(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 770 055**

(51) Int. CI.:

C12Q 1/6869 (2008.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(86) Fecha de presentación y número de la solicitud internacional: **19.11.2015 PCT/US2015/061615**

(87) Fecha y número de publicación internacional: **26.05.2016 WO16081740**

(96) Fecha de presentación y número de la solicitud europea: **19.11.2015 E 15801655 (0)**

(97) Fecha y número de publicación de la concesión europea: **15.01.2020 EP 3221469**

(54) Título: **Secuenciación sin enzima ni amplificación**

(30) Prioridad:

21.11.2014 US 201462082883 P

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
30.06.2020

(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:

SÁEZ MAESO, Ana

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 enzima ni amplificación

Antecedentes de la invención

Actualmente hay una variedad de métodos para la secuenciación de ácido nucleico, 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 de luz. Tales etapas de amplificación y polimerización son costosas y/o requieren mucho tiempo. Por lo tanto, existe una necesidad en la técnica de un método de secuenciación de ácido nucleico que esté libre de amplificación y enzimas. El documento WO 2013/055995 A2 divulga una sonda de secuenciación que tiene un dominio de unión diana hexámero (N1-N6) y una estructura de estructura de ácido nucleico no hibridante (TNH) que tiene seis sitios de hibridación de sonda (B1-B6), en donde cada uno de los B1-B6 en el TNH corresponden a uno de N1-N6. El documento US 7 745 129 B1 divulga una sonda de secuenciación que tiene uno o más dominios de sonda que se unen a un ácido nucleico diana y un dominio de ensamblaje de etiqueta que tiene un indicador de inicio, regiones codificadoras de bases y un indicador final, en el que cada una de las regiones codificadoras de bases en el dominio de ensamblaje de etiqueta corresponde a nucleótidos dentro de los dominios de sonda. Las regiones codificadoras de bases pueden contener sitios de unión a proteínas o sitios de hibridación de sonda. La presente invención aborda estas necesidades.

Resumen de la invención

La presente invención, tal como se define en las reivindicaciones adjuntas, proporciona sondas de secuenciación, métodos y kits que proporcionan secuenciación de ácido nucleico libre de enzimas, libre de amplificación y libre de bibliotecas que tiene longitudes de lectura largas y con bajo índice error. Además, los métodos y kits tienen una rápida capacidad de muestra a respuesta. Estas características son particularmente útiles para la secuenciación en un entorno clínico. En el presente documento se proporcionan sondas de secuenciación que comprenden un dominio de unión diana y un dominio de código de barras. El dominio de unión diana y el dominio del código de barras pueden estar operativamente unidos, por ejemplo, unidos covalentemente. Una sonda de secuenciación opcionalmente comprende un espaciador entre el dominio de unión diana y el dominio del código de barras. El espaciador puede ser cualquier polímero con propiedades mecánicas apropiadas, por ejemplo, un espaciador de ADN de cadena sencilla o de cadena doble (de 1 a 100 nucleótidos, por ejemplo, de 2 a 50 nucleótidos). Los ejemplos no limitantes de espaciadores de ADN de cadena dobles incluyen las secuencias cubiertas por SEQ ID NO: 25 a SEQ ID NO: 29.

El dominio de unión diana comprende al menos ocho nucleótidos (por ejemplo, 9, 10, 11, 12 o más) y es capaz de unirse a un ácido nucleico diana (por ejemplo, ADN, ARN y PNA), y en el que al menos dos nucleótidos en el dominio de unión diana son bases universales capaces de unirse a cualquier base dentro del ácido nucleico diana. El dominio de código de barras comprende una cadena principal sintética, el dominio de código de barras comprende al menos seis posiciones de unión, por ejemplo, siete, ocho, nueve, diez, once, doce o más posiciones de unión; cada posición de unión tiene al menos una o más (por ejemplo, una a cincuenta) regiones de unión; dicha región de unión comprende al menos una (es decir, una a cincuenta, por ejemplo, diez a treinta copias de una) secuencia(s) de ácido nucleico capaz de unirse por una molécula de ácido nucleico complementaria (ARN o ADN). Cada posición de unión de las al menos seis posiciones de unión corresponde a un nucleótido en el dominio de unión diana y cada una de las al menos seis posiciones de unión tiene una secuencia de ácido nucleico diferente, en el que dicha secuencia de ácido nucleico de cada posición de las al menos seis posiciones de unión determina la posición y la identidad del nucleótido correspondiente en dicho ácido nucleico diana que está unido por dicho dominio de unión diana, y en el que las bases universales preceden y siguen a los nucleótidos correspondientes a las posiciones en el dominio de código de barras. Ciertas posiciones en un dominio de código de barras pueden tener más regiones de conexión que otras posiciones; alternativamente, cada posición en un dominio de código de barras tiene el mismo número de regiones de conexión. La secuencia de ácido nucleico de una primera región de unión determina la posición y la identidad de un primer nucleótido en el ácido nucleico diana que está unido por un primer nucleótido del dominio de unión a la diana, mientras que la secuencia de ácido nucleico de una segunda región de unión determina la posición y identidad de un segundo nucleótido en el ácido nucleico diana que está unido por un segundo nucleótido del dominio de unión diana. Del mismo modo, la secuencia de ácido nucleico de una sexta región de unión determina la posición y la identidad de un sexto nucleótido en el ácido nucleico diana que está unido por un sexto nucleótido del dominio de unión diana. En realizaciones, la cadena principal sintética comprende un polisacárido, un polinucleótido (por ejemplo, ADN o ARN de cadena sencilla o doble), un péptido, un ácido nucleico peptídico o un polipéptido. El número de nucleótidos en un dominio de unión diana es igual o mayor que (por ejemplo, 1, 2, 3, 4 o más) el número de posiciones en el dominio de código de barras. Cada región de unión 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 unión incluirá una secuencia de ácido nucleico diferente que una región de unión en una posición diferente del dominio del código de barras, incluso cuando ambas regiones de unión identifican el mismo tipo de nucleótido, por ejemplo, adenina, timina, citosina, guanina, uracilo y análogos de los mismos. Una región de unión puede estar unida a un monómero modificado, por ejemplo, un nucleótido modificado, en la cadena principal sintética, creando así una rama con respecto a la cadena principal sintética. Una región de unión puede ser parte de una

secuencia de polinucleótidos de la cadena principal sintética. Una o más regiones de unión pueden estar adyacentes a al menos un polinucleótido de cadena sencilla flanqueante, es decir, una región de unión puede estar operativamente unida a un polinucleótido de cadena sencilla flanqueante 5' y/o a un polinucleótido de cadena sencilla flanqueante 3'. Una región de unión con o sin uno o dos polinucleótidos de cadena sencillas flanqueantes puede hibridarse con una molécula de ácido nucleico hibridante que carece de una etiqueta detectable. Una molécula de ácido nucleico hibridante que carece de una etiqueta detectable puede tener entre aproximadamente 4 y aproximadamente 20 nucleótidos de longitud, por ejemplo, 12 nucleótidos o más.

Una región de unión puede estar unida por un ácido nucleico complementario que comprende una etiqueta detectable. Cada ácido nucleico complementario puede comprender una etiqueta detectable.

Alternativamente, una región de unión puede estar unida por un ácido nucleico complementario que es parte de un complejo indicador (que comprende etiquetas detectables). Un ácido nucleico complementario (que comprende una etiqueta detectable o de un complejo indicador) puede tener una longitud de aproximadamente 4 y aproximadamente 20 nucleótidos, por ejemplo, aproximadamente 8, 10, 12 y 14 nucleótidos, o más. En un complejo indicador, un ácido nucleico complementario está ligado (directa o indirectamente) a una molécula de ácido nucleico primario. Un ácido nucleico complementario puede ligarse indirectamente a una molécula de ácido nucleico primario a través de un conector de ácido nucleico de cadena sencilla o doble (por ejemplo, un polinucleótido que comprende de 1 a 100 nucleótidos). Un ácido nucleico primario se hibrida con uno o más ácidos nucleicos secundarios (por ejemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más). 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 etiquetas detectables. Uno o cada ácido nucleico secundario puede comprender una región que no se hibrida con una molécula de ácido nucleico primario y no se hibrida con una molécula de ácido nucleico terciario (un “asa adicional”); esta región puede tener cuatro o más (por ejemplo, aproximadamente 6 a aproximadamente 40, por ejemplo, aproximadamente 8, 10, 12 y 14) nucleótidos de longitud. La región que no se hibrida con una molécula de ácido nucleico primario y no se hibrida con una molécula de ácido nucleico terciario puede comprender la secuencia de nucleótidos de la molécula de ácido nucleico complementaria que está unida a la molécula de ácido nucleico primario. Esta región puede estar ubicada cerca del extremo del ácido nucleico secundario distal a su extremo que se hibrida con el ácido nucleico primario. Al tener “asas adicionales” que comprenden la secuencia de nucleótidos del ácido nucleico complementario, la probabilidad y la velocidad a la que un complejo indicador se une a una sonda de secuenciación aumenta considerablemente. En cualquier realización o aspecto de la presente invención, cuando un complejo indicador comprende “asas adicionales”, el complejo indicador puede hibridarse con una sonda de secuenciación a través del ácido nucleico complementario del complejo indicador o mediante el “asa adicional”. Por lo tanto, por ejemplo, la frase “que se une a la primera región de unión ... una primera molécula complementaria de ácido nucleico de un primer complejo indicador “se entendería de acuerdo con su significado simple y también se entendería que significa” unión a la primera región de unión ... un 'manejo adicional' de un primer complejo indicador”.

En las realizaciones, los términos “dominio de código de barras” y “cadena principal sintética” son sinónimos.

En el presente documento, tal como se define en las reivindicaciones adjuntas, se encuentra un método para secuenciar un ácido nucleico utilizando una sonda de secuenciación de la presente invención. Cada uno (por ejemplo, primero, segundo, tercero, cuarto, quinto, sexto, séptimo, octavo, noveno, décimo o superior) molécula de ácido nucleico complementario (que tiene una etiqueta detectable o parte de un complejo indicador) tiene la misma secuencia de ácido nucleico que su correspondiente (es decir, primero, segundo, tercero, cuarto, quinto, sexto, séptimo, octavo, noveno, décimo o superior) hibridando la molécula de ácido nucleico que carece de una etiqueta detectable. El ácido nucleico diana se inmoviliza en un sustrato uniendo 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 etiqueta de afinidad que se une selectivamente a un sustrato. Las posiciones primera y/o segunda 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 un dispositivo microfluídico (por ejemplo, recubierto con estreptavidina). Otras posiciones que se encuentran distantes de un extremo de un ácido nucleico diana pueden unirse selectivamente al sustrato. El ácido nucleico puede alargarse aplicando una fuerza (por ejemplo, gravedad, fuerza hidrodinámica, fuerza electromagnética, estiramiento por flujo, una técnica de menisco en retroceso y combinaciones de los mismos) suficiente para extender el ácido nucleico diana.

En el presente documento se proporciona un método para secuenciar un ácido nucleico utilizando una población de sondas de la presente invención o una pluralidad de poblaciones de sondas de la presente invención. El método comprende etapas de: (1) hibridar al menos una primera población de primeras sondas de secuenciación que comprenden una pluralidad de la sonda de secuenciación de la presente invención a un ácido nucleico diana que se inmoviliza en un sustrato, en el que el ácido nucleico diana se inmoviliza a el sustrato en una o más posiciones (con cada sonda de secuenciación en la primera población deshibridando del ácido nucleico diana inmovilizado en aproximadamente las mismas condiciones, por ejemplo, nivel de agente caotrópico, temperatura, concentración de sal, pH y fuerza hidrodinámica); (2) unir una primera molécula complementaria de ácido nucleico que comprende una etiqueta detectable o una primera molécula complementaria de ácido nucleico de un primer complejo indicador que comprende una etiqueta detectable a una primera posición de unión de las al menos seis posiciones de unión; (3) detectar la etiqueta detectable de la primera molécula de ácido nucleico complementaria unida o la etiqueta 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) unirse a la primera posición de unión una primera molécula de ácido nucleico hibridante que carece de una etiqueta detectable, desuniendo así las primeras moléculas de ácido nucleico complementarias que comprenden una etiqueta detectable o la primera molécula de ácido nucleico complementario del primer complejo indicador que comprende una etiqueta detectable; (6) unir una segunda molécula complementaria de ácido nucleico que comprende una etiqueta detectable o una segunda molécula complementaria de ácido nucleico de un segundo complejo indicador que comprende una etiqueta detectable a una segunda posición de unión de las al menos seis posiciones de unión; (7) detectar la etiqueta detectable de la segunda molécula de ácido nucleico complementario unido o la etiqueta detectable de la segunda molécula de ácido nucleico complementario unido del segundo complejo indicador; (8) identificar la posición e identidad de un segundo nucleótido en el ácido nucleico diana inmovilizado; (9) repetir los pasos (5) a (8) hasta que cada posición de unión de las al menos seis posiciones de unión haya sido unida por una molécula complementaria de ácido nucleico que comprende una etiqueta detectable o una molécula complementaria de ácido nucleico de un complejo indicador que comprende una etiqueta detectable, y se ha detectado la etiqueta detectable de la molécula complementaria unida de ácido nucleico o la etiqueta detectable de la molécula complementaria unida de ácido nucleico de un complejo indicador, identificando así el orden lineal de al menos seis nucleótidos para al menos una primera región del ácido nucleico inmovilizado de diana que se hibridó mediante el dominio de unión diana de la sonda de secuenciación; y (10) deshibridar al menos una primera población de primeras sondas de secuenciación del ácido nucleico diana inmovilizado. Los pasos (5) y (6) pueden ocurrir concurrente o secuencialmente. De este modo, se identifica el orden lineal de los nucleótidos para las regiones del ácido nucleico diana inmovilizado que se hibridaron mediante el dominio de unión diana de las sondas de secuenciación en la primera población de sondas de secuenciación.

En las realizaciones, cuando se usa una pluralidad de poblaciones (es decir, más de una población) de sondas, el método comprende además etapas de: (11) hibridar al menos una segunda población de segundas sondas de secuenciación que comprenden una pluralidad de la secuenciación sondas de la presente invención a un ácido nucleico diana que se inmoviliza en un sustrato, en el que el ácido nucleico diana se inmoviliza al sustrato en una o más posiciones, en donde el dominio de unión diana de las primeras sondas de secuenciación y la segunda sonda de secuenciación son diferentes, y en el que 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; (12) unir una primera molécula complementaria de ácido nucleico que comprende una etiqueta detectable o una primera molécula complementaria de ácido nucleico de un primer complejo indicador que comprende una etiqueta detectable a una primera región de unión de al menos seis posiciones de unión; (13) detectar la etiqueta detectable de la primera molécula de ácido nucleico complementaria unida o la etiqueta 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; (15) unirse a la primera posición de unión una primera molécula de ácido nucleico hibridante que carece de una etiqueta detectable, desuniendo así las primeras moléculas de ácido nucleico complementarias (que tienen una etiqueta detectable o de un complejo indicador); (16) unir una segunda molécula complementaria de ácido nucleico que comprende una etiqueta detectable o una segunda molécula complementaria de ácido nucleico de un segundo complejo indicador que comprende una etiqueta detectable a una segunda posición de unión de las al menos seis posiciones de unión; (17) detectar la etiqueta detectable de la segunda molécula de ácido nucleico complementario unido o la etiqueta detectable de la segunda molécula de ácido nucleico complementario unido del segundo complejo indicador; (18) identificar la posición e identidad de un segundo nucleótido en el ácido nucleico diana inmovilizado; y (19) repetir los pasos (15) a (18) hasta que cada posición de unión de las al menos seis posiciones de unión haya sido unida por una molécula complementaria de ácido nucleico que comprende una etiqueta detectable o una molécula de ácido nucleico complementaria de un complejo indicador que comprende una etiqueta detectable, y la etiqueta detectable de la molécula complementaria de ácido nucleico unida o la etiqueta detectable de la molécula complementaria de ácido nucleico unida de un complejo indicador se ha detectado, identificando así 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 diana de la sonda de secuenciación; y (20) deshibridar la al menos una segunda población de segundas sondas de secuenciación del ácido nucleico diana inmovilizado. Los pasos (16) y (17) pueden ocurrir concurrente 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 más alta, un nivel más alto de agente caotrópico, una concentración de sal más alta, una velocidad de flujo más alta y un pH diferente) que la condición promedio para la cual las sondas de secuenciación en la primera población se deshibridan 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 condiciones diferentes y las sondas en poblaciones no secuenciales pueden deshibridarse en condiciones similares. Como ejemplo, las sondas en una primera población y una tercera población pueden deshibridarse en condiciones similares. En realizaciones, las poblaciones secuenciales de sondas se deshibridan en condiciones cada vez más estrictas (por ejemplo, niveles más altos de agente caotrópico, concentración de sal y temperatura). Para un dispositivo microfluídico, utilizando la temperatura como ejemplo, una primera población de sondas puede permanecer hibridada a una primera temperatura, pero deshibridarse a una segunda temperatura, que es más alta que la primera. Una segunda población de sondas puede permanecer hibridada a la segunda temperatura, pero deshibridarse a una tercera temperatura, que es más alta que la segunda. En este ejemplo, las soluciones (que

comprenden los reactivos requeridos por el presente método) que fluyen sobre un ácido nucleico diana para las poblaciones de sonda iniciales están a una temperatura más baja que las soluciones que fluyen sobre el ácido nucleico diana para las poblaciones de sonda posteriores.

En algunas realizaciones, después de que se haya 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 que una primera población de sondas ha sido hibridada, detectada y deshibridada, se hibrida una alícuota posterior de la primera población de sondas. Alternativamente, como ejemplo, una primera población de sondas puede deshibridarse y reemplazarse con una segunda población de sondas; una vez que la segunda población ha sido detectada y deshibridada, una parte alícuota posterior de la primera población de sondas se hibrida con el ácido nucleico diana. Por lo tanto, una sonda en la población posterior puede hibridarse con una región del ácido nucleico diana que se había secuenciado previamente (obteniendo así información de secuencia duplicada y/o confirmatoria) o una sonda en la población posterior puede hibridarse con una región del nucleico diana ácido que no había sido secuenciado previamente (obteniendo así nueva información de secuencia). De acuerdo con lo anterior, una población de sondas se puede volver a dividir en alícuotas cuando una lectura previa no fue satisfactoria (por cualquier motivo) y/o para mejorar la precisión de la alineación resultante de las lecturas de secuenciación.

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

Cuando se usan más de dos poblaciones de sondas, los pasos, como se describe para la primera y segunda poblaciones de sondas de secuenciación, se repiten con poblaciones adicionales de sondas (por ejemplo, 10 a 100 a 1000 poblaciones). El número de poblaciones de sondas utilizadas dependerá de una variedad de factores, que incluyen, entre otros, el tamaño del ácido nucleico diana, el número de sondas únicas en cada población, el grado de superposición entre las sondas de secuenciación deseadas y el enriquecimiento de sondas a 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 menor interés en un ácido nucleico diana.

Una población de sondas de secuenciación puede compartimentarse en grupos discretos más pequeños de sondas de secuenciación. La compartimentación puede basarse en la temperatura de fusión prevista del dominio de unión diana en las sondas de secuenciación y/o en el motivo de secuencia del dominio de unión diana en las sondas de secuenciación. La compartimentación puede basarse en reglas derivadas empíricamente. Los diferentes grupos de sondas de secuenciación pueden hacerse reaccionar con el ácido nucleico diana utilizando diferentes condiciones de reacción, por ejemplo, en función de la temperatura, la concentración de sal y/o el contenido del regulador. La compartimentación puede realizarse para cubrir el ácido nucleico diana con una 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 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, por encima del límite de resolución del aparato de secuenciación.

De manera alternativa o adicional, 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, por encima del límite de resolución del aparato de secuenciación. Los métodos para secuenciar un ácido nucleico comprenden además etapas de ensamblar cada orden lineal de nucleótidos identificado para cada región del ácido nucleico diana inmovilizado, identificando así una secuencia para el ácido nucleico diana inmovilizado. Los pasos de ensamblaje usan un medio de almacenamiento legible por ordenador no transitorio con un programa ejecutable almacenado en el mismo que instruye a un microprocesador para organizar cada orden lineal identificado de nucleótidos, obteniendo así la secuencia del ácido nucleico. El ensamblaje puede ocurrir en “tiempo real”, es decir, mientras se recopilan datos de sondas de secuenciación en lugar de después de que se hayan recopilado todos los datos.

El ácido nucleico diana, es decir, que está secuenciado, puede tener una longitud de entre aproximadamente 4 y 1,000,000 de nucleótidos. La diana puede incluir un cromosoma completo intacto o un fragmento del mismo, cualquiera de los cuales tiene más de 1,000,000 de nucleótidos de longitud.

En el presente documento se proporcionan kits que incluyen sondas de secuenciación de la presente invención y para realizar métodos de la presente invención. En las realizaciones, los kits incluyen un sustrato capaz de inmovilizar un ácido nucleico a través de una sonda de captura, una pluralidad de sondas de secuenciación de la presente invención,

al menos dos sondas de captura, en el que una sonda de captura comprende una secuencia de ácido nucleico que se hibrida a una región de un ácido nucleico diana que es diferente de la región del ácido nucleico diana que se hibrida con una sonda de secuenciación y una etiqueta de afinidad, al menos seis moléculas de ácido nucleico complementarias que comprenden una etiqueta detectable, en el que un ácido nucleico complementario que comprende una etiqueta detectable comprende una secuencia de ácido nucleico que hibrida con una de las al menos seis posiciones de unión de una sonda de secuenciación y una etiqueta detectable, al menos seis moléculas de ácido nucleico complementarias que carecen de una etiqueta detectable, en el que un ácido nucleico complementario que carece de una etiqueta detectable comprende una secuencia de ácido nucleico que hibrida con una de las al menos seis posiciones de conexión de una sonda de secuenciación, e instruye complementos para su uso. En realizaciones, el kit comprende aproximadamente o al menos 4096 sondas de secuenciación únicas. 4096 es el número mínimo de sondas únicas necesarias para incluir cada combinación hexamérica posible (es decir, para las sondas que tienen seis regiones de unión en los dominios de código de barras). Aquí, “4096” se logra ya que hay cuatro opciones de nucleótidos para seis posiciones: 4^6. Para un conjunto de sondas que tienen cuatro regiones de unión en los dominios de código de barras, solo 256 (es decir, 4^4 se necesitarán) sondas únicas. Para un conjunto de sondas que tienen ocho nucleótidos en sus dominios de unión diana, 4^8 (es decir, se necesitarán 65,536) sondas únicas. Para un conjunto de sondas que tienen diez nucleótidos en sus dominios de unión diana, 4^{10} (es decir, se necesitarán 1,048,576) sondas únicas.

En las realizaciones, el kit comprende aproximadamente o al menos veinticuatro moléculas de ácido nucleico complementarias distintas que tienen una etiqueta detectable y aproximadamente o al menos veinticuatro moléculas de ácido nucleico hibridante distintas que carecen de una etiqueta detectable. Un ácido nucleico complementario puede unirse a una región de unión que tiene una secuencia de una de las SEQ ID NO: 1 a 24, como ejemplos no limitantes. Secuencias ejemplares adicionales que pueden incluirse en un dominio de código de barras se enumeran 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, 50% a 99.9%) con una secuencia de nucleótidos conocida; Esto ayuda a evitar la hibridación indeseable de un ácido nucleico complementario y un ácido nucleico diana.

Cualquiera de los aspectos y realizaciones anteriores se puede combinar con cualquier otro aspecto o realización. A menos que se defina lo contrario, todos los términos técnicos y científicos utilizados en el presente documento tienen el mismo significado que comúnmente entiende un experto en la materia a la que pertenece esta invención. En la Especificación, las formas singulares también incluyen el plural a menos que el contexto indique claramente lo contrario; Como ejemplos, los términos “un”, “uno” y “la” se entienden en singular o plural y el término “o” se entiende que es inclusivo. A modo de ejemplo, “un elemento” significa uno o más elementos. A lo largo de la especificación, se entenderá que la palabra “que comprende” o variaciones tales como “comprende” o “que comprende” implica la inclusión de un elemento, número entero o paso, o grupo de elementos, números enteros o pasos, pero no la exclusión. de cualquier otro elemento, entero o paso, o grupo de elementos, enteros o pasos. Aproximadamente puede entenderse como dentro del 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, 1%, 0.5%, 0.1%, 0.05% o 0.01% del Valor establecido. A menos que el contexto indique lo contrario, todos los valores numéricos proporcionados en este documento se modifican con el término “aproximadamente”.

Breve descripción de los dibujos

El archivo de patente o solicitud contiene al menos un dibujo ejecutado en color. La Oficina proporcionará copias de esta patente o publicación de solicitud de patente con dibujos en color, previa solicitud y pago de la tarifa necesaria.

Las características anteriores y adicionales se apreciarán más claramente a partir de la siguiente descripción detallada cuando se toma junto con los dibujos adjuntos.

La Figura 1 a la Figura 5 muestran esquemas de sondas de secuenciación ejemplares de la divulgación.

La Figura 6A a la Figura 6D son esquemas que muestran variantes de una sonda de secuenciación de la divulgación.

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

La Figura 8A a la Figura 8E ilustran los pasos de un método de secuenciación de la divulgación.

La figura 9A muestra un paso inicial de un método de secuenciación de la divulgación.

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

La Figura 9C muestra una pluralidad de complejos indicadores, cada uno de los cuales comprende etiquetas detectables.

Las Figuras 9D a 9G muestran pasos adicionales del método de secuenciación comenzado en la Figura 9A.

La figura 10 muestra una ilustración alternativa de los pasos mostrados en la figura 9D y la figura 9E y datos ejemplares obtenidos a partir de los mismos. El fragmento de la sonda de secuencia que se muestra tiene la secuencia de SEQ ID NO: 82.

La Figura 11 ilustra una variación del método que se muestra en la Figura 10. El fragmento de la sonda de secuencia que se muestra también tiene la secuencia de SEQ ID NO: 82.

La Figura 12 ilustra un método de la presente invención.

La Figura 13 compara los pasos requeridos en un método de secuenciación de la divulgación con los pasos requeridos con otros métodos de secuenciación.

La Figura 14 y la Figura 15 ejemplifican las mediciones de rendimiento obtenibles por los métodos de la divulgación.

La Figura 16 compara la velocidad de secuenciación, el número de lecturas y la utilidad clínica para la presente invención y varios otros métodos/aparatos de secuenciación.

La Figura 17 demuestra el bajo índice de error bruto de los métodos de secuenciación de la divulgación. La secuencia de plantilla mostrada tiene la secuencia de SEQ ID NO: 83.

La Figura 18 compara los datos de secuencia obtenibles de los métodos de la divulgación con otros métodos de secuenciación.

La Figura 19 demuestra la especificidad de base única de los métodos de secuenciación de la divulgación. Las secuencias de plantilla y sonda mostradas (de arriba a abajo) tienen las secuencias de SEQ ID NO: 84 a SEQ ID NO: 88.

La Figura 20A muestra varios diseños de complejos indicadores de la divulgación.

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

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

La Figura 21A muestra diseños de complejos indicadores que comprenden “asas adicionales”.

La Figura 21B muestra los recuentos fluorescentes obtenidos de los complejos indicadores que tienen “asas adicionales”.

La Figura 22A y la Figura 22B muestran la cinética de hibridación de dos diseños ejemplares de complejos indicadores de la divulgación.

La Figura 23 muestra un esquema de una sonda de secuenciación de la divulgación utilizada en un método distinto del que se muestra en la Figura 8 a la Figura 12.

La Figura 24 muestra un esquema de una tarjeta de secuenciación consumible útil en la divulgación.

La Figura 25 muestra la detección de falta de coincidencia de un 10 mer, como se describe en el Ejemplo 3. Los nucleótidos mostrados (de arriba a abajo) 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 diana, como se describe en el Ejemplo 3. El fondo es alto debido a una concentración de indicador muy alta y no hubo purificación previa. Los nucleótidos mostrados (de arriba a abajo) tienen las secuencias de SEQ ID NO: 100 a SEQ ID NO: 104.

La Figura 27 muestra una comparación entre un punto único frente a un indicador de longitud completa. Los resultados para puntos únicos muestran que la velocidad de hibridación es 1000 veces mayor que para un código de barras de longitud completa (Condiciones objetivo de 100 nM, hibridación de 30 minutos).

Descripción detallada de la invención

La presente invención proporciona sondas de secuenciación, métodos, kits y aparatos que proporcionan secuenciación de ácido nucleico libre de enzimas, libre de amplificación y libre de bibliotecas que tiene longitudes de lectura largas y con baja tasa de error.

Sonda de secuencia

La presente invención, tal como se define en las reivindicaciones adjuntas, y la divulgación se refieren a una sonda de secuenciación que comprende un dominio de unión diana y un dominio de código de barras. En las figuras 1 a 6 se muestran ejemplos de sondas de secuenciación de la divulgación.

La figura 1 muestra un esquema de una sonda de secuenciación de la divulgación. Esta sonda de secuenciación ejemplar tiene un dominio de unión diana de seis nucleótidos, cada uno de los cuales corresponde a una posición en el dominio de código de barras (que comprende una o más regiones de unión). Se observa una primera región de unión; corresponde al nucleótido de un ácido nucleico diana unido por un primer nucleótido en el dominio de unión diana. Se anota la tercera posición en el dominio del código de barras. Se observa una quinta posición que comprende dos regiones de unión. Cada posición en un dominio de código de barras puede tener múltiples regiones de unión. Por ejemplo, una posición puede tener de 1 a 50 regiones de unión. Ciertas posiciones en un dominio de código de barras pueden tener más regiones de unión que otras posiciones (como se muestra aquí en la posición 5 en relación con las posiciones 1 a 4 y 6); alternativamente, cada posición en un dominio de código de barras tiene el mismo número de regiones de unión (ver, por ejemplo, Figuras 2, 3, 5 y 6). Aunque no se muestra, cada región de unión comprende al menos una (es decir, una a cincuenta, por ejemplo, diez a treinta) copias de una secuencia o secuencias de ácido nucleico capaces de unirse reversiblemente a una molécula complementaria de ácido nucleico (ARN o ADN). En la Figura 1, las regiones de unión son integrales a la molécula de polinucleótido lineal que forma el dominio del código de barras.

La figura 2 muestra un esquema de una sonda de secuenciación de la divulgación. Esta sonda de secuenciación ejemplar tiene un dominio de unión diana de seis nucleótidos, cada uno de los cuales corresponde a una región de unión en el dominio de código de barras. Se observa una primera región de unión; corresponde al nucleótido de un ácido nucleico diana unido por un primer nucleótido en el dominio de unión diana. La cuarta posición en el dominio del código de barras, que comprende una parte del dominio del código de barras y dos cuartas regiones de unión están rodeadas. Se observan dos sextas regiones de archivos adjuntos. Aquí, cada posición tiene dos regiones de unión; sin embargo, cada posición en un dominio de código de barras puede tener una región de unión o múltiples regiones de unión, por ejemplo, de 2 a 50 regiones de unión. Aunque no se muestra, cada región de unión comprende al menos una (es decir, una a cincuenta, por ejemplo, diez a treinta) copias de una secuencia o secuencias de ácido nucleico capaces de unirse reversiblemente a una molécula complementaria de ácido nucleico (ARN o ADN). En la Figura 2, el dominio del código de barras es una molécula de polinucleótido lineal a la que están unidas las regiones de unión; Las regiones de unión no son integrales a la molécula polinucleotídica.

La figura 3 muestra otro esquema de una sonda de secuenciación de la divulgación. Esta sonda de secuenciación ejemplar tiene un dominio de unión diana de cuatro nucleótidos, con estos cuatro nucleótidos en las cuatro posiciones correspondientes en el dominio de código de barras. Cada posición se muestra con tres regiones de unión vinculadas.

La figura 4 muestra aún otro esquema de una sonda de secuenciación de la divulgación. Esta sonda de secuenciación ejemplar tiene un dominio de unión diana de diez nucleótidos. Sin embargo, solo los primeros seis nucleótidos corresponden a seis posiciones en el dominio del código de barras. Los nucleótidos séptimo a décimo (indicados por "n_1 a n_4") se agregan para aumentar la longitud del dominio de unión a la diana, lo que afecta la probabilidad de que una sonda hibrida y permanezca hibridada con un ácido nucleico diana.

En las realizaciones, los nucleótidos "n" pueden preceder a los nucleótidos correspondientes a las posiciones en el dominio del código de barras. En realizaciones, los nucleótidos "n" pueden seguir a los nucleótidos correspondientes a posiciones en el dominio de código de barras. En la Figura 4, se muestran cuatro nucleótidos "n"; sin embargo, un dominio de unión diana puede incluir más de cuatro nucleótidos "n". Los "n" nucleótidos pueden tener bases universales (por ejemplo, Inosina, derivados de 2'-desoxiinosina (hipoxantina-desoxinucleótido), nitroindol, análogos de nitroazol y bases hidrófobas aromáticas sin enlaces de hidrógeno) que pueden emparejarse con cualquiera de las cuatro bases canónicas.

Otra sonda de secuenciación de la unión 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 secuenciación ejemplar mostrada tiene un dominio de unión diana de diez nucleótidos. Sin embargo, solo los terceros a ocho nucleótidos en el dominio de unión diana corresponden a seis posiciones (primera a sexta) en el dominio de código de barras. Los nucleótidos primero, segundo, noveno y décimo (indicados por "n_1 a n_4") se agregan para aumentar la longitud del dominio de unión diana. En la Figura 5, se muestran cuatro nucleótidos "n"; sin embargo, un dominio de unión diana puede incluir más o menos de cuatro nucleótidos "n". La Figura 6A a la Figura 6D muestran variantes de una sonda de secuenciación de la Figura 1. En la Figura 6A, el orden lineal de los nucleótidos en el dominio de unión diana y el orden lineal de las regiones de unión en el dominio del código de barras progresan de izquierda a derecha (con respecto a la ilustración) En la Figura 6B, el orden lineal de los nucleótidos en el dominio de unión diana y el orden lineal de las regiones de unión en el dominio del código de barras progresan de derecha a izquierda (con respecto a la ilustración). En la Figura 6C, el orden lineal de los nucleótidos en el dominio de unión diana se invierte con respecto al orden lineal de las regiones de unión en el dominio del código de barras. En cualquier sonda de la divulgación, puede haber una falta de orden estricto de los nucleótidos en el dominio de unión diana y de las regiones de unión en el dominio de código de barras, siempre que la sonda esté diseñada de manera que cada nucleótido en el dominio de unión diana corresponda a un dominio de unión o dominios de unión en el dominio de código de barras; la falta de

orden estricto se muestra en la Figura 6D. Cualquier sonda de la divulgación (por ejemplo, las ejemplificadas en las Figuras 1 a 5) puede tener un orden de nucleótidos y regiones de unión como se muestra en la Figura 6.

El dominio de unión 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 diana preferible es un polinucleótido. El dominio de unión diana es capaz de unir un ácido nucleico diana.

Una sonda puede incluir múltiples copias del dominio de unión diana operativamente unido a una cadena principal sintética.

Las sondas pueden diseñarse para controlar la probabilidad de hibridación y/o deshibridación y las velocidades a las que se producen. En general, cuanto más baja es la Tm de una sonda, más rápido y más probable es que la sonda se hibride a/desde un ácido nucleico diana. Por lo tanto, el uso de sondas de Tm más bajas disminuirá el número de sondas unidas a un ácido nucleico diana.

La longitud de un dominio de unión diana, en parte, afecta la probabilidad de que una sonda hibrida y permanezca hibridada con un ácido nucleico diana. En general, cuanto más largo (mayor número de nucleótidos) es un dominio de unión diana, menos probable es que una secuencia complementaria esté presente en el nucleótido diana. Por el contrario, cuanto más corto es un dominio de unión diana, más probable es que una secuencia complementaria esté presente en el nucleótido diana. Por ejemplo, existe una probabilidad de 1/256 de que una secuencia de cuatro meros se ubique en un ácido nucleico diana frente a una probabilidad de 1/4096 de que una secuencia de seis meros se ubique en el ácido nucleico diana. En consecuencia, una colección de sondas más cortas probablemente se unirá en más ubicaciones para un tramo dado de un ácido nucleico en comparación con una colección de sondas más largas.

La figura 7 muestra dominios de unión diana de 10 mer. En algunas realizaciones, el dominio de unión diana incluye cuatro bases universales (identificadas como "U_b") que se emparejan con cualquiera de los cuatro nucleótidos canónicos (A, G, C y T). En realizaciones, el dominio de unión diana incluye de una a seis (por ejemplo, 2 y 4) bases universales. Un dominio de unión diana puede incluir nucleótidos no universales. La Figura 7 señala que una población "completa" de sondas que tienen 6 nucleótidos específicos en el dominio de unión diana requerirá 4096 sondas únicas y una población "completa" de sondas que tienen 10 nucleótidos específicos requerirá ~1 millón de sondas únicas.

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

Sin embargo, puede ser preferible tener menos números de sondas unidas a un ácido nucleico diana ya que hay ocasiones en que demasiadas sondas en una región pueden causar la superposición de su etiqueta detectable, evitando así la resolución de dos sondas cercanas. Esto se explica de la siguiente manera. Dado que un nucleótido tiene una longitud de 0.34 nm y 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 sería incapaz de resolver las señales de dos sondas hibridadas con un ácido nucleico diana cuando las dos sondas están dentro de aproximadamente 588 pares de bases entre sí. Por lo tanto, dos sondas, dependiendo de la resolución del aparato de secuenciación, necesitarán estar separadas aproximadamente 600 pb antes de que su etiqueta detectable pueda resolverse como "puntos" distintos. Por lo tanto, en una separación óptima, debe haber una sola sonda por 600 pb de ácido nucleico diana. Se puede utilizar una variedad de enfoques de software (por ejemplo, valores de intensidad de fluorescencia y relaciones dependientes de la longitud de onda) para monitorizar, limitar y potencialmente desconvolucionar el número de sondas que se hibridan dentro de una región resoluble de un ácido nucleico diana y para diseñar poblaciones de sondas en consecuencia. Además, etiquetas detectables (por ejemplo, se pueden seleccionar etiquetas fluorescentes) que proporcionan señales más discretas. Además, los métodos en la literatura (por ejemplo, Small y Parthasarthy: "Superresolution localization methods". Annu. Rev. Phys Chem., 2014; 65: 107-25) describen la iluminación estructurada y una variedad de enfoques de superresolución que disminuyen el límite de resolución de un microscopio de secuenciación de hasta 10's de nanómetros. El uso de aparatos de secuenciación de mayor resolución permite el uso de sondas con dominios de unión diana más cortos.

Como se mencionó anteriormente, el diseño de la Tm de las sondas puede afectar el número de sondas hibridadas con un ácido nucleico diana. 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, por encima del límite de resolución del aparato de secuenciación.

El término "ácido nucleico diana" significará una molécula de ácido nucleico (ADN, ARN o ANP) cuya secuencia se determinará mediante las sondas, métodos y aparatos de la invención. En general, los términos "ácido nucleico diana", "molécula de ácido nucleico", "secuencia de ácido nucleico", "ácido nucleico", "fragmento de ácido nucleico",

"oligonucleótido" y "polinucleótido" se usan indistintamente y están destinados a incluir, pero no limitado a, una forma polimérica de nucleótidos que pueden tener varias longitudes, ya sea desoxirribonucleótidos o ribonucleótidos, o análogos de los mismos. Los ejemplos no limitantes de ácidos nucleicos incluyen un gen, un fragmento de gen, un exón, un intrón, ADN intergénico (que incluye, sin limitación, ADN heterocromático), ARN mensajero (ARNm), ARN de transferencia, ARN ribosómico, ribozimas, ARN de interferencia corta (ARNic), 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 métodos actuales 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 la secuenciación basada en ARN, los presentes métodos no requieren la conversión de una molécula de ARN en una molécula de ADN (es decir, mediante la síntesis de ADNc) antes de que se pueda obtener una secuencia. Como no se requiere amplificación ni conversión, un ácido nucleico secuenciado como se describe en este documento retendrá cualquier base única y/o marcador epigenético presente en el ácido nucleico cuando el ácido nucleico está en la muestra o cuando se obtuvo de la muestra. Tales bases únicas y/o marcadores epigenéticos se pierden en los métodos de secuenciación conocidos en la técnica.

El ácido nucleico diana se puede obtener de cualquier muestra o fuente de ácido nucleico, por ejemplo, cualquier célula, tejido u organismo, in vitro, sintetizador químico, etc. El ácido nucleico diana se puede obtener 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 utilizando 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 por medios enzimáticos o mecánicos. Los medios mecánicos pueden ser sonicación o cizallamiento físico. Los medios enzimáticos pueden realizarse mediante 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, se deben tomar medidas 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 pueden tener longitudes totales (que incluyen el dominio de unión diana, el dominio del 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 del código de barras comprende una cadena principal sintética. La cadena principal sintética y el dominio de unión diana están operativamente ligados, por ejemplo, están unidos covalentemente o unidos mediante un ligador. La cadena principal sintética puede comprender cualquier material, por ejemplo, polisacárido, polinucleótido, polímero, plástico, fibra, péptido, ácido nucleico peptídico o polipéptido. Preferiblemente, la cadena principal sintética es rígida. En las realizaciones, la cadena principal comprende "origami de ADN" de seis hélices dobles de ADN (véase, por ejemplo, Lin et al., "Submicrometre geometrically encoded fluorescent barcodes self-assembled from DNA". Nature Chemistry; 2012 Oct; 4 (10): 832- 9) Se puede hacer un código de barras con mosaicos de origami de ADN (Jungmann et al., "Multiplexed 3D cellular super-resolution imaging with DNA-PAINT and Exchange-PAINT", Nature Methods, Vol. 11, No. 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, igual o mayor que el número de nucleótidos en el dominio de unión diana. Es preferible incluir nucleótidos adicionales en un dominio de unión diana que el número de posiciones en el dominio principal, por ejemplo, uno, dos, tres, cuatro, cinco, seis, siete, ocho, nueve, diez o más nucleótidos. La longitud del dominio del código de barras no está limitada siempre que haya espacio suficiente para al menos cuatro posiciones, como se describió anteriormente.

Cada posición en el dominio de código de barras corresponde a un nucleótido en el dominio de unión diana y, por lo 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 diana y la sexta posición en el dominio de código de barras corresponde al sexto nucleótido en el dominio de unión diana.

Cada posición en el dominio de código de barras comprende al menos una región de unión, por ejemplo, una a 50, o más, regiones de unión. Ciertas posiciones en un dominio de código de barras pueden tener más regiones de unión que otras posiciones (por ejemplo, una primera posición puede tener tres regiones de unión mientras que una segunda posición puede tener dos posiciones de unión); alternativamente, cada posición en un dominio de código de barras tiene el mismo número de regiones de unión. Cada región de unión comprende al menos una (es decir, una a

cincuenta, por ejemplo, diez a treinta) copias de una secuencia o secuencias de ácido nucleico capaces de unirse reversiblemente por una molécula de ácido nucleico complementaria (por ejemplo, ADN o ARN). En ejemplos, la secuencia de ácido nucleico en una primera región de unión determina la posición y la identidad de un primer nucleótido en el ácido nucleico diana que está unido por un primer nucleótido del dominio de unión diana. Cada región de unión puede estar unida a un monómero modificado (por ejemplo, Nucleótido modificado) en la cadena principal sintética de tal manera que la región de unión se ramifica desde la cadena principal sintética. En realizaciones, las regiones de unión son integrales a una cadena principal polinucleotídica; es decir, la cadena principal es un polinucleótido único y las regiones de unión son partes de la secuencia del polinucleótido único. En realizaciones, los términos “dominio de código de barras” y “cadena principal sintética” son sinónimos.

La secuencia de ácido nucleico en una región de unión identifica la posición e identidad de un nucleótido en el ácido nucleico diana que está unido por un nucleótido en el dominio de unión diana de una sonda de secuenciación. En una sonda, cada región de unión tendrá una secuencia general única. De hecho, cada posición en un dominio de código de barras puede tener una región de unión que comprende una secuencia de ácido nucleico que codifica uno de cuatro nucleótidos, es decir, específico para uno de adenina, timina/uracilo, citosina y guanina. Además, la región de unión de una primera posición (y la citosina codificante, por ejemplo) incluirá una secuencia de ácido nucleico diferente de la región de unión de una segunda posición (y la citosina codificante, por ejemplo). Por lo tanto, a una secuencia de ácido nucleico en una región de unión 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 diana. Además, a una región de unión 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 diana.

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

La Tabla 1 proporciona secuencias ejemplares, solo con fines ilustrativos, para regiones de unión para secuenciar sondas que tienen hasta seis posiciones en su dominio de código de barras y etiquetas detectables en ácido nucleico complementario que se unen al mismo.

Tabla 1:

Nucleótido en dominio/posición de unión de diana en dominio de código de barras	**Nucleótido**	**Secuencia de Ácido Nucleico (5’ a 3’) Región de unión**	**Etiqueta 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
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	TAACTATGAC	RFP	22
6	C	ACTAATGTTC	CFP	23
6	U/T	TCATTGAATG	YFP	24

Como se ve en la Tabla 1, la secuencia de ácido nucleico de una primera región de unión puede ser una de SEQ ID NO: 1 a SEQ ID NO: 4 y la secuencia de ácido nucleico de una segunda unión 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 unión 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 unión tendría la secuencia de SEQ ID NO: 5.

En las realizaciones, una molécula de ácido nucleico complementaria puede estar unida por una etiqueta detectable. En realizaciones alternativas, un ácido nucleico complementario está asociado con un complejo indicador que comprende etiquetas detectables.

La secuencia de nucleótidos de un ácido nucleico complementario no está limitada; preferiblemente carece de homología sustancial (por ejemplo, 50% a 99.9%) con una secuencia de nucleótidos conocida; Esto ayuda a evitar la hibridación indeseable de un ácido nucleico complementario y un ácido nucleico diana.

Un ejemplo del complejo indicador útil en la divulgación se muestra en la Figura 9B. En este ejemplo, un ácido nucleico complementario se une a una molécula de ácido nucleico primario, que a su vez se hibrida con una pluralidad de moléculas de ácido nucleico secundario, cada una de las cuales se hibrida a una pluralidad de moléculas de ácido nucleico terciario que se unen a una o etiquetas más detectables.

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

La figura 9C muestra una población de complejos indicadores ejemplares. En el panel superior izquierdo de la Figura 9C se incluyen los cuatro complejos que se hibridan con la región de unión 1 de una sonda. Hay un tipo de complejo indicador para cada nucleótido posible que puede estar presente en la posición de nucleótido 1 del dominio de unión diana de una sonda. Aquí, mientras se realiza un método de secuencia de la divulgación, si la posición 1 del dominio indicador de una sonda está unida por un complejo indicador que tiene una etiqueta detectable de “color azul”, entonces el primer nucleótido en el dominio de unión diana se identifica como Adenina. Alternativamente, si la posición

1 está unida por un complejo indicador que tiene una etiqueta detectable “de color verde”, entonces el primer nucleótido en el dominio de unión diana se identifica como Timina.

Los complejos indicadores pueden ser de varios diseños. Por ejemplo, una molécula de ácido nucleico primario puede hibridarse con al menos una (por ejemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más) moléculas de ácido nucleico secundario. Cada molécula secundaria de ácido nucleico puede hibridarse con al menos una (por ejemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más) moléculas de ácido nucleico terciario. Ejemplos de complejos indicadores se muestran en la Figura 20A. Aquí, el complejo indicador “4x3” tiene una molécula de ácido nucleico primaria (que está unida a una molécula de ácido nucleico complementaria) hibridada con cuatro moléculas de ácido nucleico secundarias, cada una de las cuales se hibrida con tres moléculas de ácido nucleico terciario (cada una de las cuales tiene una etiqueta detectable) En esta figura, cada ácido nucleico complementario de un complejo tiene 12 nucleótidos de largo (“12 bases”); sin embargo, la longitud del nucleico complementario no está limitada y puede ser inferior a 12 o superior a 12 nucleótidos. El complejo inferior derecho incluye una región espaciadora entre su ácido nucleico complementario y su molécula primaria de ácido nucleico. El espaciador se identifica con una longitud de 20 a 40 nucleótidos; sin embargo, la longitud de un espaciador no es limitante y puede ser más corta que 20 nucleótidos o más larga que 40 nucleótidos.

La figura 20B muestra recuentos promedio variables (fluorescentes) obtenidos de los cuatro complejos indicadores ejemplares mostrados en la figura 20A. En la Figura 20B, se unieron 10 pM de plantilla diana biotinilada sobre una superficie de celda de flujo revestida con estreptavidina, se hizo fluir 10 nM de un complejo indicador sobre la celda de flujo; Después de una incubación de un minuto, se lavó la celda de flujo, se tomó una imagen de la celda de flujo y se contaron las características fluorescentes.

En realizaciones, los complejos indicadores están “preconstruidos”. Es decir, cada polinucleótido en el complejo se hibrida antes de contactar el complejo con una sonda. En la Figura 20C se muestra una receta ejemplar para preconstruir cinco complejos indicadores ejemplares.

La figura 21A muestra complejos indicadores alternativos en los que las moléculas de ácido nucleico secundario tienen “asas extra” que no se hibridan con una molécula de ácido nucleico terciario y son distales a la molécula de ácido nucleico primario. En esta figura, cada “asa extra” tiene 12 nucleótidos de largo (“12 mer”); sin embargo, sus longitudes no son limitadas y pueden ser menos de 12 o más de 12 nucleótidos. En realizaciones, las “asas adicionales” comprenden cada una la secuencia de nucleótidos del ácido nucleico complementario; por lo tanto, cuando un complejo indicador comprende “asas adicionales”, el complejo indicador puede hibridarse con una sonda de secuenciación a través del ácido nucleico complementario del complejo indicador o mediante un “controlador adicional”. Por consiguiente, la probabilidad de que un complejo indicador se una a un la sonda de secuenciación aumenta. El diseño de “asa extra” también puede mejorar la cinética de hibridación. Sin estar ligados a la teoría, las “asas adicionales” esencialmente aumentan la concentración efectiva del ácido nucleico complementario del complejo indicador.

La Figura 21B muestra recuentos promedio variables (fluorescentes) obtenidos de los cinco complejos indicadores ejemplares que tienen “asas adicionales” utilizando el procedimiento descrito para la Figura 20B.

Las Figuras 22A y 22B muestran cinética de hibridación e intensidades fluorescentes para dos complejos indicadores ejemplares. Alrededor de 5 minutos, los recuentos totales comienzan a estabilizarse, lo que indica que la mayoría del complejo reportero agregado ha encontrado una diana disponible.

Una fracción, etiqueta o indicador detectable puede unirse a un ácido nucleico complementario o a una molécula de ácido nucleico terciario en una variedad de formas, que incluyen la unión directa o indirecta de una fracción detectable tal como una fracción fluorescente, una fracción colorimétrico y similares. Un experto en la materia puede consultar referencias dirigidas a etiquetar ácidos nucleicos. Los ejemplos de restos fluorescentes incluyen, pero no se limitan a, 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. Las etiquetas fluorescentes y su unión a nucleótidos y/u oligonucleótidos se divulgan en muchas revisiones, incluyendo Haugland, Handbook of Fluorescent Probes and Research Chemicals, Novena edición (Molecular Probes, Inc., Eugene, 2002); Keller y Manak, DNA Probes, 2ª 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 los Estados Unidos Nos. 4.757.141; 5.151.507; y 5.091.519. En un aspecto, uno o más colorantes fluorescentes se usan como etiqueta para secuencias diana etiquetadas, por ejemplo, como se divulga en las patentes de los Estados Unidos Nos. 5.188.934 (colorantes de 4,7-diclorofluoresceína); 5.366.860 (colorantes de rodamina espectralmente resolubles); 5.847.162 (colorantes de 4,7-diclororhodamina); 4.318.846 (colorantes de fluoresceína sustituidos con éter); 5.800.996 (colorantes de transferencia de energía); Lee y col. 5.066.580 (colorantes de xantina); 5.688.648 (colorantes de transferencia de energía); y similares. El etiquetado también se puede llevar a cabo con puntos cuánticos, como se divulga en las siguientes patentes y publicaciones de patentes: Patentes de los Estados Unidos Números 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 el presente documento, el término “etiqueta fluorescente” comprende

una fracción de señalización que transmite información a través de las propiedades fluorescentes de absorción y/o emisión de una o más moléculas. Dichas propiedades fluorescentes incluyen intensidad de fluorescencia, vida útil de fluorescencia, características del espectro de emisión, transferencia de energía y similares.

Los análogos de nucleótidos fluorescentes disponibles comercialmente incorporados fácilmente en secuencias de nucleótidos y/u oligonucleótidos incluyen, pero no se limitan a, 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 GREENR™ 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 TM FL-14-UTP, BODIPY TMR- 14-UTP, BODIPY TM TR-14-UTP, RHODAMINE GREEN™-5-UTP, ALEXA FLUOR™ 488-5-UTP, LEXA FLUOR™ 546-14-UTP (Molecular Probes, Inc. Eugene, OR) y similares. Alternativamente, los fluoróforos anteriores y los mencionados en el presente documento pueden añadirse durante la síntesis de oligonucleótidos usando, por ejemplo, fosforoamidita o química de NHS. Los protocolos son conocidos en la técnica para la 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 de oligonucleótidos durante su síntesis. El ácido nucleico también podría teñirse, a priori, con un colorante intercalante como DAPI, YOYO-1, bromuro de etidio, tintes de cianina (por ejemplo, SYBR Green) y similares.

Otros fluoróforos disponibles para la unión post-sintética incluyen, entre otros, 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 576/581, BODIPY 576/589, BODIPY 581/591, BODIPY TR, BODIPY 630/650, BODIPY 650/665, Cascade Blue, Cascade Yellow, Dansyl, lissamine rhodamine B, Marina Blue, Oregon Green 488, Oregon Green 514, Blue Pacific, Orange Pacific, rodamina 6G, verde rodamina, rodamina roja, tetrametil rodamina, Texas Red(disponible en Molecular Probes, Inc., Eugene, OR), Cy2, Cy3, Cy3.5, Cy5, Cy5.5, Cy7 (Amersham Biosciences, Piscataway, NJ) y similares. También se pueden usar fluoróforos en tándem FRET, incluidos, entre otros, 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.

Se pueden usar partículas metálicas de plata u oro para mejorar la señal de secuencias de nucleótidos y/u oligonucleótidos etiquetados fluorescentemente (Lakowicz et al. (2003) BioTechniques 34:62).

Otras etiquetas adecuadas para una secuencia de oligonucleótidos pueden incluir fluoresceína (FAM, FITC), digoxigenina, dinitrofenol (DNP), dansilo, biotina, bromodeoxiuridina (BrdU), hexahistidina (6xHis), aminoácidos fosforosos (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 el que cada uno de los anticuerpos se deriva con una etiqueta detectable: biotina/a-biotina, digoxigenina/a-digoxigenina, dinitrofenol (DNP)/a- DNP, 5- Carboxifluoresceína (FAM)/a-FAM.

Las etiquetas detectables descritas en este documento son espectralmente resolubles. “Espectralmente resoluble” en referencia a una pluralidad de etiquetas fluorescentes significa que las bandas de emisión fluorescente de las etiquetas son suficientemente distintas, es decir, que no se solapan lo suficiente, que las etiquetas moleculares a las que se unen las etiquetas respectivas se pueden distinguir en función de señal fluorescente generada por las etiquetas respectivas mediante sistemas de fotodetección estándar, por ejemplo, empleando un sistema de filtros de paso de banda y tubos fotomultiplicadores, o similares, como se ejemplifica en los sistemas descritos en las patentes de los Estados Unidos números 4,230,558; 4,811,218; o similares, o en Wheeless et al., páginas 21-76, en Citometría de flujo: Instrumentación y análisis de datos Flow Cytometry: Instrumentation and Data Analysis (Academic Press, Nueva York, 1985). En un aspecto, los colorantes orgánicos espectralmente resolubles, tales como fluoresceína, rodamina y similares, significa que los máximos de emisión de longitud de onda están separados al menos a 20 nm, y en otro aspecto, al menos a 40 nm. En otro aspecto, los compuestos de lantánidos quelados, los puntos cuánticos y similares, que se pueden resolver espectralmente significa que los máximos de emisión de longitud de onda están separados al menos a 10 nm, y en un aspecto adicional, al menos a 15 nm.

Método de secuenciación

La presente invención, tal como se define por las reivindicaciones adjuntas, y la divulgación se refieren a métodos para secuenciar un ácido nucleico utilizando una sonda de secuenciación de la presente invención o divulgación, respectivamente. Se muestran ejemplos del método en las Figuras 8 a 12.

El método comprende hibridar reversiblemente al menos una sonda de secuenciación a un ácido nucleico diana que está inmovilizado (por ejemplo, en una, dos, tres, cuatro, cinco, seis, siete, ocho, nueve, diez o más posiciones) para un sustrato.

El sustrato puede ser cualquier soporte sólido conocido en la técnica, por ejemplo, un portaobjetos recubierto y un dispositivo microfluídico, que es capaz de inmovilizar un ácido nucleico diana. En ciertas realizaciones, el sustrato es una superficie, membrana, perla, material poroso, electrodo o matriz. El ácido nucleico diana puede inmovilizarse sobre cualquier sustrato aparente para los expertos en la materia.

En realizaciones, el ácido nucleico diana está unido por una sonda de captura que comprende un dominio que es complementario a una porción del ácido nucleico diana. La porción puede ser un extremo del ácido nucleico diana o no hacia un extremo.

Ejemplos de sustratos útiles incluyen aquellos que comprenden una fracción de unión seleccionado del grupo que consiste en ligandos, antígenos, carbohidratos, ácidos nucleicos, receptores, lectinas y anticuerpos. La sonda de captura comprende una fracción de unión capaz de unirse con el fracciones de unión del sustrato. Los sustratos útiles ejemplares que comprenden fracciones reactivas incluyen, pero no se limitan a, superficies que comprenden epoxi, aldehído, oro, hidrazida, sulfhidrilo, NHS-éster, amina, tiol, carboxilato, maleimida, hidroximetilfosfina, imidoéster, isocianato, hidroxilo, pentafluorofeniléster, psoralen, 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 estándar. Los sustratos útiles ejemplares que comprenden fracciones reactivas incluyen, pero sin limitación, el grupo OptArray-DNA NHS (Accler8), Nexterion Slide AL (Schott) y Nexterion Slide E (Schott).

En las realizaciones, la fracción de unión de la sonda de captura es biotina y el sustrato comprende avidina (por ejemplo, estreptavidina). Los sustratos útiles que comprenden avidina están disponibles comercialmente, incluidos TB0200 (Accelr8), SAD6, SAD20, SAD100, SAD500, SAD2000 (Xantec), SuperAvidin (Array-It), streptavidin slide (catálogo #MPC 000, Xenopore) y STREPTAVIDINnslide (catálogo # 439003, Greiner Bio-one).

En realizaciones, el fracciones de unión de la sonda de captura es avidina (por ejemplo, estreptavidina) y el sustrato comprende biotina. Los sustratos útiles que comprenden biotina que están disponibles comercialmente incluyen, pero no se limitan a, Optiarray-biotina (Accler8), BD6, BD20, BD100, BD500 y BD2000 (Xantec). En realizaciones, el fracciones de unión de la sonda de captura puede comprender una fracción reactivo que es capaz de unirse al sustrato por fotoactivación. El sustrato podría comprender la fracción fotorreactivo, o la primera porción del nanoinformador podría comprender la fracción fotorreactiva. Algunos ejemplos de fracciones fotorreactivas incluyen aril azidas, tales como N((2-piridilditio)etil)-4-azidosalicilamida; aril azidas fluoradas, tales como ácido 4-azido-2,3,5,6-tetrafluorobenzoico; reactivos basados en benzofenona, tales como el succinimidil éster del ácido 4-benzoilbenzoico; y 5-bromo-desoxiuridina.

En las realizaciones, el fracciones de unión de la sonda de captura puede inmovilizarse al sustrato a través de otros pares de unión aparentes para los expertos en la materia.

Después de unirse al sustrato, el ácido nucleico diana puede alargarse aplicando una fuerza (por ejemplo, gravedad, fuerza hidrodinámica, fuerza electromagnética “electroestiramiento”, estiramiento por flujo, una técnica de menisco en retroceso y combinaciones de los mismos) suficiente para extenderse el ácido nucleico diana

El ácido nucleico diana puede estar unido por una segunda sonda de captura que comprende un dominio que es complementario a una segunda porción del ácido nucleico diana. La porción puede ser un extremo del ácido nucleico diana o no hacia un extremo. La unión de una segunda sonda de captura puede ocurrir después o durante el alargamiento del ácido nucleico diana o a un ácido nucleico diana que no se ha alargado. La segunda sonda de captura puede tener una unión como se describió anteriormente.

Una sonda de captura puede comprender o estar asociada con una etiqueta detectable, es decir, un punto fiducial.

La sonda de captura es capaz de aislar un ácido nucleico diana de una muestra. Aquí, se agrega una sonda de captura a una muestra que comprende el ácido nucleico diana. La sonda de captura se une al ácido nucleico diana a través de 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 una fracción que se une a las fracciones de unión de la sonda de captura, el ácido nucleico se inmoviliza sobre el sustrato.

Para garantizar que un usuario “captura” tantas moléculas de ácido nucleico diana como sea posible de muestras altamente fragmentadas, es útil incluir una pluralidad de sondas de captura, cada una complementaria a una región diferente del ácido nucleico diana. Por ejemplo, puede haber tres grupos de sondas de captura, con un primer grupo complementario a las regiones del ácido nucleico diana cerca de su extremo 5', un segundo grupo complementario a las regiones en el medio del ácido nucleico diana y un tercer grupo cerca de su extremo 3'. Esto puede generalizarse a “n-regiones de interés” por ácido nucleico diana. En este ejemplo, cada grupo individual de ácido nucleico diana fragmentado se une a una sonda de captura que comprende o se une a una etiqueta de biotina. 1/nésimo de la muestra de entrada (donde n = el número de regiones distintas en el ácido nucleico diana) se aísla para cada cámara de grupo. La sonda de captura se une al ácido nucleico diana de interés. Luego, el ácido nucleico diana se inmoviliza, a través de 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. Todos los grupos n pueden estirarse y unirse

simultáneamente, o, para maximizar el número de moléculas completamente estiradas, el pool 1 (que captura la mayoría de la región 5') puede estirarse y unirse primero; luego el grupo 2 (que captura la región del medio del objetivo) se puede estirar y unir; finalmente, el grupo 3 se puede estirar y unir.

El número de sondas de captura distintas requeridas está inversamente relacionado con el tamaño del fragmento de ácido nucleico diana. En otras palabras, se necesitarán más sondas de captura para un ácido nucleico diana altamente fragmentado. Para los tipos de muestra con ácidos nucleicos diana altamente fragmentados y degradados (por ejemplo, Tejido Embebido en Parafina Fijada en Formalina) puede ser útil incluir múltiples grupos de sondas de captura. Por otro lado, para muestras con fragmentos de ácido nucleico diana largos, por ejemplo, ácidos nucleicos aislados obtenidos in vitro, una sola sonda de captura en un extremo 5' puede ser suficiente.

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 este documento un “espacio”. El espacio es una porción del ácido nucleico diana que está disponible para unirse mediante una sonda de secuenciación. El espacio mínimo es una longitud del dominio de unión diana (por ejemplo, 4 a 10 nucleótidos) y un espacio máximo es la mayoría de un cromosoma completo.

En la Figura 12 se muestra un ácido nucleico diana inmovilizado. 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). En la parte superior se muestra un primer grupo de ácidos nucleicos complementarios que comprende una etiqueta detectable o complejos indicadores, cada miembro del grupo tiene una etiqueta detectable diferente (por ejemplo, la timidina se identifica por una señal verde) y una secuencia de nucleótidos diferente. El primer nucleótido en el dominio de unión diana se une a la T en el ácido nucleico diana. Las primeras regiones de unión de la sonda incluyen una o más secuencias de nucleótidos que especifican que el primer nucleótido en el dominio de unión diana de la sonda se une a una timidina. Por lo tanto, solo el ácido nucleico complementario para la timidina se une a la primera posición del dominio del código de barras. Como se muestra, un primer ácido nucleico complementario que codifica timidina que comprende una etiqueta detectable o complejos indicados que comprenden etiquetas detectables están unidos a regiones de unión en la primera posición del dominio de código de barras de la sonda.

El número de grupos de ácidos nucleicos complementarios o complejos indicadores es idéntico al número de posiciones en el dominio de código de barras. Por lo tanto, para un dominio de código de barras que tiene seis posiciones, se agruparán seis grupos 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 una etiqueta detectable o un complejo indicador. Por lo tanto, cuando se pone en contacto con su ácido nucleico diana, la sonda es capaz de emitir una señal detectable desde su primera posición y no es necesario proporcionar un primer grupo de ácidos nucleicos complementarios o complejos indicadores que se dirigen a la primera posición en el dominio del código de barras .

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

En las realizaciones, los ácidos nucleicos complementarios que comprenden una etiqueta detectable o complejos indicadores pueden eliminarse de la región de unión, pero no reemplazarse con un ácido nucleico hibridante que carece de una etiqueta detectable. Esto puede ocurrir, por ejemplo, agregando un agente caotrópico, aumentando la temperatura, cambiando la concentración de sal, ajustando el pH y/o aplicando una fuerza hidrodinámica. En estas realizaciones menos reactivos (,es decir, se necesitan hibridación de ácidos nucleicos que carecen de etiquetas detectables).

La figura 8C continúa el método de la invención reivindicada. Aquí, el ácido nucleico diana tiene una citidina (C) después de su timidina (T). En la parte superior se muestra un segundo grupo de ácidos nucleicos complementarios o complejos indicadores, cada miembro del grupo tiene una etiqueta detectable diferente y una secuencia de nucleótidos diferente. Además, las secuencias de nucleótidos para los ácidos nucleicos complementarios o los ácidos nucleicos complementarios de los complejos indicadores del primer grupo son diferentes de las secuencias de nucleótidos para los del segundo grupo. Sin embargo, las etiquetas detectables específicas de la base son comunes a los grupos de ácidos nucleicos complementarios, por ejemplo, las timidinas se identifican mediante señales verdes. Aquí, el segundo nucleótido en el dominio de unión diana se une a la C en el ácido nucleico diana. Las segundas regiones de unión de la sonda tienen una secuencia de nucleótidos que especifica que el segundo nucleótido en el

dominio de unión diana de la sonda se une a una citidina. Por lo tanto, solo los ácidos nucleicos complementarios que comprenden una etiqueta detectable o complejos indicadores del segundo grupo y para la citidina se unen a la segunda posición del dominio del código de barras. Como se muestra, el segundo ácido nucleico complementario que codifica la citidina o el complejo indicador está unido en la segunda posición del dominio de código de barras de la sonda.

En las realizaciones, los pasos mostrados en la Figura 8C son posteriores a los pasos mostrados en la Figura 8B. Aquí, una vez que se ha reemplazado el primer grupo de ácidos nucleicos complementarios o complejos indicadores (de la Figura 8A) con los primeros ácidos nucleicos hibridantes que carecen de una etiqueta detectable (en la Figura 8B), se proporciona un segundo grupo de ácidos nucleicos complementarios o complejos indicadores (como se muestra en la Figura 8C). Alternativamente, los pasos que se muestran en la Figura 8C son concurrentes con los pasos que se muestran en la Figura 8B. Aquí, los primeros ácidos nucleicos hibridantes que carecen de una etiqueta detectable (en la Figura 8B) se proporcionan simultáneamente con un segundo grupo de ácidos nucleicos complementarios o complejos indicadores (como se muestra en la Figura 8C).

La figura 8D continúa el método mostrado en la figura 8C. Aquí, las posiciones primera a quinta en el dominio del código de barras estaban unidas por ácidos nucleicos complementarios que comprenden una etiqueta detectable o complejo indicador y han sido reemplazados por ácidos nucleicos hibridantes que carecen de etiquetas detectables. La sexta posición del dominio de código de barras está actualmente unida por un ácido nucleico complementario que comprende una etiqueta detectable o complejo indicador, que identifica la sexta posición en el dominio de unión diana como unida a una guanina (G).

Como se mencionó anteriormente, los ácidos nucleicos complementarios que comprenden etiquetas detectables o complejos indicadores se pueden eliminar de las regiones de unión, pero no se pueden reemplazar con ácido nucleico hibridante que carece de etiquetas detectables.

Si es necesario, la tasa de intercambio de etiqueta detectable puede acelerarse incorporando pequeños oligonucleótidos de cadena sencillas que aceleran la tasa de intercambio de etiquetas detectables (por ejemplo, sondas “Toe-Hold”; ver, por ejemplo, Seeling et al., “Catalyzed Relaxation of a Metastable DNA Fuel”; J. Am. Chem. Soc. 2006, 128(37), pp12211-12220).

Es posible reemplazar 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 una sonda de secuenciación debe ser reemplazada por otra sonda de secuenciación. De hecho, la sonda de secuenciación de la Figura 8D ahora puede deshibridarse y eliminarse del ácido nucleico diana y reemplazarse por una segunda sonda de secuenciación (superpuesta o no superpuesta) que aún no se ha unido a ningún ácido nucleico complementario, como se muestra en 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 los pasos del método; sin embargo, las Figuras 9A y 9D a 9G muestran claramente que los complejos indicadores (que comprenden etiquetas detectables) están unidos a las regiones de unión de las sondas de secuenciación. Las Figuras 9D y 9E muestran señales fluorescentes emitidas por sondas hibridadas a complejos indicadores. Las Figuras 9D y 9E muestran que el ácido nucleico diana tiene una secuencia de “TA”.

La Figura 10 resume los pasos mostrados 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 separador de ADN de doble cadena opcional entre su dominio de unión diana y su dominio de código de barras. El dominio de código de barras comprende, en orden, una porción “Flanco 1”, una porción “AR-1”, una porción “AR-1/Flanco 2”, una porción “AR-2” y una porción “AR-2/Flanco 3”. En el Paso 1, la “Detección AR-1” se hibrida con las porciones “AR-1” y “AR-1/Flanco 2” de la sonda. “Detección AR-1” corresponde a un complejo indicador o ácido nucleico complementario que comprende una etiqueta detectable que codifica una timidina de primera posición. Por lo tanto, el Paso 1 corresponde a la Figura 9D. En el Paso 2, “Carece 1” se hibrida con las porciones “Flanco 1” y “AR-1” de la sonda. “Carece 1” corresponde al ácido nucleico de hibridación que carece de una etiqueta detectable que sea específico de la primera región de unión de la sonda (como se muestra en la Figura 9E como una barra negra que cubre la primera región de unión). Al hibridarse a la posición “Flanco 1”, que es 5’ del complejo indicador o ácido nucleico complementario, el ácido nucleico hibridante desplaza más eficientemente el complejo indicador/ácido nucleico complementario de la sonda. Las porciones de “Flanco” también se conocen como “Toe-Holds”. En el Paso 3, la “Detección AR-2” se hibrida con las porciones “AR-2” y “AR-2/Flanco 3” de la sonda. “Detección AR-2” corresponde a un complejo indicador o ácido nucleico complementario que comprende una etiqueta detectable que codifica una segunda posición Guanina. Por lo tanto, el Paso 3 corresponde a la Figura 9E. En esta realización, el ácido nucleico de hibridación que carece de una etiqueta detectable y los ácidos nucleicos complementarios que comprenden etiquetas detectables/complejos indicadores se proporcionan secuencialmente.

Alternativamente, el ácido nucleico hibridante que carece de una etiqueta detectable y ácidos nucleicos complementarios que comprenden etiquetas detectables/complejos indicadores se proporcionan simultáneamente. Esta realización alternativa se muestra en la Figura 11. En el Paso 2, se proporciona “Carece 1” (ácido nucleico

hibridante que carece de una etiqueta detectable) junto con “Detección AR-2” (complejo indicador que codifica una segunda posición Guanina). Esta realización alternativa puede ser más efectiva en el tiempo que la realización ilustrada en la Figura 10 porque combina dos pasos en uno.

La figura 12 ilustra los métodos descritos en este documento. Aquí, un ácido nucleico diana es capturado e inmovilizado en dos posiciones, produciendo así un “espacio” a la que una sonda puede unirse. Una primera población de sondas se hibrida sobre el ácido nucleico diana y se detectan etiquetas detectables. Los pasos iniciales se repiten con una segunda población de sondas, una tercera población de sondas, a más de 100 poblaciones de sondas. El uso de aproximadamente 100 poblaciones de sondas proporciona una cobertura de aproximadamente 5X de cada nucleótido en un ácido nucleico diana. La Figura 12 proporciona tasas estimadas de tiempos de lectura en función del tiempo requerido para detectar señales de un Campo de Visión (FOV).

La distribución de sondas a lo largo de una longitud de ácido nucleico diana es crítica para la resolución de la señal detectable. Como se discutió anteriormente, el límite de resolución para dos etiquetas detectables es de aproximadamente 600 nucleótidos. Preferiblemente, cada sonda de secuenciación en una población de sondas se unirá a no más de 600 nucleótidos entre sí. Como se discutió anteriormente, 600 nucleótidos es el límite de resolución de un aparato de secuenciación típico. En este caso, una sonda de secuenciación proporcionará una sola lectura; esto se muestra en la Figura 12 en el punto de resolución limitada más a la izquierda.

Aleatoriamente, pero en parte dependiendo de la longitud del dominio de unión diana, la Tm de las sondas y la concentración de las sondas aplicadas, es posible que dos sondas de secuenciación distintas en una población se unan dentro de 600 nucleótidos entre sí. En este caso, se emitirán múltiples lecturas desordenadas de un único punto con resolución limitada; esto se muestra en la Figura 12 en el segundo punto con resolución limitada.

Alternativa o adicionalmente, la concentración de sondas de secuenciación en una población puede reducirse para disminuir la cobertura de las sondas en una región específica de un ácido nucleico diana, por ejemplo, por encima del límite de resolución del aparato de secuenciación, produciendo así una sola lectura desde un punto de resolución limitada.

La Figura 23 muestra un esquema de una sonda de secuenciación distinta de la utilizada en las Figuras 8 a 12. Aquí, cada posición en un dominio de código de barras está unida por ácidos nucleicos complementarios que comprenden etiquetas detectables o por complejos indicadores. Por lo tanto, en este ejemplo, se puede leer una secuencia de seis nucleótidos sin necesidad de reemplazar secuencialmente ácidos nucleicos complementarios. El uso de esta sonda de secuenciación reduciría el tiempo para obtener información de secuencia ya que se omiten muchos pasos del método descrito. Sin embargo, esta sonda se beneficiaría de las etiquetas detectables que no se superponen, por ejemplo, los fluoróforos son excitados por longitudes de onda de luz no superpuestas o los fluoróforos emiten longitudes de onda de luz no superpuestas.

El método comprende además etapas de ensamblar cada orden lineal identificado de nucleótidos para cada región del ácido nucleico diana inmovilizado, identificando así una secuencia para el ácido nucleico diana inmovilizado. Los pasos de ensamblaje utilizan un medio de almacenamiento legible por ordenador no transitorio con un programa ejecutable almacenado en él. El programa instruye a un microprocesador para organizar cada orden lineal identificado de nucleótidos para cada región del ácido nucleico diana, obteniendo así la secuencia del ácido nucleico. El ensamblaje puede ocurrir en “tiempo real”, es decir, mientras se recopilan datos de sondas de secuenciación en lugar de después de que se hayan recopilado todos los datos.

Cualquiera de los aspectos y realizaciones anteriores se puede combinar con cualquier otro aspecto o realización como se describe aquí en las secciones de Resumen y/o Descripción Detallada.

Definiciones:

En ciertas realizaciones ejemplares, los términos “hibrida” e “hibridación”, como se usan en el presente documento, se usan indistintamente para significar la formación de un dúplex estable. En un aspecto, el dúplex estable significa que una estructura dúplex no se destruye por un lavado riguroso en condiciones tales como una temperatura de aproximadamente 5°C por debajo o aproximadamente 5°C por encima de la Tm de una cadena del dúplex 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 técnica. El término “perfectamente emparejado”, cuando se usa en referencia a un dúplex significa que las cadenas de polinucleótidos y/u oligonucleótidos que forman el dúplex forman una estructura de doble cadena entre sí, de modo que cada nucleótido en cada cadena experimenta un emparejamiento de bases Watson-Crick con un nucleótido en la otra cadena. El término “dúplex” comprende, pero no se limita a, el emparejamiento de análogos de nucleósidos, tales como desoxiinosina, nucleósidos con bases de 2-aminopurina, PNA y similares, que pueden emplearse. Un “emparejamiento incorrecto” en un dúplex entre dos oligonucleótidos significa que un par de nucleótidos en el dúplex no puede someterse al enlace Watson-Crick.

Como se usa en el presente documento, el término “condiciones de hibridación” incluirá típicamente concentraciones de sal de menos de aproximadamente 1 M, más usualmente menos de aproximadamente 500 mM e incluso más

usualmente 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, superiores a aproximadamente 37°C. Las hibridaciones se realizan generalmente en condiciones rigurosas, por ejemplo, condiciones en las que una sonda hibridará específicamente con su subsecuencia diana. Las condiciones estrictas dependen de la secuencia y son diferentes en diferentes circunstancias. Los fragmentos más largos pueden requerir temperaturas de hibridación más altas para la hibridación específica. Como otros factores pueden afectar la rigurosidad de la hibridación, incluida la composición de la base y la longitud de las cadenas complementarias, la presencia de disolventes orgánicos y el grado de emparejamiento incorrecto de la base, la combinación de parámetros es más importante que la medida absoluta de cualquiera sola.

En general, las condiciones rigurosas se seleccionan para que sean aproximadamente 5°C más bajas que la Tm para la secuencia específica a una fuerza iónica y pH definidos. Las condiciones rigurosas ejemplares incluyen una concentración de sal de al menos 0.01 M a no más de 1 M de concentración de iones Na (u otras sales) a un pH de 7.0 a 8.3 y una temperatura de al menos 25°C. Por ejemplo, las condiciones de 5X SSPE (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 sonda específicas de alelo. Para condiciones estrictas, ver, por ejemplo, Sambrook, Fritsche y Maniatis, “Molecular Cloning A Laboratory Manual, 2nd Ed.” Cold Spring Harbor Press (1989) y Anderson Nucleic Acid Hybridization, 1st Ed., BIOS Scientific Publishers Limited (1999). Como se usa en el presente documento, los términos “hibridando específicamente a” o “hibridando específicamente a” o términos similares se refieren a la unión, duplicación o hibridación de una molécula sustancialmente a una secuencia o secuencias de nucleótidos particulares en condiciones rigurosas.

Las etiquetas detectables asociadas con una posición particular de una sonda pueden ser “leídas” (por ejemplo, su fluorescencia detectada) una o varias veces; una “lectura” puede ser sinónimo del término “identificación de nucleótidos”. Las lecturas múltiples mejoran la precisión. Una secuencia de ácido nucleico diana se “lee” cuando se detecta un tramo contiguo de información de secuencia derivada de una única molécula diana original; típicamente, esto se genera a través del consenso de múltiples pasadas (como se define a continuación). Como se usa en este documento, el término “cobertura” o “profundidad de cobertura” se refiere al número de veces que se ha secuenciado una región de destino (mediante lecturas discretas) y se ha alineado con una secuencia de referencia. La cobertura de lectura es el número total de lecturas que se asignan a una secuencia objetivo de referencia específica; La cobertura base es el número total de llamadas base realizadas en una posición genómica específica.

Como se usa en el presente documento, un “ciclo hybe y seq” se refiere a todos los pasos necesarios para detectar cada región de unión en una sonda particular o población de sondas. Por ejemplo, para una sonda capaz de detectar seis posiciones en un ácido nucleico diana, un “ciclo hybe y seq” incluirá, al menos, hibridación de la sonda con el ácido nucleico diana, hibridación de ácidos nucleicos complementarios/complejos indicadores a la región de unión en cada una de las seis posiciones en el dominio de código de barras de la sonda y detectar las etiquetas detectables asociadas con cada una de las seis posiciones.

El término “sonda de k-mero” es sinónimo de una sonda de la presente invención o divulgación.

Cuando dos o más secuencias de lecturas discretas están alineadas, las porciones superpuestas se pueden combinar para crear una secuencia consenso única. En las posiciones donde las porciones superpuestas tienen la misma base (una sola columna de la alineación), esas bases se convierten en el consenso. Se pueden usar varias reglas para generar el consenso para las posiciones donde hay desacuerdos entre las secuencias superpuestas. Una regla de mayoría simple usa la base más común en la columna como consenso. Un “consenso de múltiples pasos” es una alineación de todas las lecturas de sonda discretas de una sola molécula objetivo. Dependiendo del número total de ciclos de poblaciones de sonda/encuestas aplicadas, cada posición de base dentro de una sola molécula objetivo puede consultarse con diferentes niveles de redundancia o superposición; en general, la redundancia aumenta el nivel de confianza de una identificación de nucleótidos.

La “precisión bruta” es una medida de la capacidad inherente del sistema para identificar correctamente una base. La precisión bruta depende de la tecnología de secuenciación. La “precisión de consenso” es una medida de la capacidad del sistema para identificar correctamente una base con el uso de lecturas adicionales y poder estadístico. La “especificidad” se refiere al porcentaje de lecturas que se asignan a los objetivos previstos del total de lecturas por serie. “Uniformidad” se refiere a la variabilidad en la cobertura de secuencia en las regiones diana; alta uniformidad se correlaciona con baja variabilidad. Esta característica se informa comúnmente como la fracción de regiones diana cubiertas por ≥20% de la profundidad de cobertura promedio en todas las regiones diana. Los errores estocásticos (es decir, los errores químicos de secuenciación intrínseca) pueden corregirse fácilmente con la secuenciación 'multipaso' del mismo ácido nucleico diana; dado un número suficiente de pases, se puede lograr una secuenciación sustancialmente 'consenso perfecto' o 'libre de errores'. Los métodos descritos en este documento pueden implementarse y/o registrarse los resultados utilizando cualquier dispositivo capaz de implementar los métodos y/o registrar los resultados. Los ejemplos de dispositivos que se pueden usar incluyen, entre otros, dispositivos informáticos electrónicos, incluidos ordenadores de todo tipo. Cuando los métodos descritos en este documento se implementan y/o graban en un ordenador, el programa de ordenador que puede usarse para configurar el ordenador para llevar a cabo los pasos de los métodos puede estar contenido en cualquier medio legible por ordenador capaz de contener el programa de ordenador. Los ejemplos de medios legibles por ordenador que se pueden usar incluyen,

entre otros, disquetes, CD-ROM, DVD, ROM, RAM, medios legibles por ordenador no transitorios y otros dispositivos de almacenamiento de memoria y ordenador. El programa de ordenador que se puede usar para configurar el ordenador para llevar a cabo los pasos de los métodos, reunir información de secuencia y/o registrar los resultados también se puede proporcionar a través de una red electrónica, por ejemplo, a través de Internet, una intranet o otra red.

Se puede incorporar una “Tarjeta de secuencia de consumibles” (Figura 24) en un dispositivo de imagen de fluorescencia conocido en la técnica. Cualquier microscopio de fluorescencia con varias características diferentes es capaz de realizar esta lectura de secuencia. Por ejemplo: la lámpara de campo amplio, láser, LED, multifotón, iluminación de reflexión confocal o interna total se puede utilizar para la excitación y/o detección. La cámara (simple o múltiple) y/o el tubo fotomultiplicador (simple o múltiple) con resolución espectral basada en filtro o en rejilla (una o más longitudes de onda de emisión resueltas espectralmente) son posibles en el canal de detección de emisiones del microscopio de fluorescencia. Los ordenadores estándar pueden controlar tanto la tarjeta de secuencia consumible, los reactivos que fluyen a través de la tarjeta y la detección por el microscopio de fluorescencia.

Los datos de secuenciación pueden ser analizados por cualquier número de ensambladores de secuenciación estándar de próxima generación (ver, 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 única región limitada por difracción del microscopio se “ensamblan localmente” para generar una secuencia de consenso de las lecturas múltiples dentro de un punto de difracción. Las lecturas ensambladas en el punto de difracción múltiple se mapean juntas para generar secuencias contiguas que representan el conjunto de genes diana completo, o un ensamblaje de novo de genomas completos.

Se describen enseñanzas adicionales relevantes para la presente invención y divulgación en uno o más de los siguientes documentos: US 8,148,512, US 7,473,767, US 7,919,237, US 7,941,279, US 8,415,102, US 8,492,094, US 8,519,115, US 2009/0220978, US 2009/0299640, US 2010/0015607, US 2010/0261026, US 2011/0086774, US 2011/0145176, US 2011/0201515, US 2011/0229888, US 2013/0004482, US 2013/0017971, US 2013/0178372, US 2013/0230851, US 2013/0337444, US 2013/0345161, US 2014/0005067, US 2014/0017688, US 2014/0037620, US 2014/0087959, US 2014/0154681 y US 2014/0162251.

Ejemplos

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

A continuación, se describe el tiempo para los pasos en los métodos de la presente invención y como se muestra en las Figuras 8 a 12.

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

Los cálculos para una serie ejemplar se muestran en la Figura 14 y los cálculos para los tiempos de ciclo 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 diana en la cadena principal sintética. Con el dispositivo de intercambio de fluidos controlado por microfluidos, el lavado de las sondas no unidas demora aproximadamente medio segundo.

La adición de un primer grupo de ácidos nucleicos complementarios (que comprende una etiqueta detectable) y su unión a las regiones de unión en la primera posición del dominio del código de barras lleva aproximadamente quince segundos. Cada campo de visión (FOV) tiene imágenes de cuatro colores diferentes, cada color representa una base única. Las manchas fiduciales colocadas en una sonda de captura de 5’ o una sonda de captura de 3' (o ambas) pueden ser útiles para leer solo los códigos de barras ópticos en línea (de acuerdo con la presencia de ácido nucleico diana con espacio) entre las dos ubicaciones. También se pueden agregar puntos fiduciales a cada campo de visión para generar una alineación igual de las imágenes en pasos sucesivos en el proceso de secuenciación. Las cuatro imágenes se pueden obtener en un solo FOV y luego el dispositivo de lectura óptica puede pasar a un nuevo FOV, o tomar todo el FOV en un color y luego volver a crear una imagen en un segundo color. Un solo campo de visión se puede leer en aproximadamente medio segundo. Se tarda aproximadamente medio segundo en pasar al siguiente campo de visión. Por lo tanto, el tiempo para leer “n” FOV es igual a “n” por 1 segundo).

Los ácidos nucleicos complementarios que tienen etiquetas detectables se eliminan de la primera posición del dominio del código de barras mediante la adición de calor o lavado con exceso de ácidos nucleicos complementarios que carecen de etiquetas detectables. Si es necesario, la tasa de intercambio de etiqueta detectable puede acelerarse incorporando pequeños oligonucleótidos de cadena sencillas que aceleran la tasa de intercambio de etiquetas

detectables (por ejemplo, sondas “Toe-Hold”; ver, por ejemplo, Seeling et al., “Catalyzed Relaxation of a Metastable DNA Fuel”; J. Am. Chem. Soc. 2006, 128(37), pp 12211-12220). Se puede volver a crear una imagen de FOV para confirmar que todos los ácidos nucleicos complementarios que tienen etiquetas detectables se eliminan antes de continuar. Esto lleva unos quince segundos. Este paso puede repetirse hasta alcanzar los niveles de señal de fondo.

Los pasos anteriores se repiten o las posiciones restantes en el dominio de código de barras de las sondas.

El tiempo total para leer iguales m (lecturas bases) por (15 seg + n FOV veces 1 seg + 15 seg). Por ejemplo, cuando el número de posiciones en el dominio del código de barras es de 6 y 20 FOV, el tiempo de lectura es igual a 6 X (30 + 20 + 15) o 390 segundos.

Las sondas de la primera población se deshibridan. Esto lleva unos sesenta segundos.

Los pasos anteriores se repiten para las segundas y posteriores poblaciones de sondas. Si las poblaciones de sondas de secuenciación se organizan por temperatura de fusión (Tm), cada población de sondas requerirá múltiples hibridaciones para garantizar que cada base esté cubierta a la profundidad requerida (esto se debe a la tasa de error). Además, al analizar las lecturas de hibridación durante una ejecución, es posible reconocer cada gen individual que se está secuenciando mucho antes de que se determine realmente la secuencia completa. Por lo tanto, el ciclo puede repetirse hasta que se cumpla una frecuencia de error (o cobertura) deseada particular.

Utilizando el tiempo descrito anteriormente, junto con algunas estimaciones de densidad de unión de ácido nucleico con espacios, se puede estimar el rendimiento de un Secuenciador de Nanocadena de Siguiente Generación (NSTG) - Next Generation de la presente invención.

El rendimiento neto del secuenciador viene dado por:

> Ocupación de Base Fraccionaria X <longitud de espacio> X número de espacios por FOV X Número de bases por código de Barras Óptico / [60 s (sondas de hibridación para ácido nucleico diana) + 0.5 s (lavado) + m: posiciones en el dominio de código de barras X (15 s (ácidos nucleicos complementarios de unión) + nfovsX1 + 15 s (ácidos nucleicos complementarios de no unión)) + 60 s (sondas de deshibridación para ácido nucleico)]

Por lo tanto, en un ejemplo, un “ciclo” total para un solo ácido nucleico con espacios (sumando el método que se muestra en la Figura 10):

60 segundos (sondas de hibridación con ácido nucleico diana) + 0.5 segundos (lavado) + m-bases X (15 s (ácidos nucleicos complementarios de unión) + nFOVs por 1 + 15 s (ácidos nucleicos complementarios no unidos)) + 60 s (sondas de deshibridación al ácido nucleico diana). Utilizando m = 6, nFOV = 20, produce tiempo = 60 + 0.5 + 390 + 60 = 510.5 s.

Suponiendo que: 1% de ocupación de la región de ácido nucleico con espacios, 4000 bases por espacio, y 5000 fragmentos de nucleósidos con espacios por FOV y una m de 6 y nFOV de 20 (como se describió anteriormente) produce un rendimiento neto de:

0.01X 4000 X 5000 X 20 = 4,000,000 6 lecturas base por 510.5 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. Estimaciones alternativas de hasta 12 Gb por día. Ver Figura 12.

Como se muestra en la Figura 14, el tiempo de ejecución requerido para secuenciar 100 ácidos nucleicos diana diferentes (un “100-plex”) es de aproximadamente 4.6 horas; El tiempo de ejecución requerido para secuenciar 1000 ácidos nucleicos diana diferentes (un “1000-plex”) es de aproximadamente 16 horas.

La Figura 16 compara la velocidad de secuenciación, el número de lecturas y la utilidad clínica para la presente invención y varios otros métodos/aparatos de secuenciación.

Ejemplo 2: El método tiene una baja tasa de error

La figura 17 muestra que la presente invención tiene una tasa de error bruto de aproximadamente 2.1%, cuando se omiten las posiciones terminales.

Para la invención reivindicada, una tasa de error asociada con la secuenciación está relacionada con la diferencia de energía libre entre un (m + n) mer completamente emparejado y un emparejamiento erróneo de base única (m-1 + n). La suma de m + n es el número de nucleótidos en un dominio de unión diana y m representa el número de posiciones en un dominio de código de barras. Se puede hacer una estimación de la selectividad de la hibridación utilizando la ecuación (See, Owczarzy, R. (2005), Biophys. Chem., 117: 207-215 y el sitio Integrated DNA Technologies: en la

World Wide Web (www) idtdna.com/analyzer/Applications/Instructions/Default.aspx?AnalyzerDefinitions=true#Mismatc hMeltTemp):

$$\theta = 1 - \left(\frac{K_a([cadena2] - [cadena1]) - 1}{2K_a[cadena2]} + \frac{\sqrt{K_a^2([cadena1] - [cadena2])^2 + 2K_a([cadena1] + [cadena2]) + 1}}{2K_a[cadena2]} \right)$$

donde K_a es la constante de equilibrio de asociación obtenida de los parámetros termodinámicos predichos,

$$K_a = \exp\left(\frac{-(\Delta H° - T\Delta S°)}{RT} \right)$$

Theta representa el porcentaje de unión del complemento exacto y las secuencias de emparejamiento erróneo de base única, que se espera que sean recocidas para alcanzar la temperatura de hibridación especificada. La T es la temperatura de hibridación en Kelvins, ΔH° (entalpía) y ΔS° (entropía) son los parámetros de fusión calculados a partir de la secuencia y los parámetros termodinámicos vecinos más cercanos publicados, R es la constante de gas ideal (1.987 cal $K^{-1}mol^{-1}$) [cadena 1/2] es la concentración molar de un oligonucleótido, y la constante de -273.15 convierte la temperatura de Kelvin a grados Celsius. Los parámetros de vecino más cercano y más precisos se obtuvieron de las siguientes publicaciones para pares de bases de ADN/ADN (Ver, Allawi, H., SantaLucia, J. Biochemistry, 36, 10581), pares de bases de ARN/ADN (Ver, Sugimoto et al. ., Biochemistry, 34, 11211-6), pares de bases de ARN/ARN (Ver, Xia, T. Et al., Biochemistry, 37, 14719),

Como ejemplo de una estimación de la tasa de error aproximada esperada del secuenciador NSTG sigue. Para (m + n) es igual a 8' mer. Considere el siguiente código de barras de 8 unidades y su emparejamiento incorrecto de base única.

5'ATCGTACG3'

(región a secuencia)

3'TAGCATGC5'

(secuencia de código de barras óptico con coincidencia perfecta)

3'TAGTATGC5'

(secuenciación de código de barras óptico con emparejamiento incorrecto de base única (GT))

Utilizando la calculadora IDT basada en las ecuaciones anteriores se obtiene:

A 17.4°C (la Tm del caso de coincidencia perfecta), (50%/0.3%) sería la relación del código de barras óptico correcto hibridado a esa secuencia versus la secuencia código de barras incorrecto en la Tm, produciendo una tasa de error estimada para esa secuencia de 0.6%.

Un cálculo de secuenciación de contenido de GC muy alto produce:

5'CGCCGGCC3'

(región a secuencia)

3'GCGGCCGG5'

(secuenciación de código de barras óptico con coincidencia perfecta)

3'GCGGACGG5'

(secuenciación de código de barras óptico con error de emparejamiento incorrecto de una sola base (GA) - emparejamiento)

A 41.9°C (la Tm del caso de coincidencia perfecta), (50%/0.4%) sería la proporción del código de barras óptico correcto hibridado a esa secuencia frente al código de barras incorrecto en la Tm, produciendo una tasa de error estimada para que la secuencia sea del 0.8%.

El examen de un número de pares de 8 mer produce una distribución de tasas de error, en el rango de 0.2% a 1%. Si bien los cálculos anteriores no serán idénticos a las condiciones utilizadas, estos cálculos proporcionan una indicación de que el método de la presente invención tendrá una tasa de error intrínseca relativamente baja, en comparación con

otras tecnologías de secuenciación de moléculas individuales, 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 invención es mayor que otros métodos de secuenciación. Por lo tanto, la presente invención proporciona una secuencia de consenso de un solo objetivo después de menos pases que los requeridos para otros métodos de secuenciación. Además, la presente invención puede obtener un secuenciamiento de "consenso perfecto"/"libre de errores" (es decir, 99.9999%/Q60) después de 30 o más pasadas, mientras que los métodos de secuenciación PacBio (por ejemplo) no pueden lograr dicho consenso después de 70 pasadas.

Ejemplo 3: Capacidad de resolución de un solo par base

La figura 19 muestra que la presente invención tiene una resolución de base única y con tasas de error bajas (que varían de 0% a 1.5% dependiendo de una sustitución de nucleótidos específica).

Se realizaron experimentos adicionales utilizando un ARN diana hibridado con código de barras e inmovilizado en la superficie del cartucho utilizando tecnología normal de unión de expresión génica NanoString (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 de dominio de unión diana y con una coincidencia perfecta (código de barras óptico YGBYGR-2um conectado a la secuencia de coincidencia perfecta de 10 meros) para hibridarse con el ARN diana (Figura 26). Una mayor longitud del dominio de unión diana proporciona recuentos más altos. También muestra que el dominio de unión diana de 10 unidades es suficiente para registrar la secuencia sobre el fondo. Cada una de las coincidencias alteradas de una sola base individual se sintetizó con códigos de barras ópticos alternativos. Se contó la proporción de códigos de barras ópticos correctos a incorrectos (Figuras 24 y 25).

La capacidad de 10 mer para detectar un SNP, la secuencia real es >15000 cuentas sobre el fondo, mientras que las secuencias incorrectas son como máximo > 00 sobre el fondo. En presencia de la sonda correcta, se espera que las tasas de error sean <3% de la secuencia real. Tenga en cuenta que estos datos son (en esencia) un peor escenario. Tener solo una secuencia de hibridación de 10 pares de bases unida a un indicador de código de barras óptico 6.6 Kilobase (estilo Gen2). No se realizaron optimizaciones de condiciones específicas. Sin embargo, estos datos revelan que el enfoque de secuenciación NanoString Next-Generation es capaz de resolver pares de secuencia de una sola base.

Los materiales y métodos detallados utilizados en el estudio anterior son los siguientes:

Protocolo de hibridación Sonda B más conjunto de códigos

- Tomar elementos 25 ul (conjunto de códigos 194)
- Agregar 5 ul Sonda B + secuencia complementaria al objetivo (100 uM)
- Agregar 15 ul Hyb Buffer (14.56 X SSPE 0.18% Tween 20)
  SSPE (NaCl 150 mM, $NaH_2PO_4xH_2O$ 10 mM, Na_2EDTA 10 mM)
- Incubar en hielo durante 10 min
- Agregar 150 ul G bolas (40 ul G bolas a 10 mg/ml más 110 ul 5x SSPE 0.1 % Tween 20)
- Incubar durante 10 minutos a temperatura ambiente
- Lavar tres veces con 0.1SSPE 0.1% Tween 20 utilizando un colector magnético
- Eluir en 100 ul 0.1 x SSPE durante 10 minutos a 45°C.

Protocolo de hibridación objetivo (NaCl 750 mM)

- Tome 20 ul por encima de la muestra eluida
- Agregue 10 ul de regulador hyb
- Agregue 1 ul Target (ARN biotinilado 100 nM)
- Incubar en hielo durante 30 minutos

Tomar 15 ul y unir en el portaobjetos de estreptavidina durante 20 minutos, estirar el flujo con ganchos G, recuento utilizando materiales nCounter

Materiales

El conjunto de códigos Elements 194

Oligos comprado de IDT

SSPE (NaCl 150 mM, $NaH_2PO_4xH_2O$ 10 mM, Na_2EDTA 10 mM)

Tampón Hyb (14.56 X SSPE 0.18% Tween 20)

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

GBRYBG	5	GACTGTACCCACGCGATGACGTTCGTCAAGAGTCGCATAATCT	3
YRBYRG	5	AGACTGTACCACAAGAATCCCTGCTAGCTGAAGGAGGGTCAAAC	3
YGBYGR	5	GAGACTGTACCCTACGTATATATCCAAGTGGTTATGTCCGACGGC	3
GBRYGB	5	TGAGACTGTACCACCCCTCCAAACGCATTCTTATTGGCAAATGGAA	3
RYGBRG	5	CTGAGACTGTACCCGGGAATCGGCATTTCGCATTCTTAGGATCTAAA	3

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

RNA	5	CAATGTGAGTCTCTTG**GTACAGTCTCAG**TTAGTCACTCCCTAAG\Bio TEG\	3

Tabla 4: Secuencias de la sonda B para emparejamiento incorrecto de 10 mer (en negrita; SEQ ID NO: 36 a SEQ ID NO: 41)

10mermis2A	GAGAC**A**GTACCCTGGTCTAGGTATCTAATTCGTGGGTCGGGTAC T
10mermis2C	GAGAC**C**GTACCGCTCATTTTGAACATACGATTGCGATTACGGAA A
10mermis2G	GAGAC**G**GTACCTTAAAGCTATCCACGAATGTCAAAAATGTGGTT T
10mermis 1G	GAGA**G**TGTACCCAATGCTTGCAGTATGTATCCTGATCGTGCGTG C
10mermis1A	GAGA**A**TGTACCCTCATACCAATGTAAAGTATAGTTAACGCCCTG T
10mermis 1T	GAGA**T**TGTACCCTACATATATAGGAAAAGGGAAGGTAGAAGAG CT

Lista de secuencias

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

<120> SECUENCIACIÓN SIN ENZIMA NI AMPLIFICACIÓN

<130> NATE-025/001

<150> US 62/082,883

<151> 2014-11-21

<160> 104

<170> Patente en la versión 3.5

<210> 1

<211> 10

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 1

atacatctag 10

<210> 2

<211> 10

<212> ADN

<213> Secuencia Artificial

< 220>

<223> Polinucleótido sintético

<400> 2

gatctacata 10

<210> 3

<211> 10

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 3

ttaggtaaag 10

<210> 4

<211 > 10

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 4

tcttcattac 10

<210> 5

<211> 10

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 5

atgaatctac 10

<210> 6

<211> 10

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 6

tcaatgtatg 10

<210> 7

<211> 10

<212> ADN

< 213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 7

aattgagtac 10

<210> 8

<211> 10

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 8

atgttaatgg 10

< 210> 9

<211> 10

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 9

aattaggatg 10

<210> 10

<211> 10

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 10

ataatggatc 10

<210> 11

<211> 10

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 11

taataaggtg 10

<210> 12

<211> 10

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 12

tagttagagc 10

<210> 13

< 211> 10

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 13

atagagaagg 10

<210> 14

<211> 10

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 14

ttgatgatac 10

<210> 15

<211> 10

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 15

atagtgattc 10

<210> 16

<211> 10

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 16

tataacgatg 10

<210> 17

<211> 10

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 17

ttaagtttag 10

<210> 18

<211> 10

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 18

atacgttatg 10

<210> 19

<211> 10

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 19

tgtactatag 10

<210> 20

<211> 10

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 20

ttaacaagtg 10

<210> 21

<211> 10

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 21

aactatgtac 10

<210> 22

<211> 10

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 22

taactatgac 10

<210> 23

<211> 10

<212> ADN

<213> Secuencia Artificial

< 220>

<223> Polinucleótido sintético

<400> 23

actaatgttc 10

<210> 24

<211> 10

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 24

tcattgaatg 10

<210> 25

<211 > 14

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 25

ctgtctcatc tctt 14

<210> 26

<211> 26

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 26

ctgtctcatc tcttgctgca tcctgt 26

<210> 27

<211> 38

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 27

ctgtctcatc tcttgctgca tcctgtcggt tcacgttg 38

<210> 28

<211> 36

<212> ADN

<213> Secuencia Artificial

<220>

<223 > Polinucleótido sintético

<400> 28

ctgtctcatc ttgctgcatc ctgtcggttc acgttg 36

<210> 29

<211> 36

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 29

ctgtctcatt ttgctgcatc ctgtccgttc 36

<210> 30

<211> 43

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 30

gactgtaccc acgcgatgac gttcgtcaag agtcgcataa tct 43

<210> 31

<211> 44

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido

<400> 31

agactgtacc acaagaatcc ctgctagctg aaggagggtc aaac 44

<210> 32

<211> 45

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 32

gagactgtac cctacgtata tatccaagtg gttatgtccg acggc 45

<211> 46

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 33

tgagactgta ccacccctcc aaacgcattc ttatt ggcaa atggaa 46

<210> 34

<211> 47

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 34

ctgagactgt acccgggaat cggcatttcg cattcttagg atctaaa 47

<210> 35

<211> 44

<212>

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 35

caatgtgagt ctcttggtac agtctcagtt agtcactccc taag 44

<210> 36

<211> 45

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400 > 36

gagacagtac cctggtctag gtatctaatt cgtgggtcgg gtact 45

<210> 37

<211> 45

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 37

gagaccgtac cgctcatttt gaacatacga ttgcgattac ggaaa 45

<210> 38

<211> 45

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 38

gagacggtac cttaaagcta tccacgaatg tcaaaaatgt ggttt 45

<210> 39

<211> 45

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 39

gagagtgtac ccaatgcttg cagtatgtat cctgatcgtg cgtgc 45

<210> 40

<211> 45

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 40

gagaatgtac cctcatacca atgtaaagta tagttaacgc cctgt 45

<210> 41

<211> 45

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 41

gagattgtac cctacatata taggaaaagg gaaggtagaa gagct 45

<210> 42

<211> 12

<212> DNA

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 42

cacgaacgtc ag 12

<210> 43

<211> 12

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 43

catcgcatgc ct 12

<210> 44

<211> 12

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 44

gtcatctcct ac 12

<210> 45

<211> 12

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 45

gtcatccgct ac 12

<210> 46

<211> 12

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 46

gtcatcgact ac 12

<210> 47

<211> 12

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 47

gtcatcttct ac 12

<210> 48

<211> 12

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 48

gtcatcacct ac 12

<210> 49

<211> 12

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 49

gtcatcactc ac 12

<210> 50

<211> 12

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 50

gtcatcttcg ac 12

<210> 51

<211> 12

<212> ADN

<213> Secuencia Artificial

<220>

<223 > Polinucleótido sintético

<400> 51

gtcatcaact ac 12

<210> 52

<211> 12

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 52

gtcatccgta ac 12

<210> 53

<211> 12

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 53

gtcatccgaa ac 12

<210> 54

<211> 12

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 54

gtcatcacaa ac 12

<210> 55

<211> 12

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 55

gtcatcttgc ac 12

<210> 56

<211> 12

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 56

gtcatcttgc ct 12

<210 > 57

<211> 12

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 57

gtcatccgtc ct 12

<210> 58

<211> 12

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 58

cttttcacct ct 12

<210> 59

<211> 12

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 59

cttttcctct ct 12

<210> 60

<211> 12

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 60

cttttcgact ct 12

<210> 61

<211> 12

<212> ADN

<213> Secuencia Artificial

<220>

<223>

<400> 61

cttttctgct ct 12

<210> 62

<211> 12

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 62

cttttctgta ct 12

<210> 63

<211> 12

<212 > ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 63

cttttctgtg ct 12

<210> 64

<211> 12

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 64

cttttctgtc ct 12

<210> 65

<211> 12

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 65

cttttcactc ct 12

<210> 66

<211> 12

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 66

cttttcgttc ct 12

<210> 67

<211> 12

< 212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 67

cttttcgtac ct 12

<210> 68

<211> 12

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400 > 68

cttttccgtc ct 12

<210> 69

<211> 12

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 69

cttttctgac ct 12

<210> 70

<211> 12

<212> ADN

< 213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 70

aggcatgcga tg 12

<210> 71

<211> 12

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 71

aggcattgtg ct 12

<210> 72

<211> 12

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 72

aggcattgct ct 12

<210> 73

<211> 12

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 73

aggcatttct ac 12

<210> 74

<211> 12

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 74

aggcatacct ac 12

<210> 75

<211> 12

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 75

aggcatttgc ac 12

<210> 76

<211> 12

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 76

aggcatcgtc ct 12

<210> 77

<211> 12

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 77

tcctgtcggt tc 12

<210> 78

<211> 12

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 78

gttcaatgct ct 12

<210> 79

<211> 12

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 79

attcggtgct ct 12

<210> 80

<211> 12

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 80

gatgcctgct ct 12

<210> 81

<211> 12

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 81

tttgcttgct ct 12

<210> 82

<211> 100

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 82

ttcactgtag ctgtctcatt ttgctgcatc ctgtccgttc acgttggagc ttgtcatccg 60

tcctcttttc actcctaggc atttgcctat tcggcgtcct 100

<210> 83

<211> 10

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 83

cgatctggtt 10

<210> 84

<211> 10

<212> ADN

<213 > Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 84

cgatctggtt 10

<210> 85

<211> 10

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 85

gctagaccaa 10

<210 > 86

<211> 10

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 86

gctggaccaa 10

<210> 87

<211> 10

<212> ADN

<213> Secuencia Artificial

<220>

< 223> Polinucleótido sintético

<400> 87

gctcgaccaa 10

<210> 88

<211> 10

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 88

gcttgaccaa 10

<210> 89

<211> 10

< 212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 89

gagactgtac 10

<210> 90

<211> 10

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 90

gagacagtac 10

<210> 91

<211> 10

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 91

gagaccgtac 10

<210> 92

<211> 10

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 92

gagacggtac 10

<210> 93

<211> 10

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 93

gagagtgtac 10

<210> 94

<211> 10

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 94

gagaatgtac 10

<210> 95

<211> 10

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 95

gagattgtac 10

<210> 96

<211> 10

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 96

gagactgtac 10

<210> 97

<211> 10

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 97

gagattgtac 10

<210> 98

<211 > 10

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 98

gagaccgtac 10

<210> 99

<211> 10

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 99

gagagtgtac 10

<210> 100

<211> 12

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 100

ctgagactgt ac 12

<210> 101

<211> 11

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 101

tgagactgta c 11

<210> 102

<211> 10

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 102

gagactgtac 10

<210> 103

<211> 12

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 103

catgtcagag tc 12

<210> 104

<211> 11

<212> ADN

<213> Secuencia Artificial

<220>

<223> Polinucleótido sintético

<400> 104

catgtcagag t 11

REIVINDICACIONES

1. Una sonda de secuenciación que comprende un dominio de unión diana y un dominio de código de barras;

en el que dicho dominio de unión diana comprende al menos ocho nucleótidos y es capaz de unirse a un ácido nucleico diana, y en el que al menos dos nucleótidos en el dominio de unión diana son bases universales capaces de unirse a cualquier base dentro del ácido nucleico diana;

en el que dicho dominio de código de barras comprende una cadena principal sintética, dicho dominio de código de barras comprende al menos seis posiciones de unión, cada posición de unión comprende al menos una región de unión, dicha región de unión comprende al menos una secuencia de ácido nucleico capaz de unirse por una molécula de ácido nucleico complementaria,

en el que cada posición de unión de las al menos seis posiciones de unión corresponde a un nucleótido en el dominio de unión diana y cada una de las al menos seis posiciones de unión tiene una secuencia de ácido nucleico diferente, y en el que dicha secuencia de ácido nucleico de cada posición de las al menos seis posiciones de unión determinan la posición e identidad del nucleótido correspondiente en dicho ácido nucleico diana que está unido por dicho dominio de unión diana, y

en el que las bases universales preceden y siguen a los nucleótidos correspondientes a las posiciones en el dominio de código de barras.

2. La sonda de secuenciación de la reivindicación 1, en la que dicha cadena principal sintética comprende ADN de cadena sencilla.

3. La sonda de secuenciación de la reivindicación 1, en la que dicha sonda de secuenciación comprende un separador de ADN de doble cadena entre el dominio de unión diana y el dominio del código de barras.

4. La sonda de secuenciación de la reivindicación 1, en la que el número de nucleótidos en un dominio de unión diana es al menos dos más que el número de regiones de unión 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 unión; (b) una región de unión; o (c) más de una región de unión.

6. La sonda de secuenciación de la reivindicación 1, en la que cada molécula de ácido nucleico complementaria;

(a) para cada posición de unión comprende una etiqueta detectable;

(b) o está indirectamente ligado a una molécula de ácido nucleico primario a través de un espaciador de ácido nucleico; o

(c) para cada posición de unión comprende entre aproximadamente 8 nucleótidos y aproximadamente 20 nucleótidos, preferiblemente,

en el que cada molécula complementaria de ácido nucleico comprende aproximadamente 12 nucleótidos.

7. La sonda de secuenciación de la reivindicación 6 (b), en la que cada molécula de ácido nucleico primario se hibrida con al menos una, dos, tres, cuatro o cinco moléculas de ácido nucleico secundario.

8. La sonda de secuenciación de la reivindicación 7, en la que la molécula o moléculas secundarias de ácido nucleico comprenden al menos una etiqueta detectable.

9. La sonda de secuenciación de la reivindicación 7, en la que cada molécula secundaria de ácido nucleico se hibrida con al menos una, dos, tres, cuatro, cinco, seis o siete moléculas de ácido nucleico terciario que comprenden al menos una etiqueta detectable.

10. La sonda de secuenciación de la reivindicación 1, en la que una o más posiciones de unión en el dominio de código de barras son adyacentes a al menos un polinucleótido de cadena sencilla flanqueante.

11. Una población de sondas de secuenciación que comprende una pluralidad de la sonda de secuenciación de la reivindicación 1.

12. Un método para secuenciar un ácido nucleico que comprende etapas de:

(1) hibridar al menos una primera población de primeras sondas de secuenciación que comprende una pluralidad de la sonda de secuenciación de una cualquiera de las reivindicaciones 1 a 10 a un ácido nucleico diana que se inmoviliza para un sustrato, en el que el ácido nucleico diana se inmoviliza al sustrato en una o más posiciones;

(2) unir una primera molécula complementaria de ácido nucleico que comprende una etiqueta detectable o una primera molécula complementaria de ácido nucleico de un primer complejo indicador que comprende una etiqueta detectable a una primera posición de unión de las al menos seis posiciones de unión;

(3) detectar la etiqueta detectable de la primera molécula de ácido nucleico complementaria unida o la etiqueta 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) unirse a la primera posición de unión una primera molécula de ácido nucleico hibridante que carece de una etiqueta detectable, desuniendo así la primera molécula complementaria de ácido nucleico que comprende una etiqueta detectable o la primera molécula complementaria de ácido nucleico del primer complejo indicador que comprende una etiqueta detectable;

(6) unir una segunda molécula complementaria de ácido nucleico que comprende una etiqueta detectable o una segunda molécula complementaria de ácido nucleico de un segundo complejo indicador que comprende una etiqueta detectable a una segunda posición de unión de las al menos seis posiciones de unión;

(7) detectar la etiqueta detectable de la segunda molécula de ácido nucleico complementario unido o la etiqueta detectable de la segunda molécula de ácido nucleico complementario unido del segundo complejo indicador;

(8) identificar la posición e identidad de un segundo nucleótido en el ácido nucleico diana inmovilizado;

(9) repetir los pasos (5) a (8) hasta que cada posición de unión de las al menos seis posiciones de unión haya sido unida por una molécula complementaria de ácido nucleico que comprende una etiqueta detectable o una molécula complementaria de ácido nucleico de un complejo indicador que comprende una etiqueta detectable, y se ha detectado la etiqueta detectable de la molécula complementaria unida de ácido nucleico o la etiqueta detectable de la molécula complementaria unida de ácido nucleico de un complejo indicador, identificando así 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 diana de la sonda de secuenciación;

y (10) deshibridar 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 los pasos (5) y (6) ocurren secuencial o concurrentemente.

14. El método de la reivindicación 12, en el que la primera molécula de ácido nucleico complementaria y la primera molécula de ácido nucleico hibridante que carece de una etiqueta detectable comprenden la misma secuencia de ácido nucleico.

15. El método de la reivindicación 12, en el que la primera molécula de ácido nucleico hibridante que carece de una etiqueta detectable comprende una secuencia de ácido nucleico complementaria a un polinucleótido de cadena sencilla flanqueante adyacente a la primera posición de unión.

16. El método de la reivindicación 12, que comprende, además

(11) hibridar al menos una segunda población de segundas sondas de secuenciación que comprenden una pluralidad de las sondas de secuenciación de una cualquiera de las reivindicaciones 1 a 10 a 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 diana de la primera sonda de secuenciación y la segunda sonda de secuenciación son diferentes, y en el que 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 condiciones diferentes que las sondas de secuenciación en la primera población;

(12) unir una primera molécula complementaria de ácido nucleico que comprende una etiqueta detectable o una primera molécula complementaria de ácido nucleico de un primer complejo indicador que comprende una etiqueta detectable a una primera posición de unión de las al menos seis posiciones de unión;

(13) detectar la etiqueta detectable de la primera molécula de ácido nucleico complementaria unida o la etiqueta 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;

(15) unirse a la primera posición de unión una primera molécula de ácido nucleico hibridante que carece de una etiqueta detectable, desuniendo así la primera molécula complementaria de ácido nucleico que comprende una etiqueta detectable o la primera molécula complementaria de ácido nucleico del primer complejo indicador que comprende una etiqueta detectable;

(16) unir una segunda molécula complementaria de ácido nucleico que comprende una etiqueta detectable o una segunda molécula complementaria de ácido nucleico de un segundo complejo indicador que comprende una etiqueta detectable a una segunda posición de unión de las al menos seis posiciones de unión;

(17) detectar la etiqueta detectable de la segunda molécula de ácido nucleico complementario unido o la etiqueta detectable de la segunda molécula de ácido nucleico complementario unido del segundo complejo indicador;

(18) identificar la posición e identidad de un segundo nucleótido en el ácido nucleico diana inmovilizado;

(19) repetir los pasos (15) a (18) hasta que cada posición de unión de las al menos seis posiciones de unión haya sido unida por una molécula complementaria de ácido nucleico que comprende una etiqueta detectable o una molécula complementaria de ácido nucleico de un complejo indicador que comprende una etiqueta detectable, y se ha detectado la etiqueta detectable de la molécula de ácido nucleico complementario unido o la etiqueta detectable de la molécula de ácido nucleico complementario unido de un complejo indicador, identificando así el orden lineal de al menos seis nucleótidos para al menos una segunda región del inmovilizado ácido nucleico diana que se hibridó mediante el dominio de unión diana de la sonda de secuenciación; y

(20) deshibridar la al menos una segunda población de sondas de segunda secuencia del ácido nucleico diana inmovilizado.

17. El método de la reivindicación 16, que comprende además etapas de ensamblar cada orden lineal identificado de nucleótidos en al menos la primera región y al menos la segunda región del ácido nucleico diana inmovilizado, identificando así una secuencia para el ácido nucleico diana inmovilizado.

18. Un kit que comprende un sustrato, una pluralidad de sondas de secuencia de cualquiera de las reivindicaciones 1 a 10, al menos dos sondas de captura, en el que una sonda de captura comprende una secuencia de ácido nucleico que se hibrida con una región de un ácido nucleico diana que es diferente de la región del ácido nucleico diana que se hibrida con una sonda de secuenciación y una etiqueta de afinidad,

al menos seis moléculas de ácido nucleico complementarias que comprenden una etiqueta detectable, en el que un ácido nucleico complementario que comprende una etiqueta detectable comprende una secuencia de ácido nucleico que se hibrida con una de las al menos seis posiciones de unión de una sonda de secuenciación y una etiqueta detectable,

al menos seis moléculas de ácido nucleico complementarias que carecen de una etiqueta detectable, en el que un ácido nucleico complementario que carece de una etiqueta detectable comprende una secuencia de ácido nucleico que se hibrida con al menos una de las seis posiciones de fijación de una sonda de secuencia,

e instrucciones de uso.

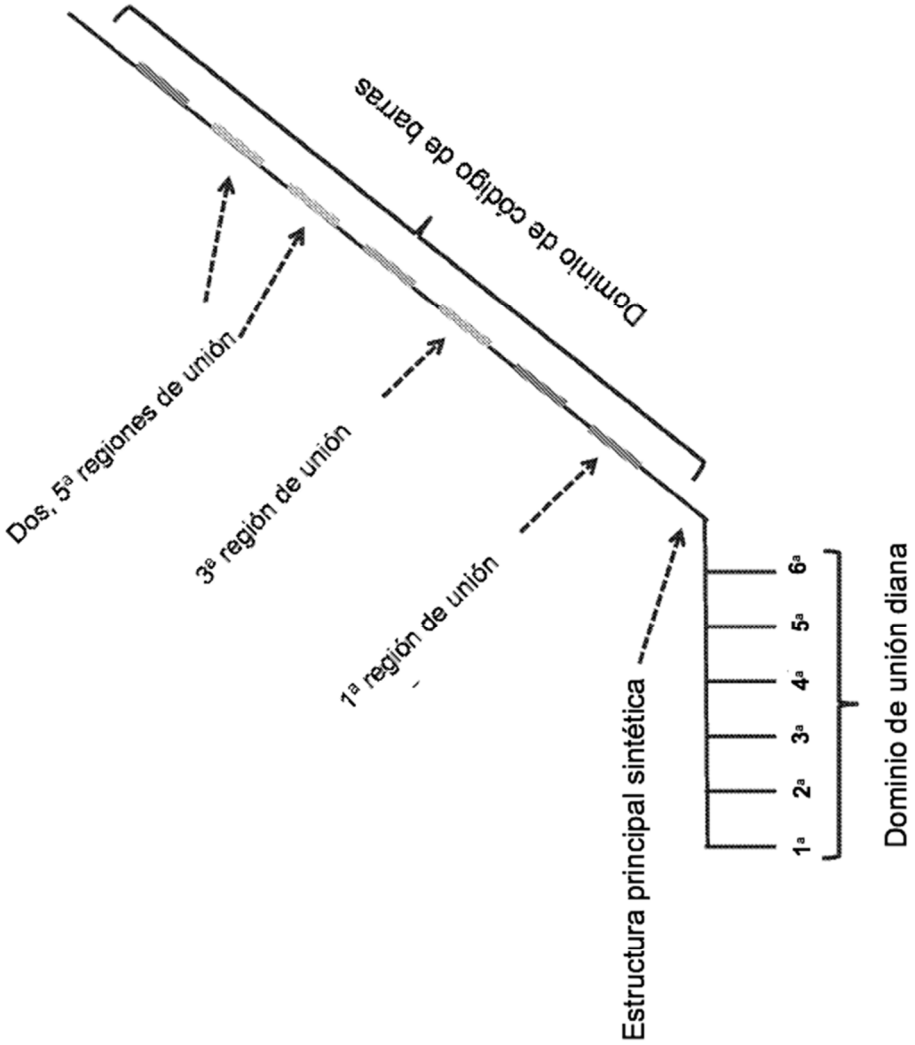

Dominio de código de barras
Dos, 5ª regiones de unión
3ª región de unión
1ª región de unión
Estructura principal sintética
1ª
2ª
3ª
4ª
5ª
6ª
Dominio de unión diana

Figura 1

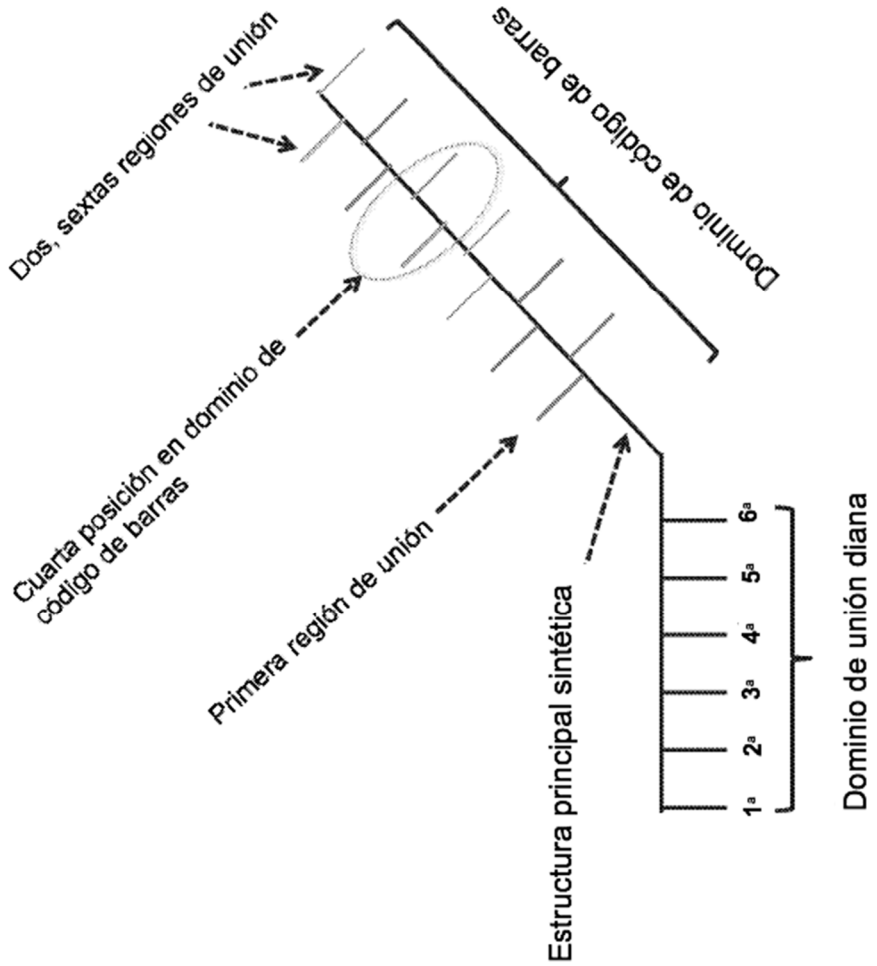

Dos, sextas regiones de unión
Dominio de código de barras
Cuarta posición en dominio de código de barras
Primera región de unión
Estructura principal sintética
1ª
2ª
3ª
4ª
5ª
6ª
Dominio de unión diana

Figura 2

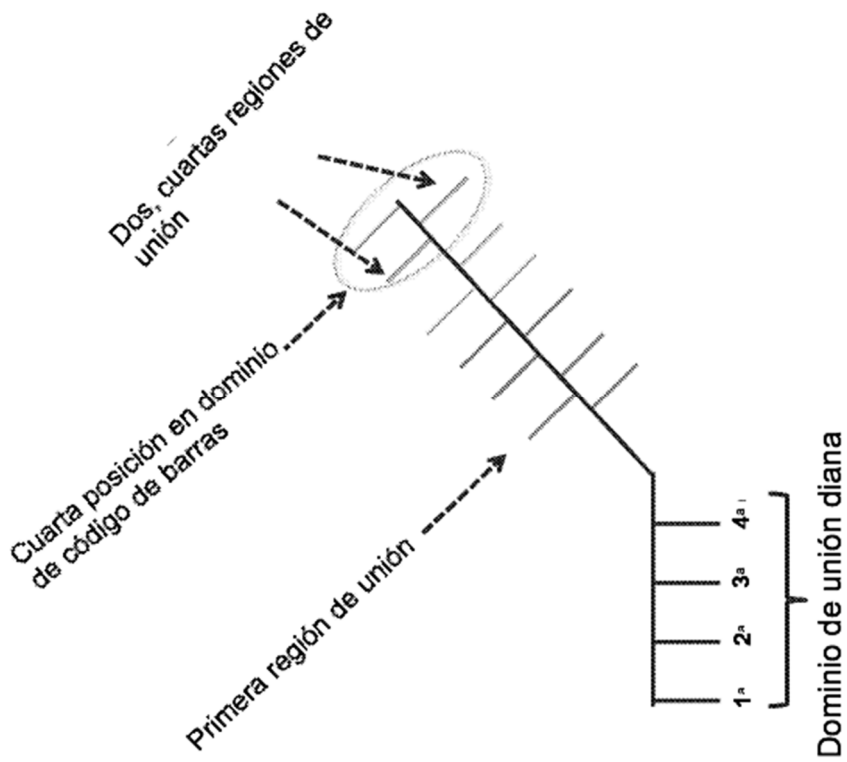

Dos, cuartas regiones de unión
Cuarta posición en dominio de código de barras
Primera región de unión
1ª
2ª
3ª
4ª
Dominio de unión diana

Figura 3

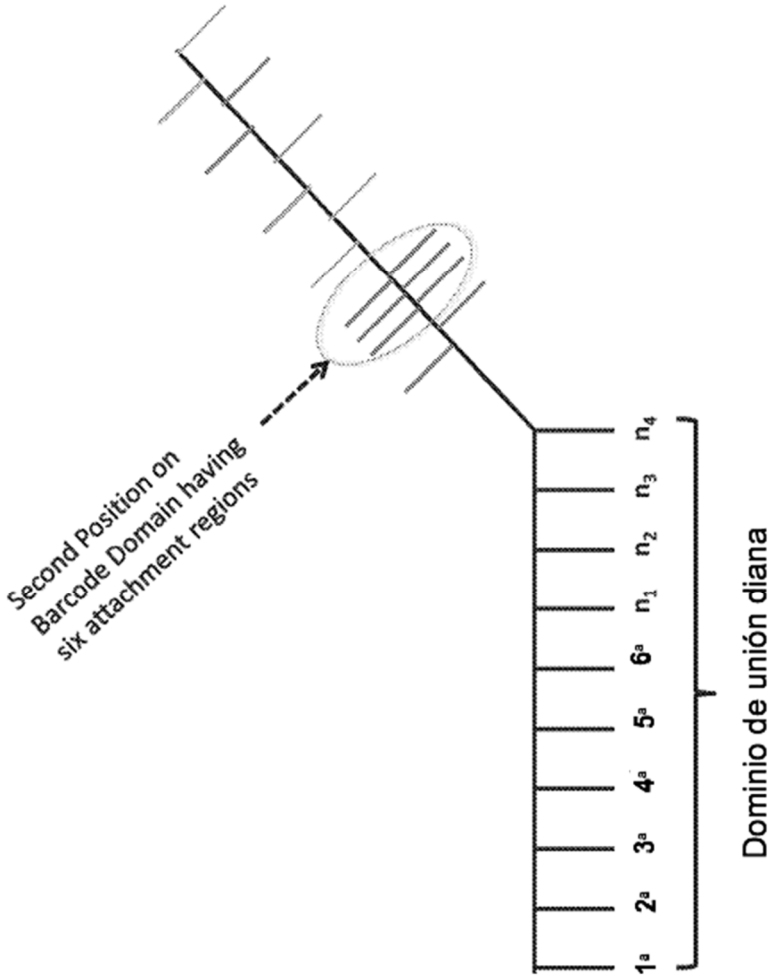

Second Position on
Barcode Domain having
six attachment regions
1ª
2ª
3ª
4ª
5ª
6ª
n1
n2
n3
n4
Dominio de unión diana

Figura 4

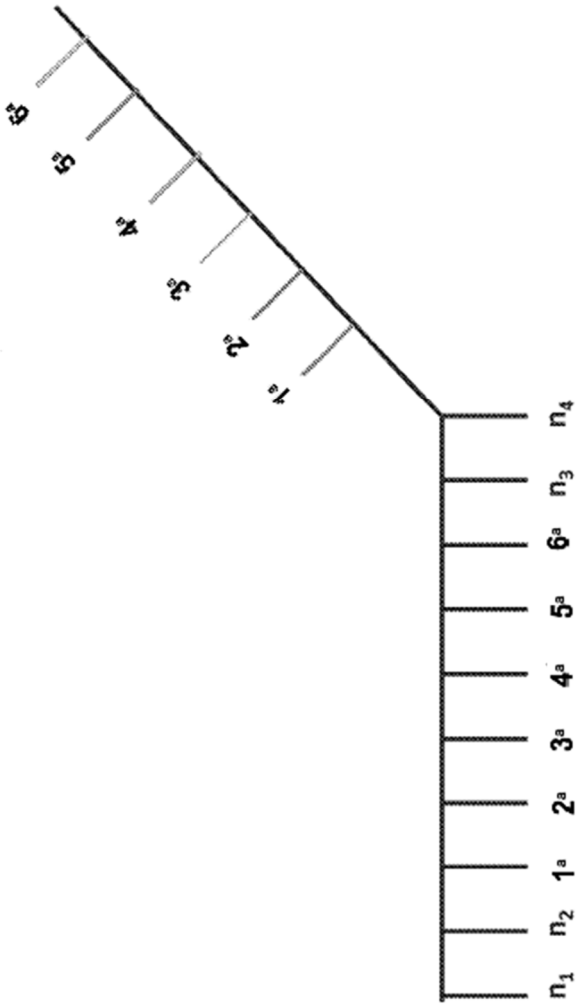


Figura 5

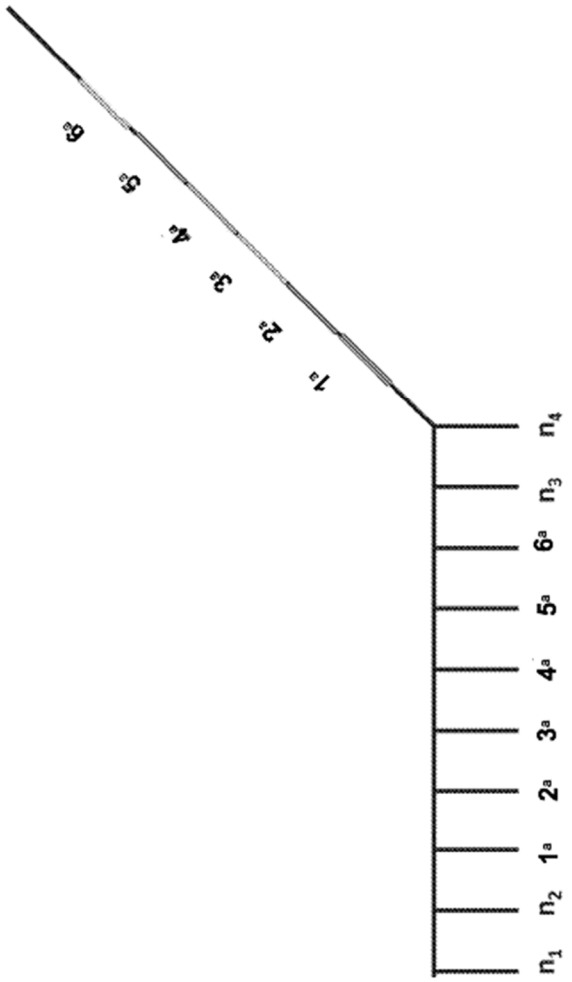


Figura 6A

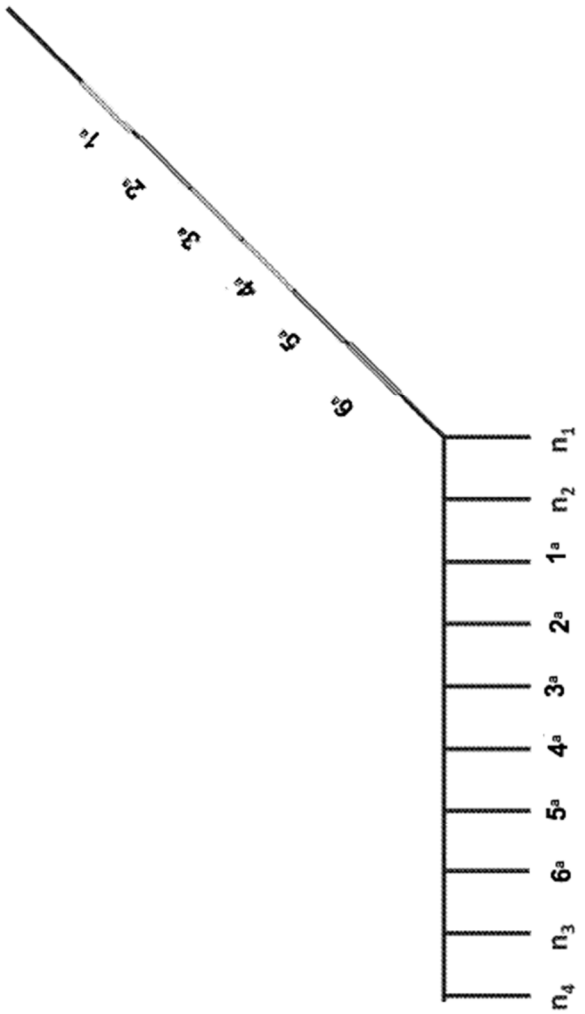


Figura 6B

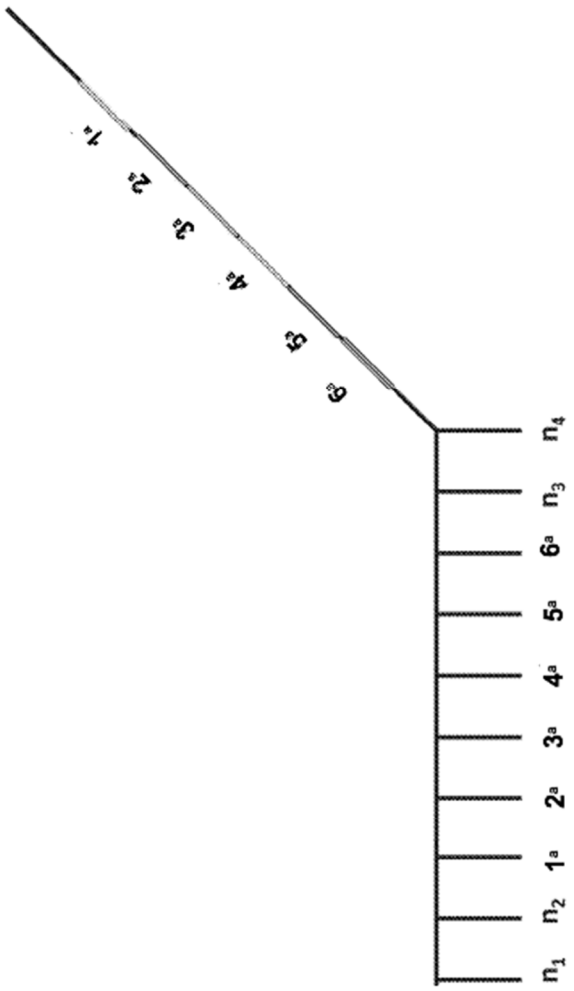


Figura 6C

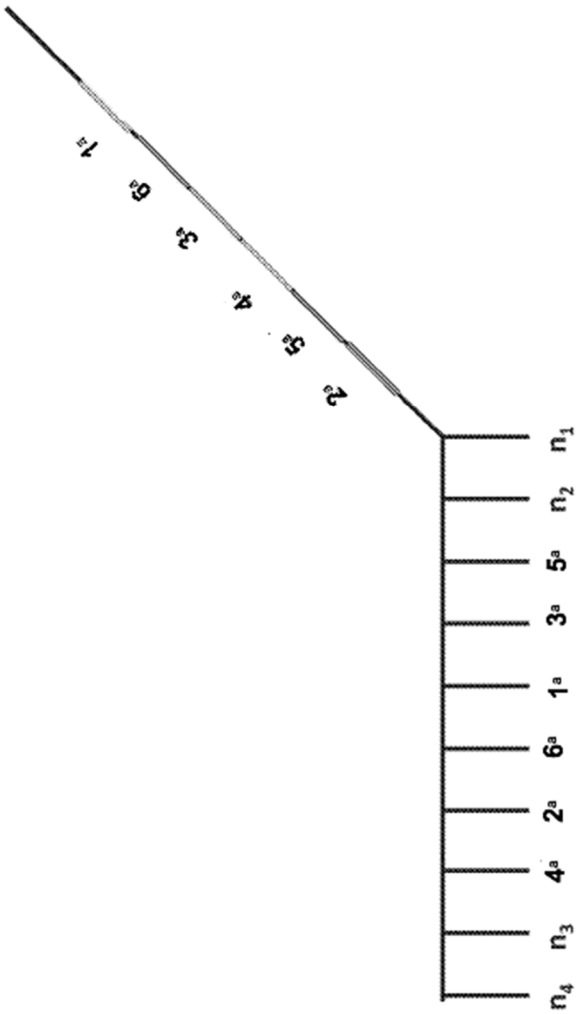
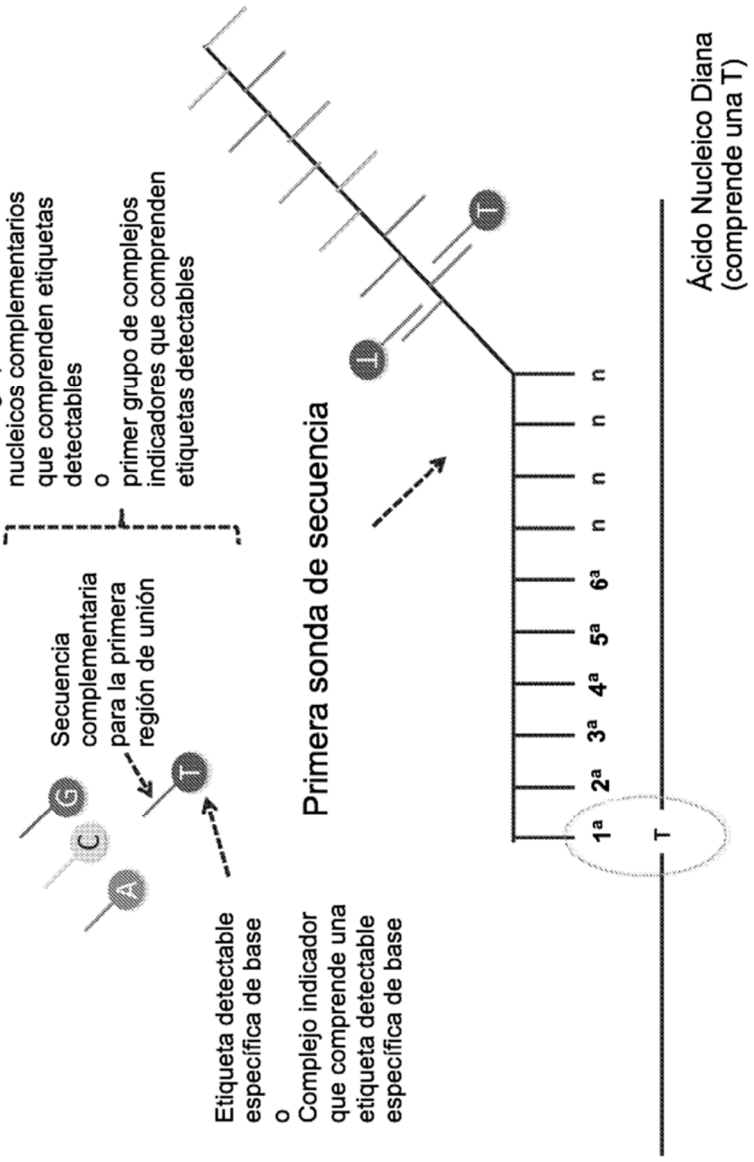


Figura 6D

Figura 7

10 bases específicas
U_b
8 bases específicas
U_b U_b
U_b U_b
8 bases específicas
10 bases específicas: 4^{10} ~ 1 millón de sondas
8 bases específicas: 4^8 ~ 65 mil sondas
8 bases específicas: 4^6 ~ 4 mil sondas
U_b = base universal (por ejemplo, Inosina y nitro-indol)

Figura 8A


Primer grupo de ácidos nucleicos complementarios que comprenden etiquetas detectables
o
primer grupo de complejos indicadores que comprenden etiquetas detectables
Secuencia complementaria para la primera región de unión
Etiqueta detectable específica de base
o
Complejo indicador que comprende una etiqueta detectable específica de base
Primera sonda de secuencia
1ª 2ª 3ª 4ª 5ª 6ª n n n n n
T
Ácido Nucleico Diana (comprende una T)

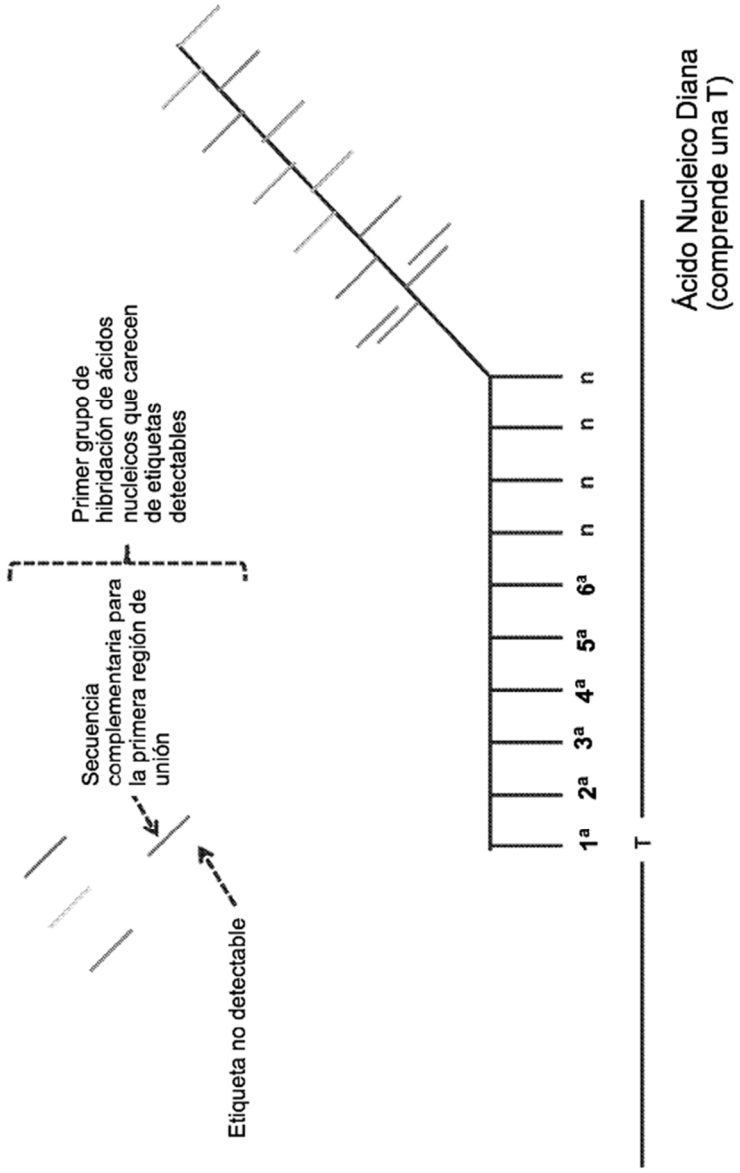

Primer grupo de hibridación de ácidos nucleicos que carecen de etiquetas detectables
Secuencia complementaria para la primera región de unión
Etiqueta no detectable
1ª 2ª 3ª 4ª 5ª 6ª n n n n n
T
Ácido Nucleico Diana (comprende una T)

Figura 8B

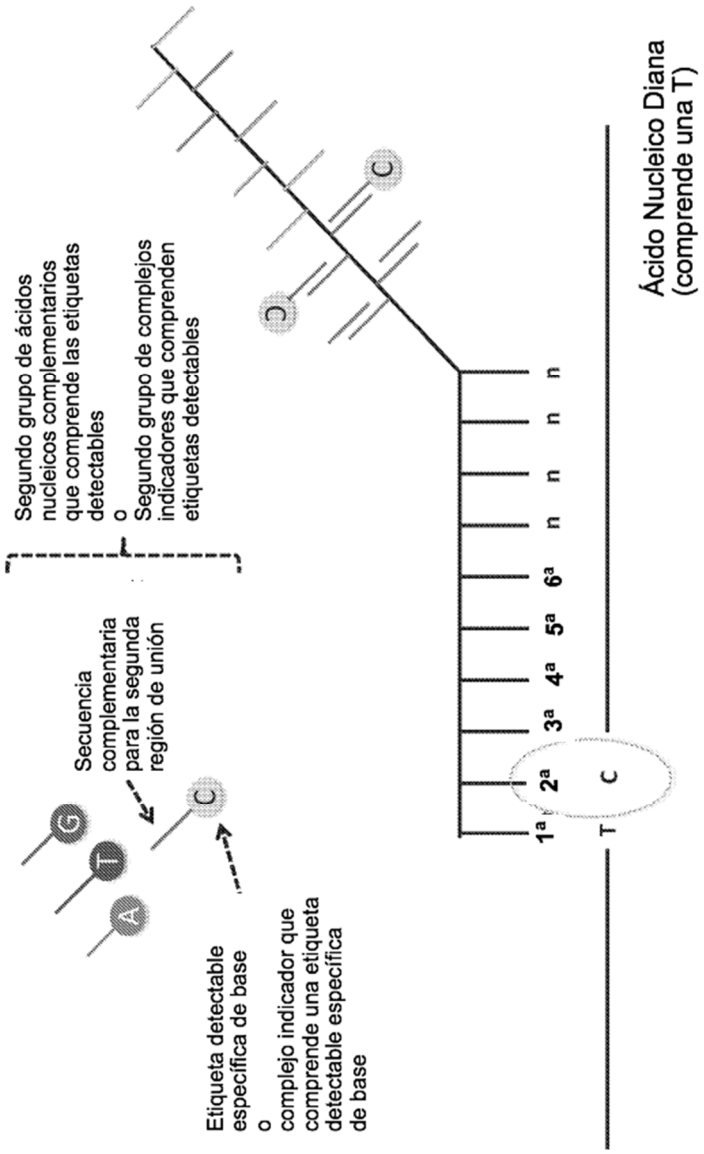
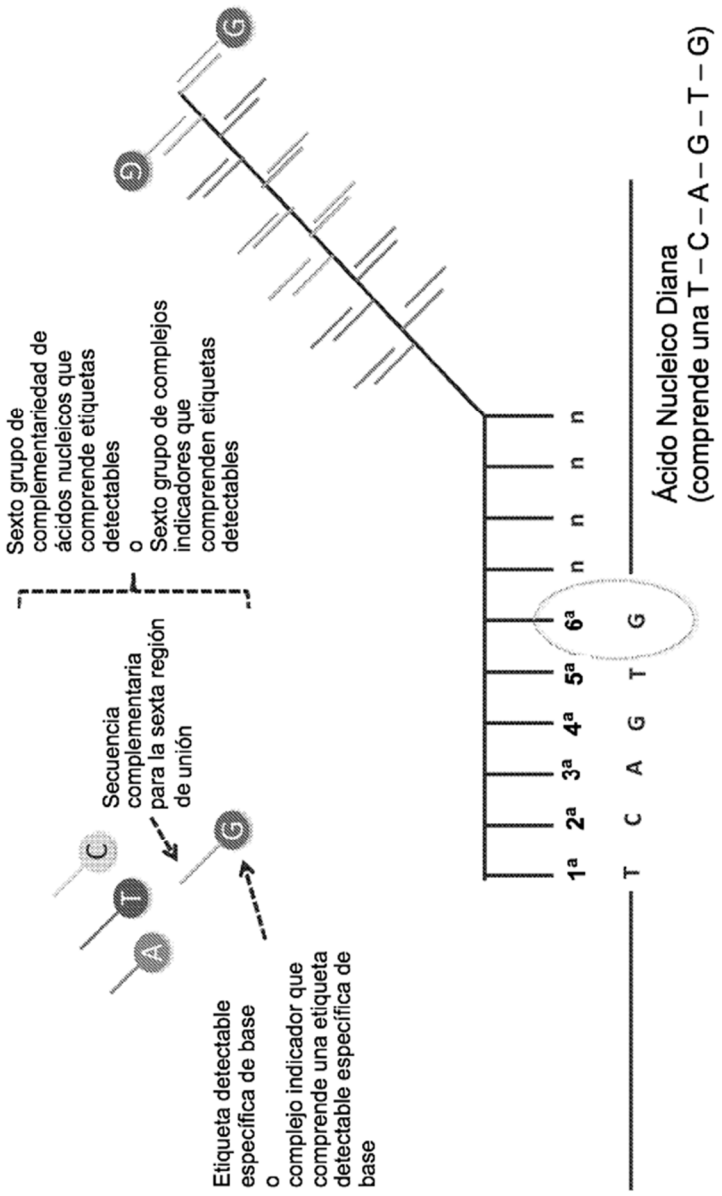

Segundo grupo de ácidos nucleicos complementarios que comprende las etiquetas detectables
o
Segundo grupo de complejos indicadores que comprenden etiquetas detectables
Secuencia complementaria para la segunda región de unión
Etiqueta detectable específica de base
o
complejo indicador que comprende una etiqueta detectable específica de base
1ª 2ª 3ª 4ª 5ª 6ª n n n n
T C
Ácido Nucleico Diana (comprende una T)

Figura 8C

Figura 8D


Sexto grupo de complementariedad de ácidos nucleicos que comprende etiquetas detectables
o
Sexto grupo de complejos indicadores que comprenden etiquetas detectables
Secuencia complementaria para la sexta región de unión
Etiqueta detectable específica de base
o
complejo indicador que comprende una etiqueta detectable específica de base
1ª 2ª 3ª 4ª 5ª 6ª
n n n n n n
T C A G T G
Ácido Nucleico Diana
(comprende una T – C – A – G – T – G)

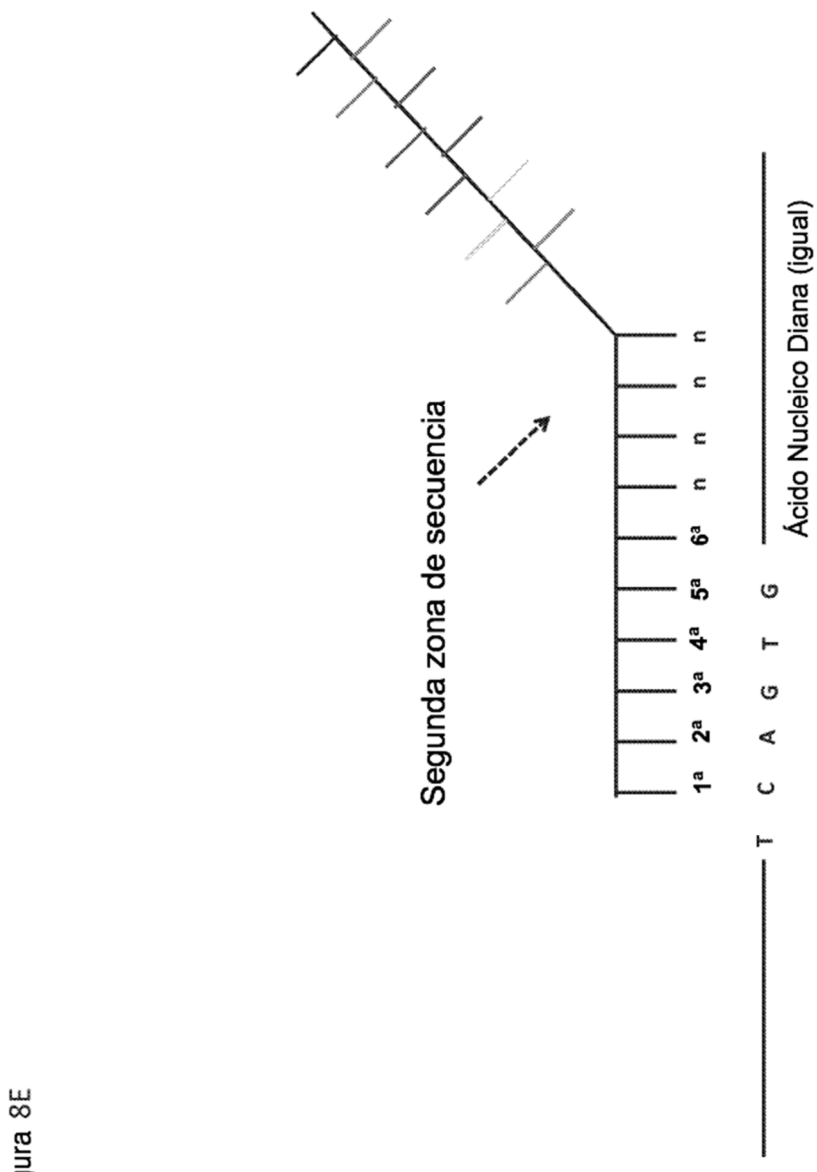

Segunda zona de secuencia
1ª 2ª 3ª 4ª 5ª 6ª n n n n
T C A G T G
Ácido Nucleico Diana (igual)

Figura 8E

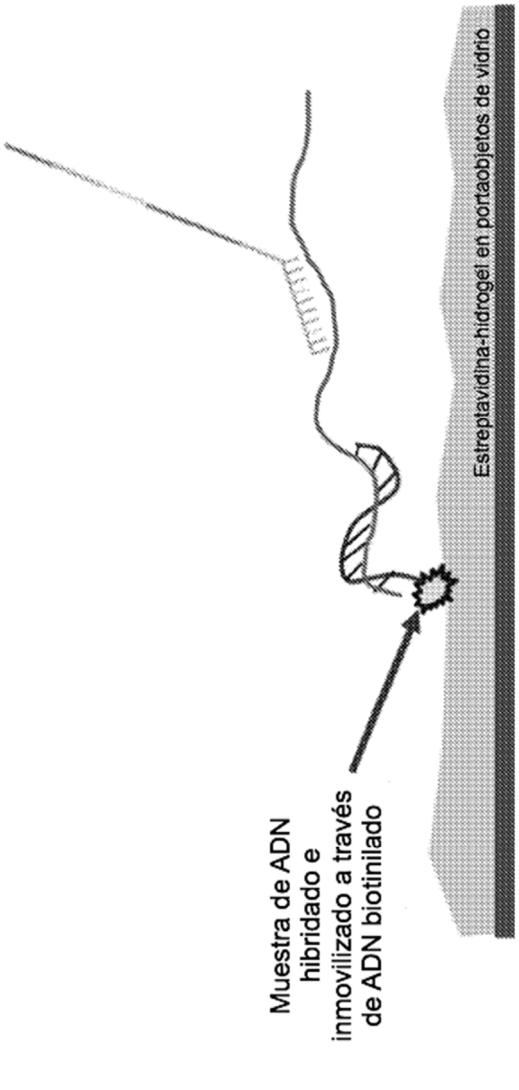

Estreptavidina-hidrogel en portaobjetos de vidrio
Muestra de ADN hibridado e inmovilizado a través de ADN biotinilado
Ciclo 1: → sonda Hyb

Figura 9A

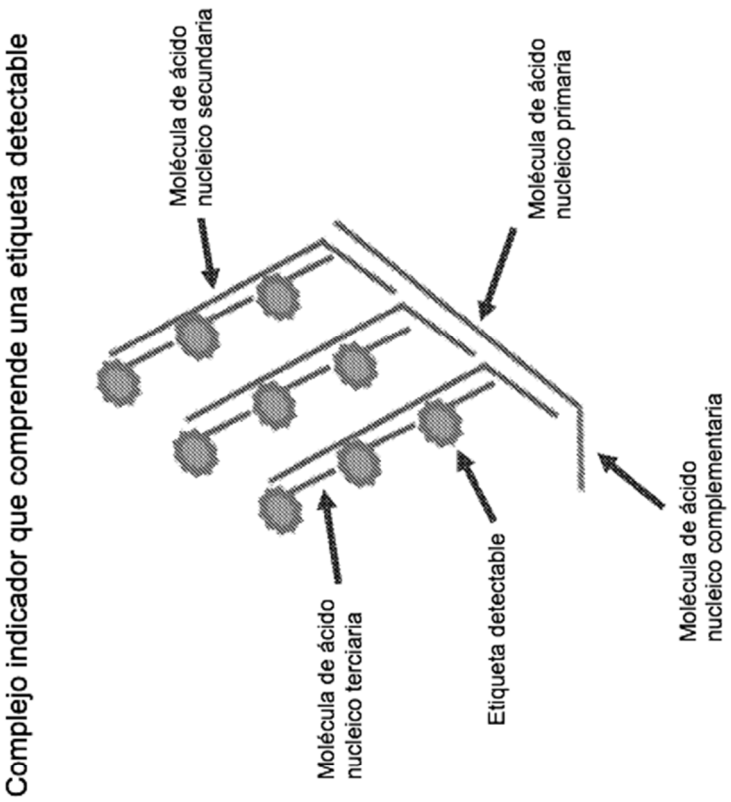
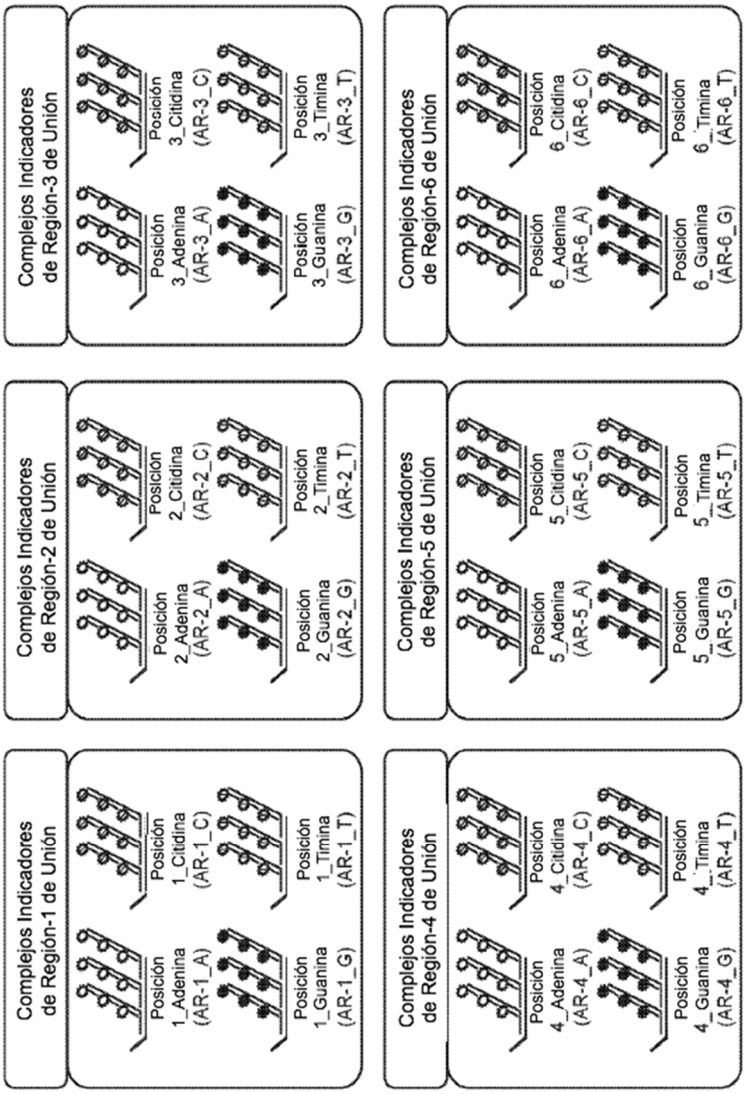

Complejo indicador que comprende una etiqueta detectable
Molécula de ácido nucleico secundaria
Molécula de ácido nucleico primaria
Molécula de ácido nucleico terciaria
Etiqueta detectable
Molécula de ácido nucleico complementaria

Figura 9B

Figura 9C


Complejos Indicadores de Región-1 de Unión
Posición 1_Adenina (AR-1_A)
Posición 1_Citidina (AR-1_C)
Posición 1_Guanina (AR-1_G)
Posición 1_Timina (AR-1_T)
Complejos Indicadores de Región-2 de Unión
Posición 2_Adenina (AR-2_A)
Posición 2_Citidina (AR-2_C)
Posición 2_Guanina (AR-2_G)
Posición 2_Timina (AR-2_T)
Complejos Indicadores de Región-3 de Unión
Posición 3_Adenina (AR-3_A)
Posición 3_Citidina (AR-3_C)
Posición 3_Guanina (AR-3_G)
Posición 3_Timina (AR-3_T)
Complejos Indicadores de Región-4 de Unión
Posición 4_Adenina (AR-4_A)
Posición 4_Citidina (AR-4_C)
Posición 4_Guanina (AR-4_G)
Posición 4_Timina (AR-4_T)
Complejos Indicadores de Región-5 de Unión
Posición 5_Adenina (AR-5_A)
Posición 5_Citidina (AR-5_C)
Posición 5_Guanina (AR-5_G)
Posición 5_Timina (AR-5_T)
Complejos Indicadores de Región-6 de Unión
Posición 6_Adenina (AR-6_A)
Posición 6_Citidina (AR-6_C)
Posición 6_Guanina (AR-6_G)
Posición 6_Timina (AR-6_T)

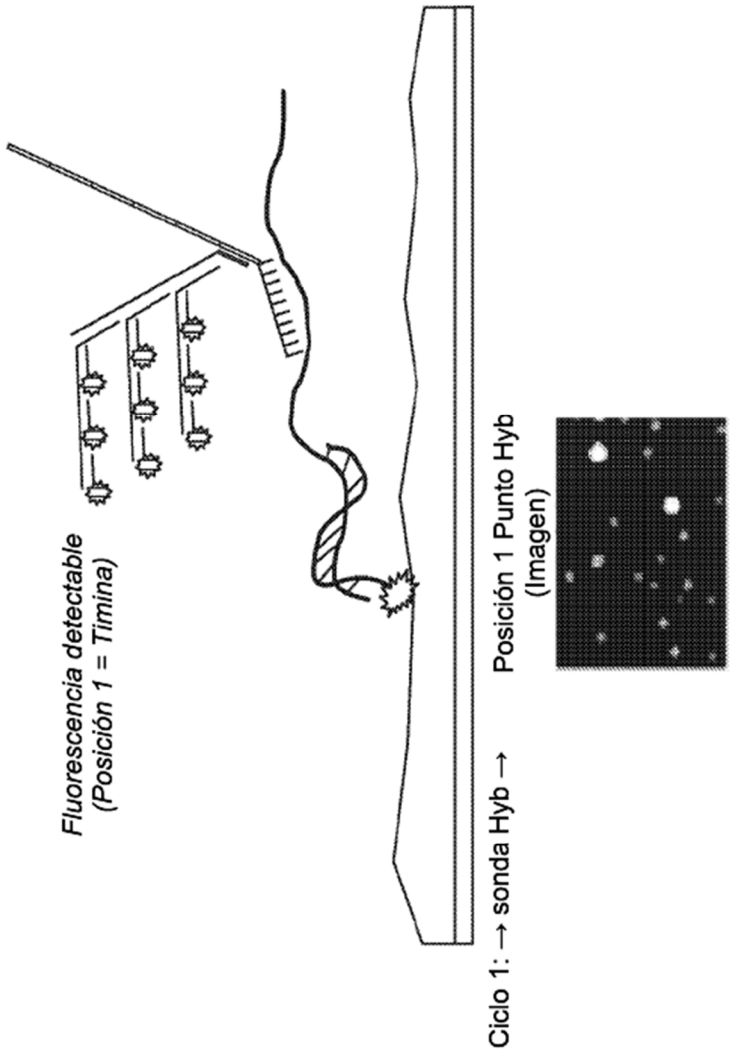

Fluorescencia detectable
(Posición 1 = Timina)
Ciclo 1: → sonda Hyb →
Posición 1 Punto Hyb
(Imagen)

Figura 9D

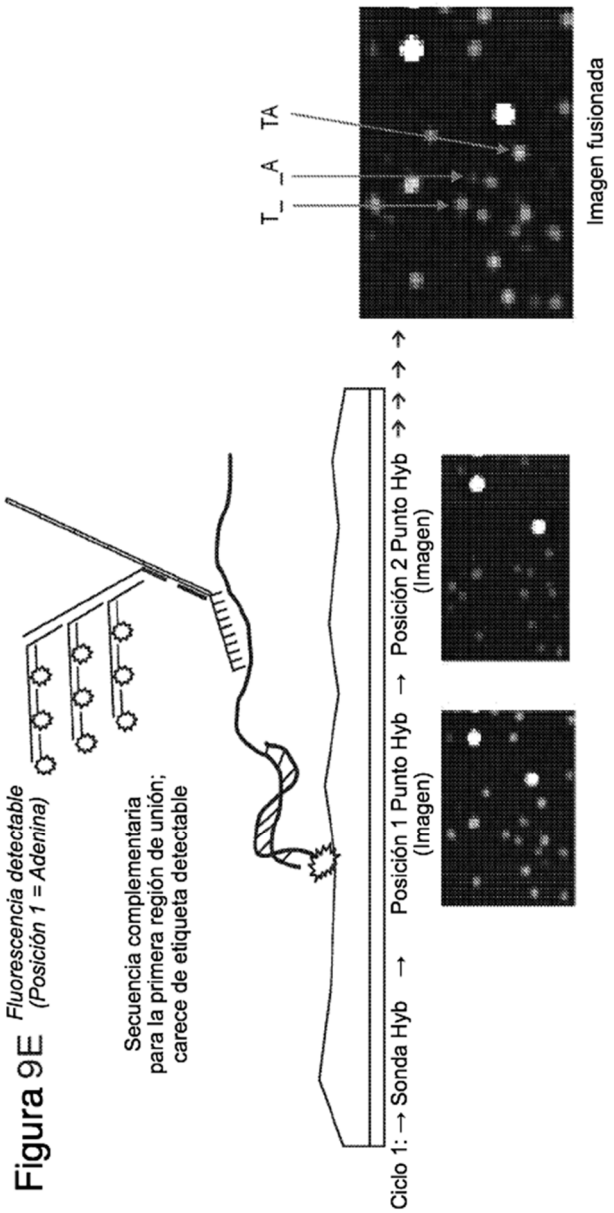

Fluorescencia detectable
(Posición 1 = Adenina)
Secuencia complementaria
para la primera región de unión;
carece de etiqueta detectable
Ciclo 1: → Sonda Hyb →
Posición 1 Punto Hyb
(Imagen)
→ Posición 2 Punto Hyb
(Imagen)
T_ _A TA
Imagen fusionada

Figura 9E

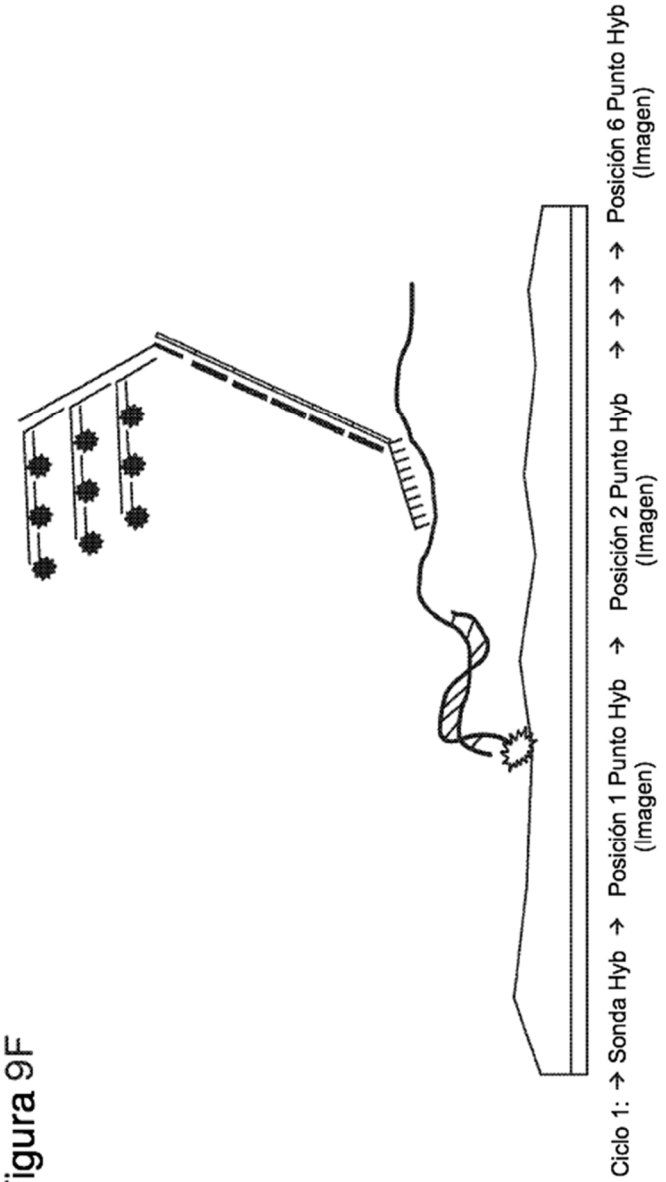

Ciclo 1: → Sonda Hyb → Posición 1 Punto Hyb (Imagen) → Posición 2 Punto Hyb (Imagen) → → → → Posición 6 Punto Hyb (Imagen)

Figura 9F

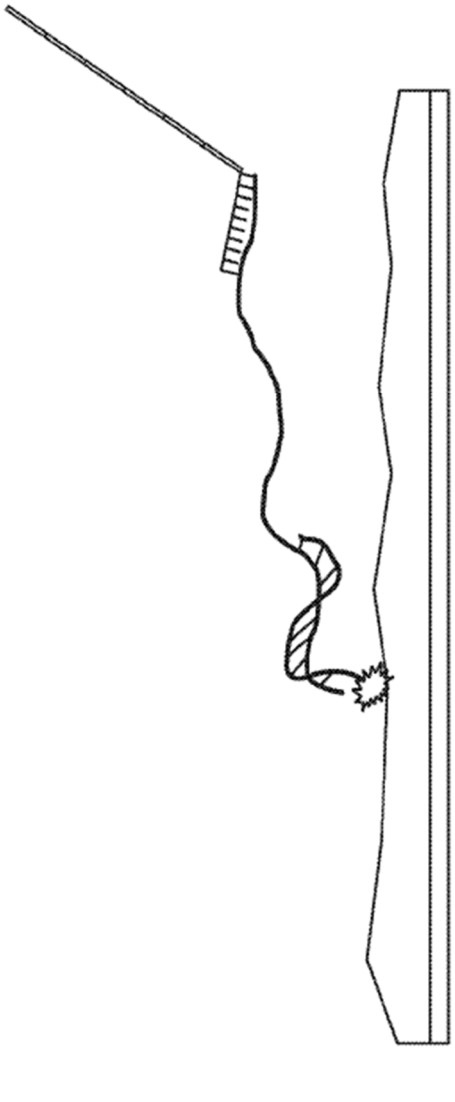


Figura 9G

Figura 10

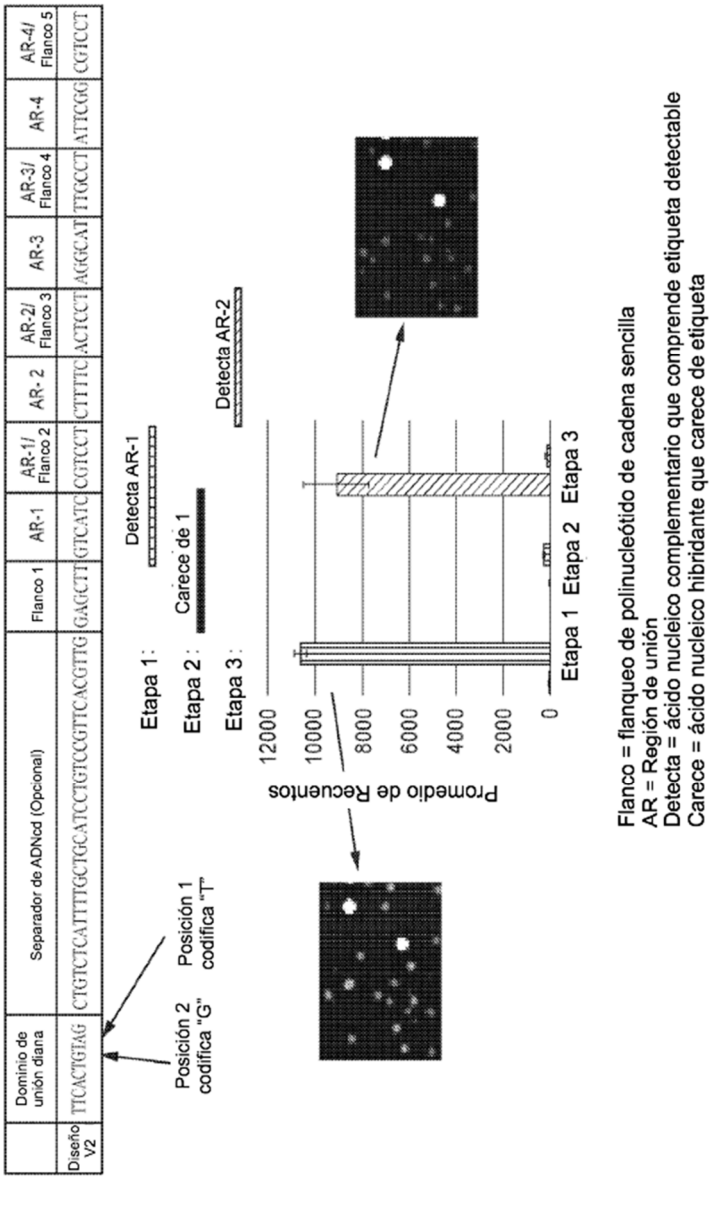

Diseño V2
Dominio de unión diana
Separador de ADNcd (Opcional)
Flanco 1
AR-1
AR-1/ Flanco 2
AR- 2
AR-2/ Flanco 3
AR-3
AR-3/ Flanco 4
AR-4
AR-4/ Flanco 5
TTCACTGTAG
CTGTCTCATTTTGCTGCATCCTGTCCGTTCACGTTG
GAGCTT
GTCATC
CGTCCT
CTTTTC
ACTCCT
AGGCAT
TTGCCT
ATTCGG
CGTCCT
Posición 2 codifica "G"
Posición 1 codifica "T"
Etapa 1:
Etapa 2:
Etapa 3:
Detecta AR-1
Carece de 1
Detecta AR-2
Promedio de Recuentos
12000
10000
8000
6000
4000
2000
0
Etapa 1
Etapa 2
Etapa 3
Flanco = flanqueo de polinucleótido de cadena sencilla
AR = Región de unión
Detecta = ácido nucleico complementario que comprende etiqueta detectable
Carece = ácido nucleico hibridante que carece de etiqueta

Figura 11

	Dominio	Separador de ADNcd (Opcional)	Flanco 1	AR-1	AR-1/ Flanco 2	AR- 2	AR-2/ Flanco 3	AR-3	AR-3/ Flanco 4	AR-4	AR-4/ Flanco 5	...
V2 Design	TTCACTGTAG	CTGTCTCATTTTGCTGCATCCTGTGTCCGTTCACGTTG	GAGCTT	GTCATC	CGTCCT	CTTTTC	ACTCCT	AGGCAT	TTGCCT	ATTCGG	CGTCCT	

Etapa 1: Detecta AR-1

Etapa 2: Carece de 1 + Detecta AR-2

Etapa 3: Carece de 2 + Detecta AR-3

Etapa 4: Carece de 3 + Detecta AR-4

.....

.....

.....

Etapa N:

Figura 12

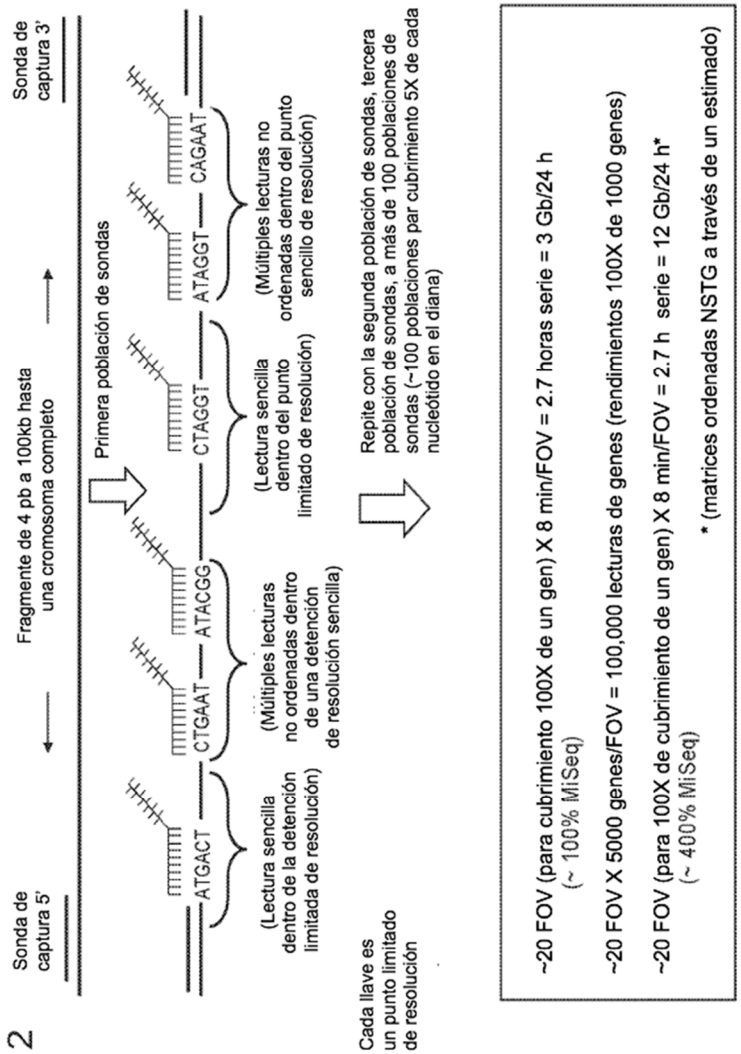
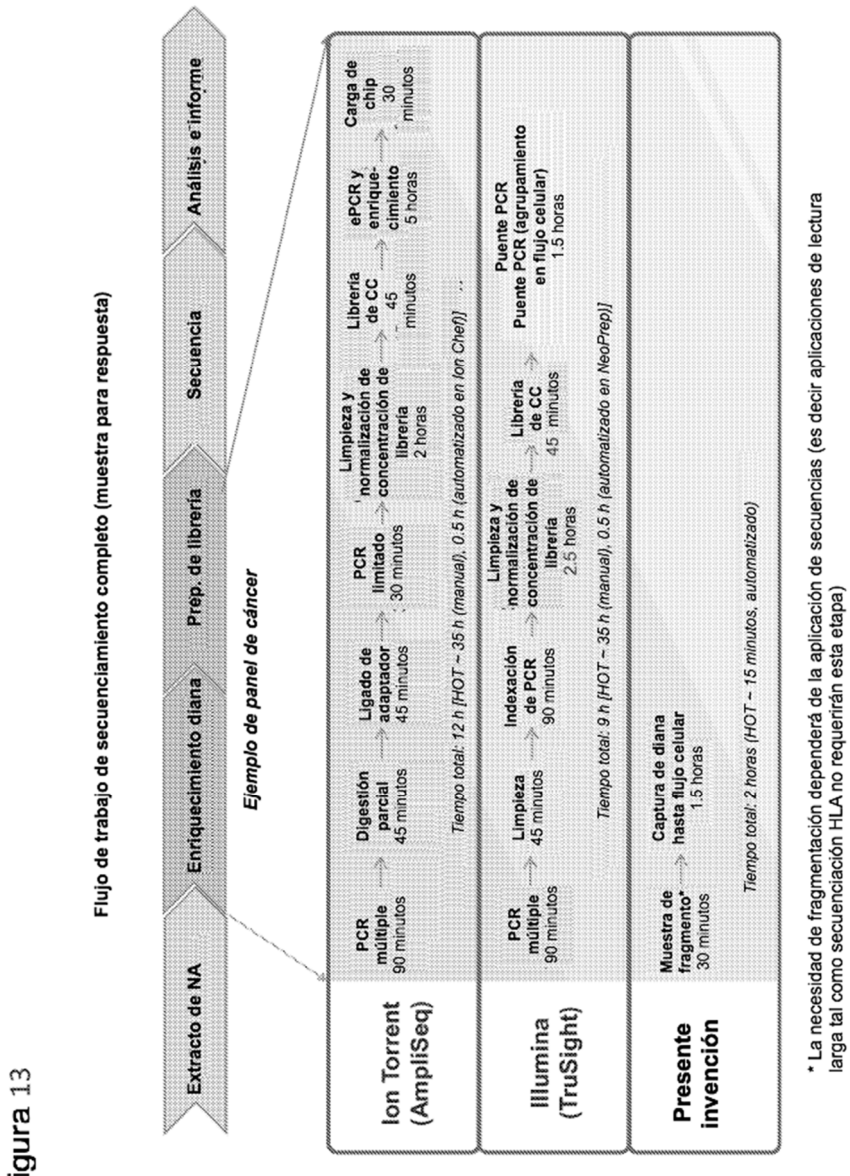
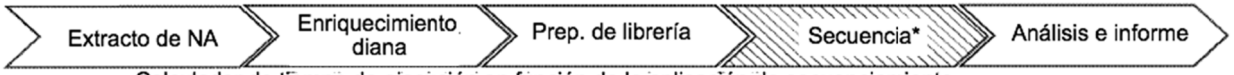

Sonda de captura 5'
Fragmente de 4 pb a 100kb hasta una cromosoma completo
Sonda de captura 3'
Primera población de sondas
ATGACT
CTGAAT
ATACGG
CTAGGT
ATAGGT
CAGAAT
(Lectura sencilla dentro de la detención limitada de resolución)
(Múltiples lecturas no ordenadas dentro de una detención de resolución sencilla)
(Lectura sencilla dentro del punto limitado de resolución)
(Múltiples lecturas no ordenadas dentro del punto sencillo de resolución)
Cada llave es un punto limitado de resolución
Repite con la segunda población de sondas, tercera población de sondas, a más de 100 poblaciones de sondas (~100 poblaciones par cubrimiento 5X de cada nucleótido en el diana)
~20 FOV (para cubrimiento 100X de un gen) X 8 min/FOV = 2.7 horas serie = 3 Gb/24 h
(~ 100% MiSeq)
~20 FOV X 5000 genes/FOV = 100,000 lecturas de genes (rendimientos 100X de 1000 genes)
~20 FOV (para 100X de cubrimiento de un gen) X 8 min/FOV = 2.7 h serie = 12 Gb/24 h*
(~ 400% MiSeq)
* (matrices ordenadas NSTG a través de un estimado)

Figura 13



* La necesidad de fragmentación dependerá de la aplicación de secuencias (es decir aplicaciones de lectura larga tal como secuenciación HLA no requerirán esta etapa)

Figura 14 Flujo de trabajo de secuenciamiento completo (muestra para respuesta)



Calculador de tiempo de ejecución en función de la aplicación de secuenciamiento

Atributos técnicos	Precisión bruta (%)	98
	Longitud de lectura (pb)	-
Aplicación de requerimientos de seq específicos	# de regiones diana	100
	Tamaño de región diana (pb)	100
	Umbral de detección de variante	0.01
	# de copias diana requeridas (X)	3000
	Total # de Lecturas requerido	0.3
Sondas que detectan 6 posiciones de nucleótidos	Total lecturas por FOV	250000
	# de FOV requerido	1.2
	Número total de ciclo hyb&seq	430
	Tiempo de movimiento de hardware (s)	0.5
	Tiempo de exposición (s)	0.2
	Tiempo de enfoque (s)	0.3
	Total de tiempo de imagen por FOV (s)	1
	Total # de puntos por ciclos de hyb&seq	6
	Tiempo de hyb de sonda k-mer (s)	5
	Tiempo de intercambio de puntos (s)	3
	Tiempo de lavado (s)	5
	Total tiempo de química por ciclo hyb&seq (s)	28
	Total tiempo de ejecución seq* (h)	4.62

Tiempos de ejecución:

100-plex = 4.6 h

1000-plex = 16 h

* Los tiempos de ejecución de secuenciamiento incluyen el tiempo para procesamiento de datos primario (es decir la identificación de nucleótidos, asignación QV, recuento y "descarte" de lectura de etiquetas de sonda)

Figura 15

Calculador de ciclos en función de ocupación de la sonda

Total # de dianas de moléculas individuales en un flujo celular	30000
Longitud promedio de diana 'con espacios' (pb)	100
Ocupación de código de barras (%) por ciclo de seq	0.75
# de ciclo de código de barras requerido para 'acertar' > 95% de los dianas en un flujo celular al menos una vez	3
Ocupación de Tinte (%) por posición	0.8
Número total de posiciones de Tinte	6
# total de lecturas codificadas de 1 pb por ciclo seq	6 35
# total de lecturas codificadas de 2 pb por ciclo seq	23 346
# total de lecturas codificadas de 3 pb por ciclo seq	92 1843
# total de lecturas codificadas de 4 pb por ciclo seq	369 5530
# total de lecturas codificadas de 5 pb por ciclo seq	1475 8847
# total de lecturas codificadas de 6 pb por ciclo seq	5898 5898
Número total de lecturas 'utilizables' por ciclo seq	22499
# de ciclo de código de barras requerido para cubrir todos los dianas de moléculas sencillas con cubrimiento ~3X	143
# de ciclo de código de barras requerido para cubrir todos los dianas de moléculas sencillas con cubrimiento ~5X	238
# de ciclo de código de barras requerido para cubrir todos los dianas de moléculas sencillas dados en flujo celular con cubrimiento ~3X	430
# de ciclo de código de barras requerido para cubrir todos los dianas de moléculas sencillas dados en flujo celular con cubrimiento ~5X	713

	Gigabases	Tiempo horas	# de lecturas	Utilidad clínica
		11 días	**3 mil millones** !	Secuencia de cáncer de secuencia de genoma
Hiseq 2500	900Gb			
Next Seq	120GB	27h	400M	Transcriptoma de exoma de secuencia de genoma
Proton PI	10Gb	2-4	70M	Transcriptoma de exoma
Presente invención	3 **a** 12 Gb	3	0.5M	Dirigido (Dx, cDx) más multianalito más muestras para respuesta, lecturas largas
MiSeq	3.5Gb	24h	25M	Transcriptoma dirigido (Dx)
PGM 318	**Hasta 2 Gb**	7.3/4.4	4M	Dirigido
PacBio RS II	2.2Gb/ **día**	3 h	50k	Nicho de genomas, lecturas largas, alto índice de error
PGM 314	0.1Gb	3.7/2.3	400k	Dirigido

Figura 16

Figura 17

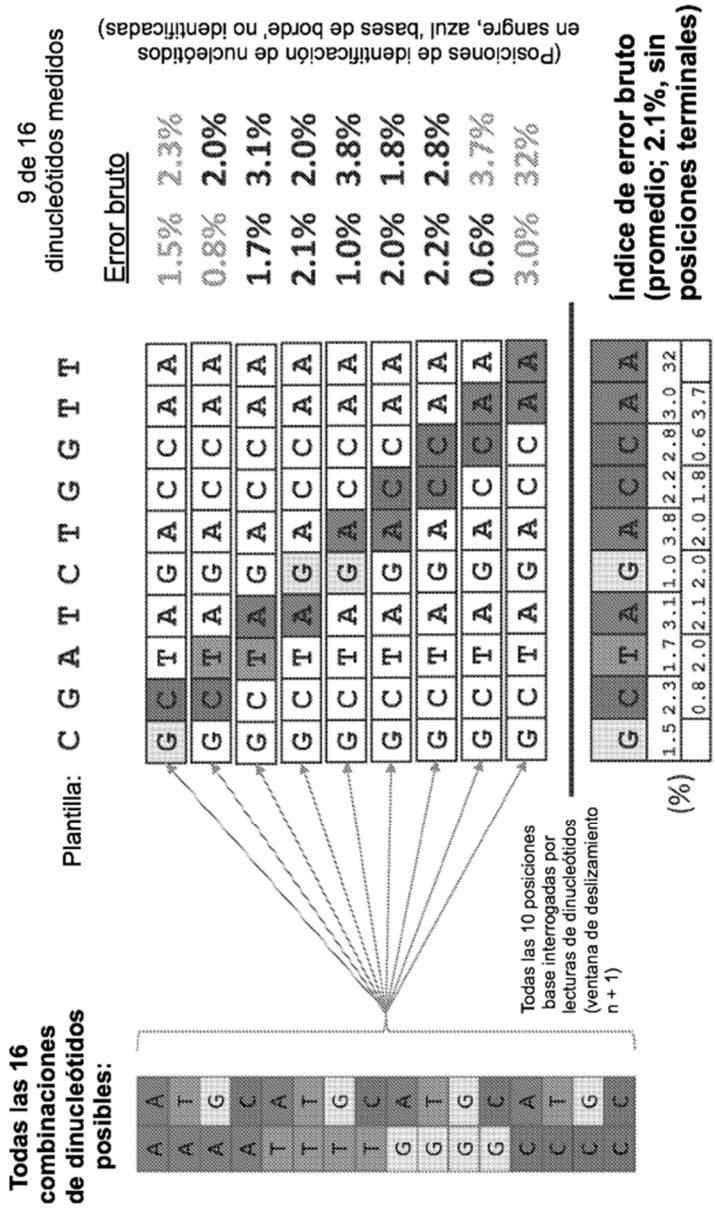

Todas las 16 combinaciones de dinucleótidos posibles:
Plantilla: C G A T C T G G T T
9 de 16 dinucleótidos medidos
Error bruto
1.5% 2.3%
0.8% 2.0%
1.7% 3.1%
2.1% 2.0%
1.0% 3.8%
2.0% 1.8%
2.2% 2.8%
0.6% 3.7%
3.0% 32%
(Posiciones de identificación de nucleótidos en sangre, azul 'bases de borde' no identificadas)
Todas las 10 posiciones base interrogadas por lecturas de dinucleótidos (ventana de deslizamiento n + 1)
(%)
1.5 2.3 1.7 3.1 1.0 3.8 2.2 2.8 3.0 32
0.8 2.0 2.1 2.0 2.0 1.8 0.6 3.7
Índice de error bruto (promedio; 2.1%, sin posiciones terminales)

Figura 18

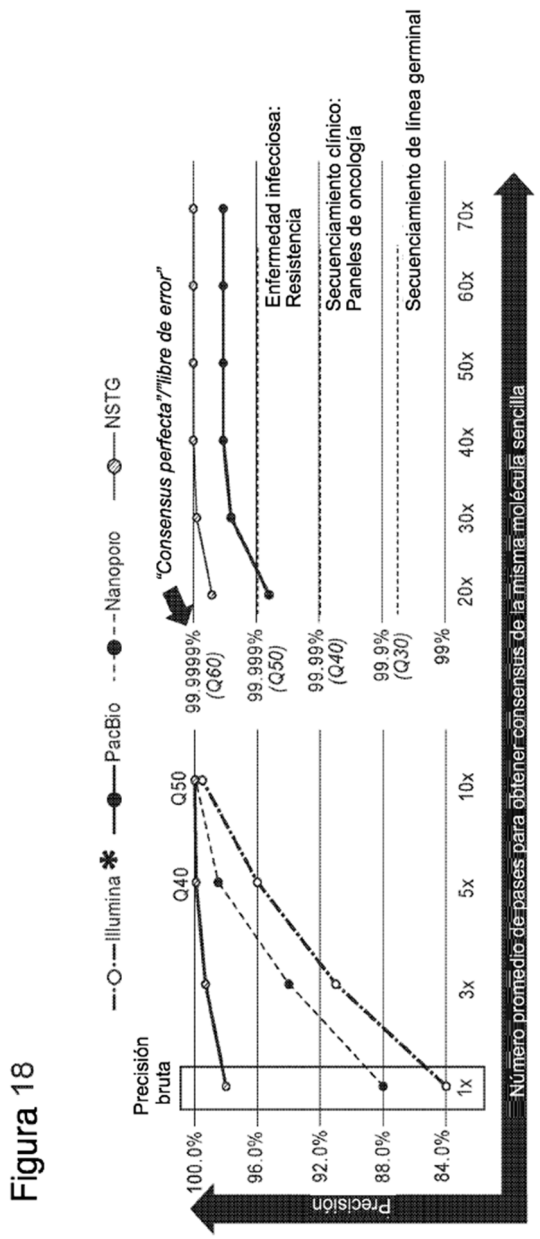

Illumina
PacBio
Nanoporo
NSTG
Precisión bruta
Q40
Q50
100.0%
96.0%
92.0%
88.0%
84.0%
1x
3x
5x
10x
20x
30x
40x
50x
60x
70x
99.9999% (Q60)
99.999% (Q50)
99.99% (Q40)
99.9% (Q30)
99%
"Consensus perfecta"/"libre de error"
Enfermedad infecciosa: Resistencia
Secuenciamiento clínico: Paneles de oncología
Secuenciamiento de línea germinal
Precisión
Número promedio de pases para obtener consensus de la misma molécula sencilla

Figura 19

	Posición: 1	2	3	4	5	6	7	8	9	10	Recuentos observados*	% de total de recuentos válidos
'Secuencia de plantilla (3' - 5'):	C	G	A	T	C	T	G	G	T	T		
'Emparejamiento perfecto' → Secuencia de sonda (5' - 3'):	G	C	T	A	G	A	C	C	A	A	2543	98%
'Emparejamiento incorrecto sencillo' Secuencia de sonda (5' - 3'):	G	C	T	G	G	A	C	C	A	A	40	1.5%
'Emparejamiento incorrecto sencillo' Secuencia de sonda (5' - 3'):	G	C	T	C	G	A	C	C	A	A	13	0.5%
'Emparejamiento incorrecto sencillo' Secuencia de sonda (5' - 3'):	G	C	T	T	G	A	C	C	A	A	1	0%
										Total:	2636	100%

* El valor observado de recuentos representa los recuentos después de la resta de recuentos de fondo
** Toda la hibridación hecha bajo condiciones de regulador: 1xSSPE/0.1% Tween20

Figura 20A

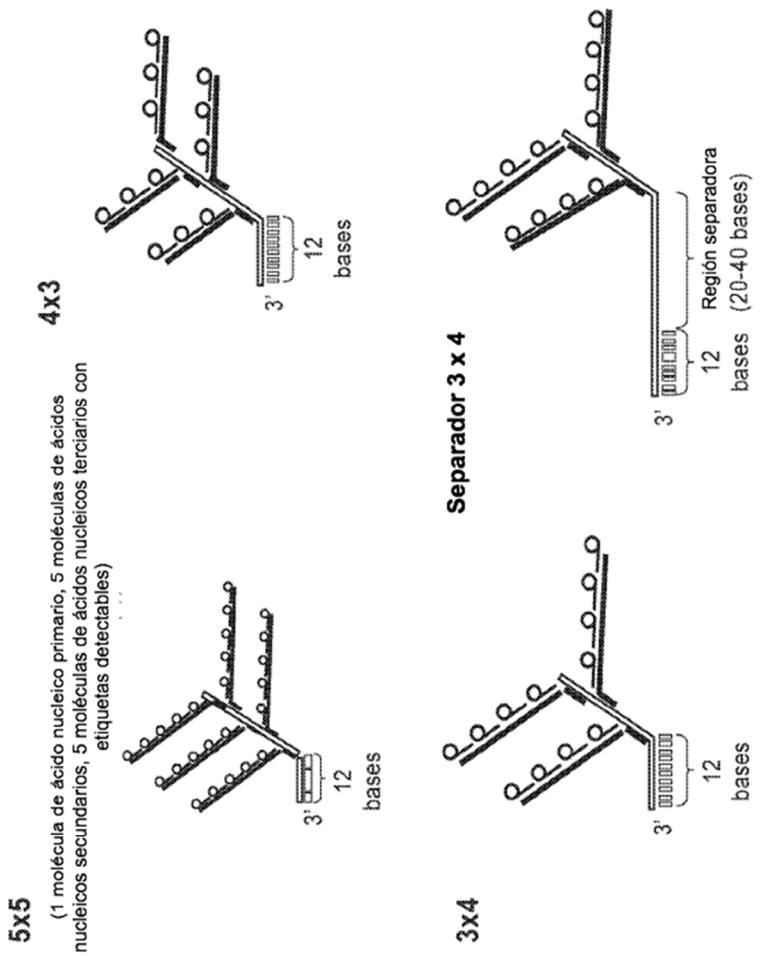

5x5
(1 molécula de ácido nucleico primario, 5 moléculas de ácidos nucleicos secundarios, 5 moléculas de ácidos nucleicos terciarios con etiquetas detectables)
4x3
3'
12 bases
3x4
3'
12 bases
Separador 3 x 4
3'
12 bases
Región separadora (20-40 bases)

Figura 20B

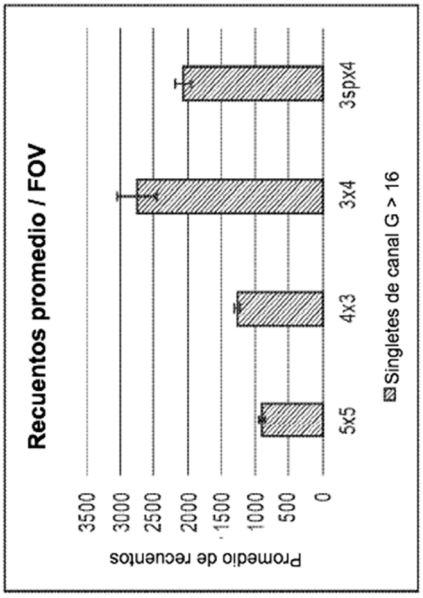


Figura 20C

Receta de ejemplo para construir complejos indicadores que comprenden etiquetas detectables

Volumen (uL)	Molécula n.a. primaria (10uM)	Molécula n.a. secundaria (10uM)	Molécula n.a. terciaria (10uM)	H20
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 mesa corresponden a colores en el diagrama de complejos mostrados a la izquierda

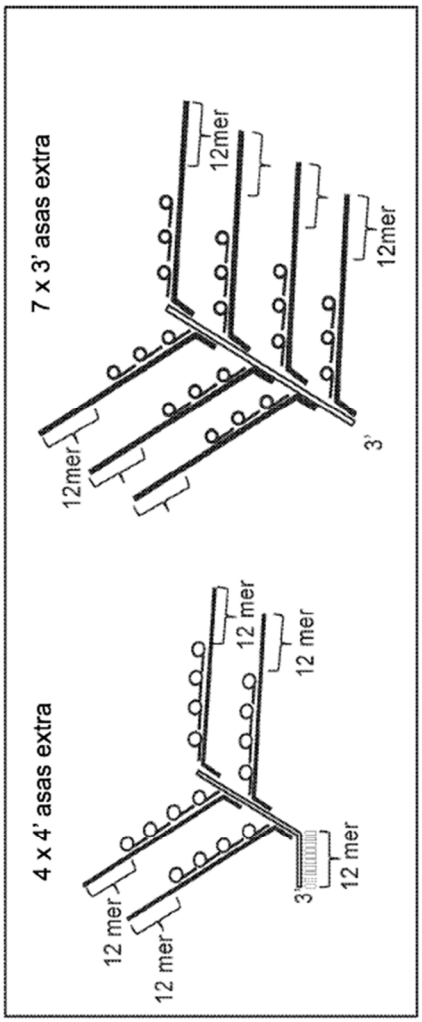


Figura 21A

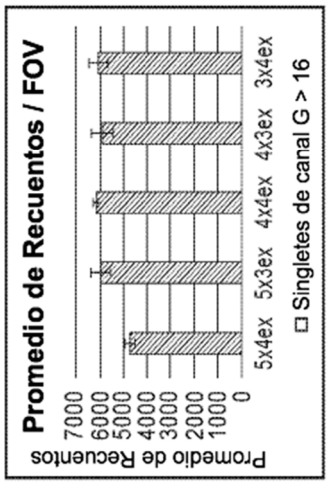


Figura 21B

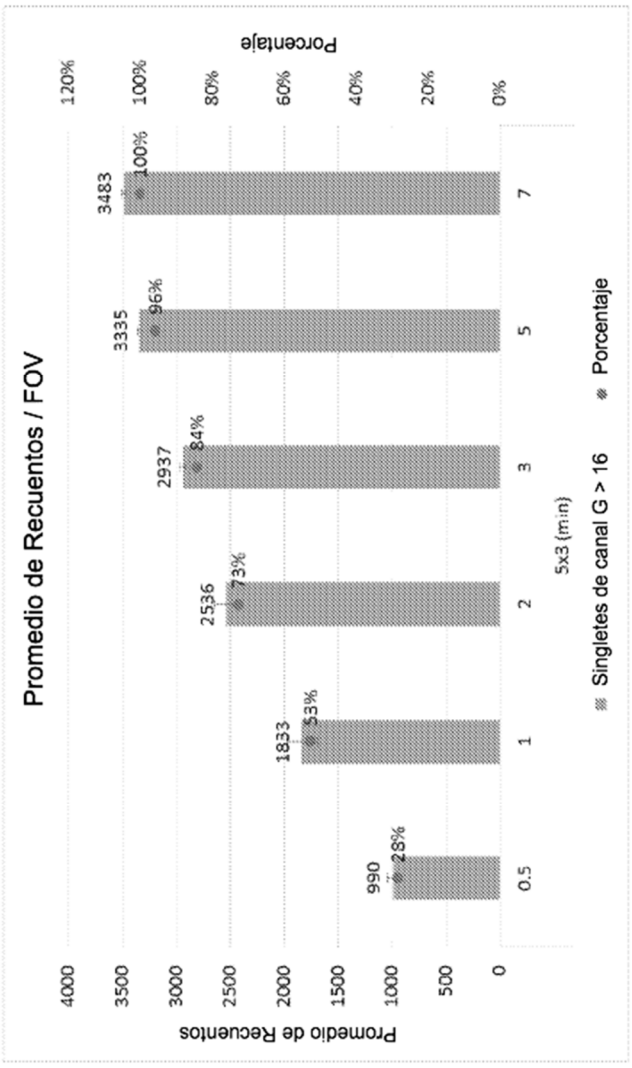

Cinéticas '5 x 3'
Promedio de Recuentos / FOV
Promedio de Recuentos
Porcentaje
4000
3500
3000
2500
2000
1500
1000
500
0
120%
100%
80%
60%
40%
20%
0%
990
28%
1833
53%
2536
73%
2937
84%
3335
96%
3483
100%
0.5
1
2
3
5
7
5x3 (min)
Singletes de canal G > 16
Porcentaje

Figura 22A

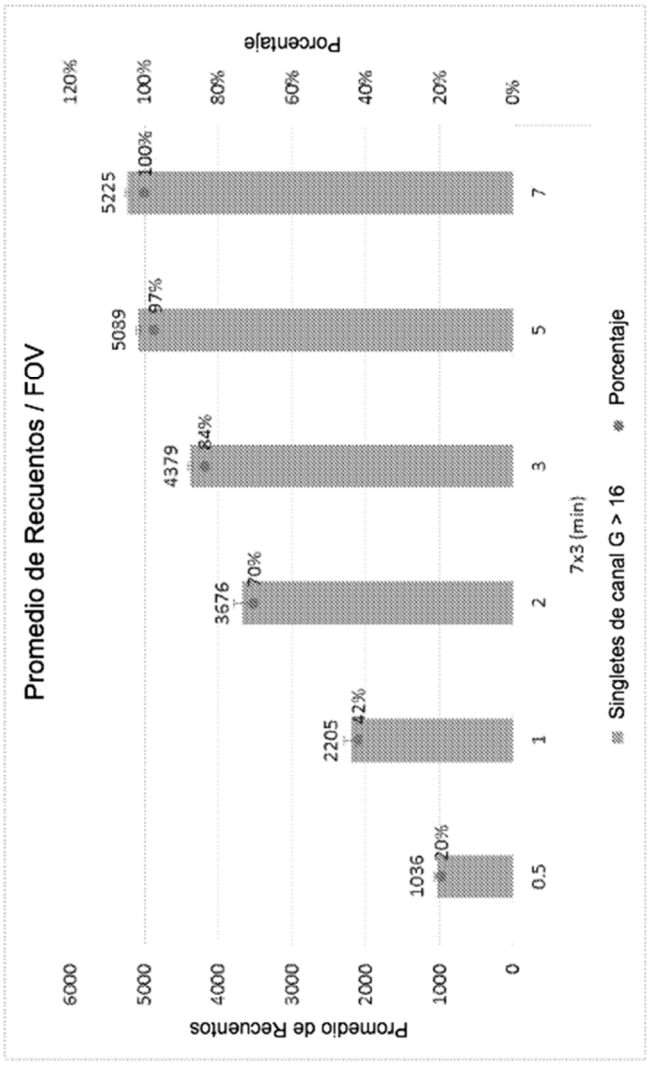

Cinéticas de '7 x 3 asas extra'
Promedio de Recuentos / FOV
Promedio de Recuentos
Porcentaje
6000
5000
4000
3000
2000
1000
0
120%
100%
80%
60%
40%
20%
0%
1036
20%
2205
42%
3676
70%
4379
84%
5039
97%
5225
100%
0.5
1
2
3
5
7
7x3 (min)
Singletes de canal G > 16
Porcentaje

Figura 22B

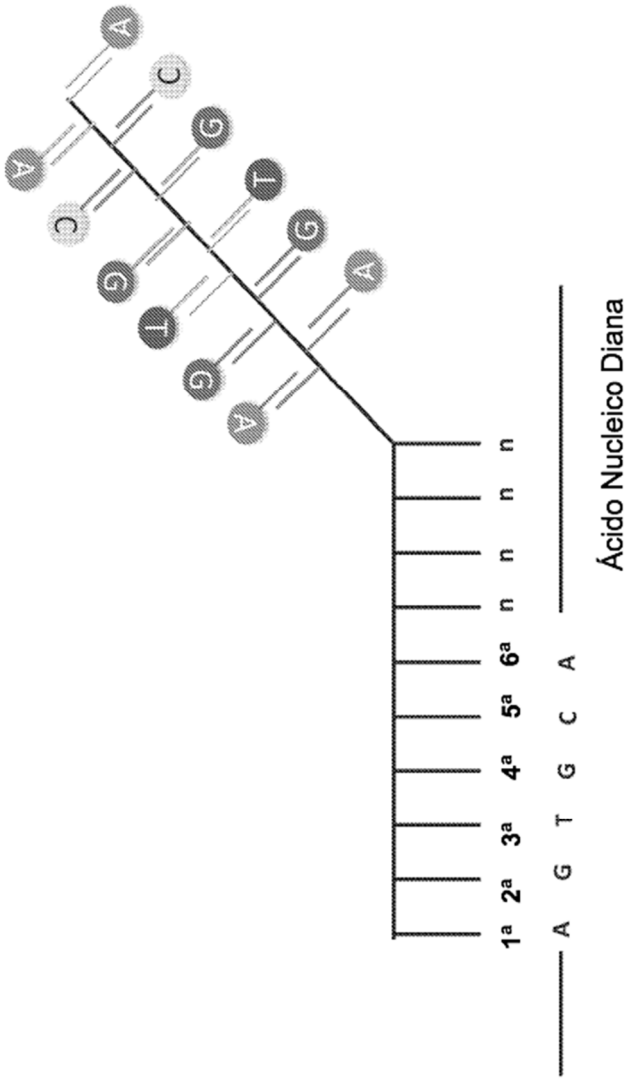

A
C
G
T
G
A
A
C
G
T
G
A
1a 2a 3a 4a 5a 6a n n n n
A G T G C A
Ácido Nucleico Diana

Figura 23

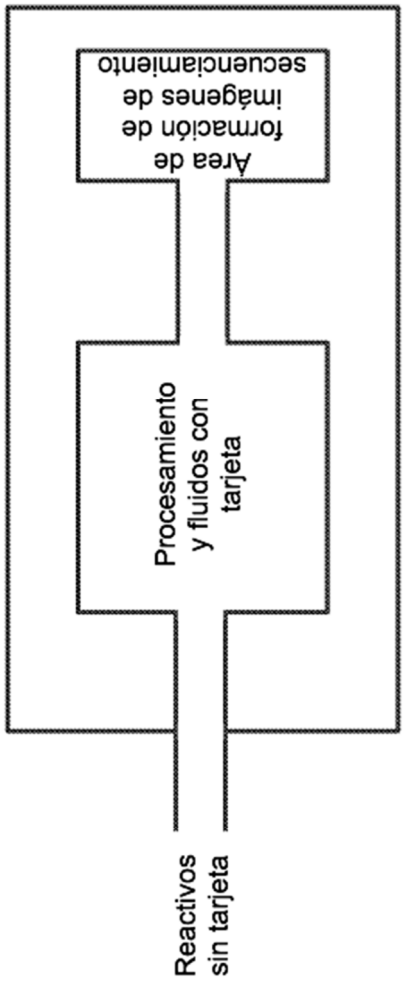

Área de formación de imágenes de secuenciamiento
Procesamiento y fluidos con tarjeta
Reactivos sin tarjeta

Figura 24

Código	Secuencias
YGBYGR	GAGACTGTAC - original
BGYRGB	GAGACAGTAC
YBRYGB	GAGACCGTAC
RGBYGR	GAGACGGTAC
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 25

Figura 26

ADN	Secuencia	Código	Recuento
12 mer	CTGAGACTGTAC	RYGBRG	80845
11 mer	TGAGACTGTAC	GBRYGB	24474
10 mer	GAGACTGTAC	YGBYGR	18339
9 mer	AGACTGTAC	YRBYRG	1667
Espejo de "12 mero"	CATGTCAGAGTC	GRBGYR	1473
Espejo de "11 mero"	CATGTCAGAGT	BGYRBG	1384
8 mer	GACTGTAC	GBRYBG	900
		Umbral de fondo	622

Figura 27

Indicador de longitud completa
Longitud completa ***20 pM*** (6 puntos) **33** recuentos/fov

Indicador de punto sencillo

Punto sencillo [fM]	Recuentos/FOV
20,000	Saturado
2,000	12000
200	308
20	***40***
2	11

Indicador NS de 20 pM ≈ punto sencillo de 20 fM