa "unir a la primera región de fijación... un "mango adicional" de un primer complejo indicador".

En realizaciones, las expresiones "dominio de código de barras" y "cadena principal sintética" son sinónimas.

En la presente memoria se divulga un método para secuenciar un ácido nucleico usando una sonda de secuenciación de la presente invención. El método comprende las etapas de: (1) hibridar al menos una sonda de secuenciación, de la presente invención, con un ácido nucleico diana que se inmoviliza (por ejemplo, en una, dos, tres, cuatro, cinco, seis, siete, ocho, nueve, diez o más posiciones) en un sustrato; (2) unir a la primera región de fijación una primera molécula de ácido nucleico complementaria (ARN o ADN) que tiene un marcador detectable (por ejemplo, un marcador fluorescente) o una primera molécula de ácido nucleico complementaria de un primer complejo indicador que comprende marcadores detectables (por ejemplo, marcadores fluorescentes); (3) detectar el uno o más marcadores detectables, y (4) identificar la posición e identidad del primer nucleótido en el ácido nucleico diana inmovilizado. Opcionalmente, el ácido nucleico diana inmovilizado se elonga antes de unirse por la sonda. El método comprende además las etapas de: (5) poner en contacto la primera región de fijación (con o sin uno o dos polinucleótidos monocatenarios flanqueantes) con una primera molécula de ácido nucleico de hibridación que carece de un marcador detectable, soltando de ese modo la primera molécula de ácido nucleico complementaria que tiene un marcador detectable o la primera molécula de ácido nucleico complementaria de un primer complejo indicador que comprende marcadores detectables y unir a, al menos, la primera región de fijación un primer ácido nucleico de hibridación que carece de un marcador detectable; (6) unir a la segunda región de fijación una segunda molécula de ácido nucleico complementaria que tiene un marcador detectable o una molécula de ácido nucleico complementaria de un segundo complejo indicador que comprende marcadores detectables; (7) detectar el uno o más marcadores detectables; y (8) identificar la posición e identidad del segundo nucleótido en el ácido nucleico diana inmovilizado. Las etapas (5) a (8) se repiten hasta que cada nucleótido en el ácido nucleico diana inmovilizado y correspondiente al dominio de unión a la diana se ha identificado. Las etapas (5) y (6) pueden producirse simultáneamente o secuencialmente. Cada (por ejemplo, primera, segunda, tercera, cuarta, quinta, sexta, séptima, octava, novena, décima o superior) molécula de ácido nucleico complementaria (que tiene un marcador detectable o parte de un complejo indicador) tiene la misma secuencia de ácido nucleico que su correspondiente (es decir, primera, segunda, tercera, cuarta, quinta, sexta, séptima, octava, novena, décima o superior) molécula de ácido nucleico de hibridación que carece de un marcador detectable. El ácido nucleico diana se inmoviliza en un sustrato por unión de una primera posición y/o segunda posición del ácido nucleico diana con una primera y/o una segunda sonda de captura; cada sonda de captura comprende una marca de afinidad que se une selectivamente a un sustrato. La primera y/o segunda posición pueden estar en o cerca de un extremo de un ácido nucleico diana. El sustrato puede ser cualquier soporte sólido conocido en la técnica, por

ejemplo, un portaobjetos recubierto y dispositivo de microfluidos (por ejemplo, recubierto con estreptavidina). Otras posiciones que están localizadas distantes de un extremo de un ácido nucleico diana pueden unirse selectivamente al sustrato. El ácido nucleico puede elongarse aplicando una fuerza (por ejemplo, gravedad, fuerza hidrodinámica, fuerza electromagnética, estiramiento fluido, una técnica de menisco menguante y combinaciones de los mismos) suficiente para prolongar el ácido nucleico diana.

En la presente memoria se divulga un método para secuenciar un ácido nucleico usando una población de sondas de la presente invención o una pluralidad de poblaciones de sondas de la presente invención. El método comprende las etapas de: (1) hibridar una primera población de sondas de secuenciación (de la presente invención) con un ácido nucleico diana que se inmoviliza en un sustrato (deshibridando, cada sonda de secuenciación en la primera población, del ácido nucleico diana inmovilizado en aproximadamente las mismas condiciones, por ejemplo, nivel de agente caótopo, temperatura, concentración salina, pH y fuerza hidrodinámica); (2) unir una pluralidad de primeras moléculas de ácido nucleico complementarias que tienen cada una un marcador detectable o una pluralidad de primeras moléculas de ácido nucleico complementarias de una pluralidad de primeros complejos indicadores, comprendiendo cada complejo marcadores detectables, a una primera región de fijación en cada sonda de secuenciación en la primera población; (3) detectar el uno o más marcadores detectables; (4) identificar la posición e identidad de una pluralidad de primeros nucleótidos en el ácido nucleico diana inmovilizado hibridado mediante sondas de secuenciación en la primera población; (5) poner en contacto cada primera región de fijación de cada sonda de secuenciación de la primera población con una pluralidad de primeras moléculas de ácido nucleico de hibridación que carecen de un marcador detectable, soltando de ese modo las primeras moléculas de ácido nucleico complementarias que tienen un marcador detectable o de un complejo indicador y unir a cada primera región de fijación una primera molécula de ácido nucleico de hibridación que carece de un marcador detectable; (6) unir una pluralidad de segundas moléculas de ácido nucleico complementarias que tienen cada una un marcador detectable o una pluralidad de segundas moléculas de ácido nucleico complementarias de una pluralidad de segundos complejos indicadores, comprendiendo cada complejo marcadores detectables, a una segunda región de fijación en cada sonda de secuenciación en la primera población; (7) detectar el uno o más marcadores detectables; y (8) identificar la posición e identidad de una pluralidad de segundos nucleótidos en el ácido nucleico diana inmovilizado hibridado mediante sondas de secuenciación en la primera población. En la etapa (9), las etapas (5) a (8) se repiten hasta que cada nucleótido en el ácido nucleico diana inmovilizado y correspondiente al dominio de unión a la diana de cada sonda de secuenciación en la primera población se ha identificado. Las etapas (5) y (6) pueden producirse simultáneamente o secuencialmente. De ese modo, el orden lineal de nucleótidos se identifica para regiones del ácido nucleico diana inmovilizado que hibridaron mediante el dominio de unión a la diana de sondas de secuenciación en la primera población de sondas de secuenciación.

En realizaciones, cuando se usa una pluralidad de poblaciones (es decir, más de una población) de sondas, el método comprende además las etapas de: (10) deshibridar cada sonda de secuenciación de la primera población del ácido nucleico; (11) retirar cada sonda de secuenciación deshibridada de la primera población; (12) hibridar al menos una segunda población de sondas de secuenciación de la presente invención, donde cada sonda de secuenciación en la segunda población se deshibrida del ácido nucleico diana inmovilizado en aproximadamente las mismas condiciones y se deshibrida del ácido nucleico diana inmovilizado en diferentes condiciones de las sondas de secuenciación en la primera población; (13) unir una pluralidad de primeras moléculas de ácido nucleico complementarias que tienen cada una un marcador detectable o una pluralidad de primeras moléculas de ácido nucleico complementarias de una pluralidad de primeros complejos indicadores, comprendiendo cada complejo marcadores detectables, a una primera región de fijación en cada sonda de secuenciación en la segunda población; (14) detectar el uno o más marcadores detectables; (15) identificar la posición e identidad de una pluralidad de primeros nucleótidos en el ácido nucleico diana inmovilizado hibridado mediante sondas de secuenciación en la segunda población; (16) poner en contacto cada primera región de fijación de cada sonda de secuenciación de la segunda población con una pluralidad de primeras moléculas de ácido nucleico de hibridación que carecen de un marcador detectable, soltando de ese modo las primeras moléculas de ácido nucleico complementarias (que tienen un marcador detectable o de un complejo indicador) y unir a cada primera región de fijación una primera molécula de ácido nucleico de hibridación que carece de marcador detectable; (17) unir una pluralidad de segundas moléculas de ácido nucleico complementarias que tienen cada una un marcador detectable o una pluralidad de segundas moléculas de ácido nucleico complementarias de una pluralidad de segundos complejos indicadores, comprendiendo cada complejo marcadores detectables, a una segunda región de fijación en cada sonda de secuenciación en la segunda población; (18) detectar el uno o más marcadores detectables; (19) identificar la posición e identidad de una pluralidad de segundos nucleótidos en el ácido nucleico diana inmovilizado hibridado mediante sondas de secuenciación en la segunda población; y (20) repetir las etapas (16) a (19) hasta que el orden lineal de nucleótidos se ha identificado para regiones del ácido nucleico diana inmovilizado que se hibridaron mediante el dominio de unión a la diana de sondas de secuenciación en la segunda población de sondas de secuenciación. Las etapas (16) y (17) pueden producirse simultáneamente o secuencialmente.

Cada sonda de secuenciación en la segunda población puede deshibridarse del ácido nucleico diana inmovilizado en una condición diferente (por ejemplo, una temperatura mayor, nivel mayor de agente caótopo, mayor concentración salina, mayor caudal y diferente pH) que la condición promedio para la que las sondas de secuenciación en la primera población se deshibridarán del ácido nucleico diana.

Sin embargo, cuando se usan más de dos poblaciones de sondas, entonces las sondas en dos poblaciones secuenciales pueden deshibridarse en diferentes condiciones y las sondas en poblaciones no secuenciales pueden deshibridarse en condiciones similares. Como un ejemplo, las sondas en una primera población y tercera población

pueden deshibridarse en condiciones similares. En realizaciones, poblaciones secuenciales de sondas deshibridadas en condiciones cada vez más rigurosas (por ejemplo, mayores niveles de agente caótopo, concentración salina y temperatura). Para un dispositivo de microfluidos, usando la temperatura como ejemplo, una primera población de sondas puede permanecer hibridada a una primera temperatura, pero se deshibridará a una segunda temperatura, que es mayor que la primera. Una segunda población de sondas puede permanecer hibridada a la segunda temperatura, pero se deshibridará a una tercera temperatura, que es mayor que la segunda. En este ejemplo, las soluciones (que comprenden reactivos requeridos por el presente método) que fluyen sobre un ácido nucleico diana para poblaciones de sondas iniciales están a una temperatura menor que las soluciones que fluyen sobre el ácido nucleico diana para posteriores poblaciones de sondas.

En algunas realizaciones, después de haber usado una población de sondas, la población de sondas se deshibrida del ácido nucleico diana y se usa una nueva alícuota de la misma población de sondas. Por ejemplo, después de haber hibridado, detectado y deshibridado una primera población de sondas, se hibrida una alícuota posterior de la primera población de sondas. Como alternativa, como ejemplo, una primera población de sondas puede deshibridarse y remplazarse con una segunda población de sondas; una vez se ha detectado y deshibridado la segunda población, se hibrida una posterior alícuota de la primera población de sondas con el ácido nucleico diana. Por tanto, una sonda en la posterior población puede hibridar con una región del ácido nucleico diana que se había secuenciado previamente (obteniendo de ese modo información de secuencia redundante y/o de confirmación) o una sonda en la posterior población puede hibridar con una región del ácido nucleico diana que no se había secuenciado previamente (obteniendo de ese modo nueva información de secuencia). Por consiguiente, puede volver a dividirse en alícuotas una población de sondas cuando una lectura previa fue insatisfactoria (por cualquier razón) y/o para mejorar la precisión de la alineación resultante de las lecturas de secuenciación.

Las sondas que hibridan y se deshibridan en condiciones similares pueden tener longitudes similares de su dominio de unión a la diana, contenido de GC o frecuencia de bases repetidas y combinaciones de los mismos. Las relaciones entre T_m y longitud de un oligonucleótido se muestran, por ejemplo, en Sugimoto *et al.*, *Biochemistry*, 34, 11211-6.

Cuando se usan más de dos poblaciones de sondas, las etapas, como se describe para la primera y segunda población de sondas de secuenciación, se repiten con poblaciones adicionales de sondas (por ejemplo, de 10 a 100 a 1000 poblaciones). El número de poblaciones de sondas usadas dependerá de una diversidad de factores, incluyendo, aunque sin limitación, el tamaño del ácido nucleico diana, el número de sondas peculiares en cada población, el grado de solapamiento entre sondas de secuenciación deseadas, y el enriquecimiento de sondas para regiones de interés.

Una población de sondas puede contener sondas de secuenciación adicionales dirigidas a una región específica de interés en un ácido nucleico diana, por ejemplo, una región que contiene una mutación (por ejemplo, una mutación puntual) o un alelo SNP. Una población de sondas puede contener menos sondas de secuenciación dirigidas a una región específica de menos interés en un ácido nucleico diana.

Una población de sondas de secuenciación puede compartimentarse en combinaciones más pequeñas diferenciadas de sondas de secuenciación. La compartimentación puede basarse en la temperatura de fusión prevista del dominio de unión a la diana en las sondas de secuenciación y/o en el motivo de secuencia del dominio de unión a la diana en las sondas de secuenciación. La compartimentación puede basarse en normas obtenidas empíricamente. Las diferentes combinaciones de sondas de secuenciación pueden hacerse reaccionar con el ácido nucleico diana usando diferentes condiciones de reacción, por ejemplo, basadas en la temperatura, la concentración salina y/o el contenido de tampón. La compartimentación puede realizarse para cubrir el ácido nucleico diana con cobertura uniforme. La compartimentación puede realizarse para cubrir el ácido nucleico diana con un perfil de cobertura conocido.

Las longitudes de los dominios de unión a la diana en una población de sondas de secuenciación pueden reducirse para aumentar la cobertura de las sondas en una región específica de un ácido nucleico diana. Las longitudes de los dominios de unión a la diana en una población de sondas de secuenciación pueden aumentarse para disminuir la cobertura de las sondas en una región específica de un ácido nucleico diana, por ejemplo, hasta por encima del límite de resolución del aparato de secuenciación.

Como alternativa o adicionalmente, la concentración de las sondas de secuenciación en una población puede aumentarse para aumentar la cobertura de las sondas en una región específica de un ácido nucleico diana. La concentración de las sondas de secuenciación puede reducirse para disminuir la cobertura de las sondas en una región específica de un ácido nucleico diana, por ejemplo, hasta por encima del límite de resolución del aparato de secuenciación.

Los métodos para secuenciar un ácido nucleico comprenden además las etapas de ensamblar cada orden lineal identificado de nucleótidos para cada región del ácido nucleico diana inmovilizado, identificando de ese modo una secuencia para el ácido nucleico diana inmovilizado. Las etapas de ensamblaje usan un medio de almacenamiento legible por ordenador no transitorio con un programa ejecutable almacenado en el mismo que envía una instrucción al microprocesador para colocar cada orden lineal identificado de nucleótidos, obteniendo de ese modo la secuencia del ácido nucleico. El ensamblaje puede producirse "de manera instantánea", es decir, mientras se están recogiendo los datos de las sondas de secuenciación en lugar de después de haber recogido todos los datos.

El ácido nucleico diana, es decir, el que se secuenciar, puede ser entre aproximadamente 4 y 1 000 000 nucleótidos de longitud. La diana puede incluir un cromosoma intacto completo o un fragmento del mismo, que es mayor de 1 000 000 nucleótidos de longitud.

En la presente memoria se divulgan aparatos para realizar un método de la presente invención.

5 En la presente memoria se divulgan kits que incluyen sondas de secuenciación de la presente invención y para realizar métodos de la presente invención. En realizaciones, los kits incluyen un sustrato que puede inmovilizar un ácido nucleico mediante una sonda de captura, una pluralidad de sondas de secuenciación de la presente invención, al menos una sonda de captura, al menos una molécula de ácido nucleico complementaria que tiene un marcador detectable, al menos una molécula de ácido nucleico complementaria que carece de un marcador detectable, e instrucciones para su uso. En realizaciones, el kit comprende aproximadamente o al menos 4096 sondas de secuenciación peculiares. 4096 es el número mínimo de sondas peculiares necesario para incluir cada posible combinación hexamérica (es decir, para sondas que tienen cada una seis regiones de fijación en el dominio de código de barras). Aquí, se consigue "4096" ya que hay cuatro opciones de nucleótidos para seis posiciones: 4⁶. Para un conjunto de sondas que tengan cuatro regiones de fijación en los dominios de código de barras, solamente se necesitarán 256 (es decir, 4⁴) sondas peculiares. Para un conjunto de sondas que tengan ocho nucleótidos en sus dominios de unión a la diana, se necesitarán 4⁸ (es decir, 65 536) sondas peculiares. Para un conjunto de sondas que tengan diez nucleótidos en sus dominios de unión a la diana, se necesitarán 4¹⁰ (es decir, 1 048 576) sondas peculiares.

20 En realizaciones, el kit comprende aproximadamente o al menos veinticuatro moléculas de ácido nucleico complementarias distintivas que tienen un marcador detectable y aproximadamente o al menos veinticuatro moléculas de ácido nucleico de hibridación distintivas que carecen de un marcador detectable. Un ácido nucleico complementario puede unirse a una región de fijación que tenga una secuencia de una de SEQ ID NO: 1 a 24, como ejemplos no limitantes. Se enumeran secuencias ejemplares adicionales que pueden incluirse en un dominio de código de barras en SEQ ID NO: 42 a SEQ ID NO: 81. De hecho, la secuencia de nucleótidos no está limitada; preferiblemente carece de homología sustancial (por ejemplo, de un 50 % a un 99,9 %) con una secuencia de nucleótidos conocida; que ayuda a evitar hibridación indeseable de un ácido nucleico complementario y un ácido nucleico diana.

Cualquiera de los aspectos y realizaciones anteriores puede combinarse con cualquier otro aspecto o realización.

Salvo que se definan de otro modo, todos los términos técnicos y científicos usados en la presente memoria tienen el mismo significado que el habitualmente comprendido por un experto en la materia a la que pertenece esta invención. 30 En la memoria descriptiva, las formas singulares también incluyen el plural, salvo que el contexto indique claramente lo contrario; como ejemplos, se entiende que los términos "un/o", "una" y "el/la" son singular o plural y se entiende que el término "o" es inclusivo. A modo de ejemplo, "un elemento" significa uno o más elementos. A lo largo de la memoria descriptiva se entenderá que la expresión "que comprende", o variaciones tales como "comprende" o "comprendiendo", implica la inclusión de un elemento, número entero o etapa, o grupo de elementos, números enteros o etapas indicado, pero no la exclusión de cualquier otro elemento, número entero o etapa, o grupo de elementos, números enteros o etapas. "Aproximadamente" puede entenderse como en un 10 %, 9 %, 8 %, 7 %, 6 %, 5 %, 4 %, 3 %, 2 %, 1 %, 0,5 %, 0,1 %, 0,05 % o 0,01 % del valor indicado. Salvo que esté claro de otro modo por el contexto, todos los valores numéricos proporcionados en la presente memoria están modificados por el término "aproximadamente".

40 Las referencias citadas en la presente memoria no se admiten como técnica anterior a la invención reivindicada. En caso de conflicto, prevalecerá la presente memoria descriptiva, incluyendo las definiciones. Además, los materiales, métodos y ejemplos son ilustrativos únicamente y no están destinados a ser limitantes.

Breve descripción de los dibujos

45 El archivo de patente o solicitud contiene al menos un dibujo ejecutado a color. Se proporcionarán copias de esta publicación de patente o solicitud de patente con dibujos a color por la Oficina tras petición y pago de la tasa necesaria.

Los rasgos característicos anteriores y adicionales se apreciarán más claramente a partir de la siguiente descripción detallada tomada conjuntamente con los dibujos adjuntos.

La figura 1 a figura 5 muestran esquemas de sondas de secuenciación ejemplares de la presente invención.

50 La figura 6A a figura 6D son esquemas que muestran variantes de una sonda de secuenciación de la presente invención.

La figura 7 muestra esquemas de dominios de unión a la diana de sondas de secuenciación de la presente invención; los dominios incluyen cero, dos o cuatro nucleótidos que tienen bases universales.

La figura 8A a figura 8E ilustran etapas de un método de secuenciación de la presente invención.

La figura 9A muestra una etapa inicial de un método de secuenciación de la presente invención.

La figura 9B muestra un esquema de un complejo indicador que comprende marcadores detectables.

La figura 9C muestra una pluralidad de complejos indicadores que comprenden cada uno marcadores detectables.

Las figuras 9D a 9G muestran etapas adicionales del método de secuenciación iniciado en la figura 9A.

5 La figura 10 muestra una ilustración alternativa de las etapas mostradas en la figura 9D y figura 9E y datos ejemplares obtenidos a partir de las mismas. El fragmento de la sonda de secuenciación mostrada tiene la secuencia de SEQ ID NO: 82.

La figura 11 ilustra una variación del método mostrado en la figura 10. El fragmento de la sonda de secuenciación mostrada, asimismo, tiene la secuencia de SEQ ID NO: 82.

La figura 12 ilustra un método de la presente invención.

10 La figura 13 compara las etapas requeridas en un método de secuenciación de la presente invención con etapas requeridas con otros métodos de secuenciación.

La figura 14 y figura 15 ejemplifican las mediciones de funcionamiento obtenibles por la presente invención.

La figura 16 compara la tasa de secuenciación, el número de lecturas y la utilidad clínica para la presente invención y otros diversos métodos/aparatos de secuenciación.

15 La figura 17 demuestra la baja tasa de error bruto de métodos de secuenciación de la presente invención. La secuencia de molde mostrada tiene la secuencia de SEQ ID NO: 83.

La figura 18 compara los datos de secuenciación obtenibles de la presente invención con otros métodos de secuenciación.

20 La figura 19 demuestra la especificidad de una sola base de los métodos de secuenciación de la presente invención. Las secuencias de molde y sonda mostradas (desde la parte superior a la parte inferior) tienen las secuencias de SEQ ID NO: 84 a SEQ ID NO: 88.

La figura 20A muestra diversos diseños de complejos indicadores de la presente invención.

La figura 20B muestra recuentos fluorescentes obtenidos de los complejos indicadores mostrados en la figura 20A.

La figura 20C muestra recetas ejemplares para construir complejos indicadores de la presente invención.

25 La figura 21A muestra diseños de complejos indicadores que comprenden "mangos adicionales".

La figura 21B muestra recuentos fluorescentes obtenidos de los complejos indicadores que tienen "mangos adicionales".

La figura 22A y figura 22B muestran la cinética de hibridación de dos diseños ejemplares de complejos indicadores de la presente invención.

30 La figura 23 muestra un esquema de una sonda de secuenciación de la presente invención usada en un método distinto del mostrado en la figura 8 a figura 12.

La figura 24 muestra un esquema de una tarjeta de secuenciación consumible útil en la presente invención.

35 La figura 25 muestra la detección de emparejamientos incorrectos de un oligómero decamérico, como se describe en el ejemplo 3. Los nucleótidos mostrados (desde la parte superior a la parte inferior) tienen las secuencias de SEQ ID NO: 89 a SEQ ID NO: 99.

La figura 26 muestra la capacidad de hibridación dependiendo del tamaño de un dominio de unión a la diana, como se describe en el ejemplo 3. El fondo es elevado debido a una elevada concentración de indicador y no hubo purificación previa. Los nucleótidos mostrados (desde la parte superior a la parte inferior) tienen las secuencias de SEQ ID NO: 100 a SEQ ID NO: 104.

40 La figura 27 muestra una comparación entre un punto individual frente a un indicador de longitud completa. Los resultados para puntos individuales muestran que la velocidad de hibridación es 1000 veces mayor que para un código de barras de longitud completa (condiciones de diana 100 nM, 30 minutos de hibridación).

Descripción detallada de la invención

45 La presente divulgación se refiere a sondas de secuenciación, métodos, kits y aparatos que proporcionan secuenciación de ácidos nucleicos sin enzimas, sin amplificación y sin colecciones que tiene longitudes de lectura largas y con baja tasa de error.

Sonda de secuenciación

La presente invención se refiere a una sonda de secuenciación que comprende un dominio de unión a la diana y un dominio de código de barras. Ejemplos no limitantes de sondas de secuenciación de la presente divulgación se muestran en las figuras 1 a 6.

5 La figura 1 muestra un esquema de una sonda de secuenciación de la presente divulgación. Esta sonda de secuenciación ejemplar tiene un dominio de unión a la diana de seis nucleótidos, de los que cada uno de ellos corresponde a una posición en el dominio de código de barras (que comprende una o más regiones de fijación). Se indica una primera región de fijación; corresponde al nucleótido de un ácido nucleico diana unido mediante un primer nucleótido en el dominio de unión a la diana. Se indica la tercera posición en el dominio de código de barras. Se indica
10 una quinta posición que comprende dos regiones de fijación. Cada posición en un dominio de código de barras puede tener múltiples regiones de fijación. Por ejemplo, una posición puede tener de 1 a 50 regiones de fijación. Determinadas posiciones en un dominio de código de barras pueden tener más regiones de fijación que otras posiciones (como se muestra aquí en la posición 5 con respecto a las posiciones 1 a 4 y 6); como alternativa, cada posición en un dominio de código de barras tiene el mismo número de regiones de fijación (véanse, por ejemplo, las figuras 2, 3, 5 y 6).
15 Aunque no se muestra, cada región de fijación comprende al menos una (es decir, de una a cincuenta, por ejemplo, de diez a treinta) copias de una o más secuencias de ácido nucleico que pueden unirse de forma reversible a una molécula de ácido nucleico complementaria (ARN o ADN). En la figura 1, las regiones de fijación están integradas en la molécula polinucleotídica lineal que compone el dominio de código de barras.

La figura 2 muestra un esquema de una sonda de secuenciación de la presente divulgación. Esta sonda de
20 secuenciación ejemplar tiene un dominio de unión a la diana de seis nucleótidos, de los que cada uno de ellos corresponde a una región de fijación en el dominio de código de barras. Se indica una primera región de fijación; corresponde al nucleótido de un ácido nucleico diana unido mediante un primer nucleótido en el dominio de unión a la diana. La cuarta posición en el dominio de código de barras, que comprende una parte del dominio de código de barras y dos cuartas regiones de fijación, está rodeada por un círculo. Se indican dos sextas regiones de fijación. Aquí, cada
25 posición tiene dos regiones de fijación; sin embargo, cada posición en un dominio de código de barras puede tener una región de fijación o múltiples regiones de fijación, por ejemplo, de 2 a 50 regiones de fijación. Aunque no se muestra, cada región de fijación comprende al menos una (es decir, de una a cincuenta, por ejemplo, de diez a treinta) copias de una o más secuencias de ácido nucleico que pueden unirse de forma reversible a una molécula de ácido nucleico complementaria (ARN o ADN). En la figura 2, el dominio de código de barras es una molécula polinucleotídica lineal a la que se ligán las regiones de fijación; las regiones de fijación no están integradas en la molécula polinucleotídica.
30

La figura 3 muestra otro esquema de una sonda de secuenciación de la presente divulgación. Esta sonda de
35 secuenciación ejemplar tiene un dominio de unión a la diana de cuatro nucleótidos, correspondiendo estos cuatro nucleótidos a cuatro posiciones en el dominio de código de barras. Cada posición se muestra con tres regiones de fijación ligadas.

La figura 4 muestra otro esquema más de una sonda de secuenciación de la presente divulgación. Esta sonda de
secuenciación ejemplar tiene un dominio de unión a la diana de diez nucleótidos. Sin embargo, solamente los seis primeros nucleótidos corresponden a seis posiciones en el dominio de código de barras. Del séptimo al décimo nucleótido (indicados por "n₁ a n₄") se añaden para aumentar la longitud del dominio de unión a la diana, afectando de
40 ese modo a la probabilidad de que una sonda hibride y permanezca hibridada con un ácido nucleico diana. En realizaciones, nucleótidos "n" pueden preceder los nucleótidos correspondientes a las posiciones en el dominio de código de barras. En realizaciones, nucleótidos "n" pueden seguir a los nucleótidos correspondientes a las posiciones en el dominio de código de barras. En la figura 4, se muestran cuatro nucleótidos "n"; sin embargo, un dominio de unión a la diana puede incluir más de cuatro nucleótidos "n". Los nucleótidos "n" pueden tener bases universales (por
45 ejemplo, inosina, derivados de 2'-desoxiinosina (desoxinucleótido de hipoxantina), nitroindol, análogos de nitroazol y bases aromáticas hidrófobas que no se unen a hidrógeno) que pueden formar pares de bases con cualquiera de las cuatro bases canónicas.

Otra sonda de secuenciación de la presente divulgación se muestra en la figura 5. Aquí, los nucleótidos "n" preceden y siguen a los nucleótidos correspondientes a las posiciones en el dominio de código de barras. La sonda de
50 secuenciación ejemplar mostrada tiene un dominio de unión a la diana de diez nucleótidos. Sin embargo, solamente del tercer al octavo nucleótido en el dominio de unión a la diana corresponden a las seis posiciones (de la primera a la sexta) en el dominio de código de barras. El primer, segundo, noveno y décimo nucleótido (indicados por "n₁ a n₄") se añaden para aumentar la longitud del dominio de unión a la diana. En la figura 5, se muestran cuatro nucleótidos "n"; sin embargo, un dominio de unión a la diana puede incluir más o menos de cuatro nucleótidos "n".

La figura 6A a figura 6D muestran variantes de una sonda de secuenciación de la figura 1. En la figura 6A, el orden lineal de nucleótidos en el dominio de unión a la diana y el orden lineal de regiones de fijación en el dominio de código de barras progresa de izquierda a derecha (con respecto a la ilustración). En la figura 6B, el orden lineal de nucleótidos en el dominio de unión a la diana y el orden lineal de regiones de fijación en el dominio de código de barras progresa de derecha a izquierda (con respecto a la ilustración). En la figura 6C, el orden lineal de nucleótidos en el dominio de
60 unión a la diana está invertido con respecto al orden lineal de regiones de fijación en el dominio de código de barras.

En cualquier sonda de la presente invención, puede haber una ausencia de orden estricto de los nucleótidos en el dominio de unión a la diana y de regiones de fijación en el dominio de código de barras siempre que la sonda se diseñe de modo que cada nucleótido en el dominio de unión a la diana corresponda a un dominio de fijación o dominios de fijación en el dominio de código de barras; las ausencias de orden estricto se muestran en la figura 6D. Cualquier sonda de la presente invención (por ejemplo, las ejemplificadas en las figuras 1 a 5) puede tener una ordenación de nucleótidos y regiones de fijación como se muestra en la figura 6.

El dominio de unión a la diana tiene al menos cuatro nucleótidos, por ejemplo, al menos, 4, 5, 6, 7, 8, 9, 10, 11, 12 o más nucleótidos. El dominio de unión a la diana preferible es un polinucleótido. El dominio de unión a la diana puede unirse a un ácido nucleico diana.

Una sonda puede incluir múltiples copias del dominio de unión a la diana unidas de forma funcional a una cadena principal sintética.

Pueden diseñarse sondas para controlar la probabilidad de hibridación y/o deshibridación y las tasas a las que estas se producen. En general, cuanto menor es la T_m de una sonda, más rápido y más probablemente se deshibridará esa sonda con/de un ácido nucleico diana. Por tanto, el uso de sondas de menor T_m disminuirá el número de sondas unidas a un ácido nucleico diana.

La longitud de un dominio de unión a la diana, en parte, afecta a la probabilidad de que una sonda hibride y permanezca hibridada con un ácido nucleico diana. En general, cuanto más largo (mayor número de nucleótidos) sea un dominio de unión a la diana, menos probable será que una secuencia complementaria esté presente en el nucleótido diana. A la inversa, cuanto más corto sea un dominio de unión a la diana, más probable será que una secuencia complementaria esté presente en el nucleótido diana. Por ejemplo, hay una probabilidad de 1/256 de que una secuencia tetramérica esté localizada en un ácido nucleico diana frente a una probabilidad de 1/4096 de que una secuencia hexamérica esté localizada en el ácido nucleico diana. Por consiguiente, una colección de sondas más cortas probablemente se unirá en menos localizaciones para un tramo dado de un ácido nucleico en comparación con una colección de sondas más largas.

La figura 7 muestra dominios de unión a la diana deca-méricos. En algunas realizaciones, el dominio de unión a la diana incluye cuatro bases universales (identificadas como "U_b") que forman pares de bases con cualquiera de los cuatro nucleótidos canónicos (A, G, C y T). En realizaciones, el dominio de unión a la diana incluye de uno a seis (por ejemplo, 2 y 4) bases universales. Un dominio de unión a la diana puede no incluir nucleótidos universales. La figura 7 indica que una población "completa" de sondas que tienen 6 nucleótidos específicos en el dominio de unión a la diana requerirá 4096 sondas peculiares y una población "completa" de sondas que tengan 10 nucleótidos específicos requerirá ~1 millón de sondas peculiares.

En circunstancias, es preferible tener sondas que tengan dominios de unión a la diana más cortos para aumentar el número de lecturas en el tramo dado del ácido nucleico, enriqueciendo de ese modo la cobertura de un ácido nucleico diana o una parte del ácido nucleico diana, especialmente una parte de interés particular, por ejemplo, cuando se detecta una mutación o alelo SNP.

Sin embargo, puede ser preferible tener números menores de sondas unidas a un ácido nucleico diana, ya que hay ocasiones en que demasiadas sondas en una región pueden provocar solapamiento de su marcador detectable, evitando de ese modo la resolución de dos sondas cercanas. Esto se explica como sigue. Dado que un nucleótido es de 0,34 nm de longitud y dado que la resolución espacial lateral (x-y) de un aparato de secuenciación es de aproximadamente 200 nm, el límite de resolución de un aparato de secuenciación es de aproximadamente 588 pares de bases (es decir, 1 nucleótido/0,34 nm x 200 nm). Es decir, el aparato de secuenciación mencionado anteriormente no podría resolver señales de dos sondas hibridadas con un ácido nucleico diana cuando las dos sondas están en aproximadamente 588 pares de bases entre sí. Por tanto, dos sondas, dependiendo de la resolución del aparato de secuenciación, tendrán que estar espaciadas aproximadamente 600 pb antes de que su marcador detectable pueda resolverse como "manchas" distintivas. De modo que, en espaciado óptimo, debe haber una sola sonda por 600 pb de ácido nucleico diana. Puede usarse una diversidad de estrategias de programa informático (por ejemplo, utilizar valores de intensidad de fluorescencia y relaciones dependientes de longitud de onda) para controlar, limitar y deconvolucionar potencialmente el número de sondas que hibridan dentro de una región resoluble de un ácido nucleico diana y para diseñar poblaciones de sondas en consecuencia. Además, pueden seleccionarse marcadores detectables (por ejemplo, marcadores fluorescentes) que proporcionen señales más diferenciadas. Además, métodos de la bibliografía (por ejemplo, Small y Parthasarthy: "Superresolution localization methods." Annu. Rev. Phys Chem., 2014; 65:107-25) describen iluminación estructurada y una diversidad de estrategias de superresolución que disminuyen el límite de resolución de un microscopio de secuenciación de hasta décimas partes de nanómetros. El uso de aparatos de secuenciación de resolución mayor permite el uso de sondas con dominios de unión a la diana más cortos.

Como se menciona anteriormente, el diseño de la T_m de las sondas puede afectar al número de sondas hibridadas con un ácido nucleico diana. Como alternativa o adicionalmente, la concentración de las sondas de secuenciación en una población puede aumentarse para aumentar la cobertura de las sondas en una región específica de un ácido nucleico diana. La concentración de las sondas de secuenciación puede reducirse para disminuir la cobertura de las

sondas en una región específica de un ácido nucleico diana, por ejemplo, hasta por encima del límite de resolución del aparato de secuenciación.

La expresión "ácido nucleico diana" significará una molécula de ácido nucleico (ADN, ARN o APN) cuya secuencia tiene que determinarse mediante las sondas, métodos y aparatos de la divulgación. En general, la expresión "ácido nucleico diana", "molécula de ácido nucleico", "secuencia de ácido nucleico", "ácido nucleico", "fragmento de ácido nucleico", "oligonucleótido" y "polinucleótido" se usan indistintamente y se pretende que incluyan, aunque sin limitación, una forma polimérica de nucleótidos que puede tener diversas longitudes, deoxirribonucleótidos o ribonucleótidos, o análogos de los mismos. Ejemplos no limitantes de ácido nucleicos incluyen un gen, un fragmento génico, un exón, un intrón, ADN intergénico (incluyendo, sin limitación, ADN heterocromático), ARN mensajero (ARNm), ARN transferente, ARN ribosómico, ribozimas, ARN interferente pequeño (ARNip), ARN no codificante (ARNnc), ADNc, polinucleótidos recombinantes, polinucleótidos ramificados, plásmidos, vectores, ADN aislado de una secuencia, ARN aislado de una secuencia, sondas de ácido nucleico y cebadores.

Los presentes métodos secuencian directamente una molécula de ácido nucleico obtenida de una muestra, por ejemplo, una muestra de un organismo y, preferiblemente, sin una etapa de conversión (o amplificación). Como ejemplo, para secuenciación basada en ARN, los presentes métodos no requieren conversión de una molécula de ARN en una molécula de ADN (es decir, mediante síntesis de ADNc) antes de que pueda obtenerse una secuencia. Como no se requiere amplificación o conversión, un ácido nucleico secuenciado en la presente invención retendrá cualquier base peculiar y/o marcador epigenético presente en el ácido nucleico cuando el ácido nucleico esté en la muestra o cuando se obtuviera de la muestra. Dichas bases peculiares y/o marcadores epigenéticos se pierden en métodos de secuenciación conocidos en la técnica.

El ácido nucleico diana puede obtenerse de cualquier muestra o fuente de ácido nucleico, por ejemplo, cualquier célula, tejido u organismo, *in vitro*, sintetizador químico y así sucesivamente. El ácido nucleico diana puede obtenerse por cualquier método reconocido en la técnica. En realizaciones, el ácido nucleico se obtiene de una muestra de sangre de un sujeto clínico. El ácido nucleico puede extraerse, aislarse o purificarse de la fuente o muestras usando métodos y kits bien conocidos en la técnica.

Una molécula de ácido nucleico que comprende el ácido nucleico diana puede fragmentarse por cualquier medio conocido en la técnica. Preferiblemente, la fragmentación se realiza mediante un medio enzimático o uno mecánico. El medio mecánico puede ser sonicación o cizallamiento físico. El medio enzimático puede realizarse por digestión con nucleasas (por ejemplo, desoxirribonucleasa I (DNasa I)) o una o más endonucleasas de restricción.

Cuando una molécula de ácido nucleico que comprende el ácido nucleico diana es un cromosoma intacto, deben adoptarse etapas para evitar la fragmentación del cromosoma.

El ácido nucleico diana puede incluir nucleótidos naturales o no naturales, que comprenden nucleótidos modificados, como es bien conocido en la técnica.

Las sondas de la presente invención pueden tener longitudes globales (incluyendo el dominio de unión a la diana, el dominio de código de barras y cualquier dominio opcional) de aproximadamente 20 nanómetros a aproximadamente 50 nanómetros. La cadena principal de una sonda puede ser una molécula polinucleotídica que comprende aproximadamente 120 nucleótidos.

El dominio de código de barras comprende una cadena principal sintética. La cadena principal sintética y el dominio de unión a la diana se unen de forma funcional, por ejemplo, se fijan covalentemente o se fijan mediante un conector. La cadena principal sintética puede comprender cualquier material, por ejemplo, polisacárido, polinucleótido, polímero, plástico, fibra, péptido, ácido peptidonucleico o polipéptido. Preferiblemente, la cadena principal sintética es rígida. En realizaciones, la cadena principal comprende "origami de ADN" de seis dobles hélices de ADN (véase, por ejemplo, Lin *et al.*, "Submicrometre geometrically encoded fluorescent barcodes self-assembled from DNA." *Nature Chemistry*; oct 2012; 4(10): 832-9). Un código de barras puede generarse de placas de origami de ADN (Jungmann *et al.*, "Multiplexed 3D cellular super-resolution imaging with DNA-PAINT and Exchange-PAINT", *Nature Methods*, vol. 11, n.º 3, 2014).

El dominio de código de barras comprende una pluralidad de posiciones, por ejemplo, una, dos, tres, cuatro, cinco, seis, siete, ocho, nueve, diez o más posiciones. El número de posiciones puede ser menor de, igual a o mayor del número de nucleótidos en el dominio de unión a la diana. Es preferible incluir nucleótidos adicionales en un dominio de unión a la diana del número de posiciones en el dominio de cadena principal, por ejemplo, uno, dos, tres, cuatro, cinco, seis, siete, ocho, nueve, diez o más nucleótidos. La longitud del dominio de código de barras no está limitada siempre que haya espacio suficiente para al menos cuatro posiciones, como se describe anteriormente.

Cada posición en el dominio de código de barras corresponde a un nucleótido en el dominio de unión a la diana y, por tanto, a un nucleótido en el ácido nucleico diana. Como ejemplos, la primera posición en el dominio de código de barras corresponde al primer nucleótido en el dominio de unión a la diana y la sexta posición en el dominio de código de barras corresponde al sexto nucleótido en el dominio de unión a la diana.

Cada posición en el dominio de código de barras comprende al menos una región de fijación, por ejemplo, de una a 50, o más, regiones de fijación. Determinadas posiciones en un dominio de código de barras pueden tener más regiones de fijación que otras posiciones (por ejemplo, una primera posición puede tener tres regiones de fijación mientras que una segunda posición puede tener dos posiciones de fijación); como alternativa, cada posición en un dominio de código de barras tiene el mismo número de regiones de fijación. Cada región de fijación comprende al menos una (es decir, de una a cincuenta, por ejemplo, de diez a treinta) copias de una o más secuencias de ácido nucleico que pueden unirse de forma reversible mediante una molécula de ácido nucleico complementaria (por ejemplo, ADN o ARN). En ejemplos, la secuencia de ácido nucleico en una primera región de fijación determina la posición e identidad de un primer nucleótido en el ácido nucleico diana que se une mediante un primer nucleótido del dominio de unión a la diana. Cada región de fijación puede ligarse a un monómero modificado (por ejemplo, nucleótido modificado) en la cadena principal sintética de modo que la región de fijación se ramifique desde la cadena principal sintética. En realizaciones, las regiones de fijación están integradas en una cadena principal polinucleotídica; es decir, la cadena principal es un polinucleótido individual y las regiones de fijación son partes de la secuencia del polinucleótido individual. En realizaciones, las expresiones "dominio de código de barras" y "cadena principal sintética" son sinónimas.

La secuencia de ácido nucleico en una región de fijación identifica la posición e identidad de un nucleótido en el ácido nucleico diana que se une mediante un nucleótido en el dominio de unión a la diana de una sonda de secuenciación. En una sonda, cada región de fijación tendrá una secuencia global peculiar. De hecho, cada posición en un dominio de código de barras puede tener una región de fijación que comprende una secuencia de ácido nucleico que codifica uno de cuatro nucleótidos, es decir, específica para uno de adenina, timina/uracilo, citosina y guanina. Además, la región de fijación de una primera posición (y que codifica citosina, por ejemplo) incluirá una secuencia de ácido nucleico diferente de la región de fijación de una segunda posición (y que codifica citosina, por ejemplo). Por tanto, en una secuencia de ácido nucleico en una región de fijación en una primera posición que codifica una timina, no habrá unión de una molécula de ácido nucleico complementaria que identifique una adenina en un ácido nucleico diana correspondiente al primer nucleótido de un dominio de unión a la diana. Además, en una región de fijación en una segunda posición, no habrá unión de una molécula de ácido nucleico complementaria que identifique una adenina en un ácido nucleico diana correspondiente al primer nucleótido de un dominio de unión a la diana.

Cada posición en un dominio de código de barras puede incluir una o más (hasta cincuenta, preferiblemente de diez a treinta) regiones de fijación; por tanto, cada región de fijación puede unirse a una o más (hasta cincuenta, preferiblemente de diez a treinta) moléculas de ácido nucleico complementarias. Como ejemplos, la sonda de la figura 1 tiene una quinta posición que comprende dos regiones de fijación y la sonda de la figura 2 tiene una segunda posición que tiene seis regiones de fijación. En realizaciones, las secuencias de ácido nucleico de regiones de fijación en una posición son idénticas; por tanto, las moléculas de ácido nucleico complementarias que se unen a esas regiones de fijación son idénticas. En realizaciones alternativas, las secuencias de ácido nucleico de regiones de fijación en una posición no son idénticas; por tanto, las moléculas de ácido nucleico complementarias que se unen a esas regiones de fijación no son idénticas, por ejemplo, cada una comprende una secuencia de ácido nucleico y/o marcador detectable diferente. Por lo tanto, en la realización alternativa, la combinación de moléculas de ácido nucleico no idénticas (por ejemplo, sus marcadores detectables) fijadas a una región de fijación conjuntamente proporciona un código para identificar un nucleótido en el ácido nucleico diana.

La tabla 1 proporciona secuencias ejemplares, con propósitos de ilustración únicamente, para regiones de fijación para sondas de secuenciación que tienen hasta seis posiciones en su dominio de código de barras y marcadores detectables en el ácido nucleico complementario que se unen al mismo.

Tabla 1:

Nucleótido en dominio de unión a la diana/posición en dominio de código de barras	Nucleótido	Secuencia de ácido nucleico (5' a 3') en región de fijación	Marcador detectable de ácido nucleico complementario	SEQ ID NO:
1	A	ATACATCTAG	GFP	1
1	G	GATCTACATA	RFP	2
1	C	TTAGGTAAAG	CFP	3
1	U/T	TCTTCATTAC	YFP	4
2	A	ATGAATCTAC	GFP	5
2	G	TCAATGTATG	RFP	6
2	C	AATTGAGTAC	CFP	7

Nucleótido en dominio de unión a la diana/posición en dominio de código de barras	Nucleótido	Secuencia de ácido nucleico (5' a 3') en región de fijación	Marcador detectable de ácido nucleico complementario	SEQ ID NO:
2	U/T	ATGTTAATGG	YFP	8
3	A	AATTAGGATG	GFP	9
3	G	ATAATGGATC	RFP	10
3	C	TAATAAGGTG	CFP	11
3	U/T	TAGTTAGAGC	YFP	12
4	A	ATAGAGAAGG	GFP	13
4	G	TTGATGATAC	RFP	14
4	C	ATAGTGATTC	CFP	15
4	U/T	TATAACGATG	YFP	16
5	A	TTAAGTTTAG	GFP	17
5	G	ATACGTTATG	RFP	18
5	C	TGTACTATAG	CFP	19
5	U/T	TTAACAAGTG	YFP	20
6	A	AACTATGTAC	GFP	21
6	G	TA ACTATGAC	RFP	22
6	C	ACTAATGTTC	CFP	23
6	U/T	TCATTGAATG	YFP	24

Como se observa en la tabla 1, la secuencia de ácido nucleico de una primera región de fijación puede ser una de SEQ ID NO: 1 a SEQ ID NO: 4 y la secuencia de ácido nucleico de una segunda fijación puede ser una de SEQ ID NO: 5 a SEQ ID NO: 8. Cuando el primer nucleótido en el ácido nucleico diana es adenina, la secuencia de ácido nucleico de la primera región de fijación tendría la secuencia de SEQ ID NO: 1 y cuando el segundo nucleótido en el ácido nucleico diana es adenina, la secuencia de ácido nucleico de la segunda región de fijación tendría la secuencia de SEQ ID NO: 5.

En realizaciones, una molécula de ácido nucleico complementaria puede unirse mediante un marcador detectable. En realizaciones alternativas, un ácido nucleico complementario se asocia con un complejo indicador que comprende marcadores detectables.

La secuencia de nucleótidos de un ácido nucleico complementario no está limitada; preferiblemente carece de homología sustancial (por ejemplo, de un 50 % a un 99,9 %) con una secuencia de nucleótidos conocida; que ayuda a evitar hibridación indeseable de un ácido nucleico complementario y un ácido nucleico diana.

Un ejemplo del complejo indicador útil en la presente invención se muestra en la figura 9B. En este ejemplo, un ácido nucleico complementario se liga a una molécula de ácido nucleico primaria que, a su vez, hibrida con una pluralidad de moléculas de ácido nucleico secundarias, de las que cada una de ellas, a su vez, hibrida con una pluralidad de moléculas de ácido nucleico terciarias que tienen fijado a las mismas uno o más marcadores detectables.

En realizaciones, una molécula de ácido nucleico primaria puede comprender aproximadamente 90 nucleótidos. Una molécula de ácido nucleico secundaria puede comprender aproximadamente 87 nucleótidos. Una molécula de ácido nucleico terciaria puede comprender aproximadamente 15 nucleótidos.

La figura 9C muestra una población de complejos indicadores ejemplares. En el panel superior de la izquierda de la figura 9C se incluyen los cuatro complejos que hibridan con la región de fijación 1 de una sonda. Hay un tipo de complejo indicador por cada posible nucleótido que puede estar presente en la posición nucleotídica 1 del dominio de unión a la diana de una sonda. Aquí, al realizar un método de secuencia de la presente invención, si la posición 1 del

dominio indicador de una sonda se une mediante un complejo indicador que tiene un marcador detectable "coloreado de azul", entonces el primer nucleótido en el dominio de unión a la diana se identifica como adenina. Como alternativa, si la posición 1 se une mediante un complejo indicador que tiene un marcador detectable "coloreado de verde", entonces el primer nucleótido en el dominio de unión a la diana se identifica como timina.

Los complejos indicadores pueden ser de diversos diseños. Por ejemplo, una molécula de ácido nucleico primaria puede hibridar con al menos una (por ejemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más) moléculas de ácido nucleico secundarias. Cada molécula de ácido nucleico secundaria puede hibridar con al menos una (por ejemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más) moléculas de ácido nucleico terciarias. Se muestran complejo indicadores ejemplares en la figura 20A. Aquí, el complejo indicador "4x3" tiene una molécula de ácido nucleico primaria (que se liga a una molécula de ácido nucleico complementaria) hibridada con cuatro moléculas de ácido nucleico secundarias, de las que cada una de ellas hibrida con tres moléculas de ácido nucleico terciarias (que comprenden cada una un marcador detectable). En esta figura, cada ácido nucleico complementario de un complejo es de 12 nucleótidos de longitud ("12 bases"); sin embargo, la longitud del ácido nucleico complementario no está limitada y puede ser de menos de 12 o de más de 12 nucleótidos. El complejo de la parte inferior a la derecha incluye una región espaciadora entre su ácido nucleico complementario y su molécula de ácido nucleico primaria. El espaciador se identifica como de 20 a 40 nucleótidos de longitud; sin embargo, la longitud de un espaciador no es limitante y puede ser más corta de 20 nucleótidos o más larga de 40 nucleótidos.

La figura 20B muestra los recuentos promedio variables (fluorescentes) obtenidos de los cuatro complejos indicadores ejemplares mostrados en la figura 20A. En la figura 20B, se fijaron 10 pM de molde diana biotinilado en una superficie de cubeta de lectura recubierta de estreptavidina, se hicieron fluir 10 nM de un complejo indicador sobre la cubeta de lectura; después de un minuto de incubación, se lavó la cubeta de lectura, se tomaron imágenes de la cubeta de lectura y se contaron los elementos fluorescentes.

En realizaciones, los complejos indicadores se "construyen previamente". Es decir, cada polinucleótido en el complejo se hibrida antes de poner en contacto el complejo con una sonda. Una receta ejemplar para preconstruir cinco complejos indicadores ejemplares se muestra en la figura 20C.

La figura 21A muestra complejos indicadores alternativos en que las moléculas de ácido nucleico secundarias tienen "mangos adicionales" que no hibridan con una molécula de ácido nucleico terciaria y están distales de la molécula de ácido nucleico primaria. En esta figura, cada "mango adicional" es de 12 nucleótidos de longitud ("dodecamérico"); sin embargo, sus longitudes no están limitadas y pueden ser de menos de 12 o de más de 12 nucleótidos. En realizaciones, los "mangos adicionales" comprenden cada uno la secuencia de nucleótidos del ácido nucleico complementario; por tanto, cuando un complejo indicador comprende "mangos adicionales", el complejo indicador puede hibridar con una sonda de secuenciación mediante el ácido nucleico complementario del complejo indicador o mediante un "mango adicional". Por consiguiente, se aumenta la probabilidad de que un complejo indicador se una a una sonda de secuenciación. El diseño de "mango adicional" también puede mejorar la cinética de hibridación. Sin limitarse a teoría alguna, los "mangos adicionales" esencialmente aumentan la concentración eficaz del ácido nucleico complementario del complejo indicador.

La figura 21B muestra los recuentos promedio variables (fluorescente) obtenidos de cinco complejos indicadores ejemplares que tienen "mangos adicionales" usando el procedimiento descrito para la figura 20B.

La figura 22A y 22B muestran la cinética de hibridación y las intensidades fluorescentes para dos complejos indicadores ejemplares. En aproximadamente 5 minutos, los recuentos totales empiezan a estancarse, lo que indica que la mayor parte del complejo indicador añadido ha encontrado una diana disponible.

Un resto, marcador o indicador detectable puede unirse a un ácido nucleico complementario o a una molécula de ácido nucleico terciaria de una diversidad de maneras, incluyendo la fijación directa o indirecta de un resto detectable tal como un resto fluorescente, resto colorimétrico y similares. Un experto en la materia puede consultar referencias referidas al marcaje de ácidos nucleicos. Ejemplos de restos fluorescentes incluyen, aunque sin limitación, proteína fluorescente amarilla (YFP), proteína fluorescente verde (GFP), proteína fluorescente cian (CFP), proteína fluorescente roja (RFP), umbeliferona, fluoresceína, isotiocianato de fluoresceína, rodamina, diclorotriazinilamina fluoresceína, cianinas, cloruro de dansilo, ficocianina, ficoeritrina y similares. Los marcadores fluorescentes y su fijación a nucleótidos y/u oligonucleótidos se describen en muchas revisiones, incluyendo Haugland, Handbook of Fluorescent Probes and Research Chemicals, novena edición (Molecular Probes, Inc., Eugene, 2002); Keller y Manak, DNA Probes, 2.^a edición (Stockton Press, Nueva York, 1993); Eckstein, editor, Oligonucleotides and Analogues: A Practical Approach (IRL Press, Oxford, 1991); y Wetmur, Critical Reviews in Biochemistry and Molecular Biology, 26:227-259 (1991). Se divulgan metodologías particulares aplicables a la invención en la siguiente muestra de referencias: patentes de Estados Unidos n.º 4.757.141; 5.151.507; y 5.091.519. En un aspecto, se usa uno o más tintes fluorescentes como marcadores para secuencias diana marcadas, por ejemplo, como se divulga por las patentes de Estados Unidos n.º 5.188.934 (tintes de 4,7-diclorofluoresceína); 5.366.860 (tintes de rodamina de resolución espectral); 5.847.162 (tintes de 4,7-diclororrodamina); 4.318.846 (tintes de fluoresceína sustituida con éter); 5.800.996 (tintes de transferencia de energía); Lee *et al.* 5.066.580 (tintes de xantina); 5.688.648 (tintes de transferencia de energía); y similares. El marcaje también puede realizarse con puntos cuánticos, como se divulga en las siguientes patentes y publicaciones de patente: patentes de Estados Unidos n.º 6.322.901; 6.576.291; 6.423.551; 6.251.303;

6.319.426; 6.426.513; 6.444.143; 5.990.479; 6.207.392; 2002/0045045; y 2003/0017264. Como se usa en la presente memoria, la expresión "marcador fluorescente" comprende un resto de señalización que transporta información a través de las propiedades de absorción y/o emisión fluorescente de una o más moléculas. Dichas propiedades fluorescentes incluyen intensidad de fluorescencia, duración de la fluorescencia, características del espectro de emisión, transferencia de energía y similares.

Los análogos nucleotídicos fluorescentes disponibles en el mercado que se incorporan fácilmente en secuencias de nucleótidos y/u oligonucleotídicas incluyen, aunque sin limitación, Cy3-dCTP, Cy3-dUTP, Cy5-dCTP, Cy5-dUTP (Amersham Biosciences, Piscataway, NJ), fluoresceína-12-dUTP, tetrametilrodamina-6-dUTP, TEXAS RED™-5-dUTP, CASCADE BLUE™-7-dUTP, BODIPY TMFL-14-dUTP, BODIPY TMR-14-dUTP, BODIPY TMTR-14-dUTP, RHODAMINE GREEN™-5-dUTP, OREGON GREEN™ 488-5-dUTP, TEXAS RED™-12-dUTP, BODIPY TM 630/650-14-dUTP, BODIPY TM 650/665-14-dUTP, ALEXA FLUOR™ 488-5-dUTP, ALEXA FLUOR™ 532-5-dUTP, ALEXA FLUOR™ 568-5-dUTP, ALEXA FLUOR™ 594-5-dUTP, ALEXA FLUOR™ 546-14-dUTP, fluoresceína-12-UTP, tetrametilrodamina-6-UTP, TEXAS RED™-5-UTP, mCherry, CASCADE BLUE™-7-UTP, BODIPY™ FL-14-UTP, BODIPY™ R-14-UTP, BODIPY™ TR-14-UTP, RHODAMINE GREEN™-5-UTP, ALEXA FLUOR™ 488-5-UTP, ALEXA FLUOR™ 546-14-UTP (Molecular Probes, Inc. Eugene, OR) y similares. Como alternativa, los fluoróforos anteriores y los mencionados en la presente memoria pueden añadirse durante síntesis de oligonucleótidos usando, por ejemplo, química de fosforoamida o NHS. Se conocen protocolos en la técnica para síntesis personalizada de nucleótidos que tienen otros fluoróforos (véase, Henegariu *et al.* (2000) *Nature Biotechnol.* 18:345). La 2-aminopurina es una base fluorescente que puede incorporarse directamente en la secuencia oligonucleotídica durante su síntesis. El ácido nucleico también podría teñirse, *a priori*, con un tinte intercalante tal como DAPI, YOYO-1, bromuro de etidio, tintes de cianina (por ejemplo, SYBR Green) y similares.

Otros fluoróforos disponibles para posfijación sintética incluyen, aunque sin limitación, ALEXA FLUOR™ 350, ALEXA FLUOR™ 405, ALEXA FLUOR™ 430, ALEXA FLUOR™ 532, ALEXA FLUOR™ 546, ALEXA FLUOR™ 568, ALEXA FLUOR™ 594, ALEXA FLUOR™ 647, BODIPY 493/503, BODIPY FL, BODIPY R6G, BODIPY 530/550, BODIPY TMR, BODIPY 558/568, BODIPY 558/568, BODIPY 564/570, BODIPY 576/589, BODIPY 581/591, BODIPY TR, BODIPY 630/650, BODIPY 650/665, Cascade Blue, Cascade Yellow, dansilo, lisamina rodamina B, Marina Blue, Oregon Green 488, Oregon Green 514, Pacific Blue, Pacific Orange, rodamina 6G, verde de rodamina, rojo de rodamina, tetrametilrodamina, Texas Red (disponible en Molecular Probes, Inc., Eugene, OR), Cy2, Cy3, Cy3.5, Cy5, Cy5.5, Cy7 (Amersham Biosciences, Piscataway, NJ) y similares. También pueden usarse fluoróforos en tándem con FRET incluyendo, aunque sin limitación, PerCP-Cy5.5, PE-Cy5, PE-Cy5.5, PE-Cy7, PE-Texas Red, APC-Cy7, tintes PE-Alexa (610, 647, 680), tintes APC-Alexa y similares.

Pueden usarse partículas metálicas de plata u oro para potenciar la señal desde secuencias de nucleótidos y/u oligonucleotídicas marcadas de forma fluorescente (Lakowicz *et al.* (2003) *BioTechniques* 34:62).

Otros marcadores adecuados para una secuencia oligonucleotídica pueden incluir fluoresceína (FAM, FITC), digoxigenina, dinitrofenol (DNP), dansilo, biotina, bromodesoxiuridina (BrdU), hexahistidina (6xHis), fosfoaminoácidos (por ejemplo, P-tyr, P-ser, P-thr) y similares. En una realización, se usan los siguientes pares de hapteno/anticuerpo para la detección, en que cada uno de los anticuerpos se derivatiza con un marcador detectable: biotina/a-biotina, digoxigenina/a-digoxigenina, dinitrofenol (DNP)/a-DNP, 5-carboxifluoresceína (FAM)/a-FAM.

Los marcadores detectables descritos en la presente memoria son de resolución espectral. "De resolución espectral" en referencia a una pluralidad de marcadores fluorescentes significa que las bandas de emisión fluorescente de los marcadores son suficientemente distintivas, es decir, suficientemente no solapantes, para que las marcas moleculares a las que se fijan los marcadores respectivos puedan distinguirse basándose en la señal fluorescente generada por los marcadores respectivos por sistemas convencionales de fotodetección, por ejemplo, empleando un sistema de filtros de paso de banda y tubos fotomultiplicadores, o similares, como se ejemplifica por los sistemas descritos en las patentes de Estados Unidos n.º 4.230.558; 4.811.218; o similares, o en Wheelless *et al.*, pág. 21-76, en *Flow Cytometry: Instrumentation and Data Analysis* (Academic Press, Nueva York, 1985). En un aspecto, tintes orgánicos de resolución espectral, tales como fluoresceína, rodamina y similares, significa que los máximos de emisión de longitud de onda están espaciados al menos 20 nm, y en otro aspecto, al menos 40 nm. En otro aspecto, compuestos lantánidos quelados, puntos cuánticos y similares, de resolución espectral, significa que los máximos de emisión de longitud de onda están espaciados al menos 10 nm, y en un aspecto adicional, al menos 15 nm.

Método de secuenciación

La presente invención se refiere a métodos para secuenciar un ácido nucleico usando una sonda de secuenciación de la presente invención. Ejemplos del método se muestran en las figuras 8 a 12.

El método comprende hibridar de forma reversible al menos una sonda de secuenciación, de la presente invención, con un ácido nucleico diana que se inmoviliza (por ejemplo, en una, dos, tres, cuatro, cinco, seis, siete, ocho, nueve, diez o más posiciones) en un sustrato.

El sustrato puede ser cualquier soporte sólido conocido en la técnica, por ejemplo, un portaobjetos recubierto y un dispositivo de microfluidos, que puede inmovilizar un ácido nucleico diana. En determinadas realizaciones, el sustrato

es una superficie, membrana, microesfera, material poroso, electrodo o matriz. El ácido nucleico diana puede inmovilizarse sobre cualquier sustrato evidente para los expertos en la materia.

En realizaciones, el ácido nucleico diana se une mediante una sonda de captura que comprende un dominio que es complementario a una parte del ácido nucleico diana. La parte puede ser un extremo del ácido nucleico diana o no hacia un extremo.

Sustratos útiles ejemplares incluyen los que comprenden un resto de unión seleccionado del grupo que consiste en ligandos, antígeno, carbohidratos, ácidos nucleicos, receptores, lectinas y anticuerpos. La sonda de captura comprende un resto de unión que puede unirse con el resto de unión del sustrato. Sustratos útiles ejemplares que comprenden restos reactivos incluyen, aunque sin limitación, superficies que comprenden epoxi, aldehído, oro, hidrazida, sulfhidrilo, NHS-éster, amina, tiol, carboxilato, maleimida, hidroximetilfosfina, imidoéster, isocianato, hidroxilo, pentafluorofenil-éster, psoraleno, disulfuro de piridilo o vinilsulfona, polietilenglicol (PEG), hidrogel o mezclas de los mismos. Dichas superficies pueden obtenerse de fuentes comerciales o prepararse de acuerdo con técnicas convencionales. Sustratos útiles ejemplares que comprenden restos reactivos incluyen, aunque sin limitación, OptArray de ADN con grupo NHS (Accler8), Nexterion Slide AL (Schott) y Nexterion Slide E (Schott).

En realizaciones, el resto de unión de la sonda de captura es biotina y el sustrato comprende avidina (por ejemplo, estreptavidina). Sustratos útiles que comprenden avidina están disponibles en el mercado, incluyendo TB0200 (Accelr8), SAD6, SAD20, SAD100, SAD500, SAD2000 (Xantec), SuperAvidin (Array-It), portaobjetos de estreptavidina (n.º de catálogo MPC 000, Xenopore) y STREPTAVIDINslide (número de catálogo 439003, Greiner Bio-one).

En realizaciones, el resto de unión de la sonda de captura es avidina (por ejemplo, estreptavidina) y el sustrato comprende biotina. Sustratos útiles que comprenden biotina que están disponibles en el mercado incluyen, aunque sin limitación, Optarray-biotina (Accler8), BD6, BD20, BD100, BD500 y BD2000 (Xantec).

En realizaciones, el resto de unión de la sonda de captura puede comprender un resto reactivo que puede unirse al sustrato por fotoactivación. El sustrato podría comprender el resto fotorreactivo, o la primera parte del nanoindicador podría comprender el resto fotorreactivo. Algunos ejemplos de restos fotorreactivos incluyen arilazidas, tales como N((2-piridilditio)etil)-4-azidosalicilamida; arilazidas fluoradas, tales como ácido 4-azido-2,3,5,6-tetrafluorobenzoico; reactivos basados en benzofenona, tales como el éster succinimidílico del ácido 4-benzoilbenzoico; y 5-bromo-desoxiuridina.

En realizaciones, el resto de unión de la sonda de captura puede inmovilizarse en el sustrato mediante otros pares de unión evidentes para los expertos en la materia.

Después de la unión al sustrato, el ácido nucleico diana puede elongarse aplicando una fuerza (por ejemplo, gravedad, fuerza hidrodinámica, fuerza electromagnética de "electroestiramiento", estiramiento fluido, una técnica de menisco menguante y combinaciones de los mismos) suficiente para prolongar el ácido nucleico diana.

El ácido nucleico diana puede unirse mediante una segunda sonda de captura que comprende un dominio que es complementario a una segunda parte del ácido nucleico diana. La parte puede ser un extremo del ácido nucleico diana o no hacia un extremo. La unión de una segunda sonda de captura puede producirse después o durante la elongación del ácido nucleico diana o en un ácido nucleico diana que no se ha elongado. La segunda sonda de captura puede tener una unión como se describe anteriormente.

Una sonda de captura puede comprender o estar asociada con un marcador detectable, es decir, una mancha de referencia.

La sonda de captura puede aislar un ácido nucleico diana de una muestra. Aquí, se añade una sonda de captura a una muestra que comprende el ácido nucleico diana. La sonda de captura se une al ácido nucleico diana mediante la región de la sonda de captura que es complementaria a una región del ácido nucleico diana. Cuando el ácido nucleico diana entra en contacto con un sustrato que comprende un resto que se une al resto de unión de la sonda de captura, el ácido nucleico queda inmovilizado sobre el sustrato.

Para garantizar que un usuario "captura" tantas moléculas de ácido nucleico diana como sea posible de muestras altamente fragmentadas, es de ayuda incluir una pluralidad de sondas de captura, cada una complementaria a una región diferente del ácido nucleico diana. Por ejemplo, puede haber tres combinaciones de sondas de captura, con una primera combinación complementaria a regiones del ácido nucleico diana cerca de su extremo 5', una segunda combinación complementaria a regiones en el centro del ácido nucleico diana y una tercera combinación cerca de su extremo 3'. Esto puede generarse a "n regiones de interés" por ácido nucleico diana. En este ejemplo, cada combinación individual de ácido nucleico diana fragmentado se unía a una sonda de captura que comprende o estaba unida a una marca de biotina. Se aísla 1 enésima parte de la muestra introducida (donde n = el número de regiones distintas en el ácido nucleico diana) por cada cámara de combinación. La sonda de captura se une al ácido nucleico diana de interés. Entonces, el ácido nucleico diana se inmoviliza, mediante la biotina de la sonda de captura, en una molécula de avidina adherida al sustrato. Opcionalmente, el ácido nucleico diana se estira, por ejemplo, mediante flujo o fuerza electrostática. Todas las n-combinaciones pueden estirarse y unirse simultáneamente o, para maximizar el número de moléculas completamente estiradas, la combinación 1 (que captura la mayor parte de la región 5') puede

estirarse y unirse en primer lugar; entonces, la combinación 2 (que captura la región central de la diana) puede estirarse y unirse; finalmente, la combinación 3 puede estirarse y unirse.

El número de sondas de captura distintas requeridas está inversamente relacionado con el tamaño del fragmento de ácido nucleico diana. En otras palabras, se requerirán más sondas de capturas para un ácido nucleico diana muy fragmentado. Para tipos de muestra con ácidos nucleicos diana muy fragmentados y degradados (por ejemplo, tejido incrustado en parafina fijado en formol) puede ser útil incluir múltiples combinaciones de sondas de captura. Por otro lado, para muestras con fragmentos de ácido nucleico diana largos, por ejemplo, ácidos nucleicos aislados obtenidos *in vitro*, puede ser suficiente una sola sonda de captura en un extremo 5'.

La región del ácido nucleico diana entre dos sondas de captura o después de una sonda de captura y antes de un extremo del ácido nucleico diana se denomina en la presente memoria "hueco". El hueco es una parte del ácido nucleico diana que está disponible para unirse mediante una sonda de secuenciación de la presente invención. El hueco mínimo es una longitud de dominio de unión a la diana (por ejemplo, de 4 a 10 nucleótidos) y un hueco máximo es la mayoría de un cromosoma completo.

Se muestra un ácido nucleico diana inmovilizado en la figura 12. Aquí, las dos sondas de captura se identifican como "sonda de captura 5'" y "sonda de captura 3'".

La figura 8A muestra un esquema de una sonda de secuenciación unida a un ácido nucleico diana. Aquí, el ácido nucleico diana tiene una timidina (T). Se muestra una primera combinación de ácidos nucleicos complementarios que comprenden un marcador detectable o complejos indicadores en la parte superior, cada miembro de la combinación tiene un marcador detectable diferente (por ejemplo, se identifica la timidina mediante una señal verde) y una secuencia de nucleótidos diferente. El primer nucleótido en el dominio de unión a la diana se une a la T en el ácido nucleico diana. Las primeras regiones de fijación de la sonda incluyen una o más secuencias de nucleótidos que especifican que el primer nucleótido en el dominio de unión a la diana de la sonda se une a timidina. Por tanto, solamente el ácido nucleico complementario para timidina se une a la primera posición del dominio de código de barras. Como se muestra, un primer ácido nucleico complementario que codifica timidina que comprende un marcador detectable o complejos indicadores que comprenden marcadores detectables se unen a regiones de fijación en la primera posición del dominio de código de barras de la sonda.

El número de combinaciones de ácidos nucleicos complementarios o complejos indicadores es idéntico al número de posiciones en el dominio de código de barras. Por tanto, para un dominio de código de barras que tiene seis posiciones, se ciclarán seis combinaciones sobre las sondas.

Como alternativa, antes de poner en contacto un ácido nucleico diana con una sonda, la sonda puede hibridarse en su primera posición con un ácido nucleico complementario que comprende un marcador detectable o un complejo indicador. Por tanto, cuando se pone en contacto con su ácido nucleico diana, la sonda puede emitir una señal detectable desde su primera posición y es innecesario proporcionar una primera combinación de ácidos nucleicos complementarios o complejos indicadores que estén dirigidos a la primera posición en el dominio de código de barras.

La figura 8B continúa con el método mostrado en la figura 8A. Aquí, los primeros ácidos nucleicos complementarios (o complejos indicadores) para timidina que se unieron a regiones de fijación en la primera posición del dominio de código de barras se han remplazado con un primer ácido nucleico de hibridación para timidina y que carece de un marcador detectable. El primer ácido nucleico de hibridación para timidina y que carece de un marcador detectable desplaza los ácidos nucleicos complementarios previamente unidos que comprenden un marcador detectable o los complejos indicadores previamente unidos. De ese modo, la posición 1 del dominio de código de barras ya no emite una señal detectable.

En realizaciones, los ácidos nucleicos complementarios que comprenden un marcador detectable o complejos indicadores pueden retirarse de la región de fijación, pero no se remplazan con un ácido nucleico de hibridación que carece de un marcador detectable. Esto puede producirse, por ejemplo, añadiendo un agente caótopo, aumentando la temperatura, cambiando la concentración salina, ajustando el pH y/o aplicando una fuerza hidrodinámica. En estas realizaciones, se necesitan menos reactivos (es decir, ácidos nucleicos de hibridación que carecen de marcadores detectables).

La figura 8C continúa con el método de la divulgación. Aquí, el ácido nucleico diana tiene una citidina (C) después de su timidina (T). Se muestra una segunda combinación de ácidos nucleicos complementarios o complejos indicadores en la parte superior, cada miembro de la combinación tiene un marcador detectable diferente y una secuencia de nucleótidos diferente. Además, las secuencias de nucleótidos para los ácidos nucleicos complementarios o ácidos nucleicos complementarios de los complejos indicadores de la primera combinación son diferentes de las secuencias de nucleótidos para los de la segunda combinación. Sin embargo, los marcadores detectables específicos de base son comunes para las combinaciones de ácidos nucleicos complementarios, por ejemplo, se identifican timidinas mediante señales verdes. Aquí, el segundo nucleótido en el dominio de unión a la diana se une a la C en el ácido nucleico diana. Las segundas regiones de fijación de la sonda tienen una secuencia de nucleótidos que especifica que el segundo nucleótido en el dominio de unión a la diana de la sonda se une a citidina. Por tanto, solamente los ácidos nucleicos complementarios que comprenden un marcador detectable o complejos indicadores de la segunda

combinación y para citidina se unen a la segunda posición del dominio de código de barras. Como se muestra, el segundo ácido nucleico complementario que codifica citidina o complejo indicador se une en la segunda posición del dominio de código de barras de la sonda.

En realizaciones, las etapas mostradas en la figura 8C son posteriores a las etapas mostradas en la figura 8B. Aquí, una vez que la primera combinación de ácidos nucleicos complementarios o complejos indicadores (de la figura 8A) se ha remplazado con los primeros ácidos nucleicos de hibridación que carecen de un marcador detectable (en la figura 8B), entonces se proporciona una segunda combinación de ácidos nucleicos complementarios o complejos indicadores (como se muestra en la figura 8C). Como alternativa, las etapas mostradas en la figura 8C son simultáneas con las etapas mostradas en la figura 8B. Aquí, los primeros ácidos nucleicos de hibridación que carecen de un marcador detectable (en la figura 8B) se proporcionan simultáneamente con una segunda combinación de ácidos nucleicos complementarios o complejos indicadores (como se muestra en la figura 8C).

La figura 8D continúa con el método mostrado en la figura 8C. Aquí, de la primera la quinta posición en el dominio de código de barras se unieron mediante ácidos nucleicos complementarios que comprenden marcadores detectables o complejos indicadores y se han remplazado con ácidos nucleicos de hibridación que carecen de marcadores detectables. La sexta posición del dominio de código de barras se une actualmente mediante un ácido nucleico complementario que comprende un marcador detectable o complejo indicador, que identifica la sexta posición en el dominio de unión a la diana como unido a una guanina (G).

Como se menciona anteriormente, los ácidos nucleicos complementarios que comprenden marcadores detectables o complejos indicadores pueden retirarse de las regiones de fijación, pero no remplazarse con ácido nucleico de hibridación que carece de marcadores detectables.

Si se necesita, puede acelerarse la tasa de intercambio de marcador detectable incorporando oligonucleótidos monocatenarios pequeños que aceleran la tasa de intercambio de marcadores detectables (por ejemplo, "Toe-Hold" Probes; véase, por ejemplo, Seeling *et al.*, "Catalyzed Relaxation of a Metastable DNA Fuel"; J. Am. Chem. Soc. 2006, 128(37), pág. 12211-12220).

Es posible remplazar los ácidos nucleicos complementarios o complejos indicadores en una posición final en un dominio de código de barras (la sexta posición en la figura 8D); sin embargo, esto puede ser innecesario cuando tiene que remplazarse una sonda de secuenciación con otra sonda de secuenciación. De hecho, la sonda de secuenciación de la figura 8D ahora puede deshibridarse y retirarse del ácido nucleico diana y remplazarse con una segunda (solapante o no solapante) sonda de secuenciación que aún no se ha unido mediante cualquier ácido nucleico complementario, como se muestra en la figura 8E. La sonda en la figura 8E puede incluirse en una segunda población de sondas.

Como las figuras 8A a 8E, las figuras 9A y 9D a 9G muestran etapas del método de la presente divulgación; sin embargo, las figuras 9A y 9D a 9G muestran claramente que los complejos indicadores (que comprenden marcadores detectables) se unen a regiones de fijación de sondas de secuenciación. Las figuras 9D y 9E muestran señales fluorescentes emitidas desde sondas hibridadas con complejos indicadores. Las figuras 9D y 9E muestran que el ácido nucleico diana tiene una secuencia de "T-A".

La figura 10 resume las etapas mostradas en las figuras 9D y 9E. En la parte superior de la figura se muestra la secuencia de nucleótidos de una sonda ejemplar e identifica dominios significativos de la sonda. La sonda incluye un espaciador de ADN bicatenario opcional entre su dominio de unión a la diana y su dominio de código de barras. El dominio de código de barras comprende, en orden, una parte "Flanc. 1", una parte "RF-1", una parte "RF-1/Flanc. 2", una parte "RF-2" y una parte "RF-2/Flanc. 3". En la etapa 1, la "RF-1 Detec." se hibrida con las partes "RF-1" y "RF-1/Flanc. 2" de la sonda. "RF-1 Detec." corresponde a un complejo indicador o ácido nucleico complementario que comprende un marcador detectable que codifica una timidina en la primera posición. Por tanto, la etapa 1 corresponde a la figura 9D. En la etapa 2, la "Car. 1" se hibrida con las partes "Flanc. 1" y "RF-1" de la sonda. "Car. 1" corresponde al ácido nucleico de hibridación que carece de un marcador detectable que es específico para la primera región de fijación de la sonda (como se muestra en la figura 9E como una barra negra que cubre la primera región de fijación). Mediante hibridación en la posición "Flanc. 1", que está 5' al complejo indicador o ácido nucleico complementario, el ácido nucleico de hibridación desplaza de forma más eficaz el complejo indicador/ácido nucleico complementario de la sonda. Las partes "Flanc." también se muestran como "puntos de apoyo". En la etapa 3, la "RF-2 Detec." se hibrida con las partes "RF-2" y "RF-2/Flanc. 3" de la sonda. "RF-2 Detec." corresponde a un complejo indicador o ácido nucleico complementario que comprende un marcador detectable que codifica una guanina en la segunda posición. Por tanto, la etapa 3 corresponde a la figura 9E. En esta realización, el ácido nucleico de hibridación que carece de un marcador detectable y los ácidos nucleicos complementarios que comprenden marcadores detectables/complejos indicadores se proporcionan secuencialmente.

Como alternativa, el ácido nucleico de hibridación que carece de un marcador detectable y los ácidos nucleicos complementarios que comprenden marcadores detectables/complejos indicadores se proporcionan simultáneamente. Esta realización alternativa se muestra en la figura 11. En la etapa 2, la "Car. 1" (ácido nucleico de hibridación que carece de un marcador detectable) se proporciona junto con la "RF-2 Detec." (complejo indicador que codifica una

guanina en la segunda posición). Esta realización alternativa puede ser más rápida que la realización ilustrada en la figura 10, porque combina dos etapas en una.

La figura 12 ilustra los métodos de la presente divulgación. Aquí, se captura un ácido nucleico diana y se inmoviliza en dos posiciones, produciendo de ese modo un "hueco" en que puede unirse una sonda. Se hibrida una primera población de sondas sobre el ácido nucleico diana y se detectan los marcadores detectables. Las etapas iniciales se repiten con una segunda población de sondas, una tercera población de sondas, hasta más de 100 poblaciones de sondas. El uso de aproximadamente 100 poblaciones de sondas proporciona una cobertura aproximadamente 5X de cada nucleótido en un ácido nucleico diana. La figura 12 proporciona tasas estimadas de tiempos de lectura basados en el tiempo requerido para detectar señales de un campo de visión (FOV - *Field of View*).

La distribución de sondas a lo largo de la longitud de un ácido nucleico diana es crucial para la resolución de la señal detectable. Como se analiza anteriormente, el límite de resolución para dos marcadores detectables es de aproximadamente 600 nucleótidos. Preferiblemente, cada sonda de secuenciación en una población de sondas se unirá a no más cerca de 600 nucleótidos entre sí. Como se analiza anteriormente, 600 nucleótidos es el límite de resolución de un aparato de secuenciación típico. En este caso, una sonda de secuenciación proporcionará una lectura individual; esto se muestra en la figura 12 en la mancha limitada por resolución que está más a la izquierda.

Aleatoriamente, pero en parte dependiendo de la longitud del dominio de unión a la diana, la T_m de las sondas, y la concentración de las sondas aplicadas, es posible que dos sondas de secuenciación distintas en una población se unan en 600 nucleótidos entre sí. En este caso, se emitirán múltiples lecturas desorganizadas desde una mancha limitada por resolución individual; esto se muestra en la figura 12 en la segunda mancha limitada por resolución.

Como alternativa o adicionalmente, puede reducirse la concentración de las sondas de secuenciación en una población para disminuir la cobertura de las sondas en una región específica de un ácido nucleico diana, por ejemplo, hasta por encima del límite de resolución del aparato de secuenciación, produciendo de ese modo una lectura individual desde una mancha limitada por resolución.

La figura 23 muestra un esquema de una sonda de secuenciación distinta de la usada en las figuras 8 a 12. Aquí, cada posición en un dominio de código de barras se une mediante ácidos nucleicos complementarios que comprenden marcadores detectables o mediante complejos indicadores. Por tanto, en este ejemplo, una secuencia de seis nucleótidos puede leerse sin necesidad de remplazar secuencialmente los ácidos nucleicos complementarios. El uso de esta sonda de secuenciación reduciría el tiempo para obtener información de secuencia ya que se omiten muchas etapas del método descrito. Sin embargo, esta sonda se beneficiaría de los marcadores detectables que no son solapantes, por ejemplo, los fluoróforos se excitan mediante longitudes de onda de luz no solapantes o los fluoróforos emiten longitudes de onda de luz no solapantes.

El método comprende además etapas de ensamblar cada orden lineal identificado de nucleótidos para cada región del ácido nucleico diana inmovilizado, identificando de ese modo una secuencia para el ácido nucleico diana inmovilizado. Las etapas de ensamblaje usan un medio de almacenamiento legible en ordenador no transitorio con un programa ejecutable almacenado en el mismo. El programa envía una instrucción al microprocesador para colocar cada orden lineal identificado de nucleótidos para cada región del ácido nucleico diana, obteniendo de ese modo la secuencia del ácido nucleico. El ensamblaje puede producirse "de manera instantánea", es decir, mientras se están recogiendo los datos de las sondas de secuenciación en lugar de después de haber recogido todos los datos.

Cualquiera de los aspectos y realizaciones anteriores puede combinarse con cualquier otro aspecto o realización como se divulga en las secciones de compendio y/o descripción detallada.

Definiciones:

En determinadas realizaciones ejemplares, los términos "asociación" e "hibridación", como se usan en la presente memoria, se usan indistintamente para indicar la formación de una estructura bicatenaria estable. En un aspecto, estructura bicatenaria estable significa que no se destruye una estructura bicatenaria mediante un lavado riguroso en condiciones tales como una temperatura de aproximadamente 5 °C por debajo o aproximadamente 5 °C por encima de la T_m de una hebra de la estructura bicatenaria y baja concentración de sal monovalente, por ejemplo, menos de 0,2 M, o menos de 0,1 M o concentraciones de sal conocidas por los expertos en la materia. La expresión "emparejado perfectamente", cuando se usa en referencia a una estructura bicatenaria significa que las hebras polinucleotídicas y/u oligonucleotídicas que componen la estructura bicatenaria forman una estructura de dos hebras entre sí de modo que cada nucleótido en cada hebra experimenta emparejamiento de bases de Watson y Crick con un nucleótido en la otra hebra. La expresión "estructura bicatenaria" comprende, aunque sin limitación, el emparejamiento de análogos nucleosídicos, tales como desoxiinosina, nucleósidos con bases de 2-aminopurina, APN y similares, que pueden emplearse. Un "emparejamiento incorrecto" en una estructura bicatenaria entre dos oligonucleótidos significa que un par de nucleótidos en la estructura bicatenaria no logra experimentar unión de Watson y Crick.

Como se usa en la presente memoria, la expresión "condiciones de hibridación", típicamente incluirá concentraciones de sal de menos de aproximadamente 1 M, más habitualmente menos de aproximadamente 500 mM e incluso más habitualmente menos de aproximadamente 200 mM. Las temperaturas de hibridación pueden ser tan bajas como 5 °C, pero típicamente son mayores de 22 °C, más típicamente mayores de aproximadamente 30 °C, y a menudo en exceso

de aproximadamente 37 °C. Las hibridaciones se realizan habitualmente en condiciones rigurosas, por ejemplo, condiciones en las que una sonda hibridará específicamente con su subsecuencia diana. Las condiciones rigurosas son dependientes de la secuencia y son diferentes en diferentes circunstancias. Fragmentos más largos pueden requerir mayores temperaturas de hibridación para hibridación específica. Como otros factores pueden afectar a la rigurosidad de la hibridación, incluyendo la composición de bases y la longitud de las hebras complementarias, la presencia de disolventes orgánicos y el grado de emparejamiento incorrecto de bases, la combinación de parámetros es más importante que la medida absoluta de uno cualquiera en solitario.

En general, las condiciones rigurosas se seleccionan para que sean aproximadamente 5 °C inferiores a la T_m para la secuencia específica a una fuerza iónica y pH definidos. Condiciones rigurosas ejemplares incluyen concentración de sal de al menos 0,01 M a no más de 1 M de concentraciones de iones de Na (u otras sales) a un pH 7,0 a 8,3 y a temperatura de al menos 25 °C. Por ejemplo, condiciones de SSPE 5X (NaCl 750 mM, fosfato de Na 50 mM, EDTA 5 mM, pH 7,4) y una temperatura de 25-30 °C son adecuadas para hibridaciones de sondas específicas de alelo. Para condiciones rigurosas, véase, por ejemplo, Sambrook, Fritsche y Maniatis, "Molecular Cloning A Laboratory Manual, 2.^a Ed." Cold Spring Harbor Press (1989) y Anderson Nucleic Acid Hybridization, 1.^a Ed., BIOS Scientific Publishers Limited (1999). Como se usa en la presente memoria, las expresiones "que hibrida específicamente con" o "que específicamente hibrida con" o expresiones similares se refieren a la unión, formación de estructura bicatenaria o hibridación de una molécula sustancialmente con una secuencia o secuencias de nucleótidos particulares en condiciones rigurosas.

Los marcadores detectables asociados con una posición particular de una sonda pueden "leerse" (por ejemplo, detectarse su fluorescencia) una vez o múltiples veces; una "lectura" puede ser sinónima de la expresión "calificación de bases". Múltiples lecturas mejoran la precisión. Una secuencia de ácido nucleico diana se "lee" cuando se detecta un tramo contiguo de información de secuencia derivado de una molécula diana original individual; típicamente, esto se genera mediante consenso multipase (como se define a continuación). Como se usa en la presente memoria, la expresión "cobertura" o "profundidad de cobertura" se refiere al número de veces que una región de la diana se ha secuenciado (mediante lecturas diferenciadas) y alineado con una secuencia de referencia. La cobertura de lectura es el número total de lecturas que cartografiaban una secuencia diana de referencia específica; la cobertura de bases es el número total de calificaciones de bases realizadas en una posición genómica específica.

Como se usa en la presente memoria, un "ciclo de hib. y sec." se refiere a todas las etapas requeridas para detectar cada región de fijación en una sonda o población de sondas particular. Por ejemplo, para una sonda que puede detectar seis posiciones en un ácido nucleico diana, un "ciclo de hib. y sec." incluirá, al menos, hibridar la sonda con el ácido nucleico diana, hibridar los ácidos nucleicos complementarios/complejos indicadores con la región de fijación en cada una de las seis posiciones en el dominio de código de barras de la sonda, y detectar los marcadores detectables asociados con cada una de las seis posiciones.

La expresión "sonda k-mérica" es sinónima de una sonda de la presente divulgación.

Cuando se alinean dos o más secuencias de lecturas diferenciadas, las partes solapantes pueden combinarse para crear una única secuencia consenso. En posiciones donde las partes solapantes tienen la misma base (una única columna de la alineación), esas bases se convierten en el consenso. Pueden usarse diversas normas para generar el consenso para posiciones donde hay desacuerdos entre secuencias solapantes. Una norma de mayoría simple usa la base más común en la columna como consenso. Un "consenso multipase" es una alineación de todas las lecturas de sonda diferenciadas desde una sola molécula diana. Dependiendo del número total de ciclos de poblaciones/combinaciones de sondas aplicadas, cada posición de base dentro de moléculas diana individuales puede consultarse con diferentes niveles de redundancia o solapamiento; en general, la redundancia aumenta el nivel de confianza de una calificación de bases.

La "precisión bruta" es una medida de la capacidad inherente del sistema de identificar correctamente una base. La precisión bruta depende de la tecnología de secuenciación. La "precisión consenso" es una medida de la capacidad del sistema de identificar correctamente una base con el uso de lecturas adicionales y la capacidad estadística. "Especificidad" se refiere al porcentaje de lecturas que cartografiaban las dianas pretendidas de las lecturas totales por ejecución. "Uniformidad" se refiere a la variabilidad en la cobertura de secuencia entre las regiones diana; una alta uniformidad se correlaciona con una baja variabilidad. Este rasgo característico se presenta comúnmente como la fracción de regiones diana cubiertas por $\geq 20\%$ de la profundidad de cobertura promedio entre todas las regiones diana. Los errores estocásticos (es decir, errores intrínsecos de la química de secuenciación) pueden corregirse fácilmente con secuenciación de "multipase" del mismo ácido nucleico diana; dado un número suficiente de pases, puede conseguirse secuenciación sustancialmente de "consenso perfecto" o "sin errores".

Los métodos descritos en la presente memoria pueden implementarse y/o los resultados registrarse usando cualquier dispositivo que pueda implementar los métodos y/o registrar los resultados. Ejemplos de dispositivos que pueden usarse incluyen, aunque sin limitación, dispositivos de computación electrónicos, incluyendo ordenadores de todo tipo. Cuando los métodos descritos en la presente memoria se implementan y/o registran en un ordenador, el programa informático que puede usarse para configurar el ordenador para realizar las etapas de los métodos puede estar contenido en cualquier medio legible por ordenador que pueda contener el programa informático. Ejemplos de medios legibles por ordenador que pueden usarse incluyen, aunque sin limitación, disquetes, CD-ROM, DVD, ROM, RAM,

medios legibles por ordenador no transitorios y otros dispositivos de memoria y almacenamiento en ordenador. El programa informático que puede usarse para configurar el ordenador para realizar las etapas de los métodos, ensamblar la información de secuencia y/o registrar los resultados también puede proporcionarse en una red electrónica, por ejemplo, en Internet, una red interna u otra red.

Puede incorporarse una "tarjeta de secuenciación consumible" (figura 24) en un dispositivo de imágenes de fluorescencia conocido en la técnica. Cualquier microscopio de fluorescencia con varios elementos variables puede realizar esta lectura de secuenciación. Por ejemplo: puede usarse lámpara de campo amplio, láser, LED, multifotón, confocal o iluminación de reflejo interno-total para excitación y/o detección. Es posible una cámara (individual o múltiple) y/o tubo fotomultiplicador (individual o múltiple) con resolución espectral basada en filtro o basada en gradilla (una o más longitudes de onda de emisión resueltas espectralmente) en el canal de emisión-detección del microscopio de fluorescencia. Los ordenadores convencionales pueden controlar tanto la tarjeta de secuenciación consumible, los reactivos que fluyen a través de la tarjeta como la detección por el microscopio de fluorescencia.

Los datos de secuenciación pueden analizarse por muchísimos ensambladores convencionales de secuenciación de última generación (véase, por ejemplo, Wajid y Serpedin, "Review of general algorithmic features for genome assemblers for next generation sequencers" Genomics, proteomics & bioinformatics, 10 (2), 58-73, 2012). Los datos de secuenciación obtenidos dentro de una sola región limitada por difracción del microscopio se "ensambla localmente" para generar una secuencia consenso a partir de las múltiples lecturas dentro de una mancha de difracción. Las múltiples lecturas ensambladas en la mancha de difracción se cartografían entonces conjuntamente para generar secuencias contiguas que representan el conjunto génico diana completo, o un ensamblaje *de novo* de uno o más genomas completos.

Se describe contenido adicional pertinente a la presente invención en uno o más de los siguientes documentos: U.S. 8.148.512, U.S. 7.473.767, U.S. 7.919.237, U.S. 7.941.279, U.S. 8.415.102, U.S. 8.492.094, U.S. 8.519.115, U.S. 2009/0220978, U.S. 2009/0299640, U.S. 2010/0015607, U.S. 2010/0261026, U.S. 2011/0086774, U.S. 2011/0145176, U.S. 2011/0201515, U.S. 2011/0229888, U.S. 2013/0004482, U.S. 2013/0017971, U.S. 2013/0178372, U.S. 2013/0230851, U.S. 2013/0337444, U.S. 2013/0345161, U.S. 2014/0005067, U.S. 2014/0017688, U.S. 2014/0037620, U.S. 2014/0087959, U.S. 2014/0154681 y U.S. 2014/0162251.

Ejemplos

Ejemplo 1: El presente método de secuenciación de un ácido nucleico diana es rápido

A continuación se describe la cronología para las etapas en los métodos de la presente divulgación y como se muestra en las figuras 8 a 12.

La presente invención requiere preparación mínima de las muestras. Por ejemplo, como se muestra en la figura 13, los ácidos nucleicos en una muestra pueden empezar a leerse después de 2 horas o menos de tiempo de preparación; esto es significativamente menos tiempo que el requerido para secuenciación con Ion Torrent (AmpliSeq™) o Illumina (TruSight) que, respectivamente, requieren aproximadamente 12 o 9 horas de tiempo de preparación.

Los cálculos para una ejecución ejemplar se muestran en la figura 14 y los cálculos para los tiempos de ciclado se muestran en la figura 15.

La unión de una población de sondas a un ácido nucleico diana inmovilizado tarda aproximadamente sesenta segundos. Esta reacción puede acelerarse utilizando múltiples copias del dominio de unión a la diana en la cadena principal sintética. Con dispositivo de intercambio de fluidos controlado por microfluidos, retirar por lavado las sondas soltadas tarda aproximadamente medio segundo.

Añadir una primera combinación de ácidos nucleicos complementarios (que comprenden un marcador detectable) y unirlos a regiones de fijación en la primera posición del dominio de código de barras tarda aproximadamente quince segundos.

Se toman imágenes de cada campo de visión (FOV) para cuatro colores diferentes, representando cada color una sola base. Las manchas de referencia colocadas en una sonda de captura 5' o sonda de captura 3' (o ambas) pueden ser de ayuda para leer solamente esos códigos de barras ópticos en una línea (coherentes con la presencia de ácido nucleico diana con huecos) entre las dos localizaciones. Las manchas de referencia también pueden añadirse a cada campo de visión para generar una alineación equitativa de imágenes tras etapas sucesivas en el proceso de secuenciación. Las cuatro imágenes pueden obtenerse en un solo FOV y después el dispositivo de lectura óptica puede moverse a un nuevo FOV, o captar todo el FOV en un color y después tomar imágenes en un segundo color. Un solo FOV puede leerse en aproximadamente medio segundo. Tarda aproximadamente medio segundo en moverse a un siguiente FOV. Por lo tanto, el tiempo hasta la lectura de "n" FOV es igual a "n" multiplicado por 1 segundo).

Los ácidos nucleicos complementarios que tienen marcadores detectables se retiran de la primera posición del dominio de código de barras mediante la adición de calor o lavado con exceso de ácidos nucleicos complementarios que carecen marcadores detectables. Si se necesita, puede acelerarse la tasa de intercambio de marcador detectable incorporando oligonucleótidos monocatenarios pequeños que aceleran la tasa de intercambio de marcadores

detectables (por ejemplo, "Toe-Hold" Probes; véase, por ejemplo, Seeling *et al.*, "Catalyzed Relaxation of a Metastable DNA Fuel"; J. Am. Chem. Soc. 2006, 128(37), pág. 12211-12220). Pueden volver a tomarse imágenes de un FOV para confirmar que todos los ácidos nucleicos complementarios que tienen marcadores detectables se retiran antes de continuar el movimiento. Esto tarda aproximadamente quince segundos. Esta etapa puede repetirse hasta que se alcanzan niveles de señal de fondo.

Las etapas anteriores se repiten para las posiciones restantes del dominio de código de barras de las sondas.

El tiempo total hasta la lectura es igual a m (bases leídas) multiplicado por $(15 \text{ s} + n \text{ FOV multiplicado por } 1 \text{ s} + 15 \text{ s})$. Por ejemplo, cuando el número de posiciones en el dominio de código de barras es 6 y 20 FOV, el tiempo hasta la lectura es igual a $6 \times (30 + 20 + 15)$ o 390 segundos.

Las sondas de la primera población se deshidridan. Esto tarda aproximadamente sesenta segundos.

Las etapas anteriores se repiten para la segunda y posteriores poblaciones de sondas. Si las poblaciones de sondas de secuenciación se organizan por la temperatura de fusión (T_m), cada población de sondas requerirá múltiples hibridaciones para garantizar que cada base queda cubierta a la profundidad requerida (esto está motivado por la tasa de error). Además, analizando las lecturas de hibridación durante una ejecución, es posible reconocer cada gen individual que se está secuenciando mucho antes de que la secuencia completa se haya determinado realmente. Por tanto, el ciclado puede repetirse hasta que se cumpla una frecuencia de error (o cobertura) deseada particular.

Usando la cronología descrita anteriormente, junto con algunas estimaciones de densidad de unión de ácido nucleico con huecos, puede estimarse el rendimiento de un secuenciador Nanosting (NSTG)-Next Generation de la presente invención.

El rendimiento neto del secuenciador se da por:

Ocupación de bases fraccionada \times longitud de hueco \times número de huecos por FOV \times número de bases por código de barras óptico / $[60 \text{ s (sondas que hibridan con ácido nucleico diana)} + 0,5 \text{ s (lavado)} + m: \text{posiciones en el dominio de código de barras} \times (15 \text{ s (ácidos nucleicos complementarios que se unen)} + n_{\text{fovs}} \times 1 + 15 \text{ s (ácidos nucleicos complementarios que se sueltan)}) + 60 \text{ s (sondas que se deshidridan de ácido nucleico diana)}]$

Por lo tanto, en un ejemplo, un "ciclo" total para un ácido nucleico con huecos individual (que se añade conjuntamente desde el método mostrado en la figura 10):

$60 \text{ s (sondas que hibridan con ácido nucleico diana)} + 0,5 \text{ s (lavado)} + m\text{-bases} \times (15 \text{ s (ácidos nucleicos complementarios que se unen)} + n_{\text{FOVs}} \text{ multiplicado por } 1 + 15 \text{ s (ácidos nucleicos complementarios que se sueltan)}) + 60 \text{ s (sondas que se deshidridan de ácido nucleico diana)}$. Usando $m = 6$, $n_{\text{FOVs}} = 20$, tiempo de producción = $60 + 0,5 + 390 + 60 = 510,5 \text{ s}$.

Suponiendo: un 1 % de ocupación de la región de ácido nucleico con huecos, 4000 bases por hueco y 5000 fragmentos de ácido nucleico con huecos por FOV y una m de 6 y n_{FOVs} de 20 (como se describe anteriormente) se produce un rendimiento neto de:

$0,01 \times 4000 \times 5000 \times 20 = 4\,000\,000$ lecturas de 6 bases por $510,5 \text{ s} = 47\,012,73$ bases/s.

Por lo tanto, en este ejemplo, un rendimiento neto por 24 horas de medición continua = 4,062 gigabases (Gb) por día. Como alternativa se estiman hasta 12 Gb por día. Véase la figura 12.

Como se muestra en la figura 14, el tiempo de ejecución requerido para secuenciar 100 ácidos nucleicos diana diferentes ("céntuple") es de aproximadamente 4,6 horas; el tiempo de ejecución requerido para secuenciar 1000 ácidos nucleicos diana diferentes ("múltiple de 1000") es aproximadamente 16 horas.

La figura 16 compara la tasa de secuenciación, el número de lecturas y la utilidad clínica para la presente divulgación y otros diversos métodos/aparatos de secuenciación.

Ejemplo 2: El presente método tiene una baja tasa de error

La figura 17 muestra que la presente divulgación tiene una tasa de error bruto de aproximadamente un 2,1 %, cuando se omiten las posiciones terminales.

Para la divulgación, una tasa de error asociada con la secuenciación está relacionada con la diferencia de energía libre entre un $(m+n)$ -mero completamente emparejado y un $(m-1+n)$ -mero con emparejamiento incorrecto de una sola base. La suma de $m+n$ es el número de nucleótidos en un dominio de unión a la diana y m representa el número de posiciones en un dominio de código de barras. Una estimación de la selectividad de hibridación puede hacerse usando la ecuación (véase, Owczarzy, R. (2005), Biophys. Chem., 117:207-215 y el sitio web de Integrated DNA Technologies: www.idtdna.com/analyzer/Applications/Instructions/Default.aspx?AnalyzerDefinitions=true#MismatchMeltTemp);

$$\theta = 1 - \left(\frac{K_a([hebra2] - [hebra1]) - 1}{2K_a[hebra2]} + \frac{\sqrt{K_a^2([hebra1] - [hebra2])^2 + 2K_a([hebra1] + [hebra2]) + 1}}{2K_a[hebra2]} \right)$$

donde K_a es la constante en equilibrio de asociación obtenida de parámetros termodinámicos previstos,

$$K_a = \exp \left(\frac{-(\Delta H^\circ - T\Delta S^\circ)}{TA} \right)$$

Teta representa el porcentaje unido del complemento exacto y las secuencias con emparejamiento incorrecto de una sola base, que se espera que hibriden con la diana a la temperatura de hibridación especificada. La T es la temperatura de hibridación en grados Kelvin, ΔH° (entalpía) y ΔS° (entropía) son los parámetros de fusión calculados a partir de la secuencia y los parámetros termodinámicos del vecino más cercano publicados, R es la constante de gas ideal (1,987 cal·K⁻¹·mol⁻¹), $[hebra1/2]$ es la concentración molar de un oligonucleótido, y la constante de -273,15 convierte la temperatura de grados Kelvin en grados Celsius. Los parámetros del vecino más cercano más precisos se obtuvieron de las siguientes publicaciones para pares de bases de ADN/ADN (véase, Allawi, H., Santa Lucia, J. Biochemistry, 36, 10581), pares de bases de ARN/ADN (véase, Sugimoto *et al.*, Biochemistry, 34, 11211-6), pares de bases de ARN/ARN (véase, Xia, T. *et al.*, Biochemistry, 37, 14719).

Como ejemplo de una estimación de la tasa de error aproximada esperada del secuenciador NSTG como sigue. Para (m + n) igual a octámero. Considerando el siguiente código de barras octamérico y su emparejamiento incorrecto de una sola base.

5'ATCGTACG3'

(región a la secuencia)

3'TAGCATGC5'

(código de barras óptico de secuenciación con emparejamiento perfecto)

3'TAGTATGC5'

(código de barras óptico de secuenciación con un par de emparejamiento incorrecto de una sola base (G-T))

Usando la calculadora de IDT basándose en las ecuaciones anteriores, produce:

A 17,4 °C (la Tm del caso de emparejamiento perfecto), (50 % / 0,3 %) sería la relación del código de barras óptico correcto hibridado con esa secuencia frente al código de barras incorrecto a la Tm, produciendo una tasa de error estimada para esa secuencia en un 0,6 %.

Un cálculo de secuenciación con contenido de GC muy alto producir:

5'CGCCGGCC3'

(región a la secuencia)

3'GCGGCCGG5'

(código de barras óptico de secuenciación con emparejamiento perfecto)

3'GCGGACGG5'

(código de barras óptico de secuenciación con un par desemparejado de emparejamiento incorrecto de una sola base (G-A))

A 41,9 °C (la Tm del caso de emparejamiento perfecto), (50 % / 0,4%) sería la relación del código de barras óptico correcto hibridado con esa secuencia frente al código de barras incorrecto a la Tm, produciendo una tasa de error estimada para esa secuencia en un 0,8%.

El examen de varios pares de octámeros produce una distribución de tasas de error, en el intervalo de un 0,2 % a un 1 %. Aunque los cálculos anteriores no serán idénticos para las condiciones usadas, estos cálculos proporcionan una indicación de que el método de la presente divulgación tendrá una tasa de error intrínseca relativamente baja, en comparación con otras tecnologías de secuenciación de una sola molécula, tales como Pacific Biosciences y Oxford Nanopore Technologies donde las tasas de error pueden ser significativas (>> 10 %).

La figura 18 demuestra que la precisión bruta de la presente divulgación es mayor que en otros métodos de secuenciación. Por tanto, la presente divulgación proporciona una secuencia consenso a partir de una sola diana después de menos pases que los requeridos para otros métodos de secuenciación. Además, la presente divulgación puede obtener una secuenciación de "consenso perfecto"/"sin errores" (es decir, 99,9999 %/Q60) después de 30 o más pases mientras que los métodos de secuenciación PacBio (por ejemplo) no pueden conseguir dicho consenso después de 70 pases.

Ejemplo 3: La presente divulgación tiene capacidad de resolución de un solo par de bases

La figura 19 muestra que la presente divulgación tiene resolución de una sola base y con tasas de error bajas (que varían de un 0 % a un 1,5 % dependiendo de una sustitución nucleotídica específica).

- Se realizaron experimentos adicionales usando un ARN diana hibridado con código de barras e inmovilizado en la superficie del cartucho usando tecnología de unión de expresión génica NanoString normal (véase, por ejemplo, Geiss *et al.*, "Direct multiplexed measurement of gene expression with color-coded probe pairs"; Nature Biotechnology, 26, 317 - 325 (2008)). Se midió la capacidad de un código de barras con diferente longitud del dominio de unión a la diana y con un emparejamiento perfecto (código de barras óptico YGBYGR-2 um conectado a una secuencia de emparejamiento deca mérico perfecto) de hibridar con una diana de ARN (figura 26). Una longitud más larga del dominio de unión a la diana da mayores recuentos. También se muestra que el dominio de unión a la diana deca mérico es suficiente para registrar la secuencia por encima del fondo. Cada uno de los emparejamientos alterados de una sola base individuales se sintetizó con códigos de barras ópticos alternativos. Se contó la relación de códigos de barras ópticos correctos con respecto a incorrectos (figuras 24 y 25).
- La capacidad de decámeros de detectar un SNP en la secuencia real es >15 000 recuentos sobre el fondo, mientras que las secuencias incorrectas son como mucho >400 sobre el fondo. En presencia de sonda correcta, se espera que las tasas de error sean <3 % de la secuencia real. Obsérvese que estos datos son (en esencia) un escenario del peor caso. Se tiene solamente una secuencia de hibridación de 10 pares de bases fijada a un indicador de código de barras óptico de 6,6 kilobases (Gen2 style). No se realizaron optimizaciones de condiciones específicas. Estos datos, sin embargo, revelan que el procedimiento de secuenciación NanoString Next-Generation puede resolver pares de bases individuales de secuencia.

Los materiales y métodos detallados utilizados en el estudio anterior son como sigue:

Protocolo de hibridación de sonda B más conjunto de códigos

- Recoger 25 ul de elementos (conjunto de códigos 194)
 - Añadir 5 ul de sonda B + secuencia complementaria a la diana (100 uM)
 - Añadir 15 ul de tampón hib. (SSPE 14,56x Tween 20 al 0,18 %)
- SSPE (NaCl 150 mM, NaH₂PO₄×H₂O 10 mM, Na₂EDTA 10 mM)
- Incubar en hielo durante 10 min
 - Añadir 150 ul de microesferas G (40 ul de microesferas G a 10 mg/ml más 110 ul de SSPE 5x Tween 20 al 0,1 %)
 - Incubar durante 10 min a TA
 - Lavar tres veces con SSPE 0,1x Tween 20 al 0,1 % usando colector de imanes
 - Eluir en 100 ul de SSPE 0,1x durante 10 min a 45 °C.

Protocolo de hibridación con la diana (NaCl 750 mM)

- Recoger 20 ul por encima de la muestra eluida
- Añadir 10 ul de tampón hib.
- Añadir 1 ul de diana (ARN biotinilado 100 nM)
- Incubar en hielo durante 30 min

Recoger 15 ul y unir a portaobjetos de estreptavidina durante 20 min, estiramiento fluido con ganchos en G, recuento usando nCounter

45 Materiales

Conjunto de códigos 194 de elementos

ES 2 910 099 T3

Oligos adquiridos de IDT

SSPE (NaCl 150 mM, NaH₂PO₄×H₂O 10 mM, Na₂EDTA 10 mM)

Tampón hib. (SSPE 14,56x Tween 20 al 0,18 %)

Tabla 2: Secuencias de sonda B para oligómeros de 12, 11, ..., 8 monómeros. (SEQ ID NO: 30 a SEQ ID NO: 34)

GBRYBG	5	GACTGTACCCACGCGATGACGTTTCGTCAAGAGTCGCATAATCT	3
YRBYRG	5	AGACTGTACCACAAGAATCCCTGCTAGCTGAAGGAGGGTCAAAC	3
YGBYGR	5	GAGACTGTACCCTACGTATATATCCAAGTGGTTATGTCCGACGGC	3
GBRYGB	5	TGAGACTGTACCACCCCTCCAAACGCATTCTTATTGGCAAATGGAA	3
RYGBRG	5	CTGAGACTGTACCCGGGAATCGGCATTTTCGCATTCTTAGGATCTAAA	3

Tabla 3: Secuencia diana (en **negrita**; SEQ ID NO: 35)

ARN	5	CAATGTGAGTCTCTTGGT ACAGTCTCAGT TAGTCACTCCCTAAG\ Bio TEG\ 	3
-----	---	--	---

Tabla 4: Secuencias de sonda B para emparejamiento incorrectos decaméricos (en **negrita**: SEQ ID NO: 36 a SEQ ID NO: 41)

10mermis2A	GAGAC AG TACCC TGGTCTAGGTATCTAA TTCTGGTGGGTGGGTACT
10mermis2C	GAGAC CG TACCGCTCATTTTGAACATACGA TTGCGATTACGGAAA
10mermis2G	GAGAC GG TACCTTAAAGCTATCCACGA TGTCAAAAATGTGGTTT
10mermis1G	GAG AG TGTACCCAA TGCTTGCAGTATGTATCCTGATCGTGGGTGC
10mermis1A	GAG A TGTACCCCTCATACCA ATGTAAAGTATAGTTAACGCCCTGT
10mermis1T	GAG ATT GTACCC TACATATATAGGAAAAGGGAAGGTAGAAGAGCT

LISTA DE SECUENCIAS

<110> NanoString Technologies, Inc. BEECHEM, Joseph KHAFIZOV, Rustem

<120> SECUENCIACIÓN SIN ENZIMAS Y SIN AMPLIFICACIÓN

<130> NATE-025/001

5 <150> US 62/082.883

<151> 21/11/2014

<160> 104

<170> PatentIn versión 3.5

<210> 1

10 <211> 10

<212> ADN

<213> Secuencia artificial

<220>

<223> Polinucleótido sintético

15 <400> 1

atacatctag 10

<210> 2

<211> 10

<212> ADN

20 <213> Secuencia artificial

<220>

<223> Polinucleótido sintético

<400> 2

gatctacata 10

25 <210> 3

<211> 10

<212> ADN

<213> Secuencia artificial

<220>

30 <223> Polinucleótido sintético

<400> 3

ttaggtaaag 10

<210> 4

<211> 10

35 <212> ADN

<213> Secuencia artificial

<220>

<223> Polinucleótido sintético
 <400> 4
 tcttcattac 10
 <210> 5
 5 <211> 10
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Polinucleótido sintético
 10 <400> 5
 atgaatctac 10
 <210> 6
 <211> 10
 <212> ADN
 15 <213> Secuencia artificial
 <220>
 <223> Polinucleótido sintético
 <400> 6
 tcaatgtatg 10
 20 <210> 7
 <211> 10
 <212> ADN
 <213> Secuencia artificial
 <220>
 25 <223> Polinucleótido sintético
 <400> 7
 aattgagtac 10
 <210> 8
 <211> 10
 30 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Polinucleótido sintético
 <400> 8
 35 atgttaatgg 10
 <210> 9
 <211> 10

<212> ADN
 <213> Secuencia artificial
 <220>
 <223> Polinucleótido sintético
 5 <400> 9
 aattaggatg 10
 <210> 10<211> 10
 <212> ADN
 <213> Secuencia artificial
 10 <220>
 <223> Polinucleótido sintético
 <400> 10
 ataatggatc 10
 <210> 11
 15 <211> 10
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Polinucleótido sintético
 20 <400> 11
 taataaggtg 10
 <210> 12
 <211> 10
 <212> ADN
 25 <213> Secuencia artificial
 <220>
 <223> Polinucleótido sintético
 <400> 12
 tagttagagc 10
 30 <210> 13
 <211> 10
 <212> ADN
 <213> Secuencia artificial
 <220>
 35 <223> Polinucleótido sintético
 <400> 13
 atagagaagg 10

<210> 14
 <211> 10
 <212> ADN
 <213> Secuencia artificial
 5 <220>
 <223> Polinucleótido sintético
 <400> 14
 ttgatgatac 10
 <210> 15
 10 <211> 10
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Polinucleótido sintético
 15 <400> 15
 atagtgattc 10
 <210> 16
 <211> 10
 <212> ADN
 20 <213> Secuencia artificial
 <220>
 <223> Polinucleótido sintético
 <400> 16
 tataacgatg 10
 25 <210> 17
 <211> 10
 <212> ADN
 <213> Secuencia artificial
 <220>
 30 <223> Polinucleótido sintético
 <400> 17
 ttaagttag 10
 <210> 18
 <211> 10
 35 <212> ADN
 <213> Secuencia artificial
 <220>

<223> Polinucleótido sintético
 <400> 18
 atacgttatg 10
 <210> 19
 5 <211> 10
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Polinucleótido sintético
 10 <400> 19
 tgtactatag 10
 <210> 20
 <211> 10
 <212> ADN
 15 <213> Secuencia artificial
 <220>
 <223> Polinucleótido sintético
 <400> 20
 ttaacaagtg 10
 20 <210> 21
 <211> 10
 <212> ADN
 <213> Secuencia artificial
 <220>
 25 <223> Polinucleótido sintético
 <400> 21
 aactatgtac 10
 <210> 22
 <211> 10
 30 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Polinucleótido sintético
 <400> 22
 35 taactatgac 10
 <210> 23
 <211> 10

<212> ADN
 <213> Secuencia artificial
 <220>
 <223> Polinucleótido sintético
 5 <400> 23
 actaatgttc 10
 <210> 24
 <211> 10
 <212> ADN
 10 <213> Secuencia artificial
 <220>
 <223> Polinucleótido sintético
 <400> 24
 tcattgaatg 10
 15 <210> 25
 <211> 14
 <212> ADN
 <213> Secuencia artificial
 <220>
 20 <223> Polinucleótido sintético
 <400> 25
 ctgtctcatc tctt 14
 <210> 26
 <211> 26
 25 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Polinucleótido sintético
 <400> 26
 30 ctgtctcatc tcttgctgca tcctgt 26
 <210> 27
 <211> 38
 <212> ADN
 <213> Secuencia artificial
 35 <220>
 <223> Polinucleótido sintético
 <400> 27

ctgtctcatc tcttgctgca tcctgtcggg tcacgttg 38

<210> 28

<211> 36

<212> ADN

5 <213> Secuencia artificial

<220>

<223> Polinucleótido sintético

<400> 28

ctgtctcatc ttgctgcatc ctgtcgggtc acgttg 36

10 <210> 29

<211> 36

<212> ADN

<213> Secuencia artificial

<220>

15 <223> Polinucleótido sintético

<400> 29

ctgtctcatt ttgctgcatc ctgtccgttc acgttg 36

<210> 30

<211> 43

20 <212> ADN

<213> Secuencia artificial

<220>

<223> Polinucleótido sintético

<400> 30

25 gactgtaccc acgcgatgac gttcgtcaag agtcgcataa tct 43

<210> 31

<211> 44

<212> ADN

<213> Secuencia artificial

30 <220>

<223> Polinucleótido sintético

<400> 31

agactgtacc acaagaatcc ctgctagctg aaggagggtc aaac 44

<210> 32

35 <211> 45

<212> ADN

<213> Secuencia artificial

<220>
 <223> Polinucleótido sintético
 <400> 32
 gagactgtac cctacgtata tatccaagtg gttatgtccg acggc 45

5 <210> 33
 <211> 46
 <212> ADN
 <213> Secuencia artificial
 <220>

10 <223> Polinucleótido sintético
 <400> 33
 tgagactgta ccaccctcc aaacgcattc ttattggcaa atggaa 46
 <210> 34
 <211> 47

15 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Polinucleótido sintético
 <400> 34

20 ctgagactgt acccggaat cggcatttcg cattcttagg atctaaa 47
 <210> 35
 <211> 44
 <212> ADN
 <213> Secuencia artificial

25 <220>
 <223> Polinucleótido sintético
 <400> 35
 caatgtgagt ctcttggtac agtctcagtt agtcactccc taag 44
 <210> 36

30 <211> 45
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Polinucleótido sintético

35 <400> 36
 gagacagtac cctggcttag gtatctaatt cgtgggtcgg gtact 45
 <210> 37

<211> 45
 <212> ADN
 <213> Secuencia artificial
 <220>
 5 <223> Polinucleótido sintético
 <400> 37
 gagaccgtac cgctcattt gaacatacga ttgcgattac ggaaa 45
 <210> 38
 <211> 45
 10 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Polinucleótido sintético
 <400> 38
 15 gagacggtac cttaaagcta tccacgaatg tcaaaaatgt ggttt 45
 <210> 39
 <211> 45
 <212> ADN
 <213> Secuencia artificial
 20 <220>
 <223> Polinucleótido sintético
 <400> 39
 gagagtgtac ccaatgcttg cagtatgtat cctgatcgtg cgtgc 45
 <210> 40
 25 <211> 45
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Polinucleótido sintético
 30 <400> 40
 gagaatgtac cctcatacca atgtaaagta tagttaacgc cctgt 45
 <210> 41
 <211> 45
 <212> ADN
 35 <213> Secuencia artificial
 <220>
 <223> Polinucleótido sintético

<400> 41
gagattgtac cctacatata taggaaaagg gaaggtagaa gagct 45
<210> 42
<211> 12
5 <212> ADN
<213> Secuencia artificial
<220>
<223> Polinucleótido sintético
<400> 42
10 cacgaacgct ag 12
<210> 43
<211> 12
<212> ADN
<213> Secuencia artificial
15 <220>
<223> Polinucleótido sintético
<400> 43
catcgcatgc ct 12
<210> 44
20 <211> 12
<212> ADN
<213> Secuencia artificial
<220>
<223> Polinucleótido sintético
25 <400> 44
gtcatctcct ac 12
<210> 45
<211> 12
<212> ADN
30 <213> Secuencia artificial
<220>
<223> Polinucleótido sintético
<400> 45
gtcatccgct ac 12
35 <210> 46
<211> 12
<212> ADN

<213> Secuencia artificial
 <220>
 <223> Polinucleótido sintético
 <400> 46
 5 gtcacgcgact ac 12
 <210> 47
 <211> 12
 <212> ADN
 <213> Secuencia artificial
 10 <220>
 <223> Polinucleótido sintético
 <400> 47
 gtcacgttct ac 12
 <210> 48
 15 <211> 12
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Polinucleótido sintético
 20 <400> 48
 gtcacgcct ac 12
 <210> 49
 <211> 12
 <212> ADN
 25 <213> Secuencia artificial
 <220>
 <223> Polinucleótido sintético
 <400> 49
 gtcacgctc ac 12
 30 <210> 50
 <211> 12
 <212> ADN
 <213> Secuencia artificial
 <220>
 35 <223> Polinucleótido sintético
 <400> 50
 gtcacgttcg ac 12

<210> 51
 <211> 12
 <212> ADN
 <213> Secuencia artificial
 5 <220>
 <223> Polinucleótido sintético
 <400> 51
 gtcacaaact ac 12
 <210> 52
 10 <211> 12
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Polinucleótido sintético
 15 <400> 52
 gtcacacgta ac 12
 <210> 53
 <211> 12
 <212> ADN
 20 <213> Secuencia artificial
 <220>
 <223> Polinucleótido sintético
 <400> 53
 gtcacacgaa ac 12
 25 <210> 54
 <211> 12
 <212> ADN
 <213> Secuencia artificial
 <220>
 30 <223> Polinucleótido sintético
 <400> 54
 gtcacacaaa ac 12
 <210> 55
 <211> 12
 35 <212> ADN
 <213> Secuencia artificial
 <220>

<223> Polinucleótido sintético
 <400> 55
 gtcacattgc ac 12
 <210> 56
 5 <211> 12
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Polinucleótido sintético
 10 <400> 56
 gtcacattgc ct 12
 <210> 57
 <211> 12
 <212> ADN
 15 <213> Secuencia artificial
 <220>
 <223> Polinucleótido sintético
 <400> 57
 gtcacccgac ct 12
 20 <210> 58
 <211> 12
 <212> ADN
 <213> Secuencia artificial
 <220>
 25 <223> Polinucleótido sintético
 <400> 58
 ctttcacac ct 12
 <210> 59
 <211> 12
 30 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Polinucleótido sintético
 <400> 59
 35 ctttcctct ct 12
 <210> 60
 <211> 12

<212> ADN
 <213> Secuencia artificial
 <220>
 <223> Polinucleótido sintético
 5 <400> 60
 cttttcgact ct 12
 <210> 61
 <211> 12
 <212> ADN
 10 <213> Secuencia artificial
 <220>
 <223> Polinucleótido sintético
 <400> 61
 cttttctgct ct 12
 15 <210> 62
 <211> 12
 <212> ADN
 <213> Secuencia artificial
 <220>
 20 <223> Polinucleótido sintético
 <400> 62
 cttttctgta ct 12
 <210> 63
 <211> 12
 25 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Polinucleótido sintético
 <400> 63
 30 cttttctgtg ct 12
 <210> 64
 <211> 12
 <212> ADN
 <213> Secuencia artificial
 35 <220>
 <223> Polinucleótido sintético
 <400> 64

ctttctgtc ct 12
 <210> 65
 <211> 12
 <212> ADN
 5 <213> Secuencia artificial
 <220>
 <223> Polinucleótido sintético
 <400> 65
 ctttcactc ct 12
 10 <210> 66
 <211> 12
 <212> ADN
 <213> Secuencia artificial
 <220>
 15 <223> Polinucleótido sintético
 <400> 66
 ctttcgttc ct 12
 <210> 67
 <211> 12
 20 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Polinucleótido sintético
 <400> 67
 25 ctttcgtac ct 12
 <210> 68
 <211> 12
 <212> ADN
 <213> Secuencia artificial
 30 <220>
 <223> Polinucleótido sintético
 <400> 68
 ctttcgctc ct 12
 <210> 69
 35 <211> 12
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> Polinucleótido sintético
 <400> 69
 cttttctgac ct 12
 5 <210> 70
 <211> 12
 <212> ADN
 <213> Secuencia artificial
 <220>
 10 <223> Polinucleótido sintético
 <400> 70
 aggcattgca tg 12
 <210> 71
 <211> 12
 15 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Polinucleótido sintético
 <400> 71
 20 aggcattgtg ct 12
 <210> 72
 <211> 12
 <212> ADN
 <213> Secuencia artificial
 25 <220>
 <223> Polinucleótido sintético
 <400> 72
 aggcattgct ct 12
 <210> 73
 30 <211> 12
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Polinucleótido sintético
 35 <400> 73
 aggcatttct ac 12
 <210> 74

<211> 12
 <212> ADN
 <213> Secuencia artificial
 <220>
 5 <223> Polinucleótido sintético
 <400> 74
 aggcatacct ac 12
 <210> 75
 <211> 12
 10 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Polinucleótido sintético
 <400> 75
 15 aggcatttgc ac 12
 <210> 76
 <211> 12
 <212> ADN
 <213> Secuencia artificial
 20 <220>
 <223> Polinucleótido sintético
 <400> 76
 aggcacgctc ct 12
 <210> 77
 25 <211> 12
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Polinucleótido sintético
 30 <400> 77
 tctgtcggc tc 12
 <210> 78
 <211> 12
 <212> ADN
 35 <213> Secuencia artificial
 <220>
 <223> Polinucleótido sintético

<400> 78
 gttcaatgct ct 12
 <210> 79
 <211> 12
 5 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Polinucleótido sintético
 <400> 79
 10 attcgggtgct ct 12
 <210> 80
 <211> 12
 <212> ADN
 <213> Secuencia artificial
 15 <220>
 <223> Polinucleótido sintético
 <400> 80
 gatgcctgct ct 12
 <210> 81
 20 <211> 12
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Polinucleótido sintético
 25 <400> 81
 ttgcttgct ct 12
 <210> 82
 <211> 100
 <212> ADN
 30 <213> Secuencia artificial
 <220>
 <223> Polinucleótido sintético
 <400> 82
 ttactgtag ctgtctcatt ttgctgcatc ctgtccgttc acgttggagc ttgtcatccg 60
 35 tcctcttttc actcctaggc attgcctat tcggcgtcct 100
 <210> 83
 <211> 10

<212> ADN
 <213> Secuencia artificial
 <220>
 <223> Polinucleótido sintético
 5 <400> 83
 cgatctggtt 10
 <210> 84
 <211> 10
 <212> ADN
 10 <213> Secuencia artificial
 <220>
 <223> Polinucleótido sintético
 <400> 84
 cgatctggtt 10
 15 <210> 85
 <211> 10
 <212> ADN
 <213> Secuencia artificial
 <220>
 20 <223> Polinucleótido sintético
 <400> 85
 gctagaccaa 10
 <210> 86
 <211> 10
 25 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Polinucleótido sintético
 <400> 86
 30 gctggaccaa 10
 <210> 87
 <211> 10
 <212> ADN
 <213> Secuencia artificial
 35 <220>
 <223> Polinucleótido sintético
 <400> 87

gctcgaccaa 10
 <210> 88
 <211> 10
 <212> ADN
 5 <213> Secuencia artificial
 <220>
 <223> Polinucleótido sintético
 <400> 88
 gcttgaccaa 10
 10 <210> 89
 <211> 10
 <212> ADN
 <213> Secuencia artificial
 <220>
 15 <223> Polinucleótido sintético
 <400> 89
 gagactgtac 10
 <210> 90
 <211> 10
 20 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Polinucleótido sintético
 <400> 90
 25 gagacagtac 10
 <210> 91
 <211> 10
 <212> ADN
 <213> Secuencia artificial
 30 <220>
 <223> Polinucleótido sintético
 <400> 91
 gagaccgtac 10
 <210> 92
 35 <211> 10
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> Polinucleótido sintético
 <400> 92
 gagacgttac 10
 5 <210> 93
 <211> 10
 <212> ADN
 <213> Secuencia artificial
 <220>
 10 <223> Polinucleótido sintético
 <400> 93
 gagagtgtac 10
 <210> 94
 <211> 10
 15 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Polinucleótido sintético
 <400> 94
 20 gagaatgtac 10
 <210> 95
 <211> 10
 <212> ADN
 <213> Secuencia artificial
 25 <220>
 <223> Polinucleótido sintético
 <400> 95
 gagattgtac 10
 <210> 96
 30 <211> 10
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Polinucleótido sintético
 35 <400> 96
 gagactgtac 10
 <210> 97

<211> 10
 <212> ADN
 <213> Secuencia artificial
 <220>
 5 <223> Polinucleótido sintético
 <400> 97
 gagattgtac 10
 <210> 98
 <211> 10
 10 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Polinucleótido sintético
 <400> 98
 15 gagaccgtac 10
 <210> 99
 <211> 10
 <212> ADN
 <213> Secuencia artificial
 20 <220>
 <223> Polinucleótido sintético
 <400> 99
 gagagtgtac 10
 <210> 100
 25 <211> 12
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Polinucleótido sintético
 30 <400> 100
 ctgagactgt ac 12
 <210> 101
 <211> 11
 <212> ADN
 35 <213> Secuencia artificial
 <220>
 <223> Polinucleótido sintético

<400> 101
 tgagactgta c 11
 <210> 102
 <211> 10
 5 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Polinucleótido sintético
 <400> 102
 10 gagactgtac 10
 <210> 103
 <211> 12
 <212> ADN
 <213> Secuencia artificial
 15 <220>
 <223> Polinucleótido sintético
 <400> 103
 catgtcagag tc 12
 <210> 104
 20 <211> 11
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Polinucleótido sintético
 25 <400> 104
 catgtcagag t 11

REIVINDICACIONES

1. Una sonda de secuenciación que comprende un dominio de unión a la diana y un dominio de código de barras;
en la que dicho dominio de unión a la diana comprende al menos 12 nucleótidos y puede unirse a un ácido nucleico diana;
- 5 en la que dicho dominio de código de barras comprende una cadena principal sintética, comprendiendo dicho dominio de código de barras al menos seis posiciones de fijación, comprendiendo cada posición de fijación al menos una región de fijación, comprendiendo dicha región de fijación al menos una secuencia de ácido nucleico que puede unirse mediante una molécula de ácido nucleico complementaria,
en la que cada posición de fijación de las al menos seis posiciones de fijación corresponde a un nucleótido en el
10 dominio de unión a la diana y cada una de las al menos seis posiciones de fijación tiene una secuencia de ácido nucleico diferente, y
en la que dicha secuencia de ácido nucleico de cada posición de las al menos seis posiciones de fijación determina la posición e identidad del nucleótido correspondiente en dicho ácido nucleico diana que se une mediante dicho dominio de unión a la diana.
- 15 2. La sonda de secuenciación de la reivindicación 1, en la que dicha cadena principal sintética comprende ADN monocatenario.
3. La sonda de secuenciación de la reivindicación 1, en la que dicha sonda de secuenciación comprende un espaciador de ADN bicatenario entre el dominio de unión a la diana y el dominio de código de barras.
- 20 4. La sonda de secuenciación de la reivindicación 1, en la que el número de nucleótidos en un dominio de unión a la diana es al menos dos más que el número de regiones de fijación en el dominio de código de barras.
5. La sonda de secuenciación de la reivindicación 1, en la que cada posición en un dominio de código de barras tiene:
(a) el mismo número de regiones de fijación; (b) una región de fijación; o (c) más de una región de fijación.
6. La sonda de secuenciación de la reivindicación 1, en la que cada molécula de ácido nucleico complementaria;
(a) por cada posición de fijación comprende un marcador detectable;
- 25 (b) o se liga indirectamente a una molécula de ácido nucleico primaria mediante un espaciador de ácido nucleico; o
(c) por cada posición de fijación comprende entre aproximadamente 8 nucleótidos y aproximadamente 20 nucleótidos, preferiblemente, en la que cada molécula de ácido nucleico complementaria comprende aproximadamente 12 nucleótidos.
7. La sonda de secuenciación de la reivindicación 6(b), en la que cada molécula de ácido nucleico primaria se hibrida
30 con al menos una, dos, tres, cuatro o cinco moléculas de ácido nucleico secundarias.
8. La sonda de secuenciación de la reivindicación 7, en la que la molécula o moléculas de ácido nucleico secundarias comprenden al menos un marcador detectable.
9. La sonda de secuenciación de la reivindicación 7, en la que cada molécula de ácido nucleico secundaria se hibrida
35 con al menos una, dos, tres, cuatro, cinco, seis o siete moléculas de ácido nucleico terciarias que comprenden al menos un marcador detectable.
10. La sonda de secuenciación de la reivindicación 1, en la que una o más posiciones de fijación en el dominio de código de barras están adyacentes a al menos un polinucleótido monocatenario flanqueante.
11. Una población de sondas de secuenciación que comprenden una pluralidad de la sonda de secuenciación de la reivindicación 1.
- 40 12. Un método para secuenciar un ácido nucleico, que comprende las etapas de:
(1) hibridar al menos una primera población de primeras sondas de secuenciación que comprenden una pluralidad de la sonda de secuenciación de una cualquiera de las reivindicaciones 1 a 10 con un ácido nucleico diana que se inmoviliza en un sustrato, en el que el ácido nucleico diana se inmoviliza en el sustrato en una o más posiciones;
- 45 (2) unir una primera molécula de ácido nucleico complementaria que comprende un marcador detectable o una primera molécula de ácido nucleico complementaria de un primer complejo indicador que comprende un marcador detectable a una primera posición de fijación de las al menos seis posiciones de fijación;
- (3) detectar el marcador detectable de la primera molécula de ácido nucleico complementaria unida o el marcador detectable de la primera molécula de ácido nucleico complementaria unida del primer complejo indicador;

- (4) identificar la posición e identidad de un primer nucleótido en el ácido nucleico diana inmovilizado;
- 5 (5) unir a la primera posición de fijación una primera molécula de ácido nucleico de hibridación que carece de un marcador detectable, soltando de ese modo la primera molécula de ácido nucleico complementaria que comprende un marcador detectable o la primera molécula de ácido nucleico complementaria del primer complejo indicador que comprende un marcador detectable;
- (6) unir una segunda molécula de ácido nucleico complementaria que comprende un marcador detectable o una segunda molécula de ácido nucleico complementaria de un segundo complejo indicador que comprende un marcador detectable a una segunda posición de fijación de las al menos seis posiciones de fijación;
- 10 (7) detectar el marcador detectable de la segunda molécula de ácido nucleico complementaria unida o el marcador detectable de la segunda molécula de ácido nucleico complementaria unida del segundo complejo indicador;
- (8) identificar la posición e identidad de un segundo nucleótido en el ácido nucleico diana inmovilizado;
- 15 (9) repetir las etapas (5) a (8) hasta que cada posición de fijación de las al menos seis posiciones de fijación se haya unido mediante una molécula de ácido nucleico complementaria que comprende un marcador detectable o una molécula de ácido nucleico complementaria de un complejo indicador que comprende un marcador detectable, y se haya detectado el marcador detectable de la molécula de ácido nucleico complementaria unida o el marcador detectable de la molécula de ácido nucleico complementaria unida de un complejo indicador, identificando de ese modo el orden lineal de al menos seis nucleótidos para al menos una primera región del ácido nucleico diana inmovilizado que se hibridó mediante el dominio de unión a la diana de la sonda de secuenciación; y
- 20 (10) deshibridar la al menos una primera población de primeras sondas de secuenciación del ácido nucleico diana inmovilizado.
13. El método de la reivindicación 12, en el que las etapas (5) y (6) se producen secuencialmente o simultáneamente.
14. El método de la reivindicación 12, en el que la primera molécula de ácido nucleico complementaria y la primera molécula de ácido nucleico de hibridación que carece de un marcador detectable comprende la misma secuencia de ácido nucleico.
- 25 15. El método de la reivindicación 12, en el que la primera molécula de ácido nucleico de hibridación que carece de un marcador detectable comprende una secuencia de ácido nucleico complementaria a un polinucleótido monocatenario flanqueante adyacente a la primera posición de fijación.
16. El método de la reivindicación 12, que comprende además
- 30 (11) hibridar al menos una segunda población de segundas sondas de secuenciación que comprenden una pluralidad de sondas de secuenciación de una cualquiera de las reivindicaciones 1 a 10 con un ácido nucleico diana que se inmoviliza en un sustrato, en el que el ácido nucleico diana se inmoviliza en el sustrato en una o más posiciones, en el que el dominio de unión a la diana de la primera sonda de secuenciación y la segunda sonda de secuenciación son diferentes;
- 35 (12) unir una primera molécula de ácido nucleico complementaria que comprende un marcador detectable o una primera molécula de ácido nucleico complementaria de un primer complejo indicador que comprende un marcador detectable a una primera posición de fijación de las al menos seis posiciones de fijación;
- (13) detectar el marcador detectable de la primera molécula de ácido nucleico complementaria unida o el marcador detectable de la primera molécula de ácido nucleico complementaria unida del primer complejo indicador;
- (14) identificar la posición e identidad de un primer nucleótido en el ácido nucleico diana inmovilizado;
- 40 (15) unir a la primera posición de fijación una primera molécula de ácido nucleico de hibridación que carece de un marcador detectable, soltando de ese modo la primera molécula de ácido nucleico complementaria que comprende un marcador detectable o la primera molécula de ácido nucleico complementaria del primer complejo indicador que comprende un marcador detectable;
- 45 (16) unir una segunda molécula de ácido nucleico complementaria que comprende un marcador detectable o una segunda molécula de ácido nucleico complementaria de un segundo complejo indicador que comprende un marcador detectable a una segunda posición de fijación de las al menos seis posiciones de fijación;
- (17) detectar el marcador detectable de la segunda molécula de ácido nucleico complementaria unida o el marcador detectable de la segunda molécula de ácido nucleico complementaria unida del segundo complejo indicador;
- (18) identificar la posición e identidad de un segundo nucleótido en el ácido nucleico diana inmovilizado;

- (19) repetir las etapas (15) a (18) hasta que cada posición de fijación de las al menos seis posiciones de fijación se haya unido mediante una molécula de ácido nucleico complementaria que comprende un marcador detectable o una molécula de ácido nucleico complementaria de un complejo indicador que comprende un marcador detectable, y se haya detectado el marcador detectable de la molécula de ácido nucleico complementaria unida o el marcador detectable de la molécula de ácido nucleico complementaria unida de un complejo indicador, identificando de ese modo el orden lineal de al menos seis nucleótidos para al menos una segunda región del ácido nucleico diana inmovilizado que se hibridó mediante el dominio de unión a la diana de la sonda de secuenciación; y
- 5 (20) deshibridar la al menos una segunda población de segundas sondas de secuenciación del ácido nucleico diana inmovilizado.
- 10 17. El método de la reivindicación 16, que comprende además etapas de ensamblar cada orden lineal identificado de nucleótidos en la al menos primera región y al menos segunda región del ácido nucleico diana inmovilizado, identificando de ese modo una secuencia para el ácido nucleico diana inmovilizado.

Figura 1

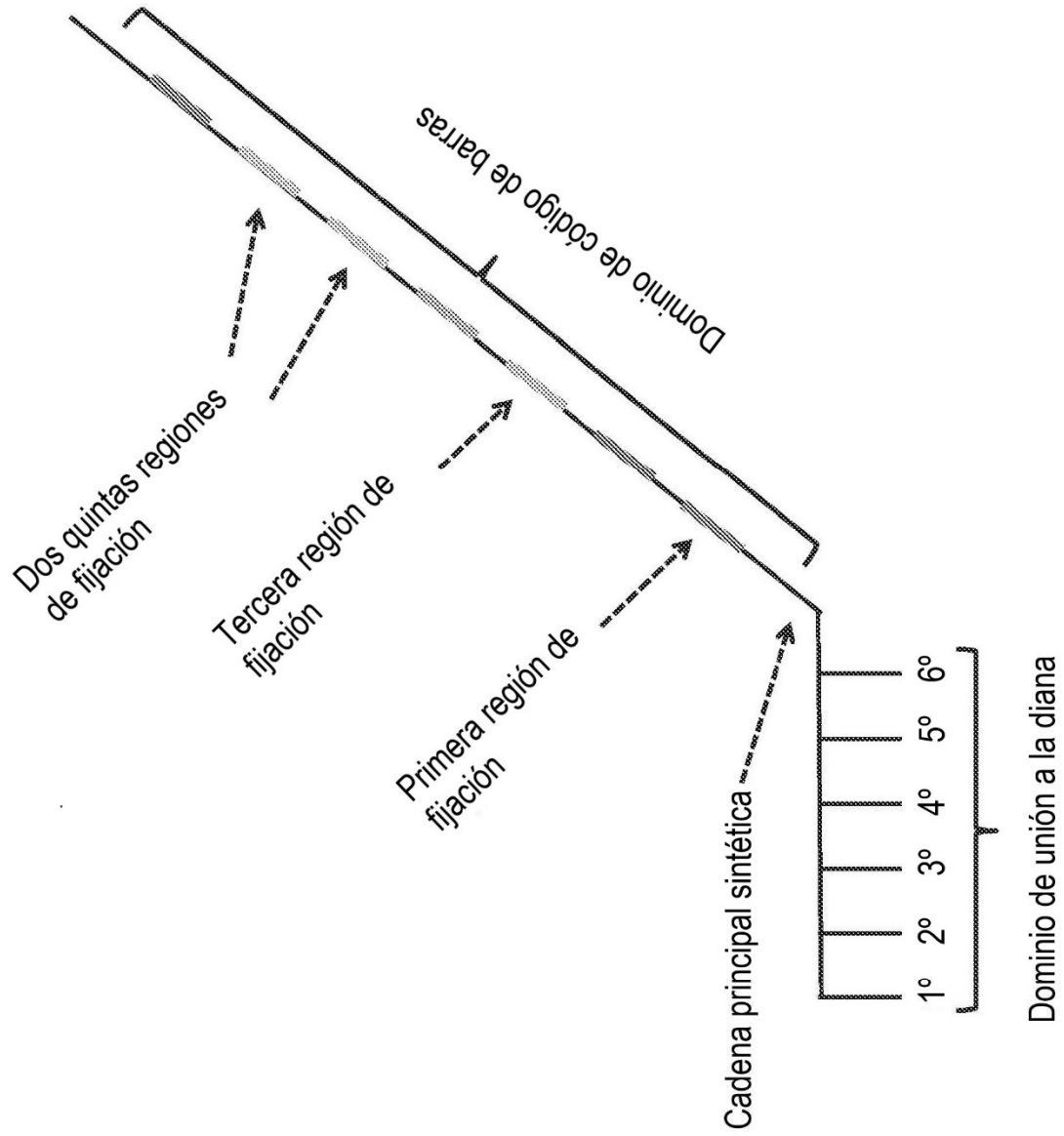


Figura 2

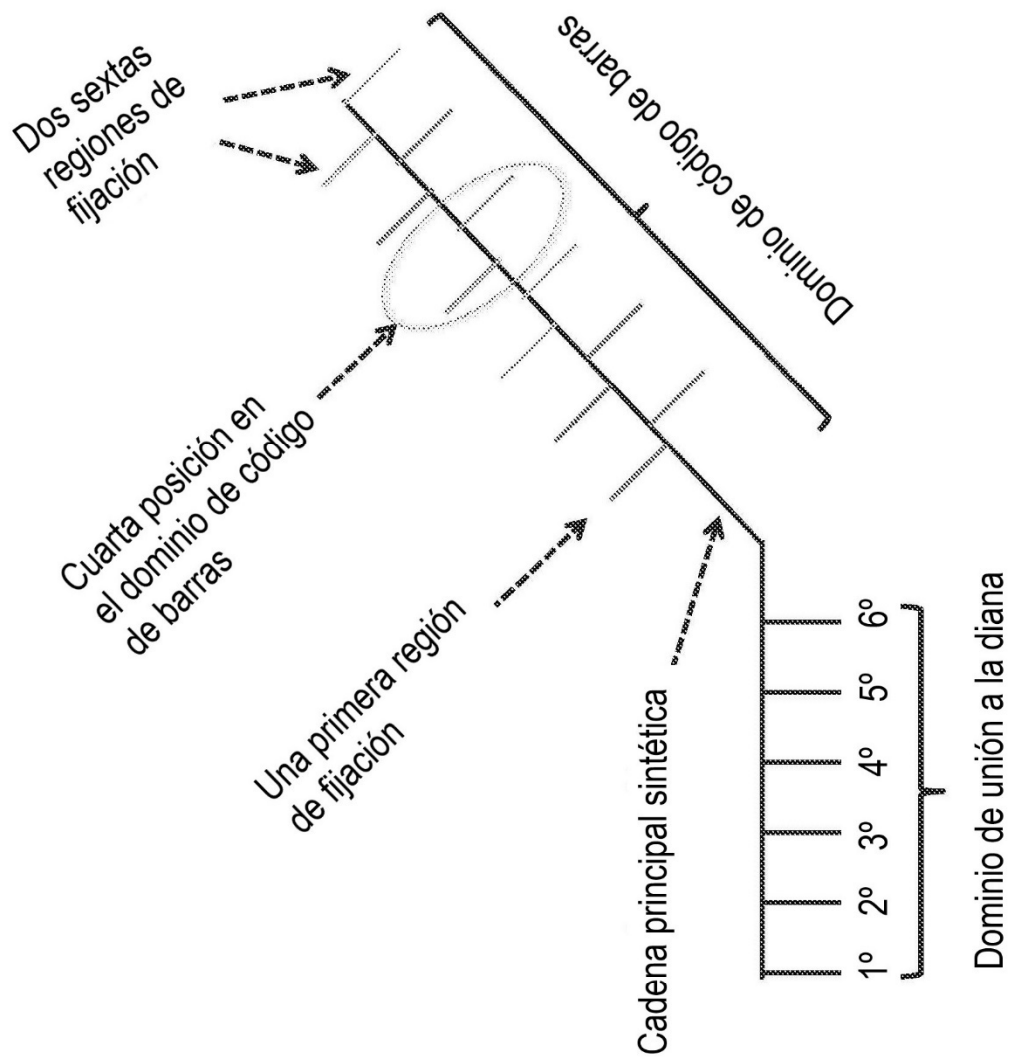


Figura 3

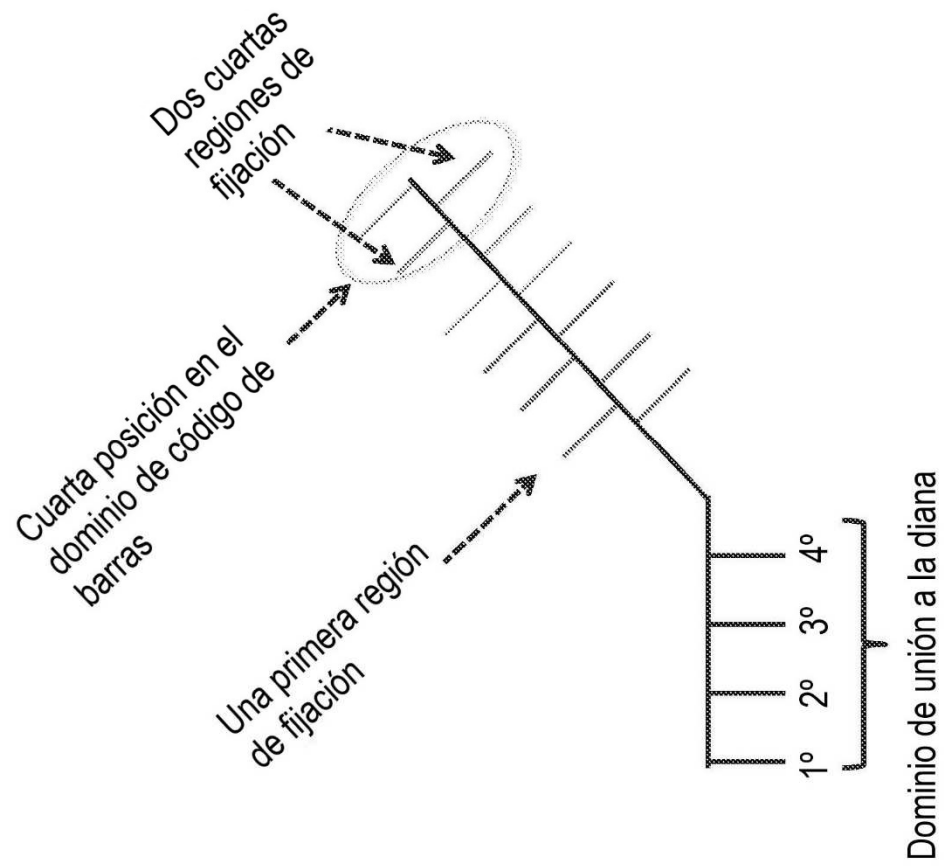


Figura 4

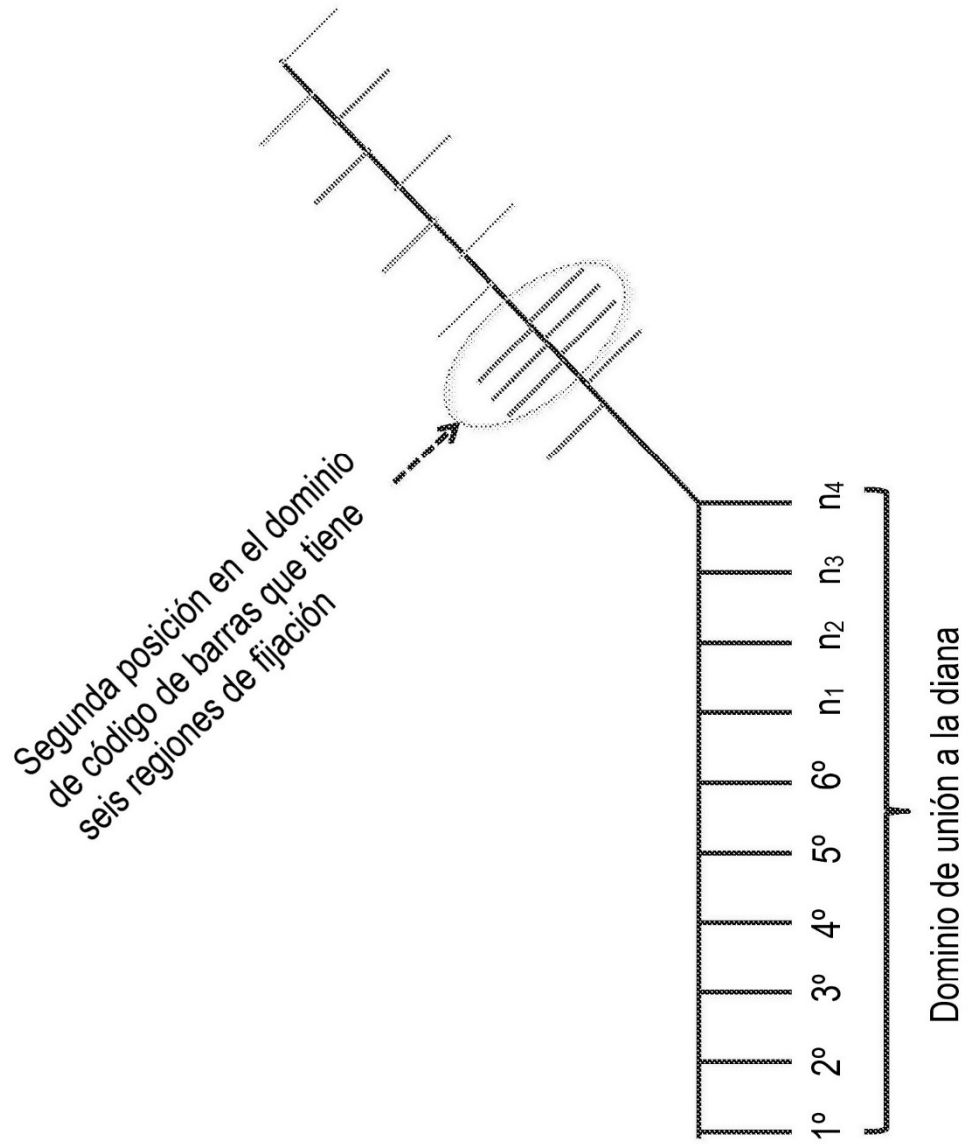


Figura 5

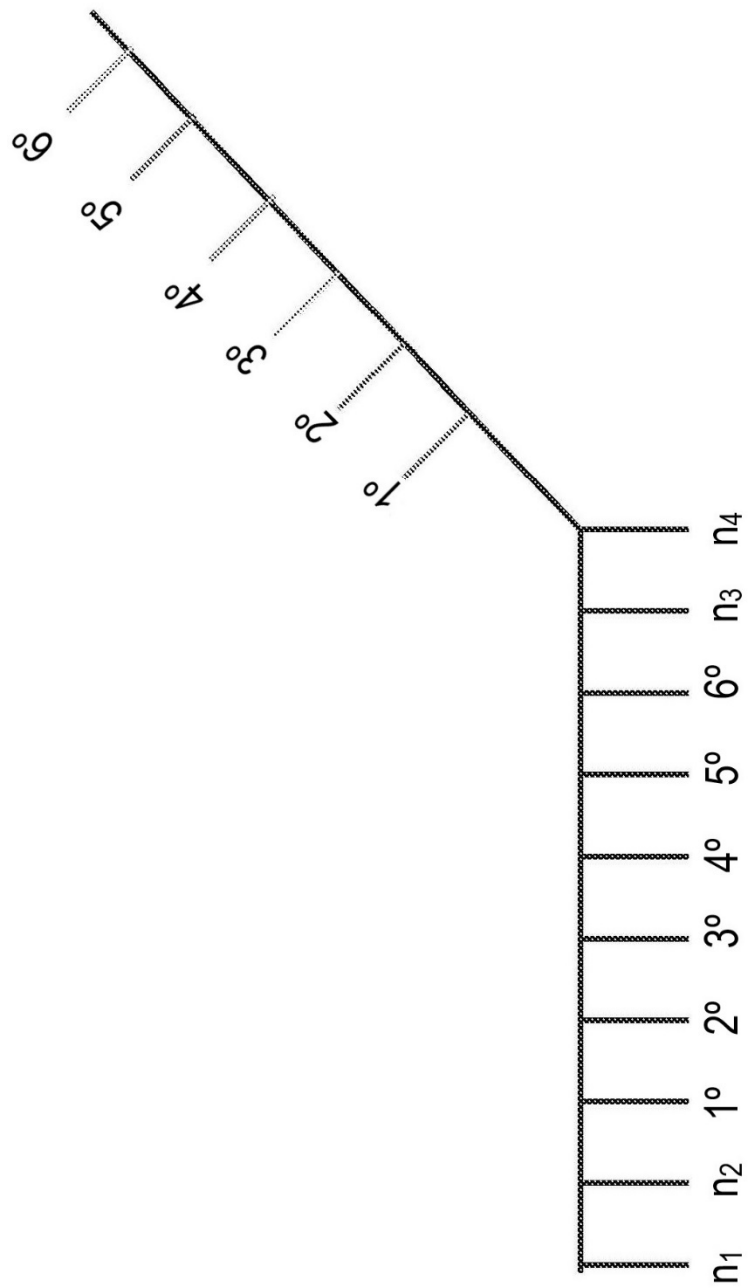


Figura 6A

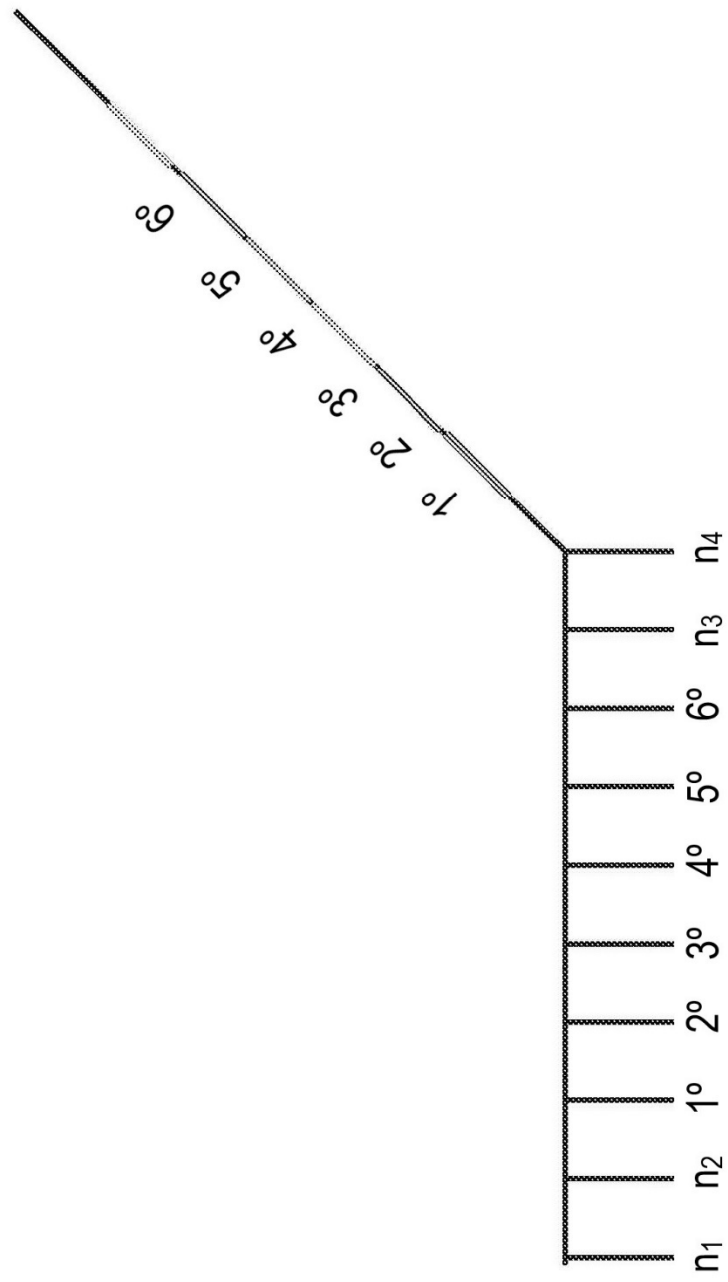


Figura 6B

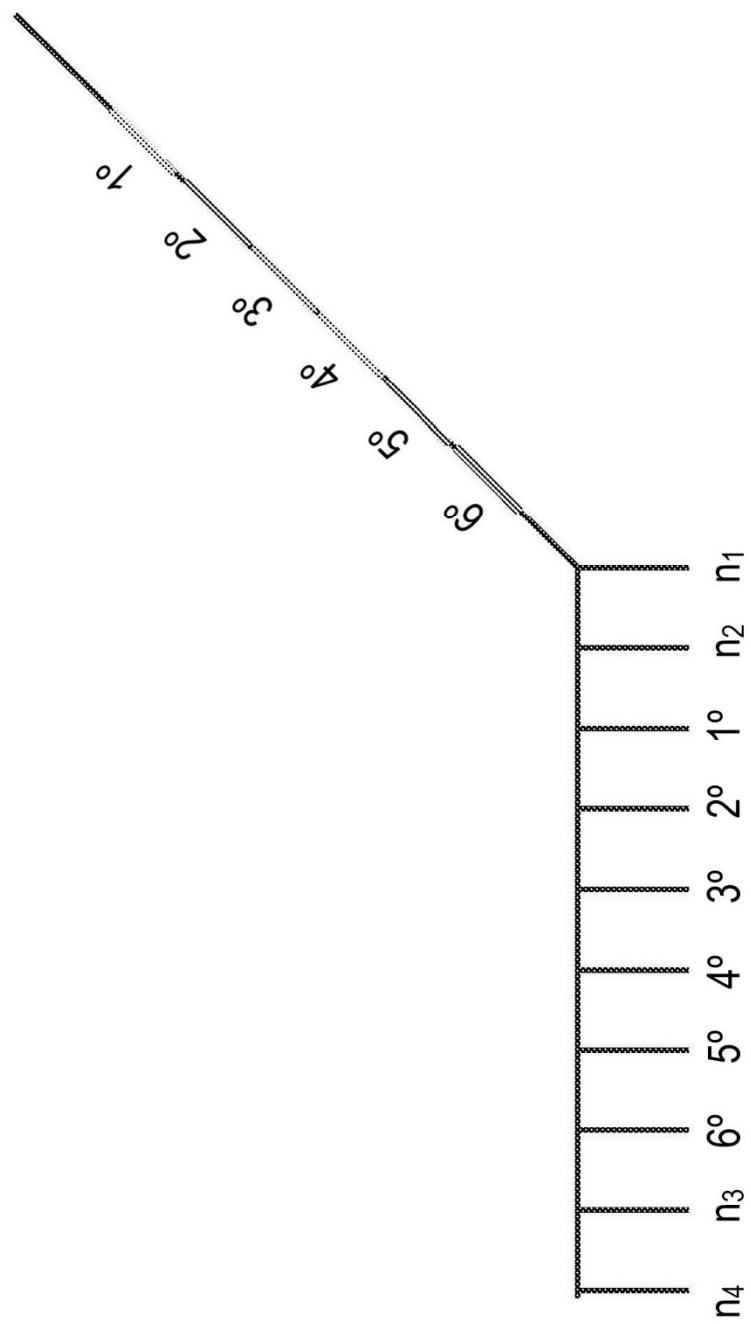


Figura 6C

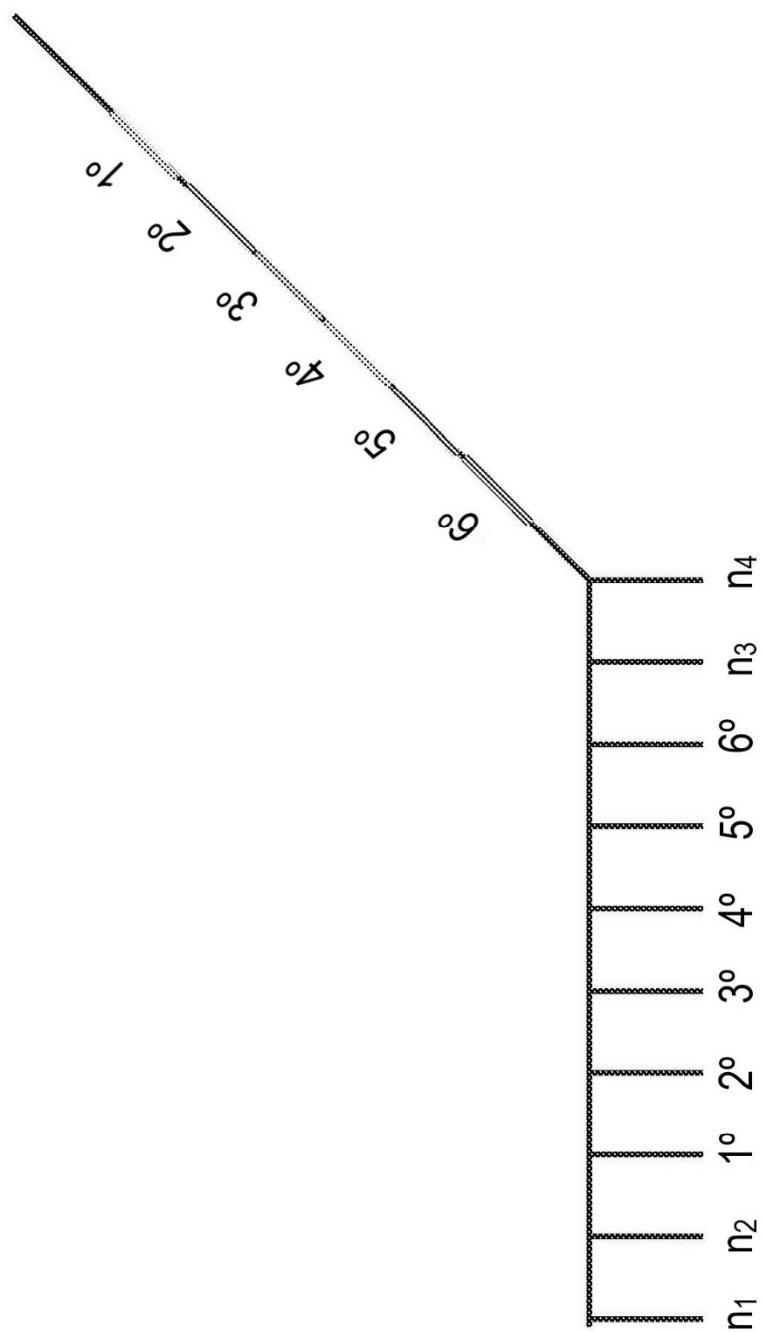


Figura 6D

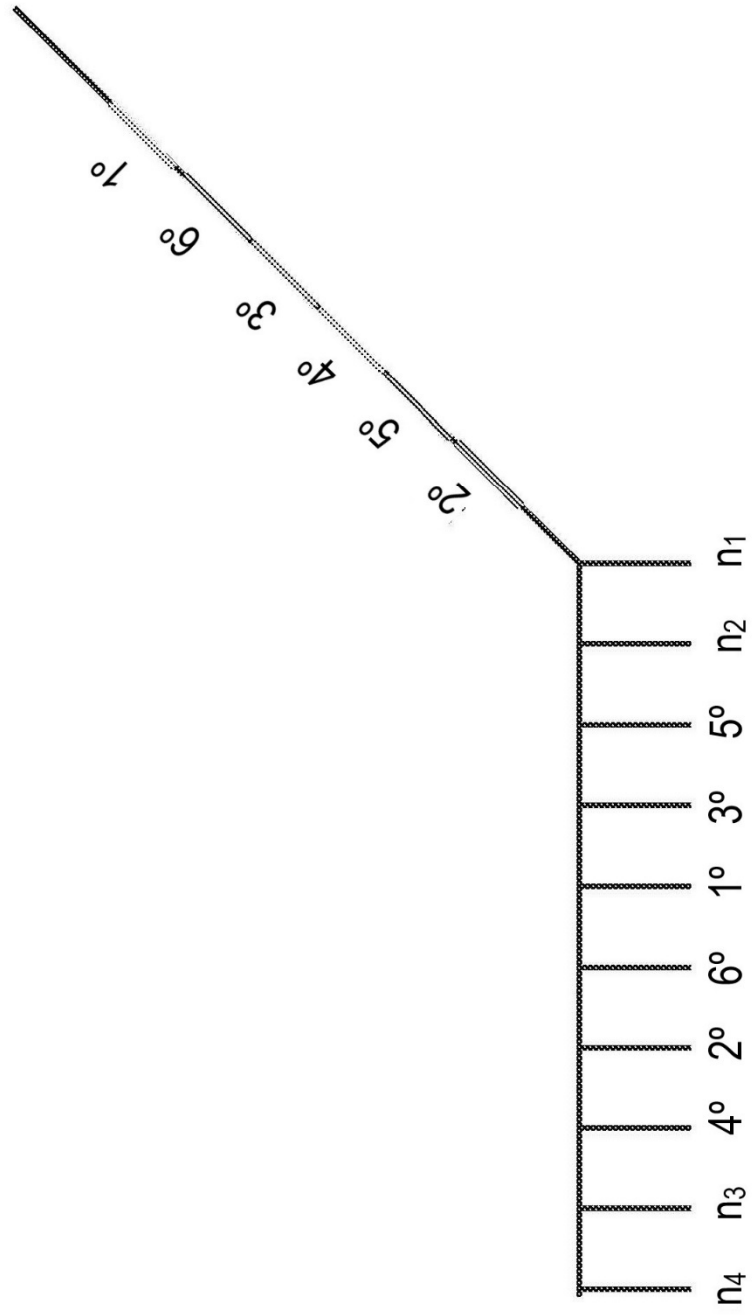


Figura 7

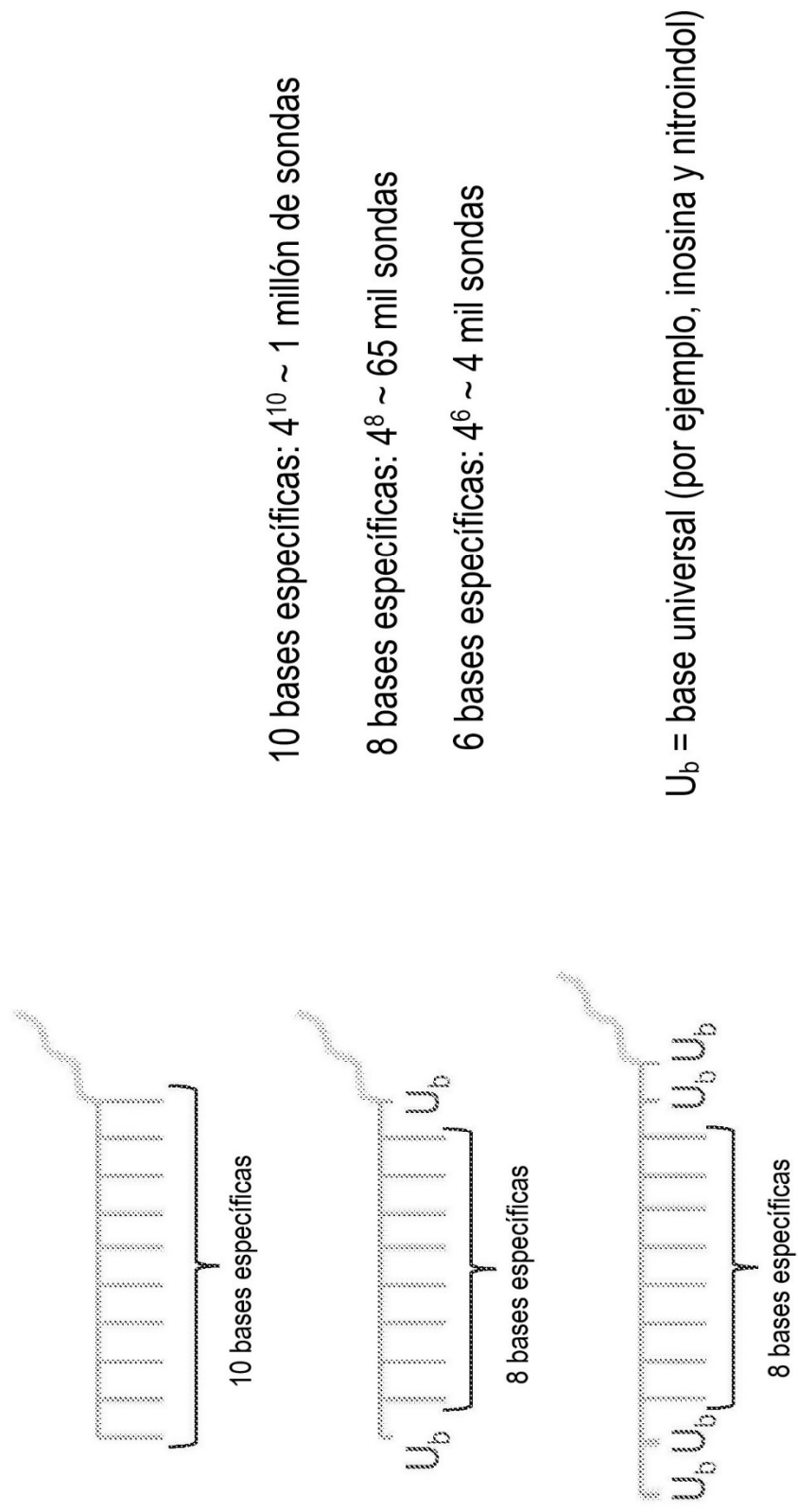


Figura 8A

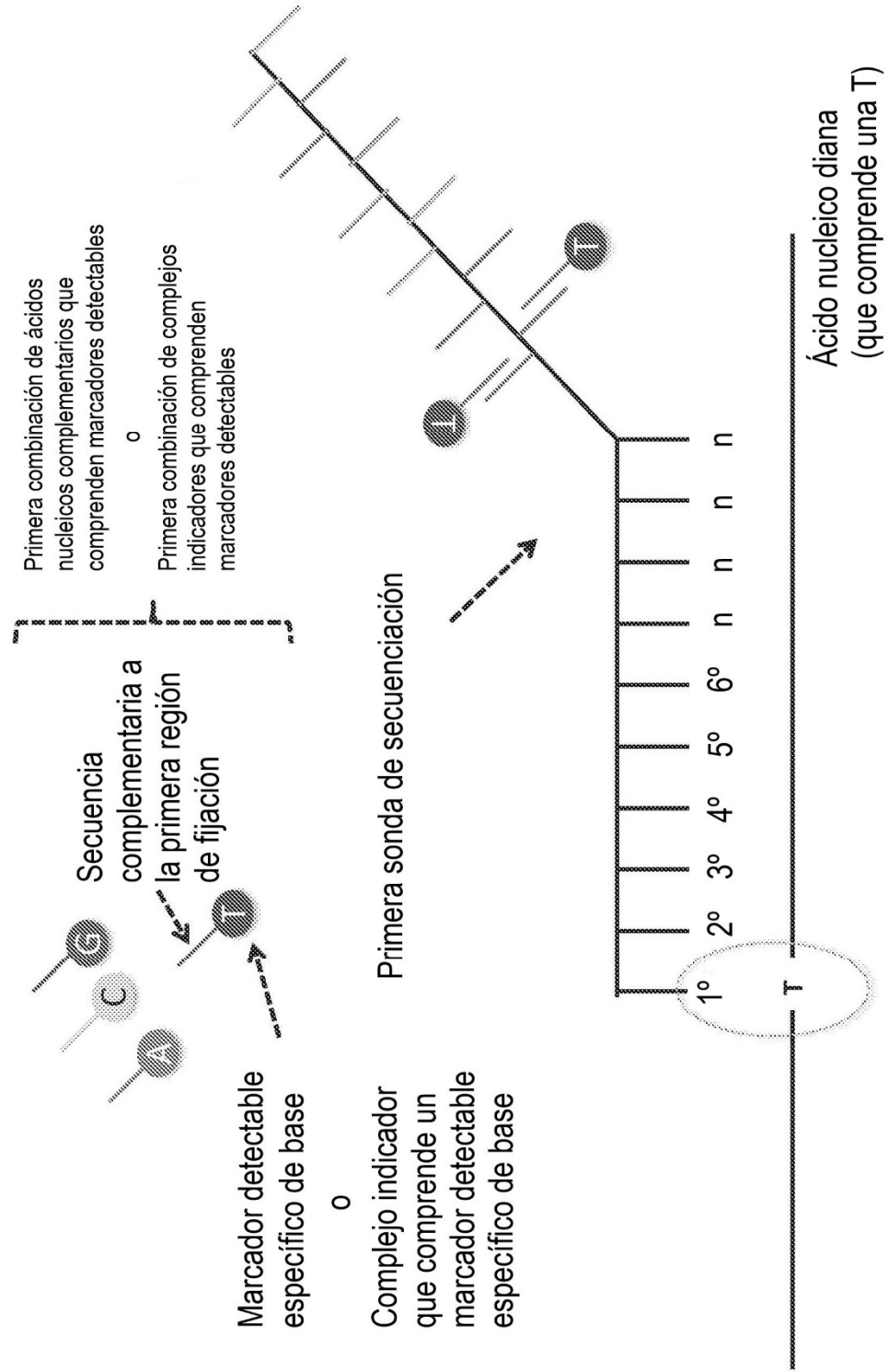


Figura 8B

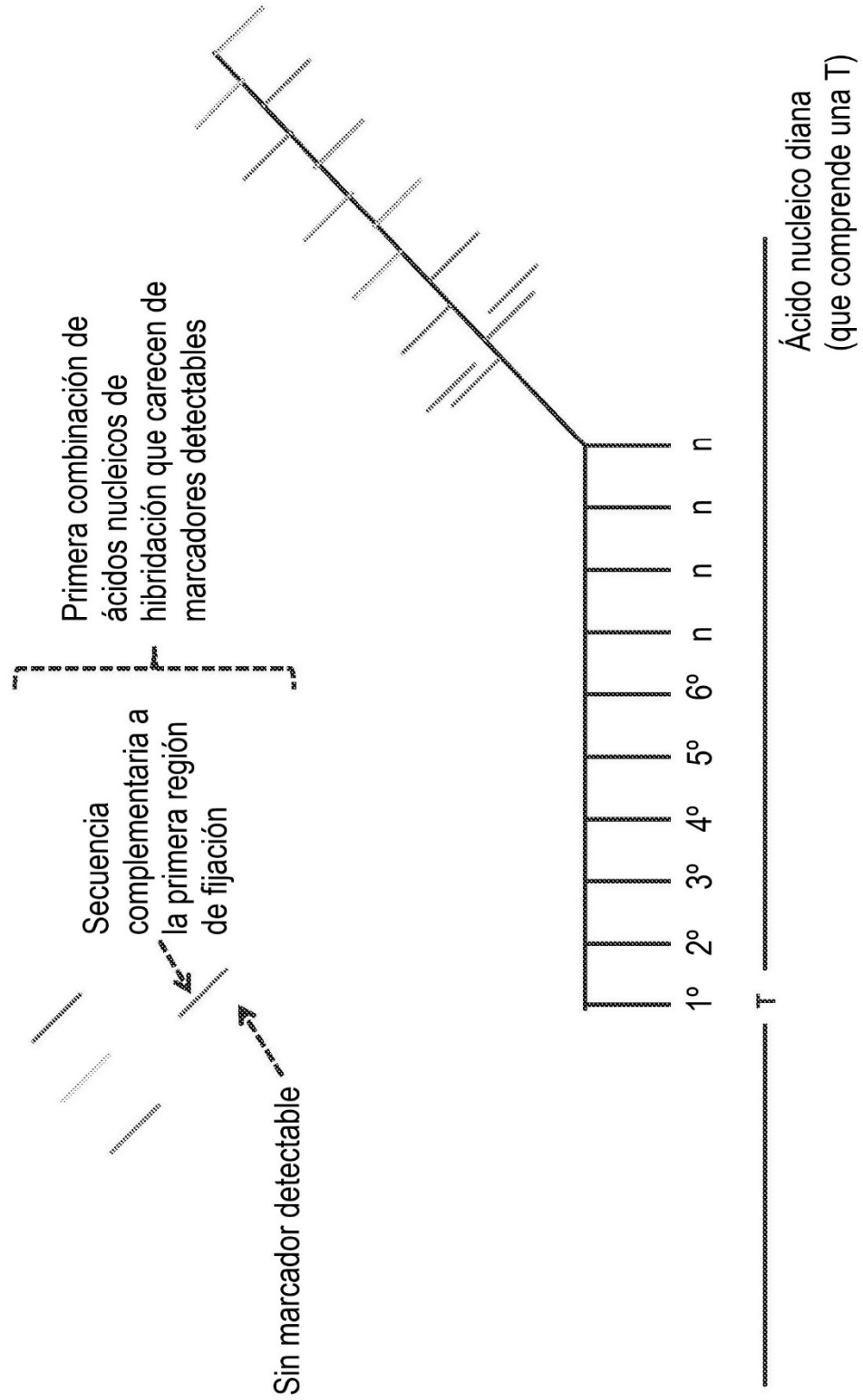


Figura 8C

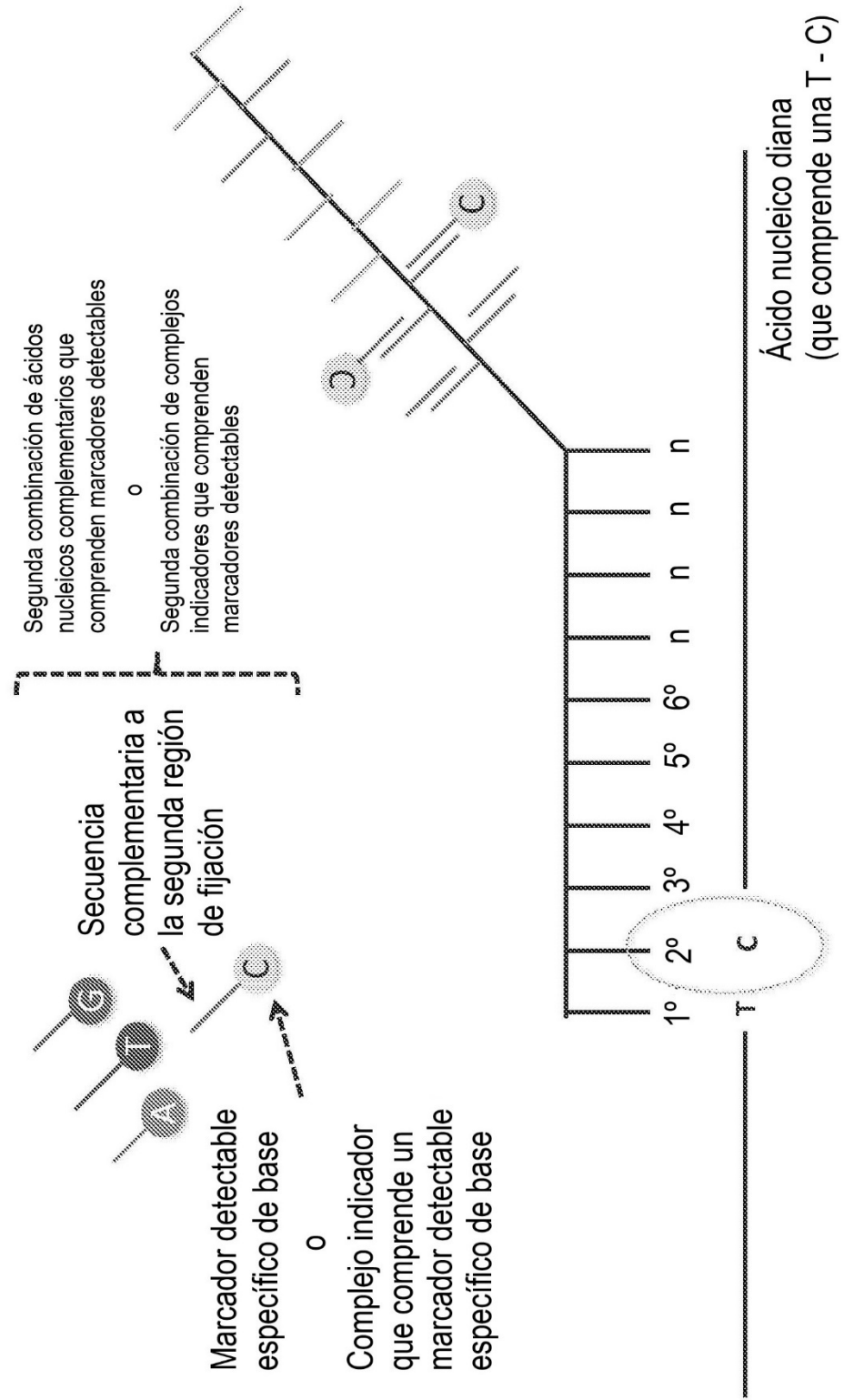


Figura 8D

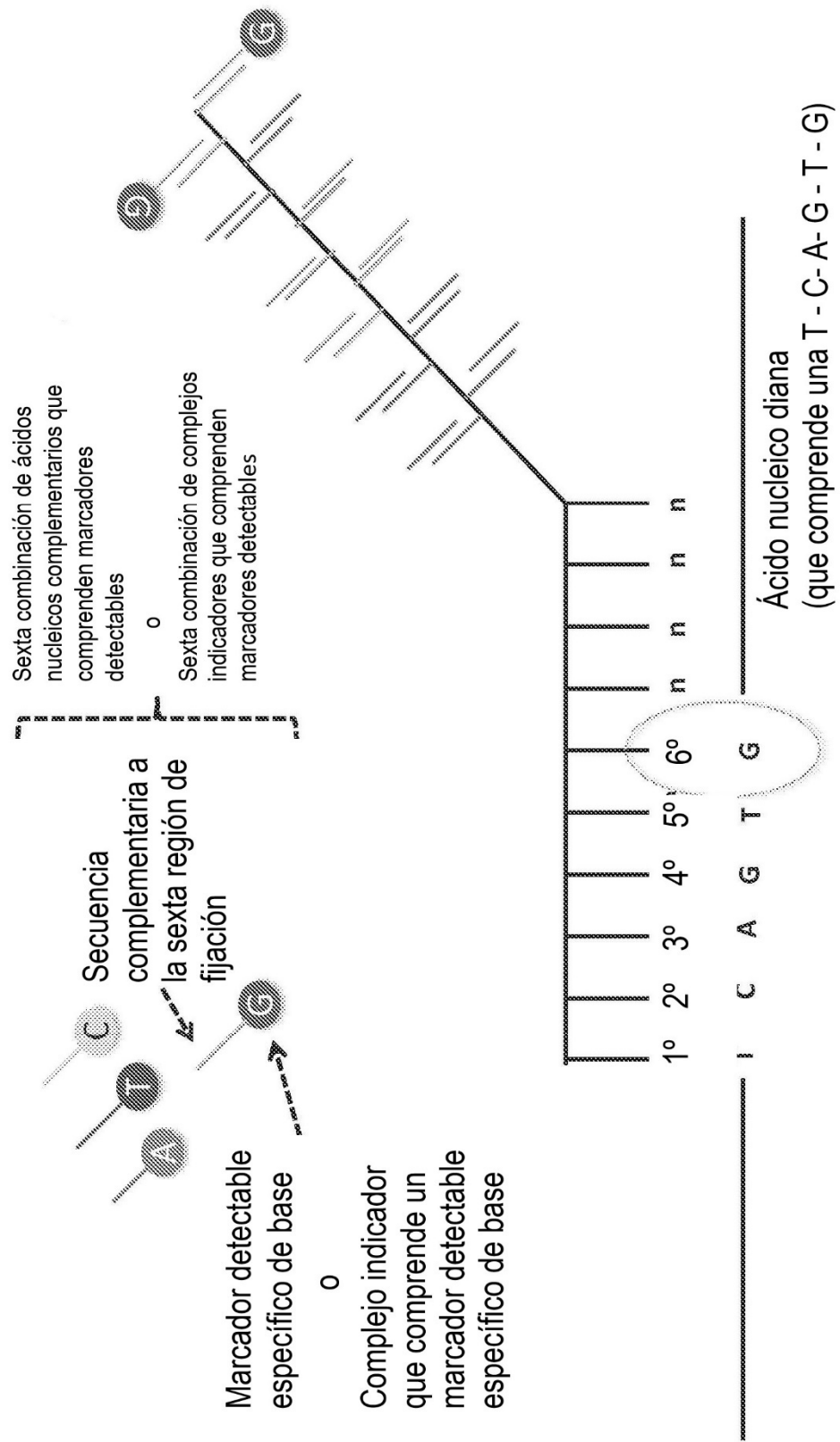


Figura 8E

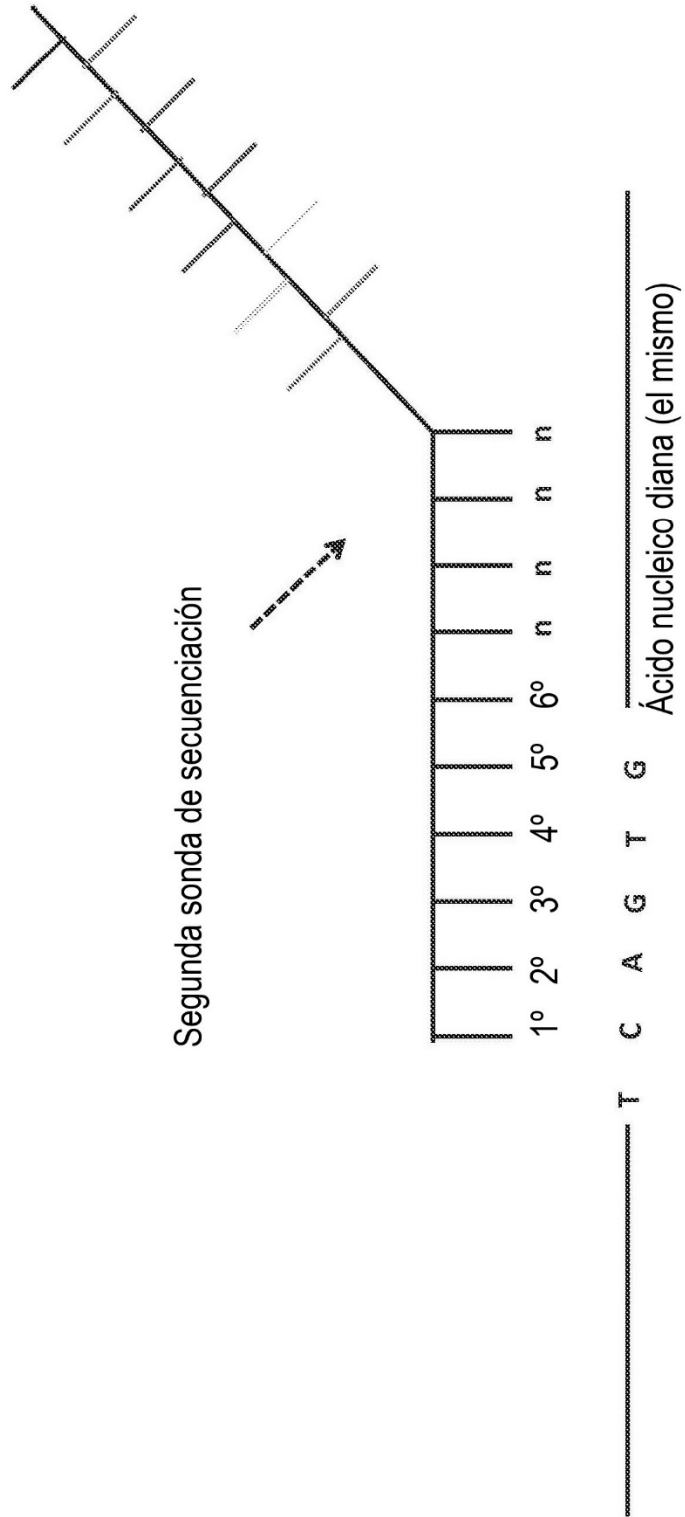
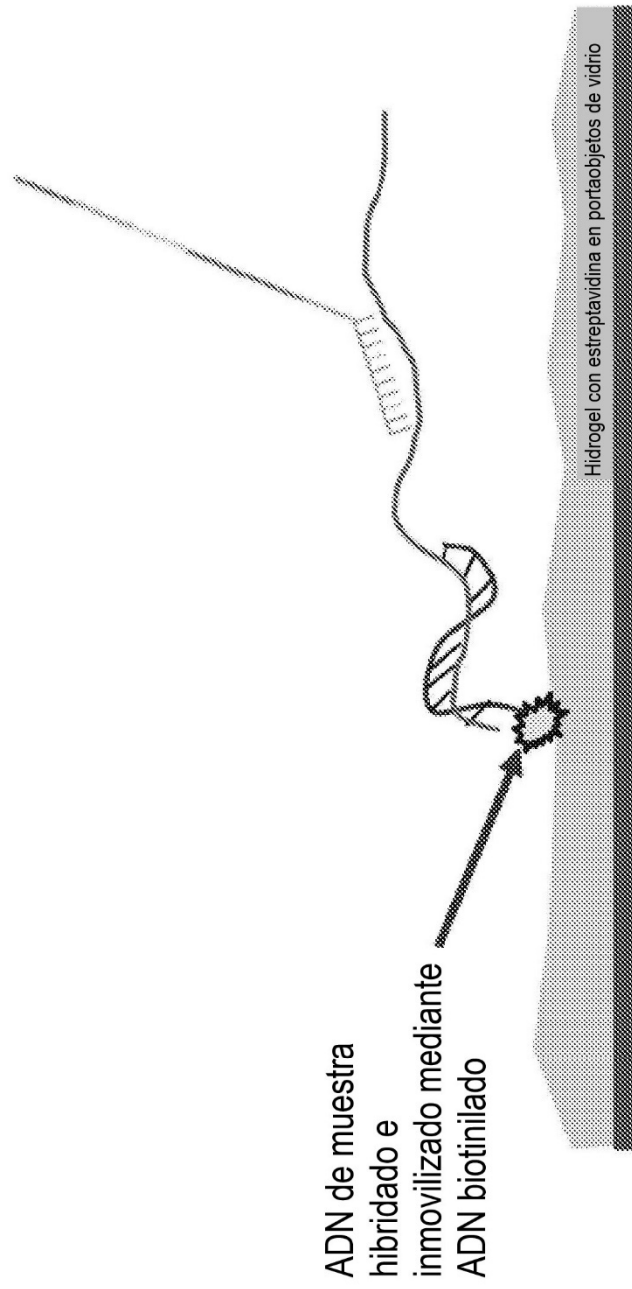


Figura 9A



Ciclo 1: → Hib. sonda

Figura 9B

Complejo indicador que comprende un marcador detectable

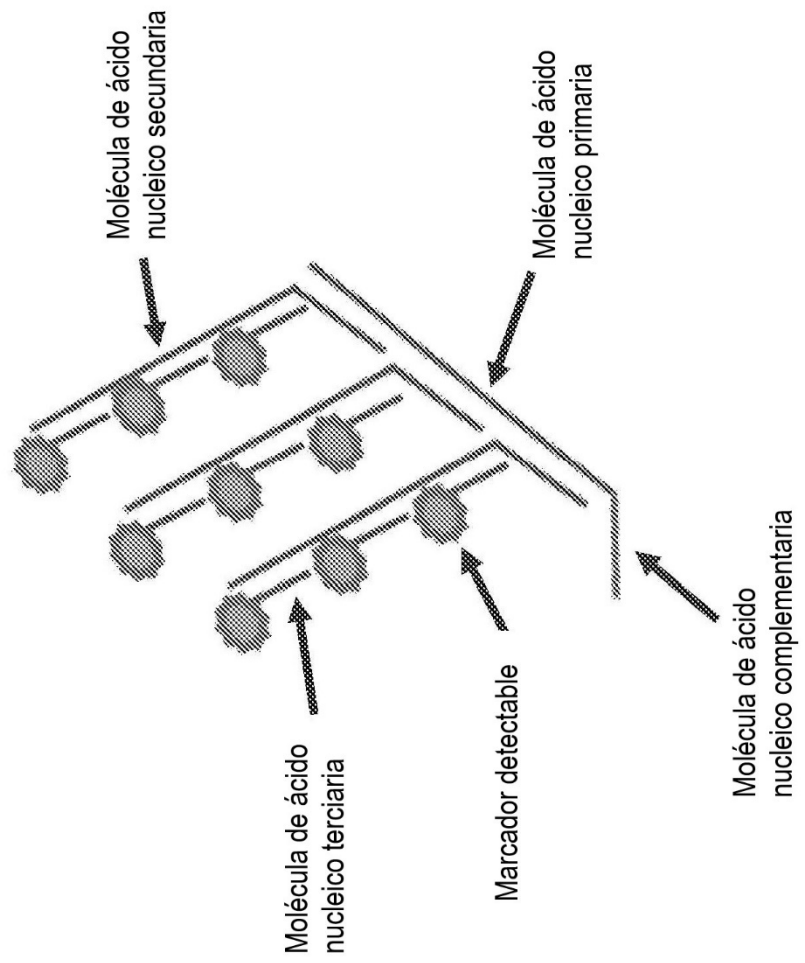


Figura 9C

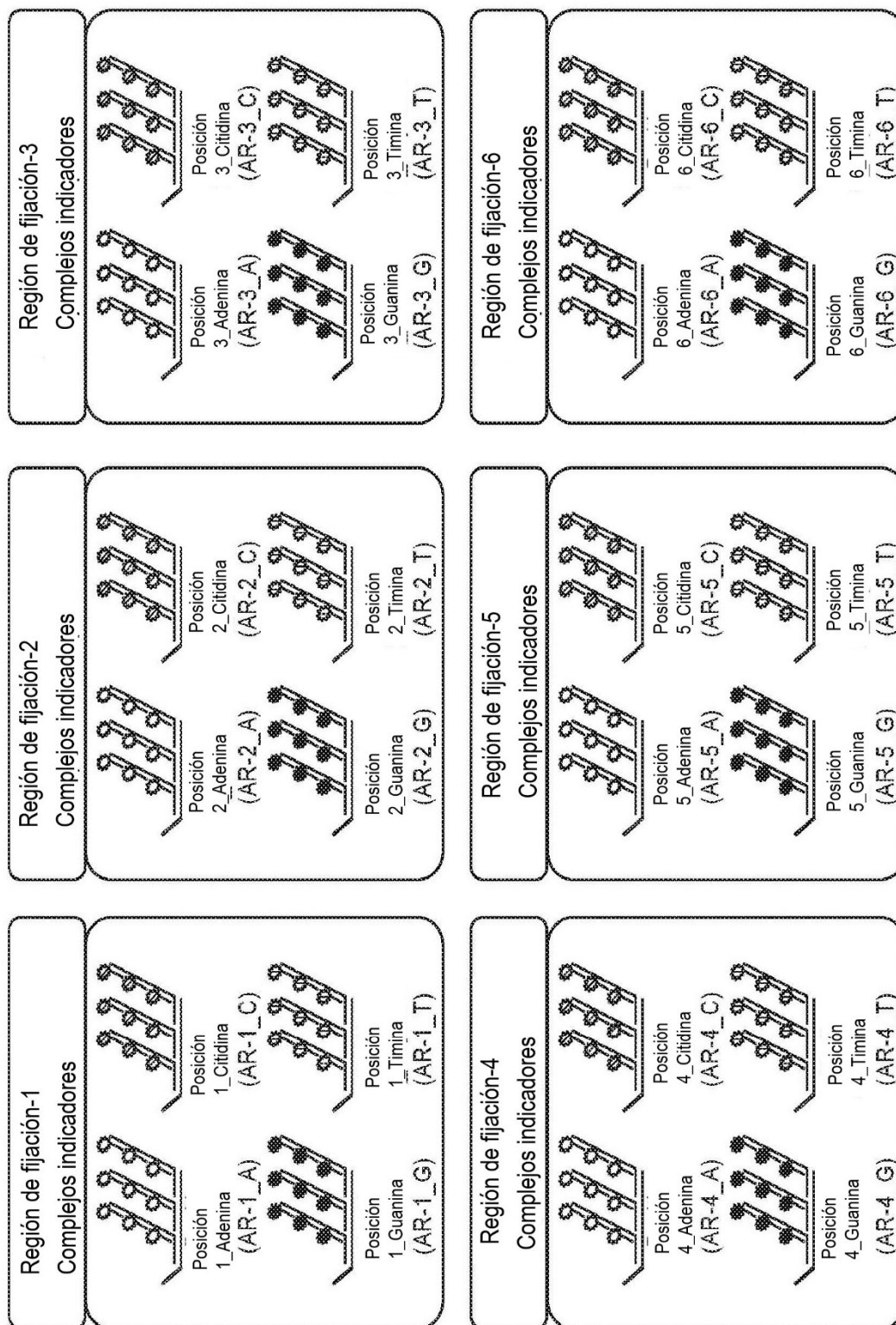


Figura 9D

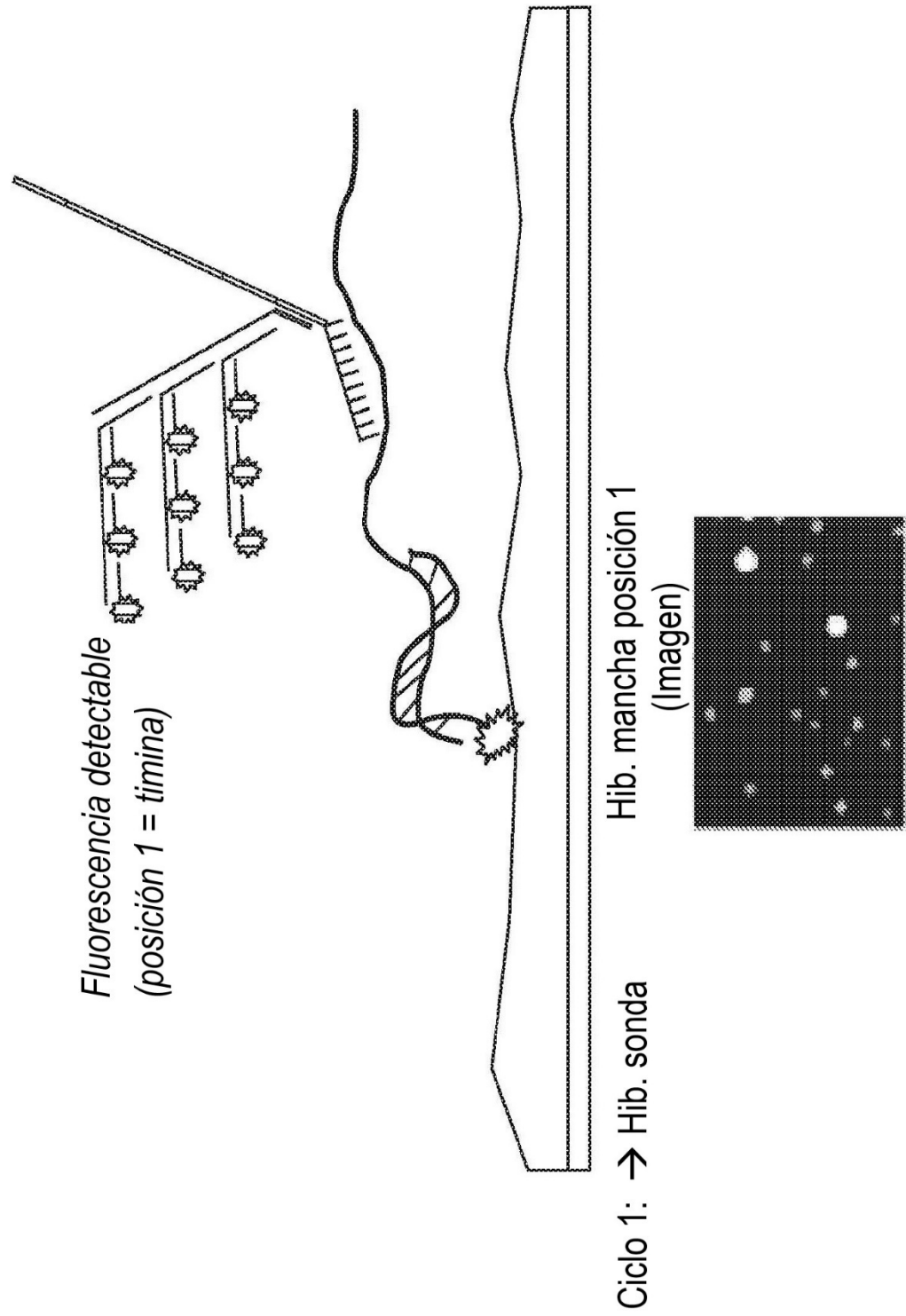


Figura 9E

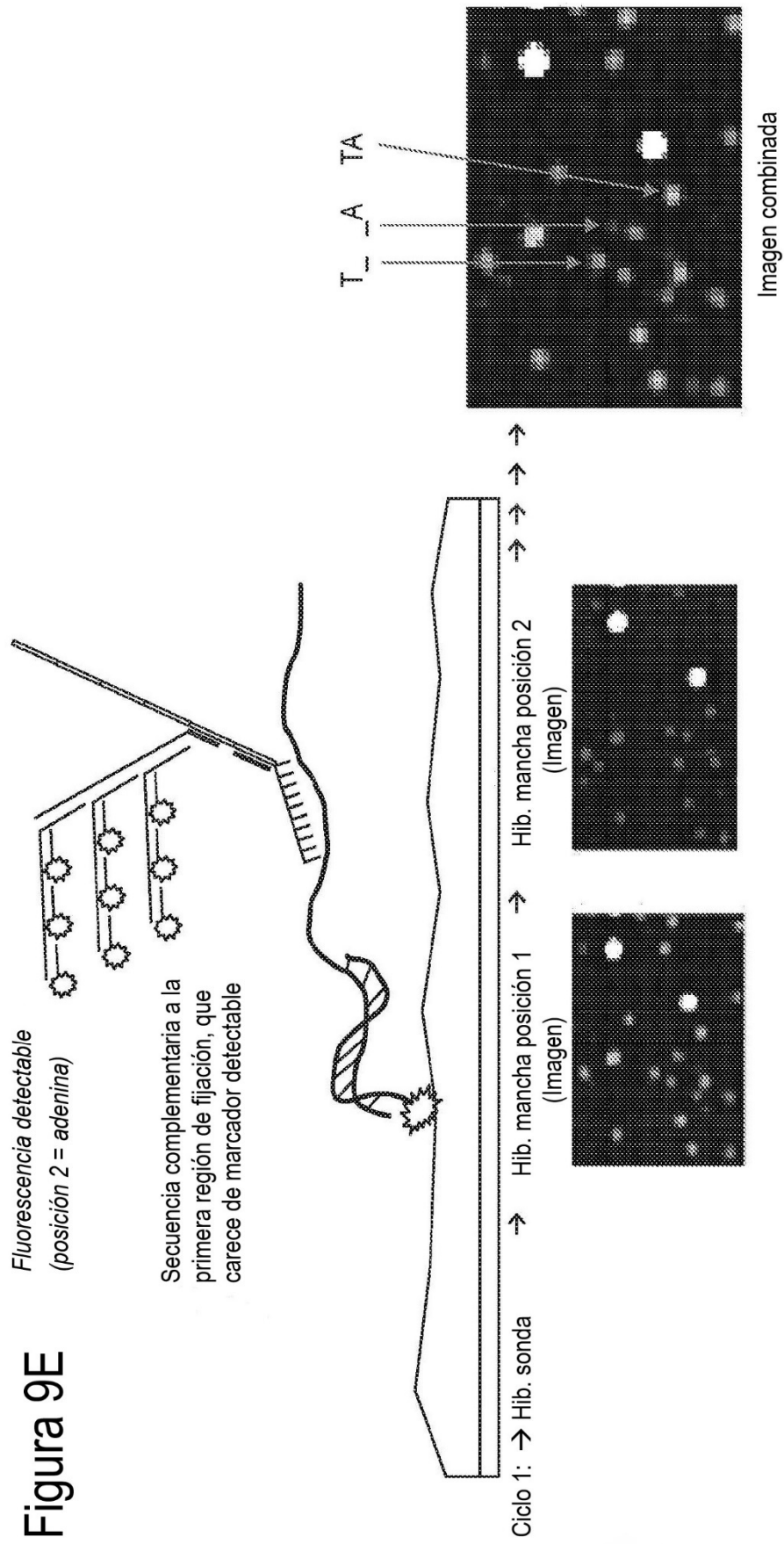


Figura 9F

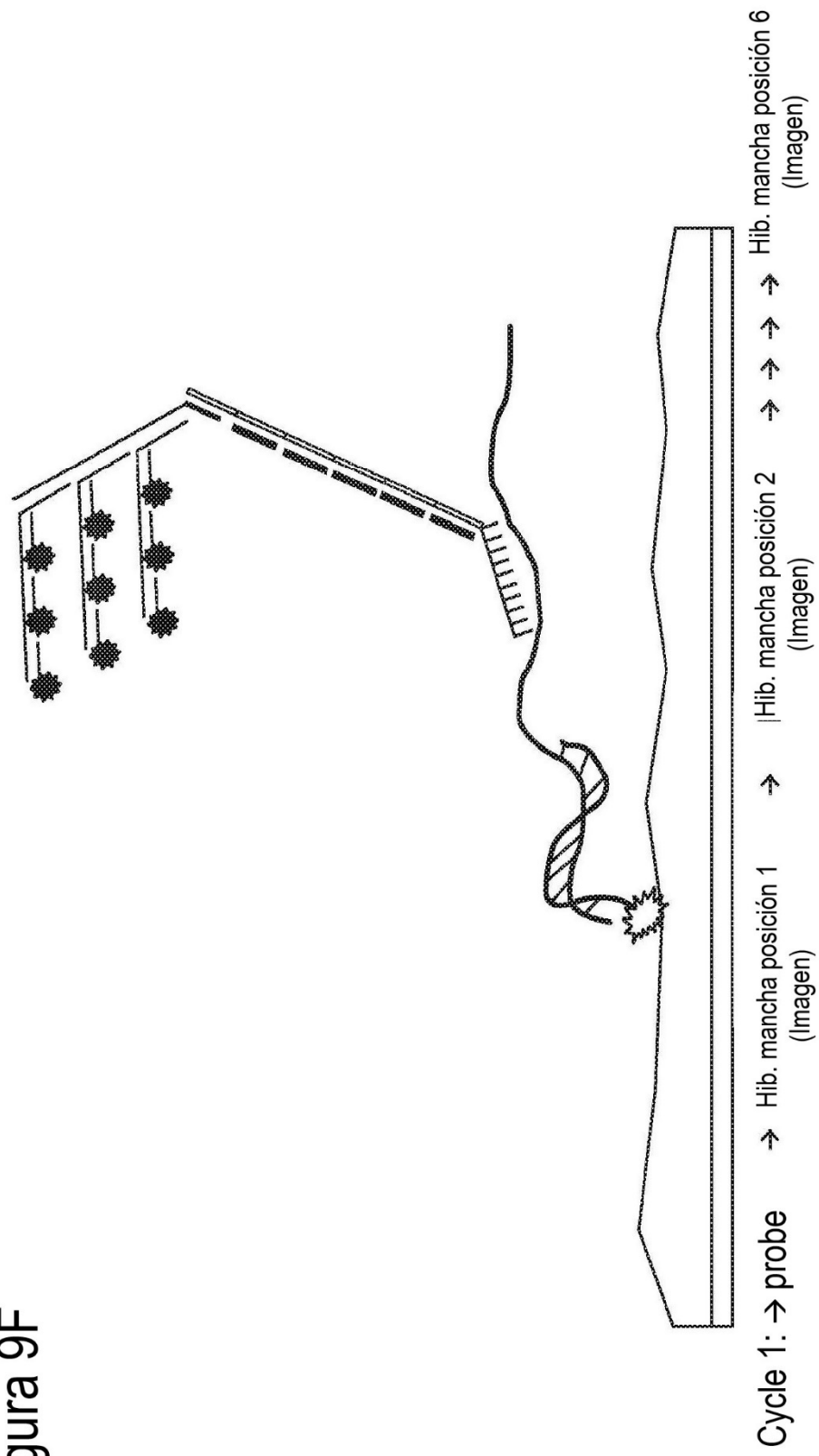
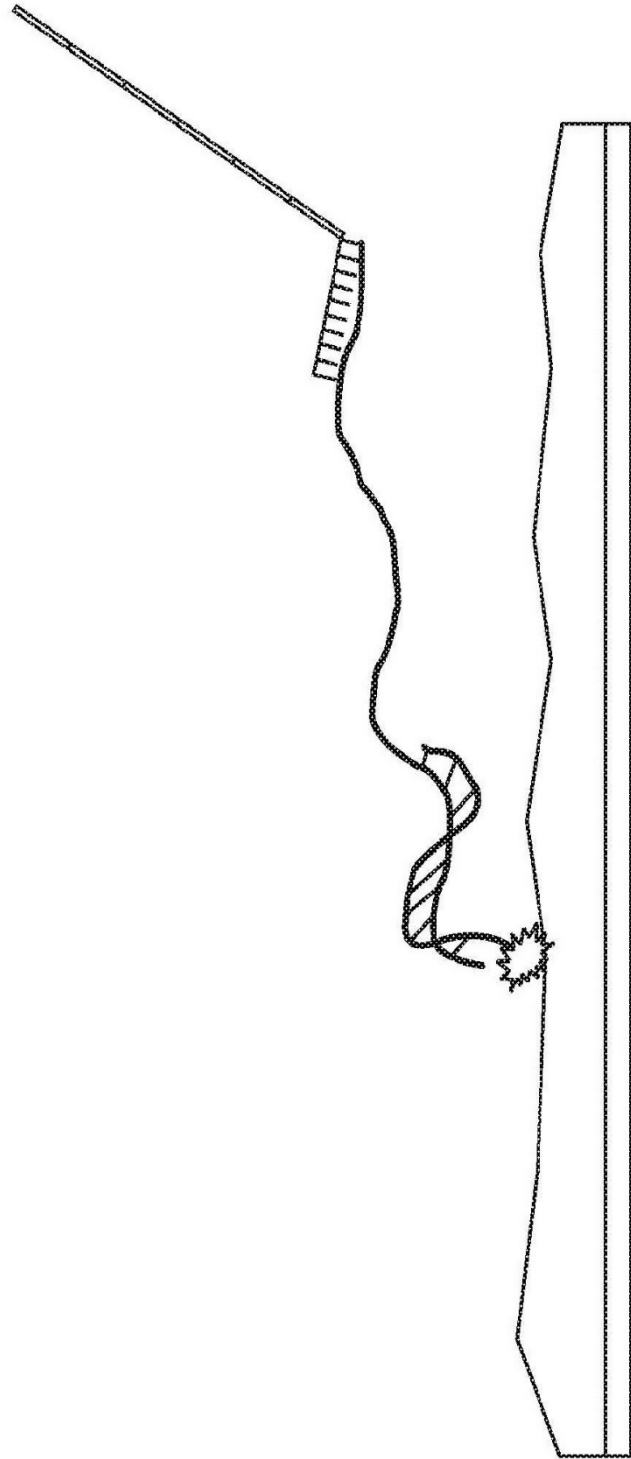
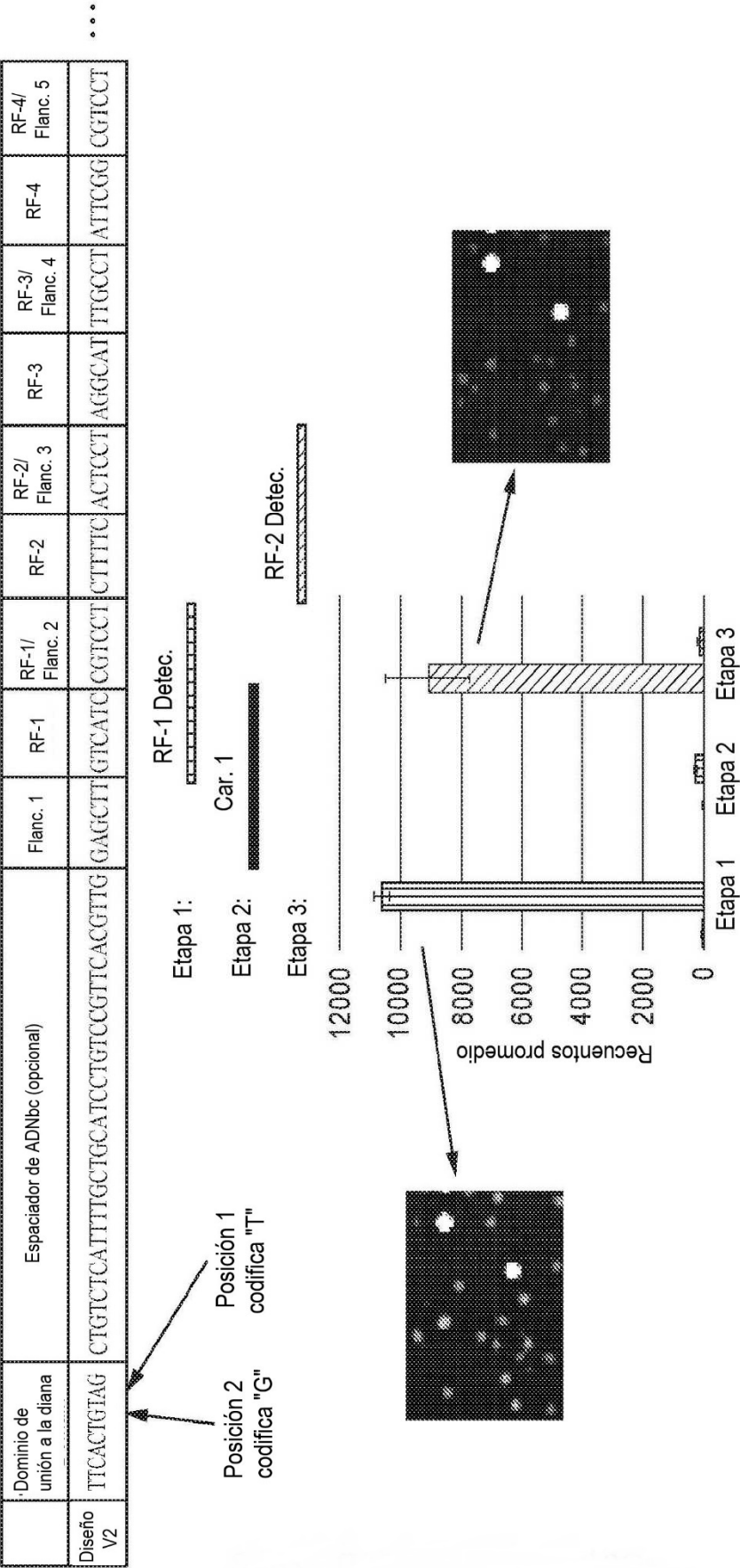


Figura 9G



Ciclo 2: → Hib. segunda sonda

Figura 10



Flanc. = polinucleótido monocatenario flanqueante
RF = región de fijación
Detec. = ácido nucleico complementario que comprende marcador detectable
Car. = ácido nucleico de hibridación que carece de marcador detectable


Figura 11


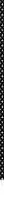

Diseño V2	Dominio	Espaciador de ADNbc (opcional)	Flanc 1	RF-1/ Flanc. 2	RF-2	RF-2/ Flanc. 3	RF-3	RF-3/ Flanc. 4	RF-4	RF-4/ Flanc. 5
	TTCACCTGTAG	CTGTCTCATTTTGTGTCATCTGTGCGGTTCAGTTG	GAGCTT	CGTCCCT	CTTTTC	ACTCCCT	AGGCAT	TTGGCT	ATTCGG	CGTCTT

...

Etapa 1: RF-1 Detec.

Car. 1	+	RF-2 Detec.
		

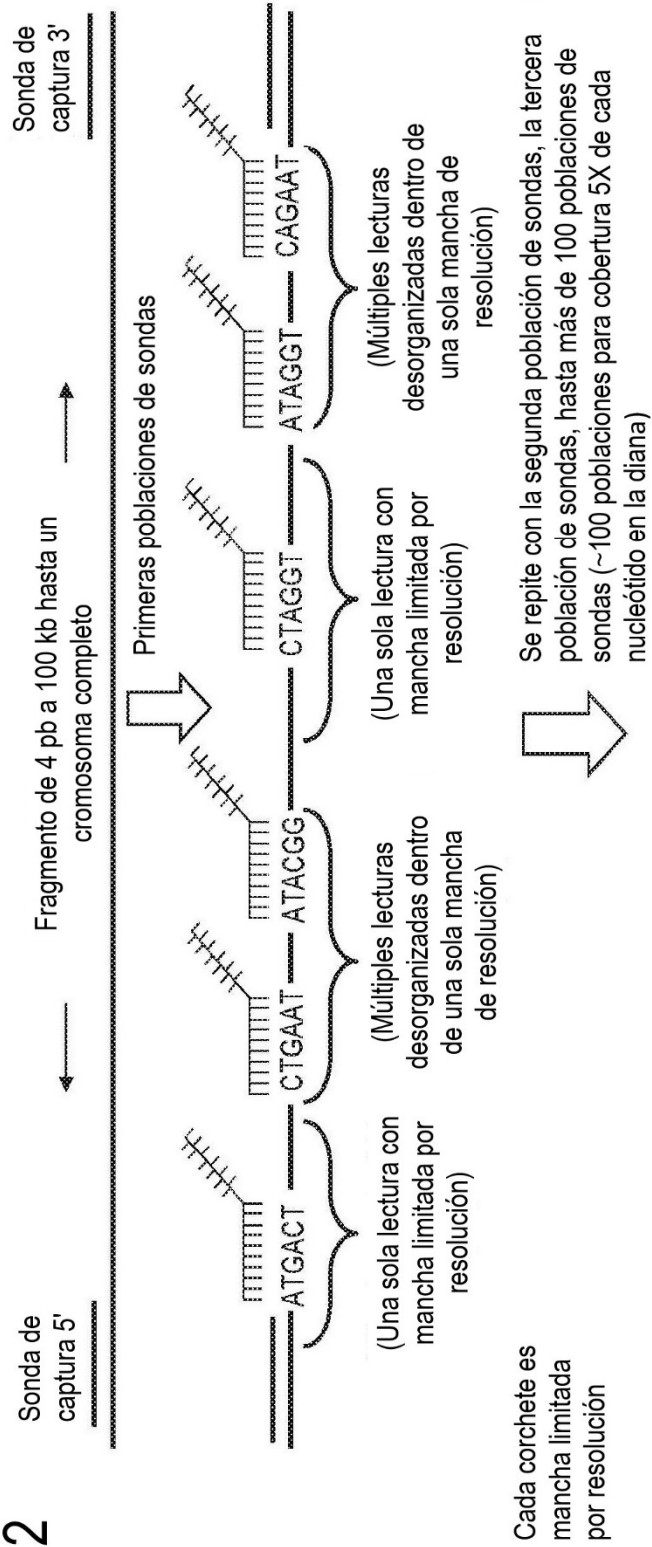
Etapa 2: 

Etapa 3:  Car. 2  RF-3 Detec. 

Etapa 4: Car. 3 + RF-4 Detec.

Etapa nº:

Figura 12



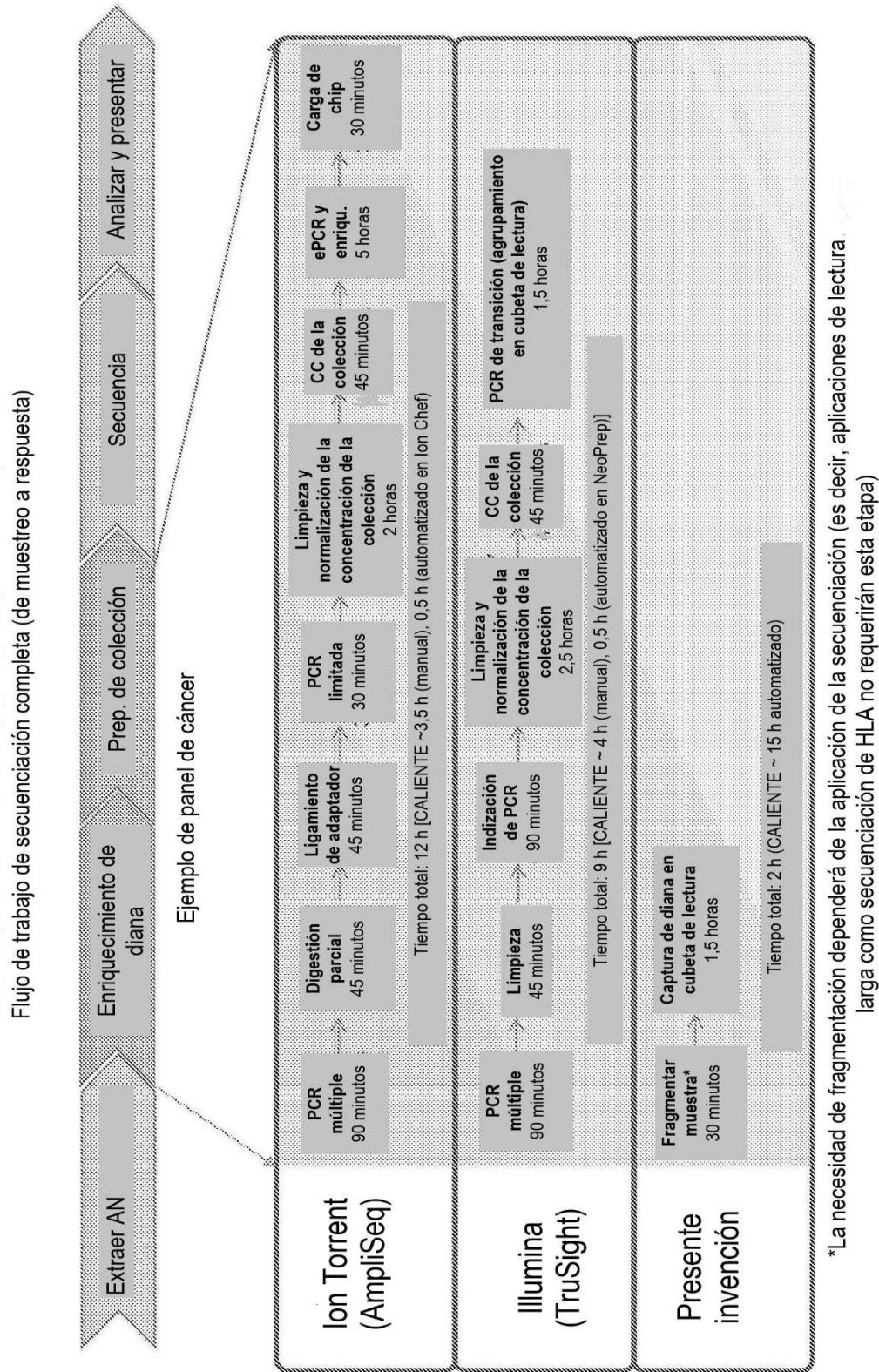
~20 FOV (para cobertura 100X de un gen) X 8 min / FOV = 2,7 horas de ejecución = 3 Gb/24 h (~100 % Miseq)

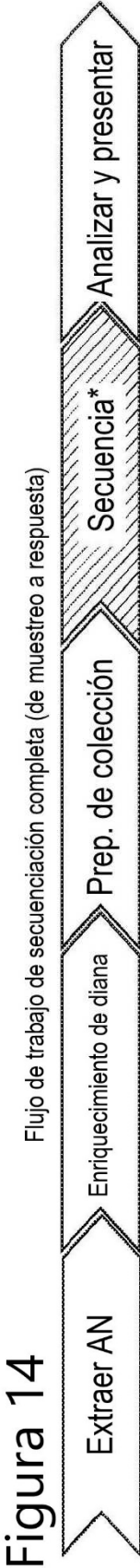
~20 FOV X 5000 genes / FOV = 100 000 lecturas de genes (produce 100X de 1000 genes)

~20 FOV (para cobertura 100X de un gen) X 8 min / FOV = 2,7 horas de ejecución = 12 Gb/24 h* (~400 % Miseq)

*(estimación del rendimiento de NSTG en matriz ordenada)

Figura 13





Calculadora de tiempo de ejecución basada en la aplicación de la secuenciación			
Atributos técnicos	Precisión bruta (%)	98	
	Longitud de lectura (pb)	-	
Requisitos de sec. específicos de la aplicación	Número de regiones diana	100	
	Tamaño de la región diana (pb)	100	
	Umbral de detección de variante	0,01	
	Número de copias de diana requeridas (X)	3000	
	Número total de lecturas requeridas (M)	0.3	
Sondas que detectan 6 posiciones nucleotídicas	Lecturas totales por FOV	250000	
	Número de FOV requeridos	1.2	
	Número total de ciclos de hib. y sec.	430	
	Tiempo en movimiento del equipo (s)	0.5	
	Tiempo de exposición (s)	0.2	
	Tiempo de enfoque (s)	0.3	
	Tiempo total para la imagen por FOV (s)	1	
	Número total de manchas por ciclo de hib. y sec.	6	
	Tiempo de hib. de sonda k-mérica (s)	5	
	Tiempo de intercambio de mancha (s)	3	
	Tiempo de lavado (s)	5	
	Tiempo de reacciones químicas total por ciclo de hib. y sec. (s)	28	
	Tiempo de ejecución de sec. total* (h)	4,62	

Tiempos de ejecución:

céntuple = 4,6 h

múltiple de 1.000 = 16 h

* El tiempo de ejecución de secuenciación incluye el tiempo para procesamiento de los datos primarios (es decir, calificación de bases, asignación de QV, recuento de lecturas de marcador de sonda y almacenamiento)

Calculadora de ciclado basada en ocupación de la sonda

Número total de dianas monomoleculares en la cubeta de lectura	30000	
Longitud promedio de diana "con huecos" (pb)	100	
Ocupación de código de barras (%) por ciclo de sec.	0.75	
Número de ciclos de código de barras requeridos para "interacciones positivas" de >95 % de las dianas en la cubeta de lectura al menos una vez	3	
Ocupación del tinte (%) por posición	0.8	
Número total de posiciones con tinte	6	
Número total de lectura codificada de 1 pb por ciclo de sec.	6	35
Número total de lectura codificada de 2 pb por ciclo de sec.	23	346
Número total de lectura codificada de 3 pb por ciclo de sec.	92	1843
Número total de lectura codificada de 4 pb por ciclo de sec.	369	5530
Número total de lectura codificada de 5 pb por ciclo de sec.	1475	8847
Número total de lectura codificada de 1 pb por ciclo de sec.	5898	5898
Número total de lecturas "utilizables" por ciclo de sec.		22499
Número de ciclos de código de barras requeridos para cubrir una diana monomolecular dada con cobertura ~3X	143	
Número de ciclos de código de barras requeridos para cubrir una diana monomolecular dada con cobertura ~5X	238	
Número de ciclos de código de barras requeridos para cubrir todas las dianas monomoleculares en la cubeta de lectura con cobertura ~3X	430	
Número de ciclos de código de barras requeridos para cubrir todas las dianas monomoleculares en la cubeta de lectura con cobertura ~5X	713	

Figura 15

Figura 16

Gigabases		Tiempo en h	Número de lecturas	Utilidad clínica
Hiseq 2500	900 Gb	11 días	¡3 mil millones!	Secuencia cancerosa en secuencia genómica
Next Seq	120 Gb	27 h	400 millones	Transcriptoma de exomas de secuencia genómica
Proton PI	10 Gb	2-4	70 millones	Transcriptoma de exomas
Presente invención	de 3 a 13 Gb	3	0,5 millones	Dirigida (Dx, cDx) más multianálito más muestreo a respuesta, lecturas largas
MiSeq	3,5 Gb	24 h	25 millones	Dirigida al transcriptoma (Dx)
PGM 318	hasta 2 Gb	7,3/4,4	4 millones	Dirigida
PacBio RS II	2,2 Gb/día	3 h	50 mil	Genomas de nicho, lecturas largas, alta tasa de error
PGM 314	0,1 Gb	3,7/2,3	400 mil	Dirigida

Figura 17

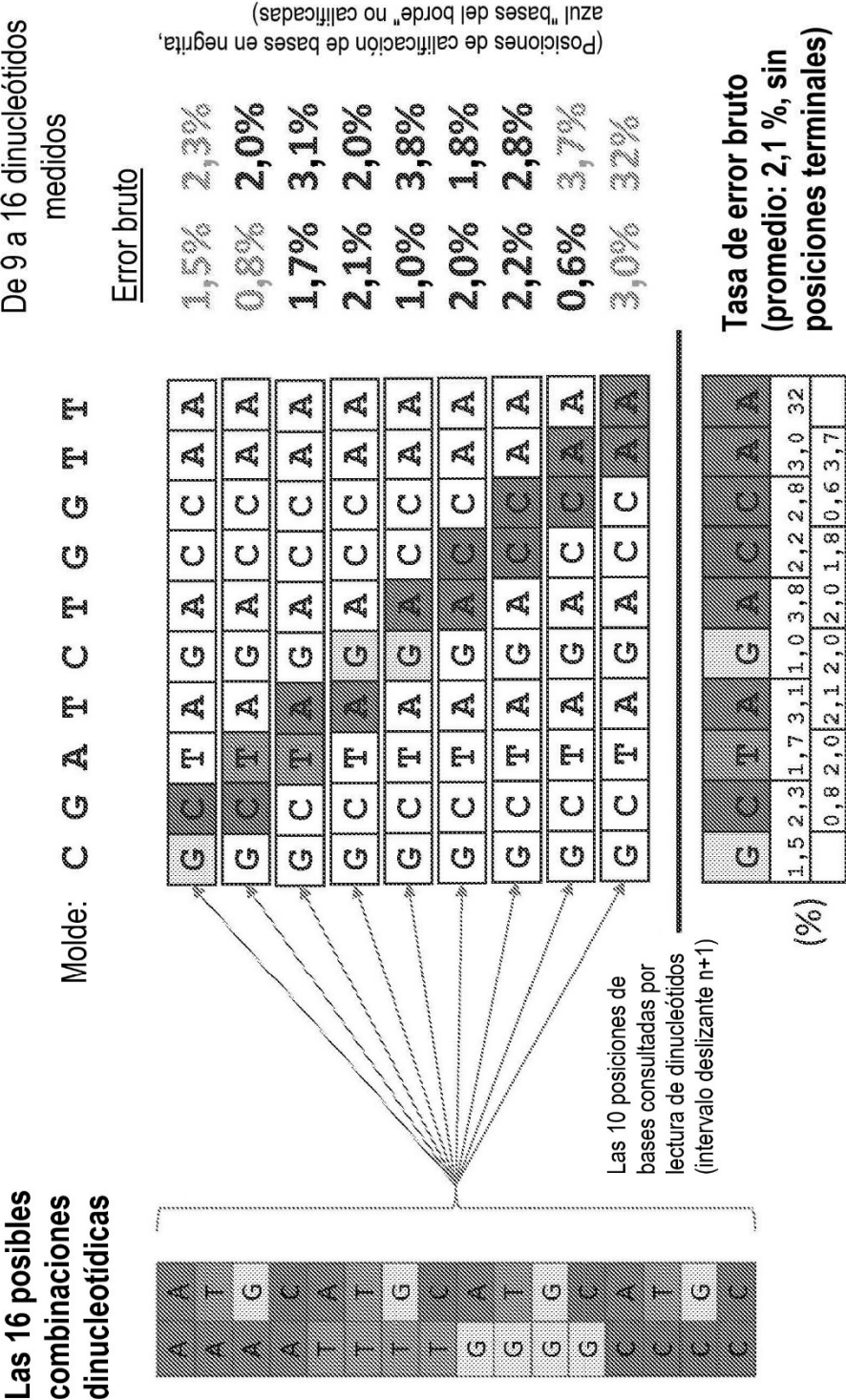


Figura 18

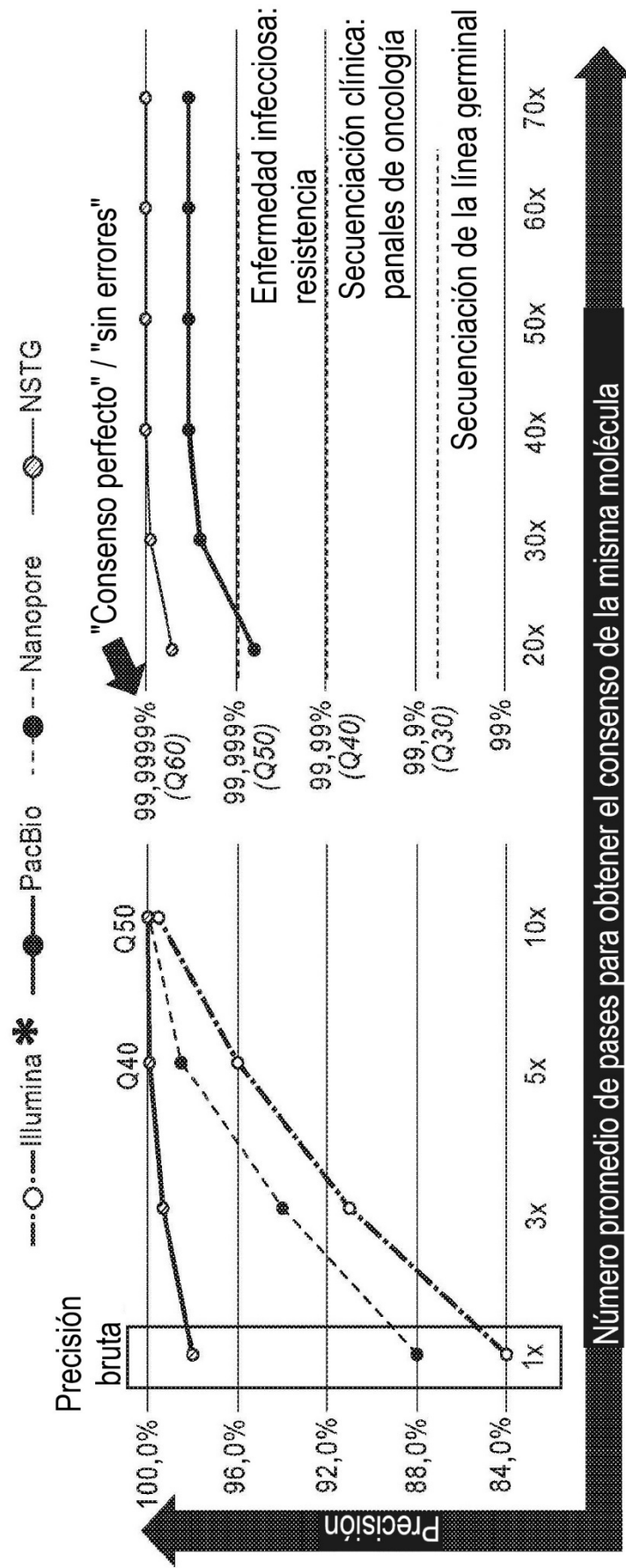
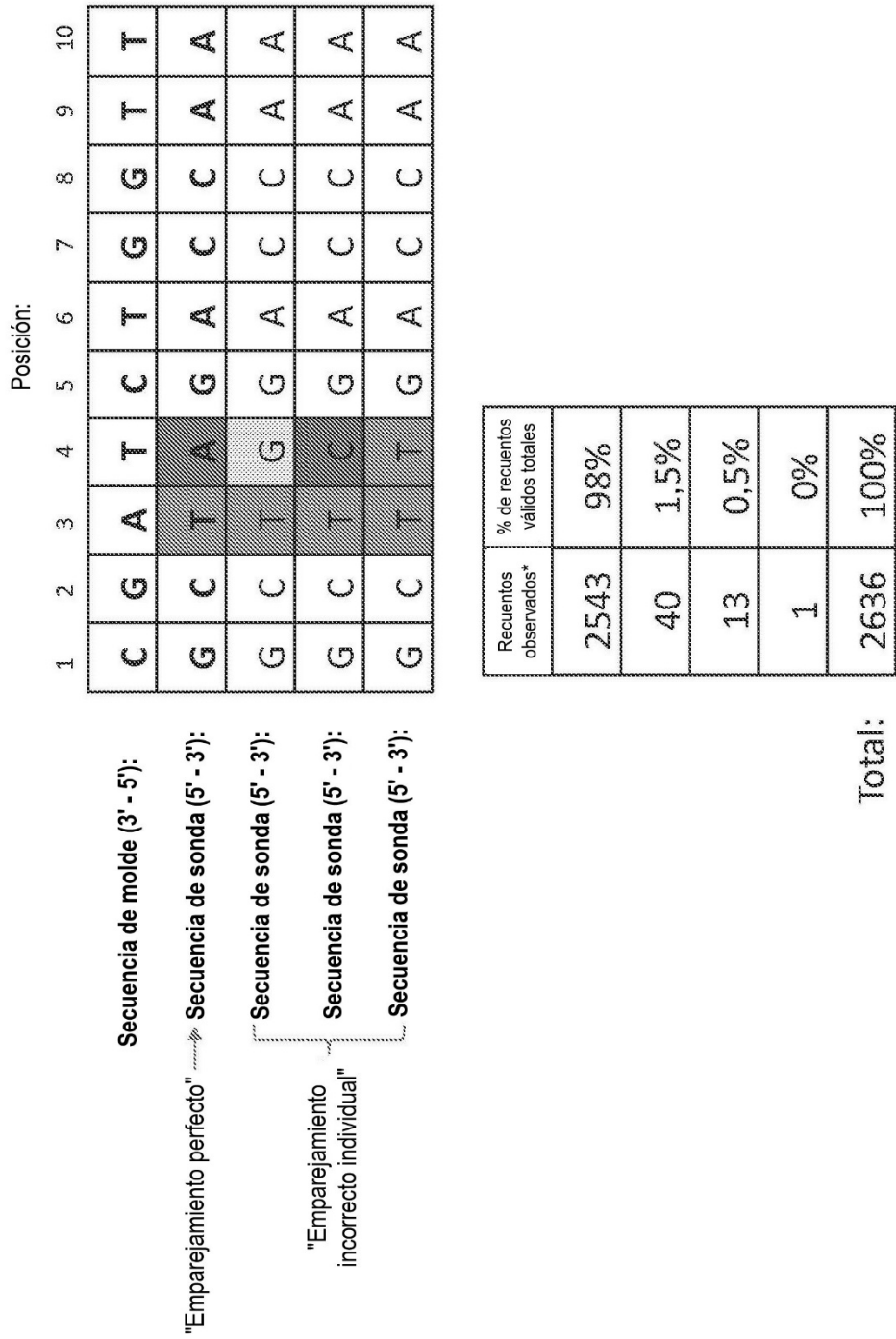


Figura 19



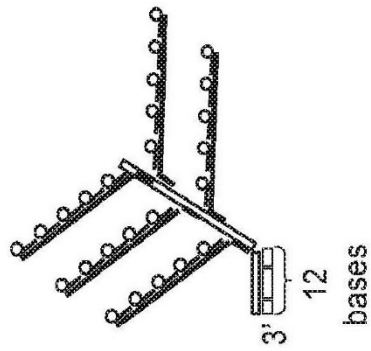
* El valor recuentos observados representa los recuentos después de restar el recuento de fondo

****** Todas las hibridaciones se hicieron en condiciones de tampón: SSPE 1x/Tween20 al 0,1 %

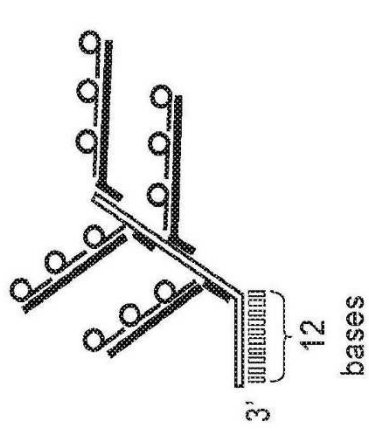
Figura 20A

5x5

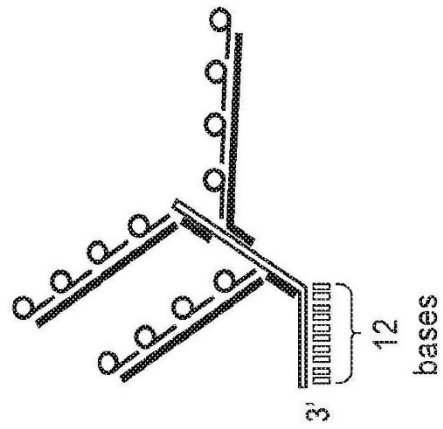
(1 molécula de ácido nucleico primaria, 5 moléculas de ácido nucleico secundarias, 5 moléculas de ácido nucleico terciarias con uno o más marcadores detectables)



4x3



3x4



Espaciador 3x4

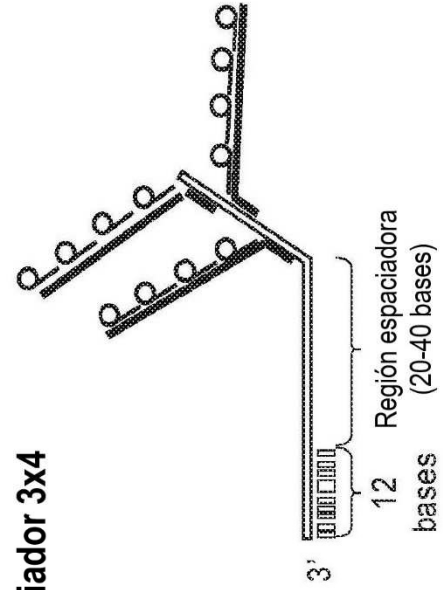


Figure 20B

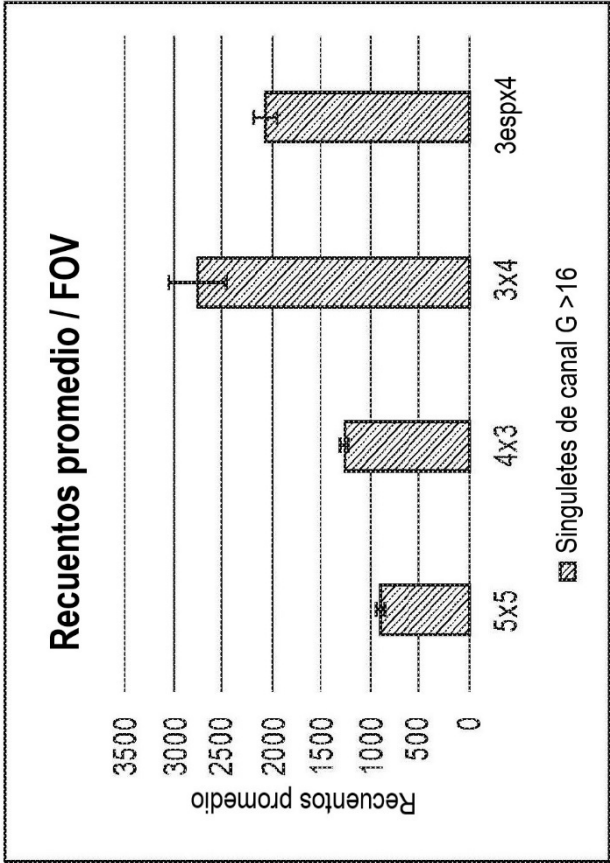


Figura 20C

Receta ejemplar para construir complejos indicadores que comprenden marcadores detectables

Volumen (ul)	Molécula de a.n. primaria (10 uM)	Molécula de a.n. secundaria (10 uM)	Molécula de a.n. terciaria (100 uM)	H2O
5x4	1	4,5	2,25	92,25
5x3	1	4,5	1,8	92,7
4x4	1,28	4,5	2,25	91,97
4x3	1,28	4,5	1,8	92,42
3x4	1,8	4,5	2,25	91,45

Los colores en la tabla corresponden a colores en los diagramas de complejos mostrados a la izquierda

Figura 21A

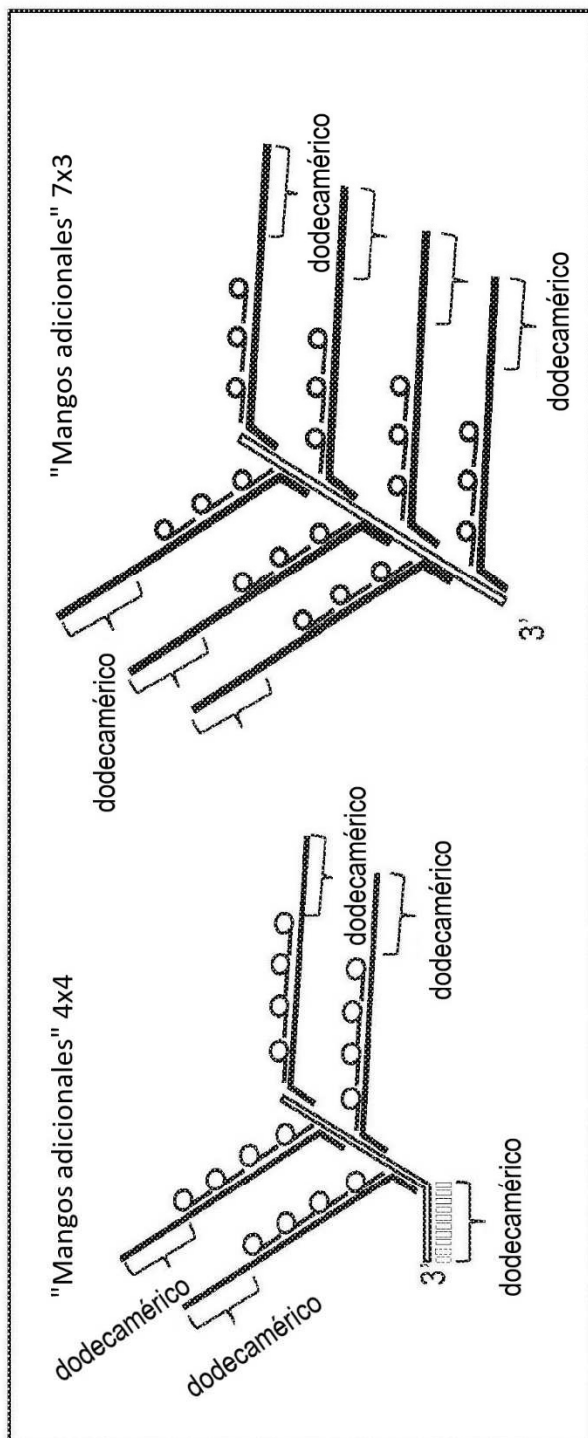


Figura 21B

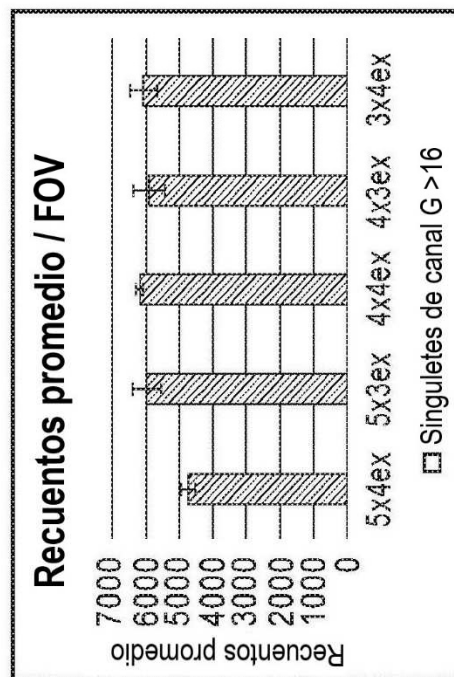


Figura 22A

Cinética de "5x3"

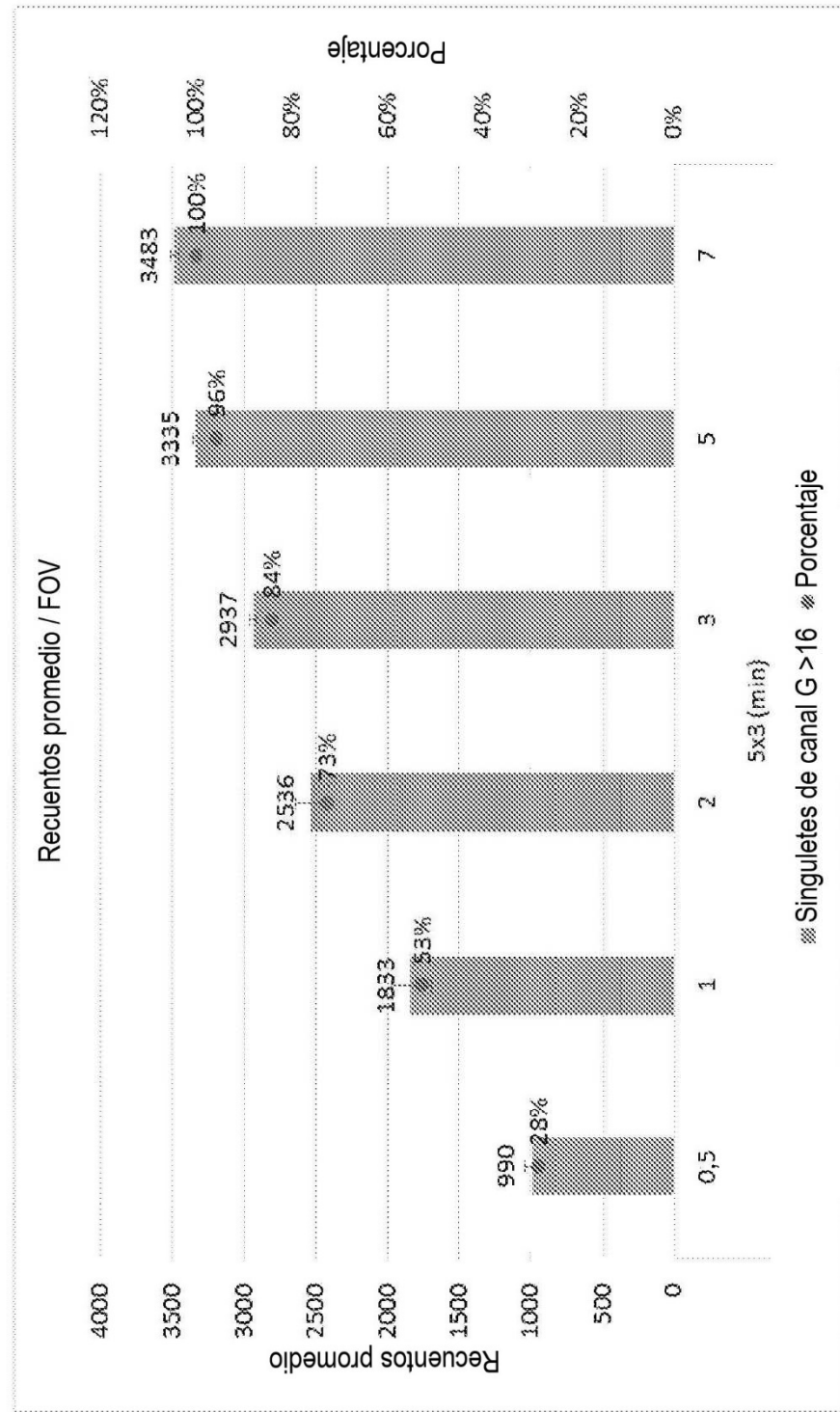


Figura 22B

Cinética de "mangos adicionales 7x3"

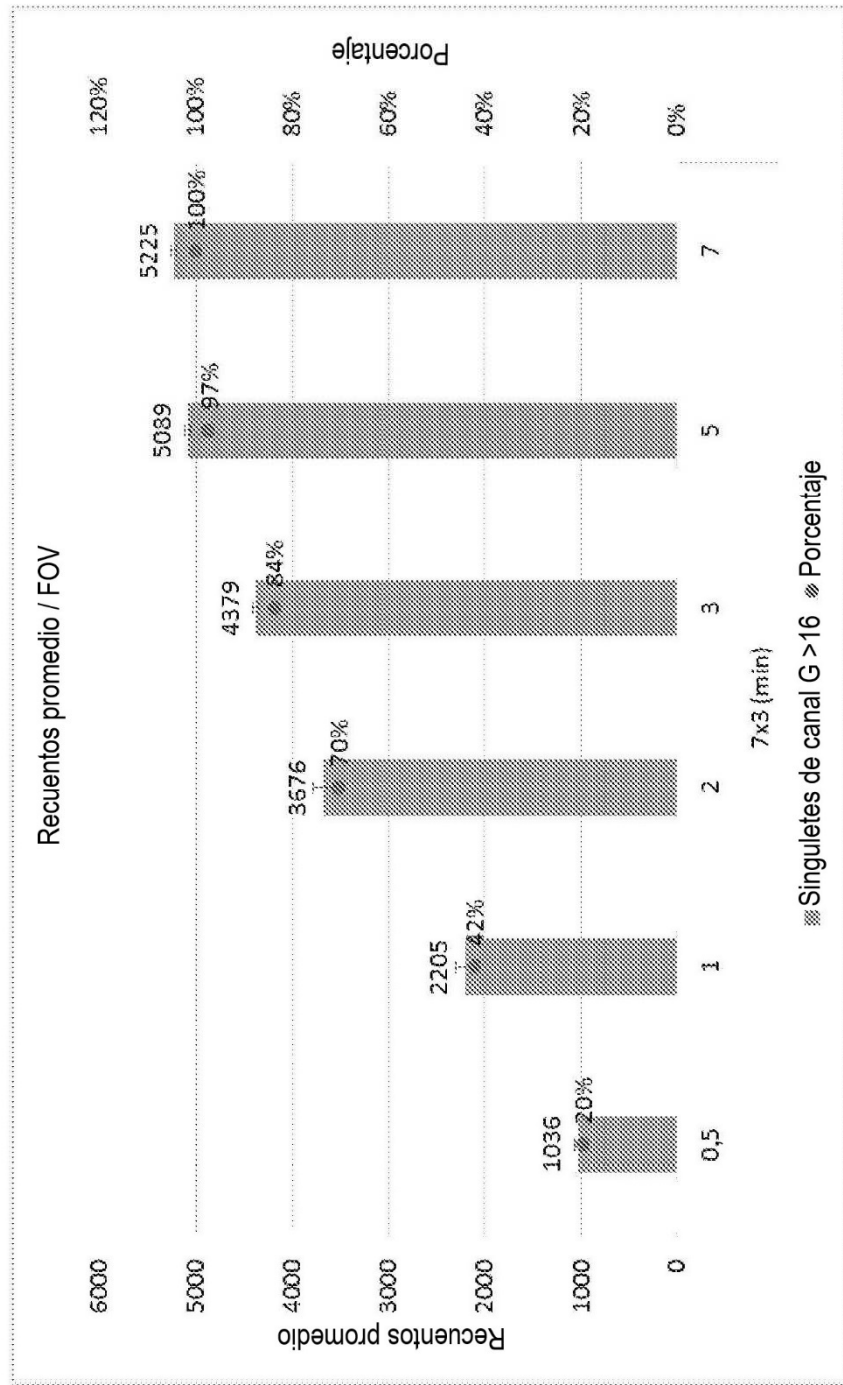


Figura 23

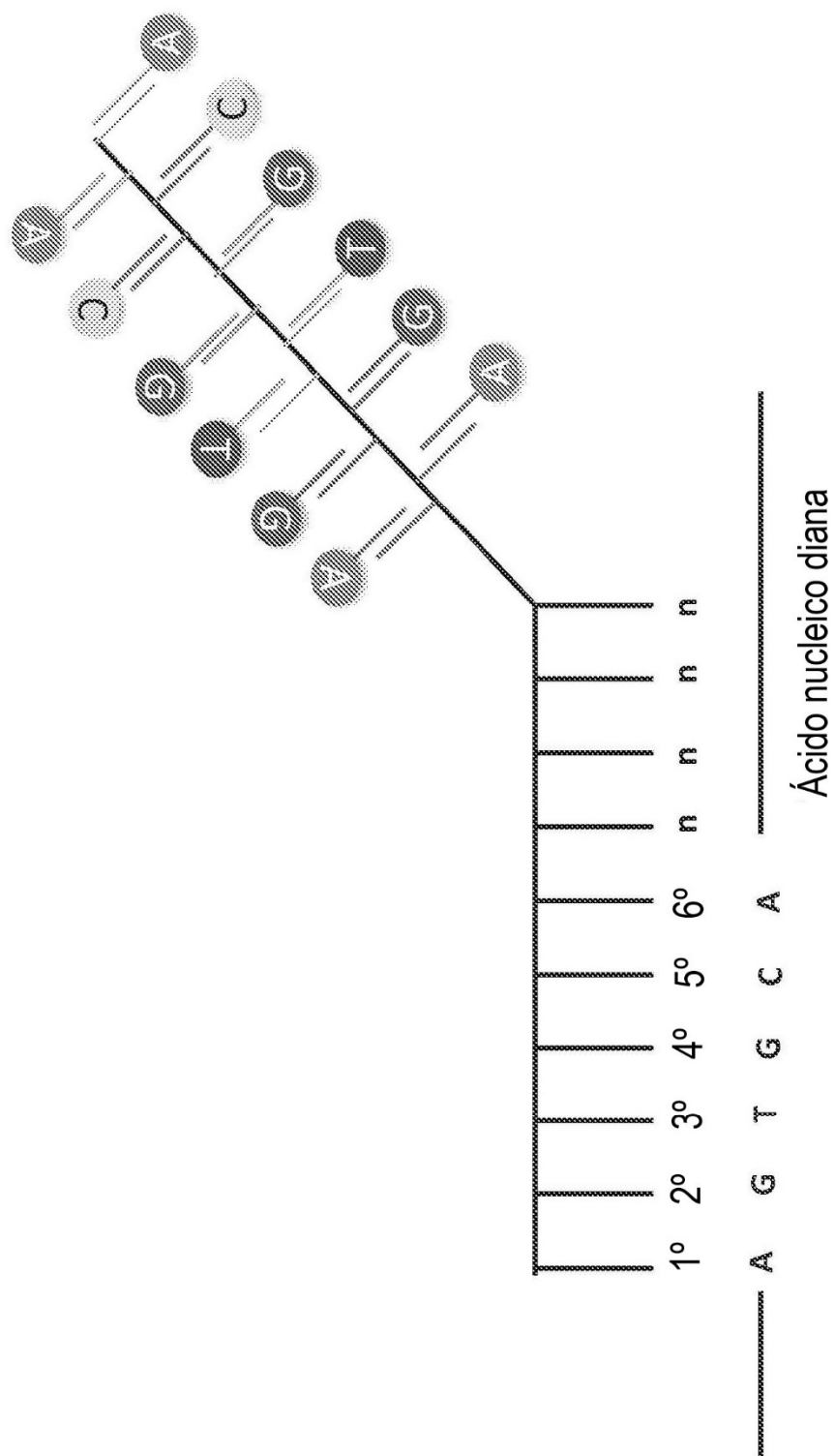


Figura 24

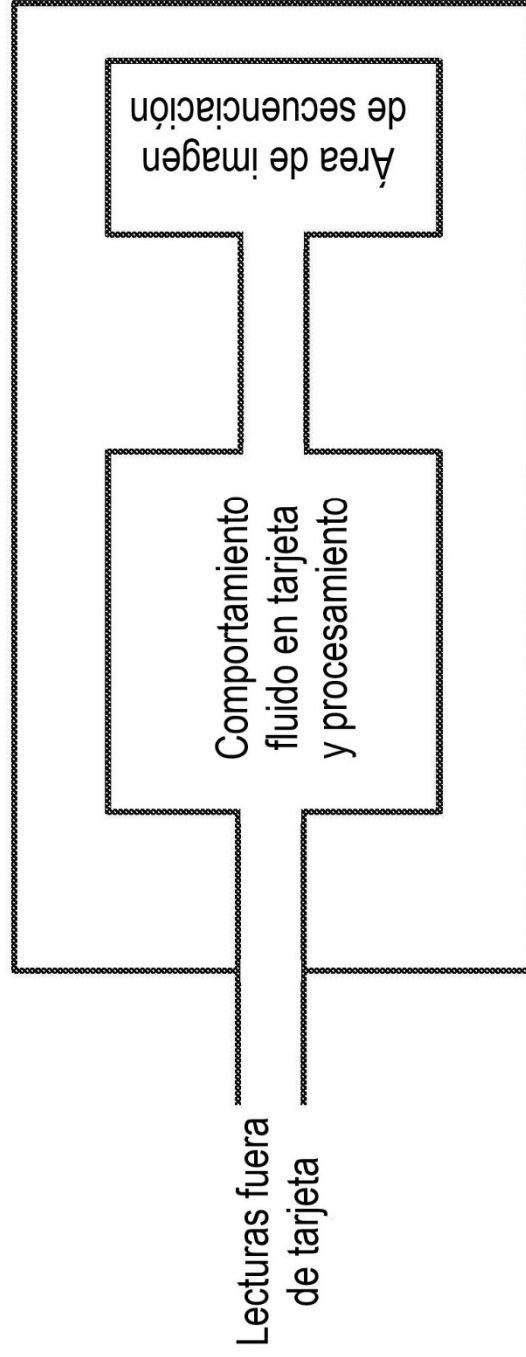


Figura 25

código	secuencias
YGBYGR	GAGACTGTAC - original
BGYRGB	GAGACAGTAC
YBRYGB	GAGACCGTAC
RGBYGR	GAGACCGTAC
YRBGYR	GAGAGTGTAC
GRYGRB	GAGAATGTAC
RGYBRG	GAGATTGTAC

código	recuentos	secuencias
YGBYGR	16074	GAGACTGTAC
RGYBRG	1132	GAGATTGTAC
YBRYGB	998	GAGACCGTAC
YRBGYR	784	GAGAGTGTAC
YBRYBG	712	
YBGRBY	699	
YGRYBG	692	
YBRGBR	674	
Umbral de fondo	717	

Figura 26

ADN	Secuencia	código	recuento
dodecamérico	CTGAGACTGTAC	RYGBRG	80845
undecamérico	TGAGACTGTAC	GBRYGB	24474
decamérico	GAGACTGTAC	YGBYGR	18339
nonamérico	AGACTGTAC	YRBYRG	1667
Especcular de dodecamérico	CATGTCAGAGTC	GRBGYR	1473
Especcular de undecamérico	CATGTCAGAGT	BGYRBG	1384
octamérico	GACTGTAC	GBRYBG	900
		Umbral de fondo	622

Figura 27

Indicador de longitud completa
 Longitud completa **20 pM** (6 manchas) **33** recuentos/FOV

Indicador de una sola mancha	
Mancha individual [fM]	Recuentos/FOV
20.000	Saturado
2.000	12000
200	308
20	40
2	11

Indicador NS 20 pM \simeq mancha individual 20 fM