

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 810 300**

51 Int. Cl.:

C12N 15/10 (2006.01)

C12N 15/66 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **02.07.2015 PCT/US2015/039080**

87 Fecha y número de publicación internacional: **07.01.2016 WO16004368**

96 Fecha de presentación y número de la solicitud europea: **02.07.2015 E 15747608 (6)**

97 Fecha y número de publicación de la concesión europea: **13.05.2020 EP 3164489**

54 Título: **Etiquetado y evaluación de una secuencia diana**

30 Prioridad:

03.07.2014 US 201462020823 P
22.04.2015 US 201562151320 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
08.03.2021

73 Titular/es:

RHODX, INC. (100.0%)
BioGenerator Labs, Suite 234 4340 Duncan Ave
St Louis, MO 63110-1110, US

72 Inventor/es:

RODI, CHARLES

74 Agente/Representante:

LEHMANN NOVO, María Isabel

ES 2 810 300 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

DESCRIPCIÓN

Etiquetado y evaluación de una secuencia diana

Campo

5 Esta divulgación se refiere al campo de las técnicas de ingeniería genética y al uso de un identificador único para marcar una secuencia de ácido nucleico diana para el análisis y la medición.

Antecedentes

10 Dawson et al, 2013, NEJM "Analysis of Circulating Tumor DNA to Monitor Metastatic Breast Cancer" informaron que niveles de ADN tumoral circulante mostraron un mayor intervalo dinámico, y una mayor correlación con cambios en la carga tumoral que CA 15-3 o células tumorales circulantes. Entre las medidas testadas, el ADN tumoral circulante proporcionó la medida de respuesta al tratamiento más temprana en 10 de 19 mujeres (53%). Estas observaciones son consistentes con la opinión de que el ADN tumoral circulante es un biomarcador informativo, inherentemente específico y altamente sensible del cáncer de mama metastásico.

15 Sin embargo, el ADN libre de células es muy corto, lo cual afecta enormemente a la sensibilidad. Tsui et al, 2012, PLOS ONE, "High Resolution Size Analysis of Fetal DNA in the Urine of Pregnant Women by Paired-End Massively Parallel Sequencing" informaron que el tamaño objetivo medio del ADN en plasma es de 168 pares de bases (pb). La mediana de los tamaños objetivo de ADN en orina fue de 29 pb a 45 pares de bases.

20 Adicionalmente, la secuenciación de nueva generación (NGS) tiene una tasa de exactitud de 99,7% y, de esta forma, una tasa de error de 0,3%, lo que afecta en gran medida a la sensibilidad y especificidad, especialmente para muestras biológicas en las que una secuencia de ácido nucleico mutante de interés está presente con un número mucho mayor de secuencias no mutantes (de tipo salvaje). Por ejemplo, hay informes de que incluso utilizando un enriquecimiento basado en la hibridación, la mejor sensibilidad es 1 mutante por cada 2500 moléculas de tipo salvaje (0,04%) que es difícil de alcanzar en la práctica y es aún demasiado alta si se combina con una tasa de error de 0,3%.

25 Damiani et al, 198, Nucleic Acids Research "Sequence analysis of heteropolymeric DNA synthesized in vitro by the enzyme terminal deoxynucleotidyl transferase and cloned in Escherichia coli" describe un método para el análisis de la secuencia de ADN heteropolimérico sintetizado in vitro por la enzima desoxinucleotidil transferasa y clonado en E. coli.

30 Adicionalmente, la secuenciación de nueva generación (NGS) tiene una tasa de exactitud de 99,7% y, de esta forma, una tasa de error de 0,3%, lo que afecta en gran medida a la sensibilidad y especificidad, especialmente para muestras biológicas en las que una secuencia de ácido nucleico mutante de interés está presente con un número mucho mayor de secuencias no mutantes (de tipo salvaje). Por ejemplo, hay informes de que incluso utilizando un enriquecimiento basado en la hibridación, la mejor sensibilidad es 1 mutante por cada 2500 moléculas de tipo salvaje (0,04%) que es difícil de alcanzar en la práctica y es aún demasiado alta si se combina con una tasa de error de 0,3%.

35 Damiani et al, 198, Nucleic Acids Research "Sequence analysis of heteropolymeric DNA synthesized in vitro by the enzyme terminal deoxynucleotidyl transferase and cloned in Escherichia coli" describe un método para el análisis de la secuencia de ADN heteropolimérico sintetizado in vitro por la enzima desoxinucleotidil transferasa y clonado en E. coli.

40 El documento US 6.114.149 describe un método para amplificar una mezcla de fragmentos de ADN de secuencia diferente que pueden formarse a partir de la transcripción de ARN, o derivarse de fragmentos de ADN de cadena sencilla o doble genómico.

El documento US 2014/0147840 A1 describe un método para marcar un ácido nucleico diana.

El documento WO01/40516 describe la generación de ADN circular de cadena sencilla a partir de segmentos lineales de auto-reasociación.

45 El documento EP131783 describe enlazadores de ácido nucleico y su uso en la síntesis génica.

Sumario

50 La divulgación se refiere a la unión de una secuencia de ácido nucleico "etiqueta" a una secuencia diana de interés. La secuencia "etiqueta" está presente en un oligonucleótido de cadena sencilla que contiene una región de secuencia de unión a la diana que es complementaria a una secuencia diana. La hibridación de la secuencia de unión a diana de la secuencia diana lleva la secuencia detectable "etiqueta" a la proximidad de la secuencia diana para la formación de un enlace covalente entre la "etiqueta" y la secuencia diana.

mezcla de fragmentos de ADN de secuencia diferente que pueden formarse a partir de la transcripción de ARN, o derivarse de fragmentos de ADN de cadena sencilla o doble genómico. El documento US 2014/0147840 A1 describe un método para marcar un ácido nucleico diana. El documento WO01/40516 describe la generación de ADN circular de cadena sencilla a partir de segmentos lineales de auto-reasociación. El documento EP131783 describe enlazadores de ácido nucleico y su uso en la síntesis génica.

Sumario

La divulgación se refiere a la unión de una secuencia de ácido nucleico "etiqueta" a una secuencia diana de interés. La secuencia "etiqueta" está presente en un oligonucleótido de cadena sencilla que contiene una región de secuencia de unión a la diana que es complementaria a una secuencia diana. La hibridación de la secuencia de unión a diana de la secuencia diana lleva la secuencia detectable "etiqueta" a la proximidad de la secuencia diana para la formación de un enlace covalente entre la "etiqueta" y la secuencia diana.

La "etiqueta" está presente opcionalmente con otras secuencias que facilitan la amplificación, detección, identificación, medición, análisis y/o evaluación de la secuencia diana y/o uno o más sitio(s) de secuencia de ácido nucleico enlazados a la secuencia diana.

En un primer aspecto, la invención proporciona un método para modificar un ácido nucleico, que comprende:

(a) poner en contacto un ácido nucleico de cadena sencilla con una actividad de transferasa terminal y una mezcla de dos o más nucleótidos diferentes en condiciones en las que los nucleótidos en la mezcla se añaden secuencial y aleatoriamente al extremo 3' del ácido nucleico mediante la actividad de transferasa terminal, añadiendo con ello un heteropolinucleótido que comprende nucleótidos en la mezcla al extremo 3' del ácido nucleico y generando un primer ácido nucleico modificado;

(b) poner en contacto el primer ácido nucleico modificado con una actividad de transferasa terminal y una composición que comprende monómeros de un solo nucleótido en condiciones en las que los monómeros se añaden al extremo 3' del primer ácido nucleico modificado mediante la actividad de transferasa terminal, añadiendo con ello un homopolinucleótido que comprende los monómeros al extremo 3' del primer ácido nucleico modificado y generando un segundo ácido nucleico modificado que comprende, 5' a 3', el heteropolinucleótido y homopolinucleótido; y

(c) poner en contacto el segundo ácido nucleico modificado con un segundo ácido nucleico de cadena sencilla que comprende una secuencia que es complementaria al, o sustancialmente complementaria al homopolinucleótido del segundo ácido nucleico modificado y una actividad de ligasa, en condiciones en las que se ligan el extremo 3' del segundo ácido nucleico modificado y el extremo 5' del segundo ácido nucleico de cadena sencilla, generando con ello un ácido nucleico ligado, en donde el extremo 5' del segundo ácido nucleico de cadena sencilla está fosforilado o adenilado.

En otro aspecto, la invención proporciona un método para producir una molécula de ácido nucleico de cadena sencilla, comprendiendo el método:

poner en contacto una primera molécula de ácido nucleico de cadena sencilla (ANcs) que comprende un extremo 5' fosforilado, una secuencia diana y un extremo 3' con una actividad de transferasa terminal y dNTPs en condiciones de reacción en las que se añade una secuencia aleatoria de residuos de heteropolinucleótidos al extremo 3' de la primera molécula de ANcs, produciendo así un ácido nucleico indexado;

añadir una secuencia de homopolímero cola 3' mediante el uso de una actividad de transferasa terminal y ribonucleótidos;

poner en contacto la primera molécula de ANcs con una actividad ligasa de ácido nucleico de cadena sencilla para ligar el extremo 3' de la primera molécula de ANcs con su resto de 5' fosfato para producir una única molécula de ANcs circular cerrada, y amplificar una porción de la única molécula de ANcs circular cerrada de tal manera que se copien el índice molecular y porciones de interés de la molécula nativa.

En un aspecto adicional, la invención proporciona un método para modificar un ácido nucleico, que comprende:

(a) poner en contacto un primer ácido nucleico con una primera proteína que tiene una actividad de transferasa terminal y una mezcla de dos o más nucleótidos diferentes en condiciones en las que los nucleótidos en la mezcla se añaden secuencial y aleatoriamente a un extremo 3' del primer ácido nucleico mediante la primera proteína, añadiendo con ello un heteropolinucleótido que comprende nucleótidos en la mezcla al extremo 3' del primer ácido nucleico y generando un primer ácido nucleico modificado;

(b) poner en contacto el primer ácido nucleico modificado con una segunda proteína que tiene una actividad de transferasa terminal y una pluralidad de monómeros de un solo nucleótido en condiciones en las que la pluralidad de monómeros de un solo nucleótido se añade al extremo 3' del primer ácido nucleico modificado por la segunda proteína, añadiendo con ello un homopolinucleótido que comprende los monómeros al extremo 3' del primer ácido

nucleico modificado y generando un segundo ácido nucleico modificado que comprende, 5' a 3', el heteropolinucleótido y el homopolinucleótido;

5 (c) poner en contacto el segundo ácido nucleico modificado con un segundo ácido nucleico de cadena sencilla que comprende una secuencia que es complementaria a o sustancialmente complementaria al homopolinucleótido del segundo ácido nucleico modificado, en condiciones que permitan la hibridación o reasociación del homopolinucleótido al segundo ácido nucleico modificado, seguido de extender el segundo ácido nucleico modificado utilizando el segundo ácido nucleico de cadena sencilla como molde, generando con ello un tercer ácido nucleico modificado que tiene, 5' a 3', el primer ácido nucleico modificado y el segundo ácido nucleico modificado, en donde opcionalmente la extensión del segundo ácido nucleico modificado para generar un equivalente
10 complementario del segundo ácido nucleico de cadena sencilla se realiza en condiciones de amplificación, opcionalmente condiciones de reacción en cadena de la polimerasa (PCR).

La divulgación incluye un método para producir una molécula de ácido nucleico de cadena sencilla a partir de dos moléculas. La primera molécula contiene una secuencia diana, con secuencias de ácido nucleico enlazadas
opcionales, y la segunda molécula contiene una "etiqueta" de secuencia de ácido nucleico.

15 La primera molécula de ácido nucleico de cadena sencilla (ANCs) puede comprender un extremo 5' desfosforilado, una secuencia diana y una o más secuencias de ácidos nucleicos opcionales enlazadas a la secuencia diana, y un extremo 3' con secuencia(s) cola 3' opcionales, mientras que la segunda molécula de ANCs comprende un resto 5'-
20 fosfato que está opcionalmente adenilado, una secuencia conductora 5', una secuencia de unión a diana que es suficientemente complementaria a una parte de la secuencia diana para permitir la hibridación a la misma, y una cola 3' opcional y un extremo 3' opcionalmente bloqueado. En el método, las moléculas de ANCs primera y segunda se hibridan o reasocian en condiciones en las que la secuencia diana y la secuencia de unión a diana se hibridan o reasocian entre sí por complementariedad de pares de bases, seguido del contacto de las moléculas hibridadas o reasociadas con una actividad de ligasa de ácido nucleico de cadena sencilla que liga el extremo 3' de la primera molécula de ANCs con el resto 5'-fosfato (opcionalmente adenilado) de la segunda molécula de ANCs para producir
25 un producto, una molécula de ANCs ligada.

En algunas realizaciones, la secuencia conductora 5' de la segunda molécula de ANCs comprende una secuencia de índice molecular y una secuencia de unión a cebador. En algunos casos en los que está presente una pluralidad de primeras moléculas de ANCs que contienen secuencias diana, cada una de las segundas moléculas de ANCs comprende una secuencia de índice molecular diferente como una "etiqueta" única para cada una de las secuencias diana, y/o uno o más sitios de secuencia de ácido nucleico enlazados a la secuencia diana, en una primera molécula de ANCs. También se puede incorporar una secuencia de índice molecular en una primera molécula de ANCs tal como se aborda en esta memoria.

30 Mientras que el método enlaza dos moléculas de ANCs, cada una de las moléculas puede contener la estructura secundaria o una o más regiones que son de doble cadena en forma, a menos que se bloquee eficazmente para que no ocurra la hibridación entre la secuencia diana y la secuencia de unión a diana o el ligamiento de las dos moléculas.

La primera molécula de ANCs puede tener una longitud de hasta aproximadamente 500 nucleótidos. La molécula puede producirse opcionalmente de forma natural tal como una molécula obtenida de un sujeto, tal como una persona, un animal o una planta. La primera molécula de ANCs también puede no producirse de forma natural y, en algunas realizaciones, la primera molécula de ANCs puede incluir una parte que se produce de forma natural y una parte que no se produce de forma natural. Si la molécula está presente por primera vez en forma de doble cadena, puede desnaturalizarse mediante métodos rutinarios conocidos por la persona experta en la técnica, para formar moléculas de cadena sencilla. En algunas realizaciones de la divulgación, se utiliza la desnaturalización por calor. Si la molécula es demasiado grande, o el tamaño medio de las moléculas obtenidas de un sujeto es demasiado grande, las moléculas pueden ser digeridas o fragmentadas por métodos conocidos por la persona experta para producir longitudes más cortas para uso en la realización de los métodos descritos.

En algunas realizaciones, la primera molécula de ANCs tiene una longitud de hasta aproximadamente 400 nucleótidos, aproximadamente 300 nucleótidos, aproximadamente 200 nucleótidos, aproximadamente 150 nucleótidos, aproximadamente 100 nucleótidos, aproximadamente 75 nucleótidos, aproximadamente 50 nucleótidos, aproximadamente 45 nucleótidos, aproximadamente 40 nucleótidos, aproximadamente 35 nucleótidos, aproximadamente 30 nucleótidos, o aproximadamente 25 nucleótidos o menos. La segunda molécula de ANCs se produce sintéticamente y puede tener una longitud de aproximadamente 25 nucleótidos, aproximadamente 30 nucleótidos, aproximadamente 35 nucleótidos, aproximadamente 40 nucleótidos, aproximadamente 45 nucleótidos, aproximadamente 50 nucleótidos, aproximadamente 60 nucleótidos, aproximadamente 70 nucleótidos, aproximadamente 80 nucleótidos o aproximadamente 90 nucleótidos o más. En un segundo aspecto, la divulgación incluye cada una de las moléculas de ANCs ligadas resultantes que contienen la secuencia diana y el ácido nucleico "etiqueta". Una molécula de ANCs ligada comprendería, de 5' a 3', un extremo 5' que está opcionalmente desfosforilado, una secuencia diana y uno o más sitios opcionales de secuencia de ácido nucleico enlazados a la

secuencia diana, una cola 3' opcional, una secuencia conductora 5' de la segunda molécula de ANcs, una secuencia de unión a diana y una cola 3' opcional y un extremo 3' opcionalmente bloqueado. Como se proporciona en esta memoria y más adelante, esta molécula marca de forma única, o "etiqueta", una secuencia diana y/o uno o más sitios de secuencia de ácido nucleico enlazados a la secuencia diana, para su posterior amplificación, detección, identificación, medición, análisis y/o evaluación de la secuencia diana o una porción de la misma.

En un tercer aspecto, la divulgación incluye un método para "etiquetar" directamente una primera molécula que contiene una secuencia diana y/o uno o más sitios de secuencias de ácidos nucleicos enlazados a la secuencia diana. La primera molécula de ácido nucleico de cadena sencilla (ANcs) puede comprender un extremo 5' desfosforilado, una secuencia diana y uno o más sitios opcionales de secuencias de ácidos nucleicos enlazados a la secuencia diana, y un extremo 3' con una secuencia de cola 3' opcional. El método incluye primero extender el extremo 3' del primer ANcs con una actividad de transferasa terminal con una mezcla de nucleótidos en condiciones apropiadas para permitir una adición aleatoria de una secuencia al extremo 3' de la primera molécula de ANcs como una secuencia de identificador, o índice molecular único. Los nucleótidos son a menudo desoxirribonucleótido trifosfatos (dNTPs), a los que también se hace referencia en esta memoria como nucleótido trifosfatos (NTPs), y ejemplos no limitantes incluyen adenosina trifosfato (dA), timidina trifosfato (dT), citidina trifosfato (dC) y guanosina trifosfato (dG).

El método niega entonces a menudo la actividad de transferasa, tal como mediante desnaturalización por calor como un ejemplo no limitante; otro ejemplo no limitante es la desfosforilación del sustrato dNTPs. Luego se introduce una segunda molécula de ANcs en la primera molécula de ANcs "etiquetada". La segunda molécula de ANcs puede comprender un resto 5'-fosfato que está opcionalmente adenilado, una secuencia conductora 5', una secuencia de unión a diana que es suficientemente complementaria a una parte de la secuencia diana para permitir la hibridación a la misma y una cola 3' y un extremo 3' opcionalmente bloqueado. En el método, las moléculas de ANcs primera y segunda se hibridan o reasocian en condiciones en las que la secuencia diana y la secuencia de unión a diana se hibridan o reasocian entre sí por complementariedad de pares de bases, seguido del contacto de las moléculas hibridadas o reasociadas con una actividad de ligasa de ácido nucleico de cadena sencilla que liga el extremo 3' de la primera molécula de ANcs con el resto 5' fosfato que está opcionalmente adenilado de la segunda molécula de ANcs para producir un producto, una molécula de ANcs ligada.

En este aspecto de la divulgación, la secuencia conductora 5' de la segunda molécula de ANcs no requiere la presencia de una secuencia de índice molecular, y así su longitud se puede reducir en consecuencia si se desea. La secuencia conductora 5' puede incluir una secuencia de unión al cebador, que estaría presente en la molécula ligada resultante.

En lo aspectos arriba descritos de la divulgación, la molécula ligada puede posteriormente ser degradada parcialmente en su extremo 3' de tal manera que se separa la secuencia de unión a diana y, opcionalmente, parte de la secuencia conductora 5'. Esta degradación o digestión puede ser mediada por el contacto con una actividad de exonucleasa de cadena sencilla apropiada tal como la conoce la persona experta. La actividad de exonucleasa también puede dar como resultado la degradación y digestión de otras moléculas de cadena sencilla. Ejemplos no limitantes incluyen cualquier primera y segunda moléculas de ANcs que no fueron ligadas tal como se describe en esta memoria.

En realizaciones adicionales, las moléculas pueden contener uno o más residuos o nucleótidos que son dianas para la degradación por parte de una endonucleasa, ejemplos no limitantes de residuos o nucleótidos son uno o más de uracilo, inosina, nucleótidos abásicos u otros nucleótidos o residuos modificados en la secuencia de unión a diana, o la secuencia conductora 5', en la segunda molécula de ANcs de modo que el contacto con una enzima que tiene una actividad endonucleasa puede utilizarse para la degradación o digestión adicional de la molécula ligada. Ejemplos no limitantes de enzimas que pueden proporcionar una actividad de endonucleasa son uracil-ADN glicosilasa (UDG), Endonucleasa V, APE 1, Endonucleasa III, TMA Endonucleasa III y Endonucleasa VIII. En muchas realizaciones, una actividad enzimática utilizada para la degradación o digestión puede terminarse por inactivación por calor. En algunas realizaciones, la secuencia conductora 5' contiene una secuencia de unión a cebador que permite la hibridación de un oligonucleótido cebador para la síntesis de una cadena de ácido nucleico que es complementaria a al menos una porción de la molécula ligada. En algunas realizaciones, el cebador puede contener una secuencia adicional en su extremo 5' que no es complementaria a la secuencia de unión al cebador.

En un cuarto aspecto, la divulgación incluye métodos de sintetizar copias adicionales de la secuencia diana mediante el uso de una molécula de ANcs ligada descrita. En algunas realizaciones, la síntesis es de copias que son complementarias a la secuencia objetivo. En algunos casos, el método comprendería el uso de una molécula de ANcs ligada, que ha sido digerida o degradada opcionalmente tal como se describe en esta memoria, y ponerla en contacto con un oligonucleótido cebador que es capaz de hibridarse a al menos una parte de la secuencia de unión al cebador de la molécula de ANcs ligada. Luego se deja que el oligonucleótido cebador y la molécula de ANcs ligada se hibriden o reasocien en condiciones apropiadas, seguido de contacto con una actividad de polimerasa y nucleótidos apropiados para sintetizar una cadena complementaria a la molécula de ANcs ligada. Esto produce una

molécula dúplex (de doble cadena) que se puede desnaturalizar mediante condiciones de reacción apropiadas que luego se devuelven a las condiciones de hibridación que permiten un ciclo adicional de hibridación entre un oligonucleótido cebador y la molécula de ANcs ligada, seguido de la síntesis de otra cadena complementaria. Por supuesto, el ciclo puede repetirse para permitir la amplificación lineal del complemento de la secuencia diana en una molécula ligada.

Para realizaciones en las que cada una de las moléculas ligadas se "etiqueta" con un identificador o índice molecular único, cada uno de los complementos sintetizados de la secuencia diana contendría el complemento del identificador único.

En algunas realizaciones, se proporciona la síntesis de la secuencia diana como un dúplex. En algunos casos, el método comprendería el uso de una molécula de ANcs ligada, que ha sido digerida o degradada opcionalmente tal como se describe en esta memoria, y la puesta en contacto con un par de cebadores para la amplificación de la secuencia diana mediante el uso de la reacción en cadena de la polimerasa (PCR) como un ejemplo no limitativo. En la PCR, un primer cebador puede ser el oligonucleótido cebador que es capaz de hibridarse con al menos una porción de la secuencia de unión al cebador de la molécula de ANcs ligada. El segundo cebador puede contener toda o parte de la secuencia diana, de modo que puede hibridarse o reasociarse con la cadena de ácido nucleico sintetizada mediante el uso del primer cebador y una actividad de polimerasa apropiada. A veces, el segundo cebador contiene una secuencia enlazada a la secuencia diana. Ciclos repetidos de PCR con los cebadores permiten la amplificación de la secuencia diana como una molécula de doble cadena. Esto amplifica la secuencia diana en una molécula ligada, así como el complemento de la secuencia diana.

Para realizaciones en las que cada una de las moléculas ligadas se "etiqueta" con un identificador o índice molecular único, cada uno de los dúplex sintetizados de la secuencia diana contendría el identificador único y su complemento.

En un quinto aspecto, la divulgación incluye métodos para la detección, identificación, medición, el análisis y/o la evaluación de la secuencia diana o una parte de la misma mediante el uso de copias sintetizadas que contienen la secuencia diana o su complemento. En realizaciones con moléculas ligadas "etiquetadas" con un identificador o índice molecular único, cada una de las secuencias diana que estaba originalmente presente en una primera molécula de ANcs puede identificarse y, opcionalmente, contarse mediante el uso del identificador o índice unido a la secuencia diana. Ejemplos no limitantes de detección, identificación, medición, análisis y/o evaluación incluyen la secuenciación del ácido nucleico de las copias sintetizadas de la secuencia diana y la detección mediante el uso del identificador único; determinación de la masa de moléculas completas o fragmentadas antes o después de la selección de una o más moléculas por hibridación.

Determinadas realizaciones se describen adicionalmente en la siguiente descripción, ejemplos, reivindicaciones y dibujos.

Breve descripción de los dibujos

Los dibujos ilustran determinadas realizaciones de la tecnología y no son limitantes. Para mayor claridad y facilidad de ilustración, los dibujos no están hechos a escala y, en algunos casos, diversos aspectos pueden mostrarse exagerados o ampliados para facilitar la comprensión de realizaciones particulares.

La Figura 1 ilustra una realización representativa no limitante de una primera molécula de ANcs de la divulgación. La molécula comienza con un extremo 5' desfosforilado (resto 5'-OH) seguido de una secuencia Conductora Diana (que es opcional), la secuencia Diana, una Cola Diana (que es opcional), un Índice Molecular (introducido por una actividad de transferasa terminal tal como la proporcionada por algunas realizaciones descritas) y una cola oligoT (introducida por una actividad de transferasa terminal tal como se describe en esta memoria). A veces, la cola oligo dT puede ser una cola oligo dA, una cola oligo dC, una cola oligo dG o una cola que incluye un oligo (dT, dA, dC o dG) junto con un segundo oligo (dT, dA, dC o dG) que es diferente del primero. Algunas veces a la cola se la puede aludir como una "indicador" o "marca". A veces, el indicador, la marca y el índice molecular pueden estar en un orden diferente. Un sitio de interés, tal como un sitio de mutación dentro de la Cola Diana que está enlazado a la secuencia diana, también se muestra como un ejemplo no limitante.

La Figura 2 ilustra una realización representativa no limitante de una segunda molécula de ANcs de la divulgación. La molécula se representa de izquierda a derecha en la orientación de 3' a 5'. De 3' a 5', la molécula contiene un extremo 3' bloqueado (que es opcional), un Sitio de Unión a Diana (o secuencia de unión a diana), un Sitio de Unión del Cebador Universal y un residuo guanina en el extremo 5' que está fosforilado.

La Figura 3 muestra una realización representativa no limitante de las moléculas de ANcs de la divulgación adecuada para el ligamiento. La primera molécula de ANcs está en la parte superior de la figura y comienza con un extremo 5' desfosforilado (resto 5'-OH) seguido de una secuencia Conductora Diana (que es opcional), la secuencia Diana, una Cola Diana (que es opcional), un Índice Molecular (introducido por una actividad de transferasa terminal

- tal como se proporciona por algunas realizaciones descritas) y una cola oligoT (introducida por una actividad de transferasa terminal tal como se describe en esta memoria). La segunda molécula de ANcs se encuentra en la parte inferior de la figura y se representa de izquierda a derecha en la orientación 3' a 5' con complementariedad entre su Sitio de Unión a Diana (o secuencia de unión a diana) y la secuencia Diana de la primera molécula de ANcs. De 3' a 5', la segunda molécula de ANcs contiene un extremo 3' bloqueado (que es opcional), un Sitio de Unión a Diana (o secuencia de unión a diana), un Sitio de Unión al Cebador Universal y un residuo de guanina en el extremo 5' que está fosforilado o adenilado (no mostrado). La flecha muestra el sitio para la actividad de ligasa de cadena sencilla entre las dos cadenas.
- La Figura 4 ilustra una molécula ligada representativa que corresponde a la Figura 3 después de la digestión de su extremo 3' con actividad UDG que se digiere dentro del Sitio de Unión a Diana. La molécula contiene un Sitio de Interés, tal como un sitio de mutación dentro de la Cola Diana que está enlazado a la secuencia diana, y es adecuado para la hibridación con un oligonucleótido cebador a través del Sitio de Unión al Cebador Universal.
- La Figura 5 ilustra una amplificación lineal representativa de dos moléculas correspondientes a la Figura 4. A la izquierda hay una ilustración en la que el Sitio de Interés no está mutado (o es de tipo salvaje tal como se indica con "W") y está enlazado al Índice Molecular 1. A la derecha está una ilustración en la que el Sitio de Interés está mutado (tal como se indica con "M") y está enlazado al Índice Molecular 2. La amplificación lineal de cada una de las moléculas con un oligonucleótido cebador que puede contener una extensión opcional en el extremo 5' produce copias del complemento de la molécula ligada. La extensión puede ser una secuencia de adaptador para facilitar la secuenciación de las moléculas amplificadas, tal como mediante el uso de técnicas de secuenciación de nueva generación (NGS). Dados los posibles errores debidos a la actividad de la polimerasa, los errores introducidos en el Sitio de Interés pueden identificarse y descartarse en función del predominio de la secuencia en el Sitio según se determina por secuenciación. A veces, otras fuentes de error pueden identificarse y descartarse de manera similar en base al predominio de la secuencia en el Sitio de Interés, según se determina por secuenciación. El predominio puede determinarse basado, en parte, en el Índice Molecular presente en las copias amplificadas.
- La Figura 6 ilustra una amplificación representativa de doble cadena de las moléculas correspondientes a la Figura 5 para la unión de una segunda secuencia de adaptador. Cebadores específicos para la diana con una secuencia de adaptador en sus extremos 5' pueden utilizarse en combinación con un oligonucleótido cebador directo para amplificar productos de la Figura 5. Las secuencias de adaptador pueden ser secuencias de adaptador de NGS.
- La Figura 7 ilustra una realización alternativa de la divulgación para la amplificación, detección, identificación, medición, el análisis y/o la evaluación de una secuencia diana y/o uno o más sitios de secuencia de ácido nucleico unidos a la secuencia diana. En determinadas realizaciones, el extremo 3' ddC del cebador puede eliminarse mediante una actividad de una polimerasa de lectura de prueba, permitiendo así la extensión.
- La Figura 8 ilustra las opciones para la detección de un sitio de interés en una secuencia nativa que ha sido fragmentada en una longitud promedio de 168 pares de bases Diana del producto ligado.
- La Figura 9 ilustra que el ligamiento tuvo éxito después de hibridar un segundo ANsc y de reasociarlo a una secuencia de unión a diana en un primer ANsc. El producto se puede procesar a una forma más corta mediante digestión con UDG que separa el resto de unión a Diana del producto ligado.
- La Figura 10 muestra que un segundo ANcs (i) se hibridó y se reasoció, (ii) se ligó y (iii) se procesó (procesamiento UDG) con éxito, para dos secuencias de unión a diana BRAF cada una en una primera realización de ANcs separada.
- La Figura 11 muestra estos segundos ANcs específicamente hibridados o reasociados y ligados a su secuencia Diana BRAF en el primer ANcs (figura superior) o específicamente reasociados y ligados a su secuencia Diana PIK3CA (figura inferior).
- La Figura 12 muestra múltiples segundas realizaciones de ANcs que fueron (i) hibridadas y reasociadas, (ii) ligadas y (iii) procesadas (procesamiento UDG) con éxito, para múltiples primeras realizaciones de ANcs en una realización de ensayo multiplex.
- La Figura 13 muestra que una segunda molécula de ANcs con una secuencia de unión a una diana específica se puede ligar a su primera diana de ANcs en un intervalo de temperaturas de al menos 60 a 75 grados Celsius.
- La Figura 14A muestra que el producto ligado de ANcs específico para la diana puede amplificarse en un intervalo de temperaturas de al menos 55 a 72 grados Celsius cuando se añade a un fondo de ADN genómico humano. La Figura 14B muestra que el control del ADN genómico humano sin el producto ligado no produjo material de producto amplificado de al menos 65,6 a 72 grados Celsius.

La Figura 15 muestra un ejemplo de cómo la indexación molecular dinámica puede conducir a una gran diversidad en el etiquetado de una secuencia.

La Figura 16 muestra un ejemplo de indexación molecular dinámica universal seguida de una Etiqueta Diana específica con un sitio de interés 3' del sitio diana y una etiqueta.

- 5 La Figura 17 muestra un ejemplo de indexación molecular dinámica universal seguida de una Etiqueta Diana específica con un sitio de interés 3' y/o 5' del sitio diana y dos etiquetas.

La Figura 18 muestra un ejemplo de indexación molecular dinámica universal, utilizando una secuencia de unión PoliA (como un ejemplo, también se puede utilizar otro homopolímero) y una Etiqueta Diana universal con una etiqueta.

- 10 La Figura 19 muestra un ejemplo de indexación molecular dinámica universal, utilizando una secuencia de unión PoliA (como un ejemplo, también se puede utilizar otro homopolímero) y una Etiqueta Diana universal con dos etiquetas.

La Figura 20 muestra un ejemplo de indexación molecular dinámica universal utilizando homopolímeros en tándem para el sitio de unión y una Etiqueta Diana universal con una etiqueta.

- 15 La Figura 21 muestra un ejemplo de indexación molecular dinámica universal utilizando homopolímeros en tándem para el sitio de unión y una Etiqueta Diana universal con dos etiquetas.

La Figura 22 muestra un ejemplo de indexación molecular dinámica universal utilizando un sitio de cebado PoliA (u otro homopolímero) y un cebador universal con una etiqueta.

- 20 La Figura 23 muestra un ejemplo de indexación molecular dinámica universal utilizando un sitio de cebado PoliA (u otro homopolímero) y un cebador universal con dos etiquetas.

La Figura 24 muestra un ejemplo de indexación molecular dinámica universal, utilizando un sitio de cebado poliA (u otro homopolímero)/poli T (u otro homopolímero) y un cebador universal con una etiqueta.

La Figura 25 muestra un ejemplo de indexación molecular dinámica universal, utilizando un sitio de cebado poliA (u otro homopolímero)/poli T (u otro homopolímero) y un cebador universal con dos etiquetas.

- 25 La Figura 26 muestra un curso en el tiempo para adiciones de dATP y dNTP por transferasa terminal.

La Figura 27 demuestra que la transferasa terminal añade aleatoriamente residuos de dNTP a la molécula de ácido nucleico.

La Figura 28 muestra un ejemplo de una realización del proceso de indexación molecular dinámica (DMI).

Descripción detallada

- 30 El tamaño fragmentada de ácidos nucleicos libres de células, tal como el ADN libre de células presente en sangre y orina, ha sido reseñado previamente. Para una secuencia de ácido nucleico de interés dada, o sitio de interés, que está presente en una secuencia nativa, la fragmentación aleatoria de la secuencia nativa da como resultado una pluralidad de posibles fragmentos que contienen el sitio de interés. Esto se ilustra en la Figura 8, en donde se muestra una secuencia nativa en la parte superior con un sitio de interés "M", por ejemplo. Se muestran doce fragmentos representativos de 168 pb de longitud y que contienen M. Para un ensayo hipotético basado en la PCR
- 35 fragmentos representativos según lo indicado por el par de flechas.

- 40 Los métodos descritos incluyen realizaciones con una secuencia diana de aproximadamente 23 nucleótidos en una primera molécula de ANcs y se indican como capaz de detectar aquellos tres fragmentos así como otros ocho. Los métodos descritos pueden utilizar la secuencia en ambos lados de M como la secuencia diana en una primera molécula de ANcs descrita en esta memoria. Cada una de las cadenas de la secuencia nativa proporciona una secuencia diana disponible enlazada a M que puede ser para hibridación a una secuencia de unión a diana correspondiente tal como se describe en esta memoria para permitir la producción de una molécula ligada.

- 45 La Tabla 1 muestra la detección esperada de pequeñas moléculas de ácido nucleico en función de su tamaño y el tamaño de la "huella" del ensayo:

Tabla 1: Porcentaje esperado de moléculas detectadas

Tamaño diana (pb)	168			45		
Tamaño del ensayo (pb)	208	120	23	208	120	23
Porcentaje detectado	0	29	87 *	0	0	51 *

El asterisco en la Tabla 1 indica que las tasas de detección global pueden aumentar significativamente cuando se incluye un ensayo no competitivo desde el otro lado del sitio de interés.

- 5 Los métodos y las moléculas descritos se basan a menudo en la secuencia más corta que hace que la secuencia sea única en el genoma humano. Solo la secuencia diana, y cualquier secuencia o sitio de interés enlazado, necesita estar presente para el etiquetado con éxito de la secuencia diana porque, en relación con un ensayo basado en la PCR, el "segundo" sitio de cebador es proporcionado por la segunda molécula de ANcs de la divulgación. El uso de las secuencias únicas más cortas permite una metodología de ensayo ultracorta.
- 10 Los métodos descritos también aseguran especificidad mediante el uso de indexación molecular, en donde cada una de las moléculas de ANcs nativas en una muestra está etiquetada con un identificador codificado por nucleótidos o único, o "código de barras". La secuencia nativa *per se*, la longitud de la secuencia nativa y la longitud de la cola oligoT opcional también contribuyen a la singularidad de cada una de las moléculas etiquetadas por un método descrito. A veces, la cola oligo dT puede ser una cola oligo dA, una cola oligo dC, una cola oligo dG o una cola
- 15 compuesta de un oligo (dT, dA, dC o dG) en tándem con un segundo oligo (dT, dA, dC o dG) que es diferente del primero, todo lo cual contribuiría a la singularidad de cada una de las moléculas etiquetadas por un método descrito.

Además de ello, la amplificación lineal de las moléculas ligadas (etiquetadas) descritas asegura que las moléculas mutantes y de tipo salvaje se puntuaran correctamente. En realizaciones en las que las secuencias y los cebadores de unión a diana descritos solo están enlazados o son adyacentes a un Sitio de Interés (p. ej., no codifican

20 secuencias mutantes o de tipo salvaje), no pueden contribuir a la puntuación de las moléculas mutantes y de tipo salvaje.

Métodos para etiquetar moléculas de ANcs

Tal como se describe en esta memoria, la divulgación incluye un método para producir una molécula de ácido nucleico de cadena sencilla a partir de dos moléculas. Por lo tanto, los métodos de etiquetado generalmente actúan

25 sobre dos NA de cadena sencilla y generan un NA de cadena sencilla. En algunas realizaciones, el método comprende ligar una primera y una segunda molécula de ANcs después de que se hayan hibridado o reasociado entre sí. La primera molécula de ácido nucleico de cadena sencilla (ANcs) comprende un extremo 5' que está opcionalmente desfosforilado, una secuencia diana y un extremo 3' con una secuencia opcional de cola 3', mientras que la segunda molécula de ANcs comprende un resto 5'-fosfato que está opcionalmente adenilado, una secuencia

30 conductora 5', una secuencia de unión a diana que es el complemento de la secuencia diana y una cola 3' opcional y un extremo 3' opcionalmente bloqueado. Las moléculas se colocan en condiciones en las que la secuencia diana y la secuencia de unión a diana se hibridan o reasocian entre sí por complementariedad de pares de bases y luego se ponen en contacto con una actividad de ligasa de ácido nucleico de cadena sencilla que liga el extremo 3' de la primera molécula de ANcs con el resto 5' fosfato que está opcionalmente adenilado de la segunda molécula de ANcs

35 para producir una sola molécula de ANcs.

El que el extremo 5' del segundo ANcs esté adenilado puede depender de la enzima ligasa de cadena sencilla utilizada para ligar el primer ANcs al segundo ANcs. Además, si adenosina trifosfato está incluida o no en las condiciones de ligamiento también depende de la enzima ligasa de cadena sencilla utilizada para el ligamiento. La siguiente Tabla 2 muestra diferentes ejemplos no limitantes de combinaciones de enzimas ligasa de cadena sencilla,

40 primer ANcs y segundo ANcs.

Tabla 2: Combinaciones de segunda molécula de ANcs adeniladas y no adeniladas

Enzima	ATP requerido	primer ANcs	segundo ANcs
CirLigase*	Sí	extremo 5' desfosforilado, extremo 3' libre	extremo 3' bloqueado, extremo 5' fosforilado o adenilado
CirLigase II*	"No"	extremo 5' desfosforilado, extremo 3' libre	extremo 3' bloqueado, extremo 5' fosforilado o adenilado
T4 ARN Ligasa 1** (ARNcs Ligasa)	Sí	extremo 5' desfosforilado, extremo 3' libre	extremo 3' bloqueado, extremo 5' fosforilado o adenilado

Enzima	ATP requerido	primer ANcs	segundo ANcs
5' AppADN/ARN Ligasa Termoestable **	No	extremo 5' fosforilado, extremo 3' libre	extremo 3' bloqueado, extremo 5' adenilado
CircLigase*	Sí	extremo 5' fosforilado, extremo 3' libre	extremo 3' bloqueado, extremo 5' fosforilado o adenilado
CircLigase II*	"No"	extremo 5' fosforilado, extremo 3' libre	extremo 3' bloqueado, extremo 5' fosforilado o adenilado
T4 ARN Ligasa 1** (ARNcs Ligasa)	Sí	extremo 5' desfosforilado, extremo 3' libre	extremo 3' bloqueado, extremo 5' fosforilado o adenilado

* p. ej., de Epicenter Biotechnologies Corporation

** p. ej., de New England BioLabs, Inc., por ejemplo

5 Un extremo 3' bloqueado puede ser un extremo 3' que tiene un extremo de nucleótidos o de no nucleótidos que previene el ligamiento; también puede estar en un estado (p. ej., de doble cadena) que no es reconocido por las ligasas de ANcs. En los casos en los que el extremo 5' del primer ANcs está fosforilado, el segundo ANcs puede tener una ventaja competitiva sobre el primer ANcs en virtud de la proximidad al extremo 3' del primer ANcs. Debido a que la enzima CircLigase II indicada en la Tabla 2 está pre-adenilada, no es necesario añadir ATP y la enzima funciona estequiométricamente. La enzima 5' AppADN/ARN ligasa termoestable indicada en la Tabla 2 es una enzima mutante que no puede adenilar el 5' fosfato de ARN o ADNcs, y requiere que el segundo ANcs tenga un extremo 5' preadenilado. El segundo ANcs a menudo incluye un extremo 3' bloqueado para evitar la circularización. El primer ANcs incluye a menudo un extremo 5' desfosforilado para evitar la circularización cuando se utilizan las enzimas CircLigase, CircLigase II y T4 ARN Ligasa 1. Cuando se utiliza la enzima 5' AppADN/ARN ligasa termoestable no hay necesidad de desfosforilar el extremo 5' del primer ANcs, ya que la enzima no puede adenilar el extremo 5'.

10 Así, en algunas realizaciones, un primer ANcs que tiene un extremo 5' desfosforilado y el extremo 3' y libre se utiliza con un segundo ANcs que tiene un extremo 3' bloqueado y un extremo 5' fosforilado. En las últimas realizaciones, las condiciones de ligamiento incluyen adenosina trifosfato (ATP) o no incluyen ATP, dependiendo de la ligasa utilizada (p. ej., véanse los ejemplos en la Tabla 2 anterior). En determinadas realizaciones, se utiliza un primer ANcs que no tiene un extremo 5' desfosforilado (p. ej., que tiene un extremo 5' fosforilado) y un extremo 3' libre con un segundo ANcs que tiene un extremo 3' bloqueado y un extremo 5' adenilado. En las últimas realizaciones, las condiciones de ligamiento no incluyen a menudo ATP. Para fines de claridad, se puede ligar el extremo 3' de un primer ANcs con el resto 5' fosfato de un segundo ANcs cuando el extremo 5' del segundo ANcs ya está adenilado (es decir, el resto adenilo terminal enlazado a un resto fosforilo adyacente).

25 Métodos pueden realizarse en determinadas realizaciones con una primera molécula de ANcs de un espécimen biológico o muestra. En algunos casos, el espécimen o muestra es de un animal, tal como un sujeto canino, felino, equino, bovino, caprino, ovino, porcino, aviar, mono o humano, o una planta. Ejemplos no limitantes de un espécimen o muestra incluyen ácidos nucleicos celulares y ácidos nucleicos libres de células (p. ej., ácido nucleico libre de células circulante), derivados de fluidos corporales que incluyen sangre, plasma, suero, saliva, líquido cefalorraquídeo y orina. En algunos casos, el ANcs se obtiene después de la fragmentación de ácidos nucleicos más grandes en el espécimen o la muestra. El ANcs también se puede preparar por desnaturalización de ácidos nucleicos de doble cadena.

35 Con una primera molécula de ANcs de un espécimen biológico o muestra, la secuencia cola 3' es una secuencia nativa que se encuentra adyacente a, o enlazada a una secuencia diana. La secuencia de cola 3' puede contener opcionalmente una posición o un sitio de interés tal como se describe en esta memoria. Como se desprende de la divulgación, una posición o sitio de interés estaría 3' de la secuencia diana en tales casos. Una primera molécula de ANcs de un espécimen o muestra biológica puede considerarse una molécula de ANcs "nativa" que se produce en la naturaleza o se deriva de una molécula natural, tal como un ADN de doble cadena libre de células, en determinadas realizaciones. En la mayoría de las realizaciones de la divulgación, una molécula de ANcs nativa puede desfosforilarse (como se aborda en esta memoria) por métodos conocidos por la persona experta. La desfosforilación proporciona el beneficio de reducir el auto-ligamiento no deseado de una primera molécula de ANcs como se aborda en esta memoria.

45 Métodos pueden realizarse en algunas realizaciones con una primera molécula de ANcs que ha sido procesada para su uso en un método descrito. En algunas realizaciones, una primera molécula de ANcs se extiende mediante el uso de la actividad de transferasa terminal para introducir una cola oligoT. En lugar de una cola oligo dT, se puede añadir al primer ANcs una cola oligo dA, una cola oligo dC, una cola oligo dG o una cola compuesta por un oligo (dT, dA, dC o dG) en tándem con un segundo oligo (dT, dA, dC o dG) que es diferente del primero. A veces, a un

homopolinucleótido se le alude como un "indicador" o "marca". Algunas veces un indicador, marca e índice molecular están en un orden diferente. En algunos casos, la cola puede utilizarse para ayudar en el ligamiento a la segunda molécula de ANcs. En estas realizaciones y otras, una primera molécula de ANcs también puede comprender una secuencia conductora diana 5'. Con una primera molécula de ANcs de un espécimen o muestra biológica, la secuencia conductora diana 5' es una secuencia que se encuentra adyacente a o enlazada a una secuencia diana. La secuencia conductora 5' diana en una primera molécula de ANcs es a menudo una secuencia nativa con respecto a la secuencia diana. En algunos casos, una posición o sitio de interés puede estar presente en la secuencia conductora 5' y, por lo tanto, estar a 5' de la secuencia diana. Adicionalmente, una secuencia en el lado 3' de la secuencia diana, si está presente, puede considerarse una secuencia de cola diana. Una secuencia de cola diana 3' en una primera molécula de ANcs comprende a menudo, o es una secuencia nativa con respecto a la secuencia diana. Si está presente, una secuencia de cola diana puede ser de cualquier longitud o secuencia que no impida el ligamiento entre una primera molécula de ANcs y una segunda molécula de ANcs. En algunas realizaciones, una cola diana, si está presente, puede tener de uno a aproximadamente 500 nucleótidos de longitud.

La secuencia diana es de una longitud y secuencia que es única entre las secuencias de ácido nucleico y restos dentro de una muestra o espécimen. A veces, la secuencia diana puede ser una secuencia de indicador o marca que ha sido añadida a una pluralidad de las moléculas en la muestra o espécimen y, por lo tanto, puede no ser única. En algunos casos, tiene una longitud de aproximadamente 18 nucleótidos o menos, aproximadamente 19 nucleótidos o menos, aproximadamente 20 nucleótidos o menos, aproximadamente 21 nucleótidos o menos, aproximadamente 22 nucleótidos o menos, aproximadamente 23 nucleótidos o menos, aproximadamente 24 nucleótidos o menos, aproximadamente 25 nucleótidos o menos, aproximadamente 26 nucleótidos o menos, aproximadamente 27 nucleótidos o menos, o aproximadamente 28 nucleótidos o más. En determinadas realizaciones, la secuencia diana puede ser de 35 nucleótidos o más, o de 40 nucleótidos o más, o de 50 nucleótidos o más, o de 70 nucleótidos o más. En la secuencia de unión a diana de una segunda molécula de ANcs, los ácidos nucleicos bloqueados (LNAs) pueden utilizarse para disminuir el tamaño de la secuencia diana asociada para la hibridación. Por consiguiente, la segunda molécula de ANcs puede contener uno o más nucleótidos modificados, residuos de uracilo, residuos de inosina, sitios abásicos u otras modificaciones de ácido nucleico, según lo desee la persona experta. En algunos casos, el extremo 5' de una segunda molécula de ANcs puede comprender ribonucleótidos de modo que la molécula sea una quimera de ADN/ARN. En algunos casos, el extremo 5' de una segunda molécula de ANcs puede comprender nucleótidos de un tipo que no se encuentra en la naturaleza.

En muchas realizaciones, una secuencia diana y la secuencia de unión a diana, cuando se hibridan o reasocian, tienen una temperatura de fusión de aproximadamente 50°C, aproximadamente 55°C, aproximadamente 60°C, aproximadamente 65°C o aproximadamente 70°C o superior. En algunas realizaciones, una secuencia diana y la secuencia de unión a diana, cuando se hibridan o reasocian, tienen una temperatura de fusión de aproximadamente 40°C o de aproximadamente 45°C.

En muchas realizaciones, el método puede realizarse con una segunda molécula de ANcs que está fosforilada u opcionalmente adenilada en el extremo 5'. Adicionalmente, una segunda molécula de ANcs puede tener un residuo guanina (G) en el extremo 5' para ayudar en el ligamiento a la primera molécula de ANcs. Además de ello, una segunda molécula de ANcs puede contener una secuencia conductora 5' que comprende una secuencia de índice molecular (ácido nucleico único) y una secuencia de unión a cebador. En realizaciones para las cuales hay una pluralidad de primeras moléculas de ANcs, tales como un espécimen biológico o muestra, cada una de las segundas moléculas de ANcs puede contener una secuencia de índice molecular diferente. La diversidad de la secuencia del índice molecular se describe con mayor detalle en esta memoria.

Además de ello, una segunda molécula de ANcs puede incluir un extremo 3' bloqueado para evitar el auto-ligamiento, un sitio de unión al cebador de aproximadamente 18 nucleótidos o más de longitud, y/o un índice molecular de aproximadamente 10 nucleótidos o más. En realizaciones descritas en esta memoria con un índice molecular añadido a la primera molécula de ANcs, la persona experta entendería que la segunda molécula de ANcs no es necesaria para proporcionar un identificador único a la primera molécula de ANcs. No obstante, la segunda molécula de ANcs aún puede comprender un índice molecular adicional para proporcionar dos identificadores a una secuencia diana.

Después de combinar la primera y segunda moléculas de ANcs, tal como en una mezcla de reacción, las mismas se pueden desnaturalizar y luego reasociar lentamente para permitir la hibridación de la secuencia diana en la primera molécula de ANcs a la secuencia de unión a diana en la segunda molécula de ANcs. Se puede considerar que la secuencia diana en la primera molécula de ANcs y la secuencia de unión a diana en la segunda molécula de ANcs forman una "pinza" que acercan el extremo 3' de la primera molécula de ANcs y el extremo 5' de la segunda molécula de ANcs para el ligamiento. Después de la hibridación o reasociación, se puede utilizar una actividad ligasa para ligar las moléculas hibridadas o reasociadas. Las moléculas de ANcs individuales ligadas resultantes son productos de la divulgación tal como se describe en esta memoria.

Después del ligamiento, las moléculas de productos de ANcs individuales pueden ser considerados como una estructura de "bucle de horquilla" con una región dúplex formada por la hibridación de la secuencia diana y la secuencia de unión a diana con un bucle formado por ligamiento. El producto de ANcs individual comprende a veces una estructura de horquilla a 25 grados Celsius. El bucle puede ser de cualquier tamaño aceptado por una actividad de ligasa de cadena sencilla. Ejemplos no limitantes incluyen de aproximadamente 15 nucleótidos o más, aproximadamente 20 nucleótidos o más, aproximadamente 30 nucleótidos o más, aproximadamente 40 nucleótidos o más, aproximadamente 50 nucleótidos o más, aproximadamente 75 nucleótidos o más, aproximadamente 100 nucleótidos o más, aproximadamente 150 nucleótidos o más, aproximadamente 200 nucleótidos o más, aproximadamente 300 nucleótidos o más, aproximadamente 400 nucleótidos o más o aproximadamente 500 nucleótidos o más. En algunas realizaciones, la estructura de horquilla comprende un bucle de aproximadamente 5 bases de nucleótidos a aproximadamente 500 bases de nucleótidos de longitud. En determinadas realizaciones, la estructura de horquilla comprende una región parcialmente de doble cadena y una región de cadena sencilla, y la región doble cadena es de aproximadamente 18 nucleótidos a aproximadamente 35 o más bases de nucleótidos de longitud. Ejemplos no limitantes de una actividad ligasa incluyen ligasas de ADNcs termoestables disponibles comercialmente como CircLigase™ y CircLigase™ II y la ligasa de 5' App ADN/ARN termoestable de *M. thermoautotrophicum*.

Después de la reacción de ligamiento, la mezcla puede ser calentada para inactivar la ligasa de cadena sencilla, en algunas realizaciones. La mezcla a veces se pone en contacto con una actividad de exonucleasa de cadena sencilla para digerir moléculas de cadena sencilla, tales como las moléculas de ANcs primera y segunda que no han reaccionado. Un ejemplo no limitante de una actividad de exonucleasa apropiada es exonucleasa I tal como conoce el experto en la técnica. Los extremos 3' de doble cadena de las moléculas ligadas las hacen resistentes a la digestión por exonucleasa, y las moléculas no ligadas están sujetas a la digestión. La digestión puede ser seguida por la inactivación por calor opcional de la actividad de exonucleasa.

En algunas realizaciones, la reacción puede entonces ponerse en contacto con una actividad de uracilo-ADN glicosilasa para digerir en posiciones que contienen uracilo presentes en la antigua segunda molécula de ANcs. En muchas realizaciones, las posiciones que contienen uracilo están dentro de la secuencia de unión a diana o muy próximas y 5' de la secuencia de unión a diana. Después del tratamiento con UDG, la actividad enzimática es opcionalmente inactivada por calor. Las moléculas procesadas resultantes también son productos de la divulgación tal como se describe en esta memoria.

Mejora de las tasas de detección generales

Tal como se ha señalado en esta memoria, las tasas de detección global pueden aumentar significativamente cuando un ensayo no competitivo desde el otro lado de un sitio de interés se incluye en un ensayo. Por lo tanto, en algunas realizaciones, la primera molécula de ANcs procede de una molécula de ácido nucleico de doble cadena (ANdc), o molécula parcialmente de ANdc y molécula parcialmente de ANcs, que comprende una primera molécula de ANcs sentido y una primera molécula de ANcs antisentido que puede ser fijada independientemente como objetivo por una segunda molécula de ANcs sentido y una segunda molécula de ANcs antisentido, respectivamente. La primera molécula de ANcs sentido y una primera molécula de ANcs antisentido pueden separarse una de otra en condiciones desnaturalizantes (p. ej., aplicación de calor, aplicación de desnaturalizantes químicos). La primera molécula de ANcs sentido y la primera molécula de ANcs antisentido pueden ponerse en contacto bajo condiciones de hibridación o reasociación con (i) una segunda molécula de ANcs sentido que comprende una secuencia de unión a diana complementaria a la secuencia diana en la primera molécula de ANcs sentido, y (ii) una segunda molécula de ANcs antisentido que comprende una secuencia de unión a diana complementaria a la secuencia diana en la primera molécula de ANcs antisentido, produciendo con ello moléculas hibridadas o reasociadas. Bajo las condiciones de hibridación o reasociación, al menos una porción de la secuencia diana y la secuencia de unión a diana en la primera molécula de ANcs sentido y la segunda molécula de ANcs sentido, y al menos una porción de la secuencia diana y la secuencia de unión a diana en la primera molécula de ANcs antisentido y la segunda molécula de ANcs antisentido se hibridan o reasocian entre sí por complementariedad de pares de bases. Las moléculas hibridadas o reasociadas se ponen en contacto a menudo con una actividad ligasa de ácido nucleico de cadena sencilla en condiciones de ligamiento. Bajo las condiciones de ligamiento, el extremo 3' de la primera molécula de ANcs sentido se liga al resto 5' fosfato (opcionalmente adenilado) de la segunda molécula de ANcs sentido, y el extremo 3' de la primera molécula de ANcs antisentido se liga al resto 5' fosfato (opcionalmente adenilado) de la segunda molécula de ANcs antisentido, generando con ello moléculas de ANcs individuales. Las características descritas en esta memoria que pertenecen generalmente a la primera molécula de ANcs y a la segunda molécula de ANcs pueden pertenecer al primer ANcs sentido, al segundo ANcs sentido, al primer ANcs antisentido y al segundo ANcs antisentido. Dado que el extremo 3' de las primeras moléculas de ANcs sentido y antisentido se liga al extremo 5' de las segundas moléculas de ANcs sentido y antisentido, la segunda molécula de ANcs sentido y la segunda molécula de ANcs antisentido se dirigen de manera efectiva a ambos lados del sitio de interés.

Incorporación y uso de secuencias de índice molecular

La divulgación incluye, además un método para producir una primera molécula de ácido nucleico de cadena sencilla (ANcs) que está etiquetada con un identificador único y luego se liga a una segunda molécula de ANcs. Una primera molécula de ANcs puede ser cualquiera de las descritas en esta memoria antes del procesamiento y de la puesta en contacto con una segunda molécula de ANcs. Una primera molécula de ANcs puede ser una molécula de cadena sencilla obtenida o preparada a partir de un espécimen biológico o muestra descrito en esta memoria, y se extiende en su extremo 3' con incorporación aleatoria de nucleótidos (también conocidos como residuos de nucleótidos) suficientes para crear un índice molecular único, generado aleatoriamente, de nucleótidos enlazados a cada una de las primeras moléculas de ANcs. A esta secuencia única se la alude como una secuencia de índice molecular, o "índice" o "polinucleótido de índice", y se añade a un ácido nucleico mediante un procedimiento al que se alude en esta memoria como "indexación molecular dinámica" (DMI). La adición de este polinucleótido único (o secuencia única) a cada uno de los primeros ANcs puede realizarse mediante el uso de una actividad de transferasa terminal en presencia de una mezcla de dNTP o NTP. Después de la fijación de un polinucleótido único (o secuencia única) como una etiqueta, la primera molécula de ANcs puede extenderse opcionalmente en su extremo 3' con una cola oligoT o una cola oligoA proporcionada por la presente divulgación.

Una primera molécula de ANcs etiquetada mediante el uso de una actividad de transferasa terminal se puede considerar que ha sido "directamente" etiquetada, mientras que una primera molécula de ANcs etiquetada mediante ligamiento a una segunda molécula de ANcs puede considerarse que ha sido "indirectamente" etiquetada. Antes o después de la preparación de una primera molécula de ANcs etiquetada directamente, puede desfosforilarse en su extremo 5' tal como se proporciona en la presente divulgación. Una primera molécula de ANcs etiquetada directamente es un producto de la divulgación como se describe en esta memoria.

La adición de 10 nucleótidos al azar, ya sea mediante etiquetado directo o como un índice molecular en un segundo ANcs, proporciona más de un millón de posibles etiquetas de secuencias únicas. El uso de 20 nucleótidos aleatorios genera más de un trillón de posibles etiquetas únicas. En el caso de un espécimen biológico o muestra, el número de posibles moléculas de ANcs a etiquetar puede ser fácilmente superado en número por el número de posibles índices moleculares.

La Figura 1 ilustra una primera molécula de ANcs directamente etiquetada de la divulgación. Se puede utilizar con una segunda molécula de ANcs tal como se ilustra en la Figura 2. La preparación y el procesamiento de una primera molécula de ANcs directamente etiquetada y una segunda molécula de ANcs se pueden realizar tal como se describe a lo largo de la presente divulgación. La Figura 3 ilustra una disposición de una primera molécula de ANcs directamente etiquetada y una segunda molécula de ANcs de la divulgación. El ligamiento entre una primera molécula de ANcs directamente etiquetada y una segunda molécula de ANcs, y el procesamiento adicional de un producto ligado se pueden realizar tal como se describe a lo largo de la presente divulgación. La Figura 4 ilustra un producto ligado y procesado, formado a partir de una primera molécula de ANcs directamente etiquetada y una segunda molécula de ANcs de la divulgación. Las moléculas ligadas y/o procesadas resultantes son productos de la divulgación tal como se describe en esta memoria.

En esta memoria, DMI es un proceso por el cual se crea un índice molecular por la adición secuencial de monómeros de nucleótidos de un ácido nucleico (p. ej., ácido nucleico en un espécimen o muestra; un primer ANcs; un segundo ANCS) a través de una actividad de transferasa terminal. Si se añade secuencialmente un solo tipo de monómero de nucleótidos, entonces se añade al ácido nucleico un homopolímero, al que también se alude como "homopolinucleótido". Si se añaden secuencialmente dos, tres o cuatro tipos de monómeros de nucleótidos, entonces se añade al ácido nucleico un heteropolímero, al que también se alude como "heteropolinucleótido". Los homopolímeros (p. ej., indicadores y marcas) y los heteropolímeros (p. ej., índices moleculares) pueden añadirse en cualquier orden a ANcs o ANdc. Además, se puede utilizar un homopolinucleótido para reasociar un segundo ANcs a un primer ANcs. En determinadas realizaciones, estas dos moléculas de ANcs pueden ligarse posteriormente juntas, y en algunas realizaciones, el segundo ANcs puede extenderse. El primer ANcs puede ser naturalmente de cadena sencilla o puede derivarse de un ANdc o parcialmente ANdc.

Un índice molecular incluye a menudo dos o más nucleótidos diferentes (es decir, un heteropolímero) que por sí mismo o en combinación con un indicador (es decir, un homopolímero) o una marca (es decir, un homopolímero de un tipo diferente de un indicador) o un ácido nucleico en un espécimen o muestra que puede variar en secuencia y/o longitud, da como resultado una secuencia general que es única o virtualmente única para la molécula a la que está contigua. Como un ejemplo no limitativo, refiriéndose a la Figura 15, una molécula en un espécimen de ADN libre de células que tiene una cola (o indicador) homopolimérica de 25 a 35 residuos de dA, seguido de un índice molecular de 10 a 20 nucleótidos aleatorios y un sitio de interés que puede aparecer en cualquier lugar dentro de un espacio de 145 nucleótidos, tendría una diversidad potencial de aproximadamente $1,5 \times 10^9$ a aproximadamente $1,5 \times 10^{15}$. Dado que una muestra de 100 ng de ADN libre de células de un espécimen de sangre de 20 mL contendría solo aproximadamente 3×10^4 moléculas de ADN de cualquier secuencia particular, sería virtualmente imposible que dos moléculas originales diferentes de un tipo particular tengan el mismo índice molecular.

La divulgación incluye productos producidos por los métodos descritos de la creación de una molécula de AN con un índice molecular añadida por un proceso de DMI. Estos incluyen moléculas ligadas y versiones procesadas de las mismas, tal como se describe. Las Figuras 5 y 6 ilustran los posibles usos de las moléculas ligadas para una amplificación, detección, identificación, medición, análisis y/o evaluación adicionales de una secuencia diana y/o uno o más sitios de secuencia de ácido nucleico enlazados a la secuencia diana. Un ácido nucleico incluye a veces un heteropolinucleótido indicador (un ácido nucleico indexado), un homopolinucleótido indicador (un ácido nucleico con inidcador), un homopolinucleótido de marca (un ácido nucleico marcado) o una combinación de los mismos. Un homopolinucleótido indicador o un homopolinucleótido de marca a veces es un homopolinucleótido en tándem que comprende un primer polinucleótido que consiste en un primer nucleótido (p. ej., de aproximadamente 5 a aproximadamente 100 nucleótidos de longitud) y un segundo polinucleótido directamente unido al primer polinucleótido que consiste en un segundo nucleótido (p. ej., también de aproximadamente 5 a aproximadamente 100 nucleótidos de longitud). Un proceso DMI para generar un ácido nucleico etiquetado, con indicador y/o marcado a veces (i) actúa en un AN de cadena sencilla y genera un AN de cadena sencilla, (ii) a veces actúa en un AN de doble cadena y genera un AN parcialmente de doble cadena y parcialmente de cadena sencilla, y (iii) a veces actúa sobre un AN parcialmente de cadena sencilla y parcialmente de doble cadena y genera un AN parcialmente de doble cadena/parcialmente de cadena sencilla.

En algunas realizaciones, la divulgación incluye una molécula ligada tal como se ilustra en la Figura 7, que está en contacto con un cebador que contiene un extremo 3' bloqueado (residuo 3' ddC como un ejemplo ilustrativo). El extremo 3' bloqueado se separa a menudo por una actividad de exonucleasa apropiada u otra actividad que puede estar asociada con una polimerasa y luego el cebador es extendido por una polimerasa apropiada para crear una molécula que es complementaria a la molécula ligada.

La cadena complementaria se circulariza entonces a menudo con una actividad de una ligasa de cadena sencilla y para su posterior amplificación, detección, identificación, medición, análisis y/o evaluación de una secuencia diana y/o uno o más nucleico sitios de secuencia de ácido nucleico enlazados a la secuencia diana, mediante el uso de cebador 1 de PCR para producir una primera cadena, y un cebador 2 de PCR que es complementario a la primera cadena para la producción de una segunda cadena complementaria.

Opcionalmente, la molécula circularizada se lineariza dentro de la región complementaria a la secuencia de unión al cebador para permitir que se produzcan reacciones basadas en PCR sobre un sustrato lineal. Un ejemplo no limitativo de linearización es con el uso de endonucleasa V para un residuo de inosina dentro de la región complementaria a la secuencia de unión al cebador.

Moléculas molecularmente indexadas permiten el recuento preciso de moléculas y una determinación precisa de la secuencia en un espécimen o una muestra. Dado que cada una de las moléculas en un espécimen o una muestra tiene un índice molecular único o virtualmente único, cada una de las incidencias de un índice molecular particular resulta de la misma molécula original. La amplificación lineal o exponencial no distorsiona el número de moléculas originales. Con referencia a la Figura 5, por ejemplo, la amplificación lineal de dos moléculas indexadas molecularmente da como resultado 20 copias de cada una de las moléculas originales. Aunque posteriormente se secuencian 40 moléculas, solo se encuentran dos índices moleculares (Índice Molecular 1 e Índice Molecular 2 en este ejemplo) y, por lo tanto, se determina que solo había dos moléculas originales en la muestra o el espécimen. Además, dado que cada una de las copias de cada una de las moléculas originales es un evento independiente, se puede determinar una secuencia de consenso para cada una de las moléculas originales que revela su secuencia verdadera. En la Figura 5, por ejemplo, el Índice Molecular 1 está asociado con una molécula de tipo salvaje y el Índice Molecular 2 está asociado con una molécula mutante. Un recuento preciso similar y una determinación precisa de la secuencia son posibles para la amplificación exponencial, tal como se ilustra en la Figura 6, por ejemplo. El recuento preciso y la determinación precisa de la secuencia se pueden aplicar a la detección de diferencias genéticas, incluidas, pero no limitadas a mutaciones puntuales, sustituciones, inserciones, deleciones, inversiones, duplicaciones, variaciones en el número de copias, translocaciones, polimorfismos de un solo nucleótido (SNPs) y transcripciones por fusión. El recuento preciso y la determinación precisa de la secuencia también se pueden aplicar a la detección de moléculas que no difieren en el sitio de interés, permitiendo con ello el recuento preciso de moléculas que incluyen, pero no se limitan a especies de ARNm, ARNm_i, ARNm_{inc} y ADN. El recuento preciso de moléculas de ARN es útil para determinadas aplicaciones, tales como la determinación de un pronóstico particular para un paciente o el diagnóstico de ciertos estados de enfermedad, por ejemplo. El recuento preciso de moléculas de ADN es útil para determinadas aplicaciones, tales como la determinación de determinadas aneuploidías utilizando ácido nucleico libre de células, por ejemplo.

En determinadas realizaciones, se proporciona un método para modificar un ácido nucleico, que incluye: poner en contacto un ácido nucleico con una actividad de transferasa terminal y una mezcla de dos o más nucleótidos diferentes, en condiciones en las que los nucleótidos en la mezcla son añadidos de forma secuencial y aleatoria al extremo 3' del ácido nucleico por la actividad de la transferasa terminal, añadiendo con ello un heteropolinucleótido índice que comprende nucleótidos en la mezcla al extremo 3' del ácido nucleico y generando un ácido nucleico indexado. También se proporciona en determinadas realizaciones un método para modificar un ácido nucleico, que

incluye: poner en contacto un ácido nucleico con una actividad de transferasa terminal y una mezcla de dos o más nucleótidos diferentes en condiciones en las que los nucleótidos en la mezcla se añaden secuencial y aleatoriamente al extremo 3' del ácido nucleico por la actividad de la transferasa terminal, añadiendo con ello un heteropolinucleótido que comprende nucleótidos en la mezcla al extremo 3' del ácido nucleico y generando un ácido nucleico modificado. El ácido nucleico modificado o producido por métodos de este tipo es a veces un primer ANcs o un segundo ANcs descrito en esta memoria.

Una actividad de transferasa terminal se puede proporcionar de cualquier forma adecuada. A menudo, una enzima que cataliza la adición de nucleótidos al extremo 3' de un ácido nucleico proporciona una actividad de transferasa terminal. Una enzima de este tipo puede añadir nucleótidos al extremo 3' de un ácido nucleico (p. ej., ADN o ARN). En determinadas realizaciones, una enzima añade nucleótidos al extremo 3' de un ácido nucleico de cadena sencilla, y algunas veces añade nucleótidos al ácido nucleico de doble cadena que tiene un colgante 3', extremos romos o extremos empotrados. Una enzima que tiene actividad de transferasa terminal es a menudo una polimerasa especializada que no requiere un molde. Una enzima que tiene actividad de transferasa terminal utiliza a veces cobalto, magnesio o manganeso como un co-factor. Un ejemplo no limitante de una enzima que tiene actividad de transferasa terminal está disponible comercialmente (p. ej., New England BioLabs, nº de catálogo M0315S). A una enzima que tiene actividad de transferasa terminal se la alude a veces como desoxinucleotidil transferasa terminal (TdT) o nucleotidixlotransferasa de ADN (DNNT). Ejemplos no limitantes de otras enzimas que pueden tener actividad de transferasa terminal incluyen Q β replicasa, que tiene una actividad intrínseca de 3'-adenilación del ARN independiente del molde; POLQ, que posee una actividad de ADN polimerasa independiente del molde y permite la extensión de ADN de cadena sencilla, así como de ADN dúplex, con extremos 3'-OH de emparejamiento incorrecto sobresalientes o multiplicados; Dpo1, que muestra una actividad de desoxinucleótido transferasa terminal (TdT) competidora diferente de cualquier otra ADN polimerasa de la familia B; EcPAP (Escherichia coli PAP (EcPAP)), que comparte una estructura con las de otras ARN polimerasas independientes del molde sugiere e incluye cambios de dominio(s) fuera del dominio central catalítico conservado que altera las especificidades de sustrato de las ARN polimerasas independientes del molde; y mutantes dirigidos al sitio de Polmu humano (p. ej., Arg387 a Lys) que proporcionan una actividad de transferasa terminal potenciada. Una actividad de transferasa terminal puede añadir desoxirribonucleótidos o ribonucleótidos a los extremos 3' de ARN y ADN. Por lo tanto, a lo largo de esta divulgación, cuando se refiere a la actividad de transferasa terminal, los desoxirribonucleótidos y los ribonucleótidos pueden sustituirse entre sí. De manera similar, dNTP puede ser sustituido por NTP; y NTP puede ser sustituido por dNTP. De manera similar, dATP y ATP; dCTP y CTP; dGTP y GTP; y dTTP y TTP o dUTP; pueden ser sustituidos el uno por el otro.

Una mezcla puede incluir cualquier combinación de dos o más tipos diferentes de nucleótidos (p. ej., 2, 3 o 4 nucleótidos) que se pueden incorporar en el extremo 3' de un ácido nucleico. Una mezcla incluye a menudo dos o más desoxinucleótido trifosfatos diferentes, y los desoxinucleótido trifosfatos a veces se eligen entre adenosina trifosfato (ATP), guanosina trifosfato (GTP), citidina trifosfato (CTP) y timidina trifosfato (TTP). El ácido nucleico a veces es un primer ANcs y/o un segundo ANcs tal como se describe en esta memoria. El ácido nucleico incluye a veces un polinucleótido candidato, en donde el polinucleótido candidato comprende o consiste en una región de interés (p. ej., un nucleótido de interés, una posición de interés de nucleótido, un polinucleótido de interés o una posición de interés de polinucleótido).

Además de la adición de un Índice molecular de heteropolinucleótido (al que también se alude como un heteropolímero) a un ácido nucleico, un método incluye a veces añadir un homopolinucleótido (al que también se alude como un homopolímero) al ácido nucleico en la proximidad del índice. Un ácido nucleico modificado por un índice de heteropolinucleótido añadido y opcionalmente un homopolinucleótido añadido, se alude en esta memoria como un "ácido nucleico modificado". Un homopolinucleótido añadido a veces se enlaza directamente a un índice de heteropolinucleótido, y a veces se separa del índice de heteropolinucleótido en aproximadamente 2 a aproximadamente 50 bases de nucleótidos consecutivas. Un homopolinucleótido añadido (i) a veces sirve como un indicador, que a veces se utiliza como una referencia en la secuenciación de aplicaciones de la tecnología; (ii) a veces sirve como una marca, que a veces se utiliza como una segunda referencia en la secuenciación de aplicaciones de la tecnología; (iii) a veces sirve como un sitio de unión para un segundo ANcs que puede ser reasociado al ácido nucleico modificado, en donde el extremo 5' del segundo ANcs está ligado a veces al extremo 3' del ácido nucleico modificado y/o en donde el segundo ANcs puede extenderse y opcionalmente amplificar el ácido nucleico modificado (p. ej., amplificación lineal o amplificación exponencial); o (iv) una combinación de dos o tres de (i), (ii) y (iii).

Para realizaciones en las que un índice de homopolinucleótido y un índice de heteropolinucleótido se añaden a un ácido nucleico mediante una actividad de transferasa terminal (a lo que se alude en lo sucesivo indistintamente como un "primer elemento" y un "segundo elemento"), (a) la actividad transferasa terminal a veces se termina después de que se añada el primer elemento y la actividad de la transferasa terminal se puede introducir luego para añadir el segundo elemento (p. ej., la actividad de la transferasa terminal a veces se termina por elevación de temperatura o introducción de un desnaturizante químico); (b) a veces la actividad de la transferasa terminal no termina después de que se añada el primer elemento; (c) nucleótidos a veces se inactivan o neutralizan (p. ej.,

desfosforilan) después de que se añada el primer elemento y antes de que se añada el segundo elemento; (d) los nucleótidos a veces no se inactivan o neutralizan (p. ej., no se desfosforilan) después de que se añada el primer elemento y/o antes de que se añada el segundo elemento; (e) los nucleótidos a veces se separan del ácido nucleico (p. ej., mediante separación en fase sólida) después de que se añada el primer elemento y se introducen nuevos nucleótidos antes de que se añada el segundo elemento; (f) los nucleótidos a veces no se separan del ácido nucleico (p. ej., el ácido nucleico no se separa por separación en fase sólida) después de que se añada el primer elemento y/o antes de que se añada el segundo elemento; y (g) donde el primer elemento es un homopolinucleótido añadido antes de un segundo elemento heteropolinucleótido, la actividad de la transferasa terminal a veces no termina, y una mezcla de uno o más nucleótidos (p. ej., una mezcla de 1, 2, 3 o 4 tipos de nucleótidos) se añade después de añadir el primer elemento y antes de añadir el segundo elemento.

Un homopolinucleótido añadido consiste generalmente en una pluralidad de monómeros de un solo nucleótido (p. ej., poliA, poliG, poliT o poliC). Un homopolinucleótido añadido a veces es un homopolinucleótido en tándem que comprende un primer homopolinucleótido y un segundo homopolinucleótido directamente enlazados a o separados de (p. ej., de aproximadamente 2 a aproximadamente 10 bases de nucleótidos enlazadas consecutivamente) del primer homopolinucleótido, en donde el primer homopolinucleótido consiste en un primer nucleótido y el segundo homopolinucleótido consiste en un segundo nucleótido diferente del primer nucleótido. A veces se genera un homopolinucleótido en tándem (i) poniendo en contacto un ácido nucleico con la actividad de transferasa terminal y una primera composición que comprende monómeros de un solo nucleótido en condiciones en las que se genera el primer homopolinucleótido, y (ii) poniendo en contacto el ácido nucleico con la actividad de transferasa terminal y una segunda composición que comprende monómeros de un solo nucleótido diferentes de los monómeros de un solo nucleótido en la primera composición en condiciones en las que se genera el segundo homopolinucleótido. La actividad de la transferasa terminal se termina a veces después de la parte (i) y luego se añade a la parte (ii), y a veces la actividad de transferasa terminal no se termina entre (i) y (ii), en el método anterior. A veces, los nucleótidos se inactivan o neutralizan (p. ej., desfosforilan) y/o a veces el ácido nucleico se separa de los nucleótidos (p. ej., separación en fase sólida), después de la parte (i) en el método anterior.

Un índice de heteropolinucleótido añadido por un proceso DMI a veces es de aproximadamente 3 bases nucleotídicas consecutivas a aproximadamente 100 bases nucleotídicas consecutivas en longitud, a veces es de aproximadamente 5 bases nucleotídicas consecutivas a aproximadamente 50 bases nucleotídicas consecutivas de longitud, a veces es de aproximadamente 5 bases nucleotídicas consecutivas a aproximadamente 40 bases nucleotídicas consecutivas de longitud, a veces es de aproximadamente 5 bases nucleotídicas consecutivas a aproximadamente 35 bases nucleotídicas consecutivas de longitud, y a veces es de aproximadamente 5 bases nucleotídicas consecutivas a aproximadamente 30 bases nucleotídicas consecutivas de longitud. Un índice de heteropolinucleótido añadido por un proceso DMI a veces tiene una longitud promedio, media, mediana, nominal o máxima de aproximadamente 5 a aproximadamente 50, de aproximadamente 5 a aproximadamente 40, de aproximadamente 5 a aproximadamente 35, de aproximadamente 5 a aproximadamente 30, o de aproximadamente 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95 o 100, bases nucleotídicas consecutivas. La longitud de los índices moleculares añadidos por un proceso DMI (p. ej., de longitud promedio, media, mediana, mínima o máxima) puede predeterminarse por el período de tiempo en el que una mezcla de dos o más tipos de nucleótidos se incuban con actividad de transferasa terminal en determinadas realizaciones. Un homopolinucleótido añadido, si está presente, es a veces de aproximadamente 3 bases nucleotídicas consecutivas a aproximadamente 100 bases nucleotídicas consecutivas de longitud, a veces es de aproximadamente 5 bases nucleotídicas consecutivas a aproximadamente 50 bases nucleotídicas consecutivas de longitud, a veces es de aproximadamente 10 bases nucleotídicas consecutivas a aproximadamente 40 bases nucleotídicas consecutivas de longitud, y a veces es de aproximadamente 10 bases nucleotídicas consecutivas a aproximadamente 35 bases nucleotídicas consecutivas de longitud. Un homopolinucleótido añadido, si está presente, tiene a veces una longitud promedio, media, mediana, nominal o máxima de aproximadamente 5 a aproximadamente 50, de aproximadamente 10 a aproximadamente 40, de aproximadamente 10 a aproximadamente 35 o de aproximadamente 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95 o 100 bases nucleotídicas consecutivas. Cada uno del primer homopolinucleótido y el segundo homopolinucleótido, independientemente, en un homopolinucleótido en tándem añadido, si está presente, tiene a veces aproximadamente 3 bases nucleotídicas consecutivas a aproximadamente 50 bases nucleotídicas consecutivas de longitud, a veces es de aproximadamente 5 bases nucleotídicas consecutivas a aproximadamente 40 bases nucleotídicas consecutivas de longitud, y a veces es de aproximadamente 10 bases nucleotídicas consecutivas a aproximadamente 35 bases nucleotídicas consecutivas de longitud. Cada uno del primer homopolinucleótido y el segundo homopolinucleótido independientemente, en un homopolinucleótido en tándem añadido, si está presente, tiene a veces una longitud promedio, media, mediana, nominal o máxima de aproximadamente 5 a aproximadamente 50, de aproximadamente 5 a aproximadamente 40, de aproximadamente 10 a aproximadamente 35 o de aproximadamente 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 45 o 50 bases nucleotídicas consecutivas.

La longitud de un índice heteropolinucleótido y/o homopolinucleótido (p. ej., homopolinucleótido en tándem) añadido a ácido nucleico mediante la actividad transferasa terminal en un proceso DMI (p. ej., longitud promedio, mediana, media, nominal, mínima o máxima) se puede determinar en parte por el período de tiempo bajo el cual el ácido nucleico se pone en contacto con la actividad de transferasa terminal antes de inactivar la actividad de transferasa terminal. Se pueden utilizar diferentes períodos de tiempo, antes de inactivar la actividad de transferasa terminal, para modular la longitud del índice de heteropolinucleótido u homopolinucleótido añadido a un ácido nucleico mediante un proceso DMI. En algunas realizaciones puede determinarse un período de tiempo predeterminado adecuado para proporcionar una longitud particular (p. ej., una longitud promedio, media, mediana, nominal, mínima o máxima particular) después de testar diferentes períodos de tiempo. En realizaciones para las cuales se añade un índice de heteropolinucleótido y un homopolinucleótido a un ácido nucleico mediante la actividad de transferasa terminal, el índice de heteropolinucleótido puede añadirse independientemente durante un primer período de tiempo y el homopolinucleótido puede añadirse independientemente durante un segundo período de tiempo, en donde el primer período de tiempo puede ser igual o diferente al segundo período de tiempo. En realizaciones para las cuales un índice de heteropolinucleótido y un homopolinucleótido se añaden a un ácido nucleico mediante la actividad de transferasa terminal, la actividad de transferasa terminal a veces se inactiva y a veces no se activa entre la adición del índice de heteropolinucleótido y la adición de homopolinucleótido. En determinadas realizaciones, un período de tiempo es de aproximadamente 1 minuto a aproximadamente 120 minutos, de aproximadamente 2 minutos a aproximadamente 90 minutos, de aproximadamente 3 minutos a aproximadamente 60 minutos, de aproximadamente 5 minutos a aproximadamente 30 minutos, de aproximadamente 10 minutos a aproximadamente 30 minutos, de aproximadamente 15 minutos a aproximadamente 25 minutos o de aproximadamente 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44 o 45 minutos, antes de inactivar la actividad de transferasa terminal. La actividad de transferasa terminal se puede inactivar mediante cualquier proceso adecuado, que incluye, sin limitación, inactivación por calor (p. ej., elevar a una temperatura que inactiva la enzima transferasa terminal), inactivación química (p. ej., añadir un desnaturizante de proteínas que desnaturiza la enzima transferasa terminal), inactivar el o los nucleótidos no añadidos por la actividad de transferasa terminal y secuestrar un co-factor para la actividad de transferasa terminal.

En determinadas realizaciones, un método incluye: (a) antes de poner en contacto el ácido nucleico con la mezcla de dos o más nucleótidos diferentes, poner en contacto el ácido nucleico con una actividad de transferasa terminal y una composición que comprende monómeros de un solo nucleótido en condiciones en las que los monómeros se añaden al extremo 3' del ácido nucleico por la actividad de transferasa terminal, añadiendo con ello un homopolinucleótido indicador que comprende los monómeros al extremo 3' del ácido nucleico y generando un ácido nucleico con indicador; y (b) poner en contacto el ácido nucleico con indicador con una actividad de transferasa terminal y la mezcla de dos o más nucleótidos diferentes en condiciones en las que los nucleótidos de la mezcla se añaden secuencial y aleatoriamente al extremo 3' del ácido nucleico con indicador por la actividad de transferasa terminal, añadiendo así un índice de heteropolinucleótido que comprende nucleótidos en la mezcla al extremo 3' del ácido nucleico con indicador y generando un ácido nucleico indexado que comprende, 5' a 3', el homopolinucleótido con indicador y el heteropolinucleótido índice. En algunas realizaciones, un método incluye: (a) antes de poner en contacto el ácido nucleico con la mezcla de dos o más nucleótidos diferentes, poner en contacto el ácido nucleico con una actividad de transferasa terminal y una composición que comprende monómeros de un solo nucleótido en condiciones en las que los monómeros son añadido al extremo 3' del ácido nucleico por la actividad de transferasa terminal, añadiendo con ello un homopolinucleótido que comprende los monómeros al extremo 3' del ácido nucleico y generando un primer ácido nucleico modificado; y (b) poner en contacto el primer ácido nucleico modificado con una actividad de transferasa terminal y la mezcla de dos o más nucleótidos diferentes en condiciones en las que los nucleótidos de la mezcla se añaden secuencial y aleatoriamente al extremo 3' del primer ácido nucleico modificado por la actividad de transferasa terminal, añadiendo con ello un heteropolinucleótido que comprende nucleótidos en la mezcla al extremo 3' del primer ácido nucleico modificado y generando un segundo ácido nucleico modificado que comprende, 5' a 3', el homopolinucleótido y el heteropolinucleótido.

En algunas realizaciones, un método incluye: (a) antes de poner en contacto el ácido nucleico con la mezcla de dos o más nucleótidos diferentes, poner en contacto el ácido nucleico con una actividad de transferasa terminal y una primera composición que comprende monómeros de un solo nucleótido en condiciones en las que los monómeros se añaden al extremo 3' del ácido nucleico mediante la actividad de transferasa terminal, añadiendo con ello un homopolinucleótido de marca que comprende los monómeros al extremo 3' del ácido nucleico y generando un ácido nucleico marcado; (b) poner en contacto el ácido nucleico marcado con una actividad de transferasa terminal y la mezcla de dos o más nucleótidos diferentes en condiciones en las que los nucleótidos de la mezcla se añaden secuencial y aleatoriamente al extremo 3' del ácido nucleico con indicador por la actividad de transferasa terminal, añadiendo con ello un índice de heteropolinucleótido que comprende nucleótidos en la mezcla al extremo 3' del ácido nucleico marcado y generando un ácido nucleico indexado que comprende, 5' a 3', el homopolinucleótido con marca y el heteropolinucleótido índice; y (c) poner en contacto el ácido nucleico indexado con una actividad de transferasa terminal y una segunda composición que comprende monómeros de un solo nucleótido diferentes a los monómeros de un solo nucleótido en la primera composición en condiciones en las que los monómeros de la segunda composición se añaden al extremo 3' del ácido nucleico indexado por la actividad de transferasa terminal, añadiendo con ello un homopolinucleótido indicador que comprende los monómeros en la segunda composición al

extremo 3' del ácido nucleico indexado y generando un ácido nucleico modificado que comprende, 5' a 3', el homopolinucleótido con marca, el heteropolinucleótido índice y el homopolinucleótido indicador.

5 En determinadas realizaciones, un método incluye: (a) antes de poner en contacto el ácido nucleico con la mezcla de dos o más nucleótidos diferentes, poner en contacto el ácido nucleico con una actividad de transferasa terminal y una primera composición que comprende monómeros de nucleótido único en condiciones en las que los monómeros se añaden al extremo 3' del ácido nucleico mediante la actividad de transferasa terminal, añadiendo con ello un primer homopolinucleótido que comprende los monómeros al extremo 3' del ácido nucleico y generando un primer ácido nucleico modificado; (b) poner en contacto el primer ácido nucleico modificado con una actividad de transferasa terminal y la mezcla de dos o más nucleótidos diferentes en condiciones en las que los nucleótidos de la
10 mezcla se añaden secuencial y aleatoriamente al extremo 3' del primer ácido nucleico modificado por la actividad de transferasa terminal, añadiendo con ello un heteropolinucleótido que comprende nucleótidos en la mezcla al extremo 3' del primer ácido nucleico modificado y generando un segundo ácido nucleico modificado que comprende, 5' a 3', el primer homopolinucleótido y el heteropolinucleótido ; y (c) poner en contacto el segundo ácido nucleico modificado con una actividad de transferasa terminal y una segunda composición que comprende monómeros de un solo nucleótido diferentes a los monómeros de un solo nucleótido en la primera composición en condiciones en las que los monómeros de la segunda composición se añaden al extremo 3' del segundo ácido nucleico modificado mediante la actividad de transferasa terminal, añadiendo con ello un segundo homopolinucleótido que comprende los monómeros en la segunda composición al extremo 3' del segundo ácido nucleico modificado y generando un tercer ácido nucleico modificado que comprende, 5' a 3', el primer homopolinucleótido, el heteropolinucleótido y el
15 segundo homopolinucleótido.
20

En determinadas realizaciones, un método incluye poner en contacto el ácido nucleico indexado (o ácido nucleico modificado) con una actividad de transferasa terminal y una composición que comprende monómeros de un solo nucleótido en condiciones en las que los monómeros se añaden al extremo 3' del ácido nucleico indexado (o ácido nucleico modificado) mediante la actividad de transferasa terminal, añadiendo con ello un homopolinucleótido
25 indicador que comprende los monómeros al extremo 3' del ácido nucleico indexado (o ácido nucleico modificado) y generando un ácido nucleico con indicador que comprende 5' a 3', el heteropolinucleótido índice y el homopolinucleótido indicador. En algunas realizaciones, un método incluye: poner en contacto el ácido nucleico indexado (o ácido nucleico modificado) con una actividad de transferasa terminal y una composición que comprende monómeros de un solo nucleótido en condiciones en las que los monómeros se añaden al extremo 3' del ácido nucleico indexado (o ácido nucleico modificado) mediante la actividad de transferasa terminal, añadiendo con ello un homopolinucleótido que comprende los monómeros al extremo 3' del ácido nucleico indexado (o ácido nucleico modificado) y generando un ácido nucleico adicionalmente modificado que comprende, 5' a 3', el heteropolinucleótido y el homopolinucleótido.
30

Un segundo ANcs incluye a veces un polinucleótido de unión complementario a un polinucleótido diana en un primer ANcs al que se ha añadido un índice heteropolinucleótido. El primer ANcs incluye a veces un polinucleótido diana nativo al que puede reasociarse un polinucleótido de unión complementario. El primer ANcs incluye a veces un homopolinucleótido no nativo que ha sido añadido y sirve como un polinucleótido diana al que puede reasociarse un polinucleótido de unión complementario en el segundo ANcs.
35

Un segundo ANcs a veces está reasociado a un primer ANcs que contiene un índice de heteropolinucleótido. En algunas realizaciones, se realiza uno o más de lo siguiente después de que un primer ANcs se reasocie a un segundo ANcs: (i) ligamiento del segundo ANcs y el primer ANcs mediante una ligasa, (ii) extensión del segundo ANcs en condiciones de extensión; (iii) circularización de ácido nucleico mediante una ligasa, (iv) escisión de ácido nucleico de un ácido nucleico mediante un agente de escisión, y (v) amplificación de ácido nucleico (p. ej., amplificación lineal o amplificación exponencial). Combinaciones de dos o tres o cuatro o cinco de (i), (ii), (iii), (iv) y
40 (v) en la frase anterior se pueden realizar en cualquier orden adecuado. En determinadas realizaciones, no se realiza uno o más de lo siguiente: (i) ligamiento del segundo ANcs y el primer ANcs mediante una ligasa, (ii) circularización de ácido nucleico mediante una ligasa, y (iii) escisión de ácido nucleico de un ácido nucleico mediante un agente de escisión.
45

Para formas realizaciones en las que un heteropolinucleótido se añade índice a un ácido nucleico, y un segundo ANcs se reasocia a un primer ANcs que comprende el índice de heteropolinucleótido, el segundo ANcs se extiende a veces en condiciones de extensión. Tal como se utiliza en esta memoria, las "condiciones de extensión" permiten la adición dependiente del molde de una base nucleotídica o dos o más bases nucleotídicas consecutivas a un ácido nucleico en un complejo que a menudo es parcialmente de doble cadena y parcialmente de cadena sencilla. Las condiciones de extensión incluyen a menudo una polimerasa, una mezcla de bases nucleotídicas trifosfato y, opcionalmente, incluyen otros elementos, tales como tampón, sal y/o co-factores, por ejemplo, tal como se conoce en la técnica. Las condiciones de extensión son a veces isotérmicas, y a veces incluyen termociclos tal como se conoce en la técnica.
50
55

El ácido nucleico se amplifica opcionalmente bajo condiciones de amplificación en determinados métodos. El segundo ANcs incluye a veces al menos un polinucleótido de etiqueta de cebado que es complementario o sustancialmente complementario a un oligonucleótido cebador o una porción del mismo. El segundo ANcs incluye a veces al menos un polinucleótido de etiqueta de cebado configurado de manera que su complemento sea complementario o sustancialmente complementario a un oligonucleótido cebador o una porción del mismo. Una etiqueta de cebado puede ser útil para amplificar ácido nucleico en condiciones de amplificación tal como se describe en esta memoria. Las condiciones de amplificación son a menudo similares a las condiciones de extensión, a menudo incluyen uno o más cebadores de amplificación, a menudo permiten la amplificación isotérmica o termociclada, a veces permiten la amplificación lineal y a veces permiten la amplificación exponencial tal como se conoce en la técnica.

En algunas realizaciones, un método incluye la extensión de un segundo ANcs reasociado a un primer ANcs, y uno o más de los siguientes se lleva a cabo después de la extensión: (i) ligamiento del segundo ANcs y el primer ANcs mediante una ligasa, (ii) circularización de ácido nucleico mediante una ligasa, (iii) escisión de un ácido nucleico mediante un agente de escisión y (iv) amplificación de ácido nucleico en condiciones de amplificación. Combinaciones de dos, tres o cuatro de (i), (ii), (iii) y (iv) en la frase anterior se pueden realizar en cualquier orden adecuado. En determinadas realizaciones, no se realiza uno o más de lo siguiente: (i) ligamiento del segundo ANcs y el primer ANcs mediante una ligasa, (ii) circularización de ácido nucleico mediante una ligasa y (iii) escisión de un ácido nucleico mediante un agente de escisión.

En determinadas realizaciones, un método incluye: (a) poner en contacto un ácido nucleico con una actividad de transferasa terminal y una mezcla de dos o más nucleótidos diferentes, en condiciones en las que se añaden secuencial y aleatoriamente nucleótidos de la mezcla al extremo 3' del ácido nucleico mediante la actividad de transferasa terminal, añadiendo con ello un heteropolinucleótido que comprende nucleótidos en la mezcla al extremo 3' del ácido nucleico y generando un primer ácido nucleico modificado; (b) poner en contacto el primer ácido nucleico modificado con una actividad de transferasa terminal y una composición que comprende monómeros de un solo nucleótido en condiciones en las que los monómeros se añaden al extremo 3' del primer ácido nucleico modificado mediante la actividad de transferasa terminal, añadiendo con ello un homopolinucleótido que comprende los monómeros en el extremo 3' del primer ácido nucleico modificado y generando un segundo ácido nucleico modificado que comprende, 5' a 3', el heteropolinucleótido y el homopolinucleótido; y (c) poner en contacto el segundo ácido nucleico modificado con un segundo ácido nucleico de cadena sencilla que comprende un polinucleótido de unión complementario, o sustancialmente complementario, al homopolinucleótido del segundo ácido nucleico modificado en condiciones de extensión y, opcionalmente, en condiciones de amplificación.

En algunas realizaciones, un método incluye: (a) poner en contacto un ácido nucleico con una actividad de transferasa terminal y una mezcla de dos o más nucleótidos diferentes, en condiciones en las que se añaden secuencial y aleatoriamente nucleótidos de la mezcla al extremo 3' del ácido nucleico mediante la actividad de transferasa terminal, añadiendo con ello un heteropolinucleótido que comprende nucleótidos en la mezcla al extremo 3' del ácido nucleico y generando un primer ácido nucleico modificado; (b) poner en contacto el primer ácido nucleico modificado con una actividad de transferasa terminal y una composición que comprende monómeros de un solo nucleótido en condiciones en las que los monómeros se añaden al extremo 3' del primer ácido nucleico modificado mediante la actividad de transferasa terminal, añadiendo con ello un homopolinucleótido que comprende los monómeros en el extremo 3' del primer ácido nucleico modificado y generando un segundo ácido nucleico modificado que comprende, 5' a 3', el heteropolinucleótido y el homopolinucleótido; y (c) poner en contacto el segundo ácido nucleico modificado con un segundo ácido nucleico de cadena sencilla que comprende un polinucleótido de unión complementario o sustancialmente complementario al homopolinucleótido en el segundo ácido nucleico modificado y una actividad ligasa en condiciones en las que el extremo 3' del segundo ácido nucleico modificado y el extremo 5' del segundo ligando de ácido nucleico de cadena sencilla, generando con ello un ácido nucleico ligado.

En determinadas realizaciones, un método incluye: (a) poner en contacto un ácido nucleico con una actividad de transferasa terminal y una mezcla de dos o más nucleótidos diferentes en condiciones en las que los nucleótidos en la mezcla se añaden secuencial y aleatoriamente al extremo 3' del ácido nucleico mediante la actividad de transferasa terminal, añadiendo con ello un heteropolinucleótido que comprende nucleótidos en la mezcla al extremo 3' del ácido nucleico y generando un primer ácido nucleico modificado; (b) poner en contacto el primer ácido nucleico modificado con una actividad de transferasa terminal y una composición que comprende monómeros de un solo nucleótido en condiciones en las que los monómeros se añaden al extremo 3' del primer ácido nucleico modificado mediante la actividad de transferasa terminal, añadiendo con ello un homopolinucleótido que comprende los monómeros en el extremo 3' del primer ácido nucleico modificado y generando un segundo ácido nucleico modificado que comprende, 5' a 3', el heteropolinucleótido y el homopolinucleótido; (c) poner en contacto el segundo ácido nucleico modificado con nucleótidos, una polimerasa y un segundo ácido nucleico de cadena sencilla que comprende un polinucleótido de unión complementario o sustancialmente complementario al homopolinucleótido en el segundo ácido nucleico modificado en condiciones de amplificación lineal, generando con ello primero amplicones

y (d) poner en contacto los primeros amplicones con una actividad ligasa en condiciones en las que el extremo 3' y el extremo 5' de los primeros amplicones se ligan y se generan los primeros amplicones circularizados.

En determinadas realizaciones, el ácido nucleico lineal se puede generar a partir del ácido nucleico en forma de horquilla o circularizada. El ácido nucleico en forma de horquilla y circularizada puede generarse poniendo en contacto ácido nucleico con una actividad ligasa, tal como se conoce e ilustra en esta memoria. El ácido nucleico lineal se genera a veces a partir del ácido nucleico en forma de horquilla o circularizada mediante la escisión del ácido nucleico en horquilla o circularizado. La escisión es realizada a veces por un agente que tiene actividad de endonucleasa, y el ácido nucleico puede ser escindido por dicho agente en un sitio que comprende una base nucleotídica escindible (p. ej., desoxiinosina) o un sitio que comprende, o está cerca de un sitio de reconocimiento de endonucleasa de restricción, por ejemplo. Un agente de escisión es a menudo una enzima endonucleasa (p. ej., Endonucleasa V). En determinadas realizaciones, un sitio escindible está presente en un segundo ANcs y está localizado 5' de un polinucleótido de unión complementario a un polinucleótido en el primer ANcs (p. ej., un polinucleótido candidato (p. ej., un polinucleótido candidato nativo) o un homopolinucleótido añadido). En algunas realizaciones, un sitio escindible está presente en un segundo ANcs y está localizado 5' de un polinucleótido de unión y 3' de un polinucleótido etiqueta (es decir, entre el polinucleótido de unión y el polinucleótido etiqueta). Un segundo ANcs proporciona a veces un sitio escindible, a veces el segundo ANcs incluye un primer polinucleótido etiqueta de cebado y un segundo polinucleótido etiqueta de cebado, y el sitio escindible a veces se encuentra entre el primer polinucleótido etiqueta de cebado y el segundo polinucleótido etiqueta de cebado.

En algunas realizaciones, el ácido nucleico lineal se genera a partir de ácido nucleico en forma de horquilla o circularizada por extensión y amplificación opcional. Determinados métodos incluyen la reasociación de un cebador que comprende un polinucleótido complementario a un polinucleótido etiqueta en el ácido nucleico (p. ej., a menudo contribuido por un segundo ANcs), extendiendo el cebador y opcionalmente amplificando el ácido nucleico en condiciones de amplificación, en donde el ácido nucleico incluye una región que la polimerasa que realiza la extensión/amplificación no lee completamente. Dicha región se denomina en esta memoria una región espaciadora, la cual no puede ser atravesada por una polimerasa. A veces, un segundo ANcs proporciona una región espaciadora, a veces el segundo ANcs incluye un primer polinucleótido etiqueta de cebado y un segundo polinucleótido etiqueta de cebado, y el espaciador a veces se encuentra entre el primer polinucleótido etiqueta de cebado y el segundo polinucleótido etiqueta de cebado.

En determinadas realizaciones, se amplifica el ácido nucleico linearizado (p. ej., amplificación lineal o amplificación exponencial). El ácido nucleico linearizado, que opcionalmente se amplifica, se analiza a veces mediante un proceso de secuenciación.

En determinadas realizaciones, el ácido nucleico que ha de ser modificado por heteropolinucleótidos añadidos se proporciona como una pluralidad de moléculas de ácido nucleico, una de los heteropolinucleótidos se añade a cada una de las moléculas de ácido nucleico, y existe una posibilidad de (i) aproximadamente 4^3 a aproximadamente 4^{50} heteropolinucleótidos diferentes generados por las condiciones de reacción de la mezcla de transferasa terminal/nucleótido, (ii) aproximadamente 4^4 a aproximadamente 4^{40} diferentes heteropolinucleótidos generados por las condiciones de reacción de la mezcla de transferasa terminal/nucleótido, (iii) aproximadamente 4^4 a aproximadamente 4^{35} heteropolinucleótidos diferentes generados por las condiciones de reacción de la mezcla de transferasa terminal/nucleótido, o (iv) aproximadamente 4^5 a aproximadamente 4^{30} diferentes heteropolinucleótidos generados por las condiciones de reacción de la mezcla de transferasa terminal/nucleótido, en donde la mezcla incluye cuatro tipos de nucleótidos. En algunas realizaciones, existe la posibilidad de (i) aproximadamente 2^3 a aproximadamente 2^{50} heteropolinucleótidos diferentes generados por las condiciones de reacción de la mezcla de transferasa terminal/nucleótido, (ii) aproximadamente 2^4 a aproximadamente 2^{40} heteropolinucleótidos diferentes generados por las condiciones de reacción de la mezcla de transferasa terminal/nucleótido, (iii) aproximadamente 2^4 a aproximadamente 2^{35} heteropolinucleótidos diferentes generados por las condiciones de reacción de la mezcla de transferasa terminal/nucleótido, o (iv) aproximadamente 2^5 a aproximadamente 2^{30} heteropolinucleótidos diferentes generados por las condiciones de reacción de la mezcla de transferasa terminal/nucleótido, en donde la mezcla incluye dos tipos de nucleótidos. En determinadas realizaciones, existe la posibilidad de (i) aproximadamente 3^3 a aproximadamente 3^{50} heteropolinucleótidos diferentes generados por las condiciones de reacción de la mezcla de transferasa terminal/nucleótido, (ii) aproximadamente 3^4 a aproximadamente 3^{40} heteropolinucleótidos diferentes generados por las condiciones de reacción de la mezcla de transferasa terminal/nucleótido, (iii) aproximadamente 3^4 a aproximadamente 3^{35} heteropolinucleótidos diferentes generados por las condiciones de reacción de la mezcla de transferasa terminal/nucleótido, o (iv) aproximadamente 3^5 a aproximadamente 3^{30} diferentes heteropolinucleótidos generados por las condiciones de reacción de la mezcla de transferasa terminal/nucleótido, en donde la mezcla incluye tres tipos de nucleótidos.

De acuerdo con una longitud x promedio, media o mediana de un índice de heteropolinucleótido añadido por un proceso DMI en determinadas realizaciones existe una posibilidad de (i) aproximadamente 4^x heteropolinucleótidos diferentes generados cuando una mezcla de cuatro tipos de nucleótidos diferentes (es decir, cuatro bases nucleotídicas diferentes) se pone en contacto con una actividad de transferasa terminal en el proceso DMI; (ii)

aproximadamente 3^x heteropolinucleótidos diferentes generados cuando una mezcla de tres tipos de nucleótidos diferentes (es decir, tres bases nucleotídicas diferentes) se pone en contacto con una actividad de transferasa terminal en el proceso DMI; o (iii) aproximadamente 2^x heteropolinucleótidos diferentes generados cuando una mezcla de dos tipos de nucleótidos diferentes (es decir, dos bases nucleotídicas diferentes) se pone en contacto con una actividad de transferasa terminal en el proceso DMI. Una longitud x promedio, media o mediana de un índice de heteropolinucleótido añadido por un proceso DMI es a veces de aproximadamente 5 a aproximadamente 100, de aproximadamente 5 a aproximadamente 50, de aproximadamente 5 a aproximadamente 40, de aproximadamente 5 a aproximadamente 35, de aproximadamente 5 a aproximadamente 30 o de aproximadamente 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95 o 100 bases nucleotídicas consecutivas.

Determinados ejemplos no limitativos de los procesos DMI, y que resultan en un ácido nucleico indexado, con indicador y/o marcado se describen en el Ejemplo 6 y se ilustran en las Figuras 16 a 25 y en la Figura 28. Como se ilustra en la Figura 28, determinados procesos DMI incluyen: (a) poner en contacto un ácido nucleico con actividad de transferasa terminal y una mezcla de dos o más nucleótidos diferentes en condiciones en las que los nucleótidos en la mezcla se añaden secuencial y aleatoriamente al extremo 3' del ácido nucleico mediante la actividad de transferasa terminal, añadiendo con ello un heteropolinucleótido que comprende nucleótidos en la mezcla al extremo 3' del ácido nucleico y generando un primer ácido nucleico modificado (es decir, un ácido nucleico indexado); (b) inactivar la actividad de transferasa terminal y los nucleótidos en la mezcla después de (a) después de un primer período de tiempo predeterminado; (c) poner en contacto el primer ácido nucleico modificado con actividad de transferasa terminal y una composición que comprende monómeros de un solo nucleótido en condiciones en las que los monómeros se añaden al extremo 3' del primer ácido nucleico modificado por la actividad de transferasa terminal, añadiendo con ello un homopolinucleótido que comprende los monómeros al extremo 3' del primer ácido nucleico modificado y generando un segundo ácido nucleico modificado; (d) inactivar la actividad de transferasa terminal después de (c) después de un segundo período de tiempo predeterminado; (e) poner en contacto el segundo ácido nucleico modificado con un primer oligonucleótido cebador que puede reasociarse al homopolinucleótido y que comprende en su extremo 5' una primera secuencia, una mezcla de nucleótidos y una polimerasa en condiciones de amplificación lineal, amplificando con ello el segundo ácido nucleico modificado y generando primeros amplicones que comprenden complementos del heteropolinucleótido (es decir, índices moleculares dinámicos) y secuencias nativas originales; (f) poner en contacto los primeros amplicones con un segundo oligonucleótido cebador que comprende una segunda secuencia que se une de una manera específica para la secuencia (p. ej., específica para el locus) a los primeros amplicones, un tercer oligonucleótido cebador que comprende en su extremo 3' la primera secuencia, una mezcla de nucleótidos y una polimerasa en condiciones de amplificación exponencial, generando con ello segundos amplicones (p. ej., segundos amplicones específicos para el locus); (g) inactivar o separar los nucleótidos que no se utilizan y los oligonucleótidos cebadores que no se utilizan después de (f); y (h) determinar secuencias de los segundos amplicones. Los segundos amplicones de diferentes reacciones pueden agruparse opcionalmente después de (g) y antes de (h). En determinadas realizaciones, (a) a (g) se realizan en un único recipiente (p. ej., un solo entorno de reacción; un solo tubo). El segundo oligonucleótido cebador puede reasociarse, en algunas realizaciones, con un locus polinucleotídico en un subconjunto de los primeros amplicones (p. ej., el segundo cebador es un cebador específico para el locus) y, en algunas realizaciones, dos o más segundas especies cebadoras (es decir, cada una de las especies tiene diferentes secuencias) en una colección que se reasocian a diferentes loci de polinucleótidos se utilizan en un proceso (p. ej., en un proceso de multiplexación). A veces, el locus polinucleotídico es parte o está adyacente a una región de interés que se analizará mediante secuenciación (p. ej., el locus es una variación, mutación, polimorfismo de la secuencia). El segundo cebador en algunas realizaciones incluye uno o más polinucleótidos que facilitan la secuenciación. El tercer oligonucleótido cebador puede reasociarse, en algunas realizaciones, al complemento de todos los primeros amplicones y, a veces, se considera un cebador universal. Un cebador universal comprende a menudo la primera secuencia o consiste en la primera secuencia. El tercer oligonucleótido cebador, en algunas realizaciones, incluye uno o más polinucleótidos que facilitan la secuenciación. La secuenciación en (h) se realiza a veces mediante un proceso de secuenciación altamente multiplexada (HMS) o proceso NGS.

Los procesos DMI descritos en esta memoria proporcionan varias ventajas para indexar, identificar y cuantificar el ácido nucleico. Una ventaja es que los procesos DMI generan un alto grado de diversidad del índice molecular. Sin estar limitados por la teoría, este alto grado de diversidad del índice molecular se debe en parte a (i) la incorporación aleatoria de bases nucleotídicas de una mezcla de bases mediante la actividad de transferasa terminal, y (ii) diferentes longitudes de índice añadidas al ácido nucleico mediante la actividad de transferasa terminal en un período de tiempo dado. Ya que con los procesos DMI cada una de las moléculas originales y nativas sirve como sustrato para la síntesis del índice molecular, cada uno de los índices moleculares dinámicos es único. En la práctica, a menudo se añaden al azar de 10 a 70 o más nucleótidos a cualquier molécula nativa original particular, produciendo un intervalo de diversidades de 4^{10} ($> 10^6$) a 4^{70} ($> 10^{42}$) o más DMIs, con una diversidad promedio de $\sim 4^{40}$ ($> 10^{24}$) que supera con creces la diversidad de índices moleculares de otros métodos e incluso es ligeramente más que el número de nucleótidos en el cuerpo humano.

El etiquetado directo de ácido nucleico de la muestra, y el alto grado de diversidad del índice molecular son características que hacen que procesos DMI sean ventajosos para detectar y cuantificar alelos relativamente raros en una muestra de ácido nucleico. Los procesos de secuenciación masiva paralela o "secuenciación NexGen" (NGS) se utilizan a menudo para analizar muestras. Sin embargo, en los procesos de NGS, no es posible a menudo determinar qué moléculas son originales, nativas y cuáles son copias, ya que se produce una cierta amplificación antes de la separación espacial de las moléculas. Además, la precisión de NGS es ~ 99,7%, lo que significa una tasa de error de ~ 0,3% o 1 falta en la secuencia por cada 333 bases secuenciadas. Estos dos problemas con los procesos NGS combinados pueden hacer que la detección de secuencias raras, p. ej., la detección de mutaciones de genes del cáncer en un contexto elevado de secuencias de tipo salvaje, sea problemática y, por lo tanto, su cuantificación precisa no sea fiable. En la práctica comercial, el límite de detección de mutación o frecuencia de alelos menores generalmente no es mejor del 2% al 3%. Los procesos DMI abordan estos problemas etiquetando directamente cada una de las moléculas nativas con un marcador único antes de que se copie. La actividad de transferasa terminal utilizada en los procesos de DMI añade DMI directamente al ácido nucleico a analizar, y el alto grado de diversidad de etiquetas potencia la probabilidad de que se añada una etiqueta única al ácido nucleico a analizar.

Otra ventaja proporcionada por la alta diversidad del índice molecular proporcionada por procesos DMI es la identificación de contaminación de la muestra. Incluso en entornos de alto rendimiento, no sería de esperar ver los mismos DMI en dos muestras. Un evento de este tipo puede ser la causa de buscar contaminación de una muestra a otra. Con una diversidad mucho más baja vista con otros métodos, sería de esperar el mismo índice molecular en múltiples muestras de forma rutinaria.

Otras ventajas son que las realizaciones del proceso DMI no hacen uso de etiquetas de índice de oligonucleótidos pre-sintetizadas (PSOTs). Los índices moleculares dinámicos (DMIs) añadidos por los procesos DMI no se pre-sintetizan porque la actividad de transferasa terminal añade bases de nucleótidos de una vez directamente al ácido nucleico a analizar y con ello forma secuencialmente diversos DMIs in situ. El uso obvio de las PSOTs ofrece varias ventajas. Ya que los DMI no se pre-sintetizan, no existen oligonucleótidos para hacer, no hay problemas de garantía de calidad o control de calidad con respecto a los mismos, no hay variación de un lote a otro, inventarios o gastos asociados. Las PSOTs se crean sobre un soporte sólido, lo que da como resultado muchas copias idénticas, e incluso tan poco como un picomol de una PSOT particular consiste en $6,022 \times 10^{11}$ moléculas, creando una agrupación prácticamente no agotadora. Los procesos que utilizan PSOTs proporcionan típicamente una diversidad de etiquetas significativamente menor que los procesos DMI. Por ejemplo, determinados procesos PSOT proporcionan una diversidad promedio de solo 10^6 , o menos de 10^4 , en comparación con una diversidad promedio de 10^{24} proporcionada por los procesos DMI.

Las PSOTs se fijan a menudo al ácido nucleico mediante el ligamiento de adaptadores o mediante la extensión del cebador, y los procesos DMI proporcionan ventajas sobre cada una de estas metodologías. Un método comercialmente disponible que utiliza PSOT ("Método A") utiliza un código de barras molecular que consiste en una región de base degenerada (DBR) de 10 nucleótidos. Aunque la DBR produce teóricamente una diversidad de aproximadamente 10^6 combinaciones, requiere adaptadores de doble cadena específicos para la diana con costos adicionales. El uso de oligonucleótidos pre-sintetizados también significa que no existen índices moleculares verdaderamente únicos: como se describe arriba, incluso tan poco como un picomol de un índice molecular particular aplicado a una muestra consiste en $6,022 \times 10^{11}$ moléculas, creando una agrupación virtualmente no agotadora. Ambos métodos también requieren que se retiren los adaptadores no utilizados más un cierto número de cambios de tampón y transferencias de tubos, todo lo cual aumenta los costos, la pérdida de material y aumenta el tiempo y la mano de obra necesarios. Otro método comercialmente disponible que utiliza PSOTs ("Método B") hace uso de un conjunto definido de 96 adaptadores de doble cadena, cada uno con un único colgante de nucleótido dT 3' para facilitar el ligamiento, que puede ser ineficiente. Esto requiere no solo la síntesis bien controlada de oligonucleótidos antes mencionada (192 de ellos), sino también su reasociación por pares para formar los 96 adaptadores. Este número limitado de adaptadores puede crear una diversidad máxima de solo 9.216 combinaciones, por lo tanto, las muestras que contienen más de ese número de moléculas no pueden contarse con precisión. Por consiguiente, los procesos DMI proporcionan un mayor grado de diversidad de etiquetas con respecto a los métodos PSOT, y tampoco existen problemas para separar el ligamiento ineficiente o cualquier adaptador no utilizado antes de la secuenciación (por ejemplo, NGS).

En lugar de utilizar adaptadores, las PSOTs se pueden fijar mediante extensión de cebador, por ejemplo, durante la PCR o, en el caso del ARN, durante la síntesis de ADNc. En estos casos, se crean copias de las moléculas de ácido nucleico durante la adición de índices moleculares, y las moléculas nativas originales nunca se marcan realmente. Por ejemplo, si los ADNc de la primera cadena se hacen utilizando un cebador indexado molecularmente, el molde de ARN original nunca se indexa molecularmente, y solo se indexa el ADNc de la primera cadena. Lo mismo ocurre si se añade un índice molecular mediante el cambio de molde o durante la creación específica de genes de un ADNc de la segunda cadena. Si se cometen errores durante estas etapas, estos se propagarán a lo largo del análisis. Si el análisis de ADNc se dirige a detectar cambios raros en la secuencia codificante (p. ej., detección de mutaciones), estos enfoques no DMI pueden comprometer los resultados. En los métodos en los que ARN de poliA+ es el material

de partida, no se incluirán ARN de poliA- incluido ARNm_i y la mayoría de los fragmentos de ARN. Por lo tanto, una ventaja de los procesos DMI es que, en contraste con los procesos que fijan PSOTs por extensión del cebador (p. ej., durante la síntesis de ADN_c de primera o segunda cadena o PCR fijada como objetivo), el DMI siempre se añade a la molécula nativa original, no a una mera copia, eliminando así una fuente de errores.

- 5 Otra ventaja de los procesos DMI es que pueden utilizarse de manera flexible en una amplia diversidad de muestras de ácido nucleico. Los procesos DMI se pueden realizar en condiciones idénticas para añadir índices moleculares al ADN de cadena sencilla, ADN de doble cadena, ARN de cadena sencilla o ARN de doble cadena o cualquier mezcla de moléculas de ácido nucleico. Se pueden utilizar ribonucleótidos o desoxirribonucleótidos para crear DMIs. El sustrato puede tener un extremo 3' empotrado, sobresaliente o romo, y puede ser de cualquier longitud.
- 10 Otra ventaja es que determinadas realizaciones del proceso DMI se pueden realizar en un solo recipiente (p. ej., recipiente de reacción único, tubo único). Otras ventajas es que los procesos DMI reducen el número de manipulaciones del usuario en comparación con determinados procesos que utilizan PSOTs. Los procesos DMI también proporcionan ventajas de reducir el tiempo de incubación, reducir el costo por muestra y ser adecuados para analizar el ácido nucleico libre de células o el ácido nucleico libre de células circulantes (p. ej., ADN libre de células (ADN_{cf} ; p. ej., de sangre, orina o saliva)) en comparación con determinados métodos basados en PSOT.
- 15

Determinadas ventajas de una realización de un proceso DMI ilustrada en la Figura 28 y descrita en el Ejemplo 6 sobre determinados métodos PSOT se resumen en la siguiente tabla.

	DMI	PSOT Método A	PSOT Método B
Diversidad promedio	10 ²⁴	10 ⁶	menos de 10 ⁴
Adaptadores	No	Sí	Sí
Tubo único	Sí	No	No
Manipulaciones (añadir/separar/transferir)	6 (6/0/0)	56 (34/8/14)	44 (24/15/5)
Incubaciones (aproximadas)	4 h	4,5 h a 18,5 h	6 h
Adecuado para AN _{cf}	+++	No	+
Costo/muestra (estimado)	inferior	mayor	Mayor

Kits

- 20 La divulgación incluye kits para el desempeño de todos o parte de los métodos descritos y/o para la producción de los productos descritos. Los kits comprenden a menudo uno o más recipientes que contienen uno o más componentes descritos en esta memoria. Un kit puede incluir uno o más componentes en cualquier cantidad de recipientes, paquetes, tubos, viales, placas de múltiples pocillos y similares, o los componentes pueden combinarse en diversas combinaciones en dichos recipientes.

- 25 En algunas realizaciones, uno o más de los siguientes componentes pueden incluirse en un kit: (i) una molécula que puede proporcionar una actividad de transferasa terminal (p. ej., una enzima); (ii) una mezcla que comprende dos o más nucleótidos diferentes (p. ej., diferentes desoxinucleótidos trifosfato; dos o más de dATP, dGTP, dCTP y dTTP); (iii) una sal; (iv) un tampón; (v) un co-factor de la enzima transferasa terminal (p. ej., cobalto, manganeso, magnesio); (vi) una composición que comprende monómeros de un solo nucleótido (p. ej., un tipo de desoxinucleótido trifosfato; uno de dATP, dGTP, dCTP o dTTP); (vii) un segundo AN_cs descrito en esta memoria (p. ej., un segundo AN_cs que comprende un polinucleótido de unión y al menos un polinucleótido de etiqueta de cebado); (viii) una polimerasa y otros componentes adecuados para la extensión de ácidos nucleicos (p. ej., mezcla de desoxinucleótidos); (ix) una polimerasa y otros componentes adecuados para la amplificación de ácidos nucleicos (p. ej., mezcla de desoxinucleótidos y cebador(es)); (x) una endonucleasa y otros componentes adecuados para escindir un ácido nucleico (p. ej., Endonucleasa V); y (xi) ligasa de ácido nucleico de cadena sencilla y otros componentes para ligar extremos de ácido nucleico de cadena sencilla.
- 30
- 35

- 40 En determinadas realizaciones, uno o más de los siguientes componentes pueden incluirse en un kit: (i) un oligonucleótido de cadena sencilla que comprende un resto 5'-fosfato (opcionalmente adenilado), una secuencia conductora 5', una secuencia de unión a diana que es complementaria a al menos una porción de una secuencia diana en un primer AN_cs y un extremo 3', tal como se describe en esta memoria; (ii) una enzima ligasa de cadena sencilla (por ejemplo, T4 ARN Ligasa 1; o una enzima termoestable; una ligasa de cadena sencilla de *M. thermoautotrophicum*); (iii) una polimerasa y otros componentes adecuados para la extensión de ácido nucleico (p. ej., mezcla de desoxinucleótidos); (iv) una polimerasa y otros componentes adecuados para la amplificación de ácidos nucleicos (p. ej., mezcla de desoxinucleótidos y cebador(es)); y (x) una endonucleasa y otros componentes adecuados para escindir un ácido nucleico (p. ej., Endonucleasa V).

Un kit se utiliza a veces junto con, o para llevar a cabo un método descrito en esta memoria, y puede incluir instrucciones o descripciones para realizar uno o más métodos y/o una descripción de una o más composiciones descritas en esta memoria. Un kit puede incluir una descripción de una ubicación de Internet que proporcione dichas instrucciones y/o descripciones (p. ej., una URL para la World-Wide-Web). Las instrucciones y/o descripciones pueden estar en forma tangible (p. ej., papel y similares) o en formato electrónico (p. ej., archivo legible por computadora en un medio tangible (p. ej., disco compacto) y similares) y/o pueden incluirse en una inserción del kit, en algunas realizaciones.

Un kit puede incluir una tabla de conversión, tabla de búsqueda, software, instrucciones ejecutables y/o una ubicación de Internet que proporcione lo anterior, en determinadas realizaciones. Se puede utilizar una tabla de conversión, tabla de búsqueda, software y/o instrucciones ejecutables para determinar la configuración de la reacción, manipular y/o interpretar datos resultantes de moléculas de ANcs individuales y/o moléculas de ANcs digeridas descritas en esta memoria.

Habiendo proporcionado ahora generalmente la divulgación, la misma se entenderá más fácilmente mediante la referencia a los siguientes ejemplos que se proporcionan a modo de ilustración, y no pretenden limitar la divulgación, a menos que se especifique.

Ejemplos

Ejemplo 1: Ligamiento de un segundo ANcs a una secuencia diana

Se testó un segundo ANcs con una secuencia de unión a diana contra una secuencia diana BRAF humana sintética para determinar su capacidad para ligarse correctamente a su diana. La Figura 9 ilustra el ligamiento correcto seguido de la digestión con UDG de la molécula ligada a una longitud prevista de 80 nucleótidos (nt).

El segundo ANcs se identifica como "Etiqueta Diana" y el procesamiento con UDG se indica en la parte superior. La pista 1 es Etiqueta Diana solo, mientras que los marcadores de peso molecular se identifican por la "M" en la pista 6. Las pistas 2-3 y 4-5 son duplicados utilizando una diana sintética con un producto ligado de 80 nt predicho. Las pistas 7-8 utilizaron una diana sintética con un producto ligado de 70 nt predicho. Para ser claros, la Pista 4 es un duplicado de la Pista 2; estas pistas muestran los productos ligados antes del tratamiento con UDG. La pista 5 es un duplicado de la pista 3; estas pistas muestran los productos ligados después del tratamiento con UDG que dan como resultado productos predichos de 80 nt. La pista 7 muestra un producto ligado antes del tratamiento con UDG y la pista 8 muestra un producto después del procesamiento con UDG con un producto predicho de 70 nt en este caso.

Ejemplo 2: Las segundas moléculas de ANcs son específicas para sus secuencias diana correspondientes

En la Figura 10 se testó una segunda molécula de ANcs con un sitio de unión a diana BRAF utilizando un diana BRAF (DS-B600) frente a cuatro dianas EGFR (US-E790, DS-E790 US-E858 y DS-E858) diferentes; dos dianas KRAS diferentes (US-K12 y DS-K12); y dos dianas PIK3CA diferentes (US-P542 y DS-P542). El ligamiento y el procesamiento (UDG) de la segunda molécula de ANcs a los dos dianas BRAF da los productos predichos de 80 nt o 70 nt tal como se muestra en las partes superior e inferior de la Figura 10, respectivamente. Ninguna de las otras dianas proporciona productos ligados con la molécula ANcs específica para BRAF. Las reacciones se realizaron por duplicado. US designa un diana aguas arriba del número de codón designado; DS designa la diana aguas abajo. Los marcadores de peso molecular se indican con una "M".

En un ensayo adicional, se testaron moléculas de ANcs con secuencias de unión a diana específicas para BRAF o PIK3CA. Cada uno de los ANcs solo se ligó a su diana afín tal como se muestra en la Figura 11.

Ejemplo 3: Uso multiplex de moléculas de ANcs

Una reacción que contiene cinco (5) moléculas de ANcs con secuencias de unión a diana específicas para una secuencia diana BRAF (DS-B600), una secuencia diana PIK3CA (US-P542), una secuencia diana KRAS (US-K12) y dos secuencias diana EGFR diferentes (DS-R790 y US-E858) se testaron contra una de las cinco secuencias diana. Los resultados se muestran en la Figura 12.

Cada uno de los cinco productos ligados correctamente se produjo para cada una de las dianas. Los puntos indican productos de los tamaños predichos cuando fueron completamente procesados mediante UDG. Se muestran reacciones duplicadas. El control no es una diana añadida, y M son marcadores de peso molecular.

Ejemplo 4: El ligamiento es estable a la temperatura

Un ensayo de las condiciones de temperatura demuestra que la molécula de ANcs con la secuencia de unión a diana específica para BRAF se puede ligar a su diana ANcs en un intervalo de temperaturas de al menos 60 a 75

grados Celsius. Esto permite un intervalo de temperaturas de fusión para la selección de secuencias de unión a diana en una molécula de ANcs para hibridación a una secuencia diana tal como se muestra en la Figura 13.

Ejemplo 5: Amplificación específica para moléculas ligadas

5 Un gradiente de temperaturas de la PCR demostró que el producto ligado de un ANcs específico para BRAF puede amplificarse en un intervalo de temperaturas de al menos 55 a 72 grados Celsius cuando se añade a un fondo de ADN genómico humano, tal como se ilustra en la Figura 14A. Tal como se ilustra en la Figura 14B, el control del ADN genómico humano sin el producto ligado no produjo material de producto amplificado de al menos 65,6 a 72 grados Celsius.

Ejemplo 6: Procesos de Indexación Molecular Dinámica

10 Ejemplos no limitantes de determinados procesos de Indexación Molecular Dinámica (DMI) se ilustran en las Figuras 16 a 25 y en la Figura 28. La Figura 16 ilustra la Indexación Molecular Dinámica (DMI) de un primer ANcs utilizando transferasa terminal y una mezcla de nucleótidos. El primer ANcs se reasocia a un segundo ANcs de una manera específica para el locus. En presencia de una ligasa de ANcs apropiada, el extremo 3' del primer ANcs y el extremo 5' del segundo ANcs están ligados entre sí. Se puede añadir y extender un cebador complementario a la Etiqueta. Se pueden hacer una o dos o tres o más copias en condiciones de amplificación (p. ej., amplificación lineal o amplificación exponencial). Se puede añadir y extender un cebador específico de la Secuencia Diana para crear ácidos nucleicos de doble cadena. Ciclos adicionales resultan en una amplificación exponencial. Los cebadores pueden tener secuencias adicionales en sus extremos 5'. Se pueden utilizar diferentes combinaciones de los extremos 5' y 3' del primer ANcs y el segundo ANcs tal como se describe en la Tabla 2 según sea apropiado para el ligamiento.

25 La Figura 17 ilustra la DMI de un primer ANcs utilizando transferasa terminal y una mezcla de nucleótidos. El primer ANcs es luego reasociado a un segundo ANcs de una manera específica para el locus. En presencia de una ligasa de ANcs apropiada, el extremo 3' del primer ANcs y el extremo 5' del segundo ANcs se ligan para formar un nuevo ANcs. En este ejemplo no limitativo, el producto se lineariza con endonucleasa V, que reconoce desoxiinosina (In) y escinde el segundo y/o tercer enlace fosfodiéster 3' a desoxiinosina, dejando un 3'-hidroxilo y 5'-fosfato. Después de la linearización, el nuevo extremo 3' se une intramolecularmente al extremo 5' del nuevo ANcs en condiciones que favorecen la circularización mediante una ligasa de ANcs apropiada. De esta manera, ambos extremos del nuevo ANcs están etiquetados. El o los sitios de interés pueden estar aguas arriba y/o aguas abajo de la secuencia Diana. Se puede añadir cebador 2, que es complementario a la Etiqueta 2, y se pueden hacer una o dos o tres o más copias. El cebador 1, que reconoce el complemento de Etiqueta 1 (cEtiqueta1), puede añadirse y extenderse para crear ácidos nucleicos de doble cadena y ciclos adicionales darían como resultado una amplificación exponencial. Un espaciador (Sp) que no puede ser atravesado por la polimerasa utilizada puede insertarse entre la Etiqueta 1 y la Etiqueta 2 de modo que solo resulten copias lineales de la acción de la polimerasa. Se pueden utilizar diferentes combinaciones de los extremos 5' y 3' del primer ANcs y el segundo ANcs tal como se describe en la Tabla 2, según sea apropiado para cada ligamiento. Los cebadores pueden tener secuencias adicionales en sus extremos 5'.

40 La Figura 18 muestra cómo se puede realizar la DMI utilizando transferasa terminal y un solo nucleótido, dando como resultado un homopolímero de longitud variable fijado al primer ANcs. Opcionalmente, se pueden añadir uno, dos, o tres tipos adicionales de nucleótidos, cada uno diferente del primer nucleótido único, para crear una mezcla de dos, o tres o cuatro nucleótidos diferentes, lo que resulta en un índice molecular más complejo y diversificado de lo que es posible con un solo nucleótido. El homopolímero (en este ejemplo, poli dA) también puede servir como un Sitio Diana para el segundo ANcs (en este ejemplo, con una Secuencia de Unión a Diana oligo dT). Después de la reasociación y en presencia de una ligasa de ANcs apropiada, el extremo 3' del primer ANcs y el extremo 5' del segundo ANcs se ligan juntos. Se puede añadir y extender un cebador complementario a la Etiqueta. Se pueden hacer una o dos o tres o más copias. Si se incluye un nucleótido escindible (p. ej., un residuo desoxiinosina) en el segundo ANcs, se puede añadir un agente de escisión (p. ej., Endonucleasa V) antes de la extensión para separar el resto oligo dT utilizado en este ejemplo. Se puede añadir y extender un cebador específico para el locus para crear ácidos nucleicos de doble cadena y ciclos adicionales darían como resultado una amplificación exponencial. Los cebadores pueden tener secuencias adicionales en sus extremos 5'. Se pueden utilizar diferentes combinaciones de los extremos 5' y 3' del primer ANcs y el segundo ANcs como se describe en la Tabla 2 según sea apropiado para el ligamiento.

55 La Figura 19 muestra cómo se puede realizar la DMI con la transferasa terminal y un solo nucleótido, lo que da como resultado un homopolímero de longitud variable que se une al primer ANcs. Luego se añaden tres tipos adicionales de nucleótidos que dan como resultado un índice molecular complejo y diversificado. El homopolímero (en este ejemplo, poli dA) sirve como un Sitio Diana para el segundo ANcs (en este ejemplo, con una Secuencia de Unión a Diana oligo dT). Después de la reasociación y en presencia de una ligasa de ANcs apropiada, el extremo 3' del primer ANcs y el extremo 5' del segundo ANcs se ligan juntos. Después de la linearización, el nuevo extremo 3' se une intramolecularmente al extremo 5' del nuevo ANcs en condiciones que favorecen la circularización mediante una

ligasa de ANcs apropiada. De esta manera, ambos extremos del nuevo ANcs están etiquetados. Se puede añadir el cebador 2, que es complementario a la Etiqueta 2, y se pueden hacer una o dos o tres o más copias. El cebador 1, que reconoce el complemento de Etiqueta 1 (cEtiqueta1), puede añadirse y extenderse para crear ácidos nucleicos de doble cadena y ciclos adicionales darían como resultado una amplificación exponencial. Un espaciador (Sp), que no puede ser atravesado por la polimerasa utilizada, puede insertarse entre la Etiqueta 1 y la Etiqueta 2, de modo que solo resulten copias lineales de la acción de la polimerasa. Se pueden utilizar diferentes combinaciones de los extremos 5' y 3' del primer ANcs y el segundo ANcs tal como se describe en la Tabla 2, según sea apropiado para cada uno de los ligamientos. Los cebadores pueden tener secuencias adicionales en sus extremos 5'.

La Figura 20 muestra cómo se puede realizar la DMI con transferasa terminal y un solo nucleótido, dando como resultado un homopolímero (en este ejemplo, poli dA). Después de que se neutraliza el dATP en la reacción, se añade un segundo nucleótido único para crear un Homopolímero en Tándem (en este ejemplo, poli dA/dT). Luego se añaden dos tipos adicionales diferentes de nucleótidos que dan como resultado un índice molecular complejo y diversificado. El Homopolímero en Tándem sirve como un Sitio Diana para el segundo ANcs (en este ejemplo, con una Secuencia de Unión a Diana oligo dT/dA). Después de la reasociación y en presencia de una ligasa de ANcs apropiada, el extremo 3' del primer ANcs y el extremo 5' del segundo ANcs se ligan juntos. Se puede añadir y extender un cebador complementario a la Etiqueta. Se pueden hacer una o dos o tres o más copias. Si se incluye un nucleótido escindible (p. ej., residuo desoxiinosina) en el segundo ANcs, entonces se puede añadir un agente de escisión (p. ej., Endonucleasa V) antes de la extensión para separar el resto oligo dT utilizado en este ejemplo. Se puede añadir y extender un cebador específico para el locus para crear ácidos nucleicos de doble cadena y ciclos adicionales darían como resultado una amplificación exponencial. Se pueden utilizar diferentes combinaciones de extremos 5' y 3' del primer ANcs y el segundo ANcs tal como se describe en la Tabla 2, según sea apropiado para cada uno de los ligamientos. Los cebadores pueden tener secuencias adicionales en sus extremos 5'.

La Figura 21 muestra cómo se puede realizar la DMI con transferasa terminal y un solo nucleótido, dando como resultado un homopolímero (en este ejemplo, poli dA). Después de que el dATP en la reacción se neutraliza, se añade un segundo nucleótido único para crear un Homopolímero en Tándem (en este ejemplo, poli dA/dT). Luego se añaden dos tipos adicionales diferentes de nucleótidos que dan como resultado un índice molecular complejo y diversificado. El Homopolímero en Tándem sirve como un Sitio Diana para el segundo ANcs (en este ejemplo, con una Secuencia de Unión a Diana oligo dT/dA). Después de la reasociación y en presencia de una ligasa de ANcs apropiada, el extremo 3' del primer ANcs y el extremo 5' del segundo ANcs se ligan juntos. Después de la linearización, el nuevo extremo 3' se une intramolecularmente al extremo 5' del nuevo ANcs en condiciones que favorecen la circularización mediante una ligasa de ANcs apropiada. De esta manera, ambos extremos del nuevo ANcs están etiquetados. Se puede añadir cebador 2, que es complementario a Etiqueta 2, y se pueden hacer una o dos o tres o más copias. El cebador 1, que reconoce el complemento de Etiqueta 1 (cEtiqueta1), puede añadirse y extenderse para crear ácidos nucleicos de doble cadena y ciclos adicionales darían como resultado una amplificación exponencial. Un espaciador (Sp) que no puede ser atravesado por la polimerasa utilizada puede insertarse entre Etiqueta 1 y Etiqueta 2 de modo que solo resulten copias lineales de la acción de la polimerasa. Se pueden utilizar diferentes combinaciones de los extremos 5' y 3' del primer ANcs y el segundo ANcs tal como se describe en la Tabla 2, según sea apropiado para cada uno de los ligamientos. Los cebadores pueden tener secuencias adicionales en sus extremos 5'.

La Figura 22 muestra cómo se puede realizar la DMI con la transferasa terminal y una mezcla de nucleótidos, dando como resultado un índice molecular complejo y diversificado. Después de que se neutraliza la mezcla de dNTP en la reacción, se añade un solo nucleótido para crear una cola de Homopolímero (en este ejemplo, poli dA) 3' del Índice Molecular. Se añade y extiende un cebador complementario a la cola poli dA (en este ejemplo, un oligo dT) con una extensión 5' con la Etiqueta 1. Se pueden hacer una o dos o tres o más copias. Se puede añadir y extender un cebador específico para el locus para crear ácidos nucleicos de doble cadena y ciclos adicionales darían como resultado una amplificación exponencial. Los cebadores pueden tener secuencias adicionales en sus extremos 5'.

La Figura 23 muestra cómo se puede realizar la DMI con la transferasa terminal y una mezcla de nucleótidos, dando como resultado un índice molecular complejo y diversificado. Después de que se neutraliza la mezcla de dNTP en la reacción, se añade un solo nucleótido para crear una cola de Homopolímero (en este ejemplo, poli dA) 3' del Índice Molecular. Se añade y extiende un cebador complementario a la cola poli dA (en este ejemplo, un oligo dT) con una extensión 5' que contiene Etiqueta 1, un residuo de desoxiinosina y Etiqueta 2. Se pueden hacer una o dos o tres o más copias. Los productos de extensión pueden circularizarse utilizando Ligasa de ANcs, y el ácido nucleico puede tener diferentes combinaciones de extremos 5' y 3' tal como se describe en la Tabla 2, por ejemplo. El ANcs circularizado se puede linearizar y, en un ejemplo no limitante, la desoxiinosina se escinde por la Endonucleasa V. Se puede incluir un espaciador que no puede ser atravesado por las polimerasas en el lado 5' de la inserción, y en estos casos, la linearización no es necesaria. El uso de cebadores apropiados para Etiqueta 1 y Etiqueta 2 se puede utilizar para crear ácidos nucleicos de doble cadena y ciclos adicionales darían como resultado una amplificación exponencial. Los cebadores pueden tener secuencias adicionales en sus extremos 5'.

La Figura 24 muestra cómo se puede realizar la DMI con transferasa terminal y una mezcla de nucleótidos, dando como resultado un índice molecular complejo y diversificado. Después de que se neutraliza la mezcla de dNTP en la reacción, se añade un solo nucleótido para crear una cola de Homopolímero (en este ejemplo, poli dA) 3' del Índice Molecular. Después de que se neutraliza el dATP en la reacción, se añade un segundo nucleótido único para crear un Homopolímero en Tándem (en este ejemplo, poli dA/dT). Se añade y extiende un cebador complementario a la cola poli dA/dT (en este ejemplo, un oligo dT/dA) con una extensión 5' que contiene Etiqueta 1. Se pueden hacer una o dos o tres o más copias. Se puede añadir y extender un cebador específico para el locus para crear ácidos nucleicos de doble cadena y ciclos adicionales darían como resultado una amplificación exponencial. Los cebadores pueden tener secuencias adicionales en sus extremos 5'.

La Figura 25 muestra cómo se puede realizar la DMI con transferasa terminal y una mezcla de nucleótidos, dando como resultado un índice molecular complejo y diversificado. Después de que se neutraliza la mezcla de dNTP en la reacción, se añade un solo nucleótido para crear una cola de Homopolímero (en este ejemplo, poli dA) 3' del Índice Molecular. Después de que se neutraliza el dATP en la reacción, se añade un segundo nucleótido único para crear un Homopolímero en Tándem (en este ejemplo, poli dA/dT). Se añade y extiende un cebador complementario a la cola de poli dA/dT (en este ejemplo, un oligo dT/dA) con una extensión 5' que contiene Etiqueta 1, un residuo desoxiinosina y Etiqueta 2. Se pueden hacer una o dos o tres o más copias. Los productos de extensión pueden circularizarse utilizando Ligasa de ANCs y el ácido nucleico puede incluir diferentes combinaciones de extremos 5' y 3', tal como se describe en la Tabla 2, por ejemplo. El ANCs circularizado puede ser linearizado, y en un ejemplo no limitante, la desoxiinosina es escindida por la Endonucleasa V. Un espaciador que no puede ser atravesado por polimerasas puede incluirse en el lado 5' de la inserción, y en tales casos, la linearización no es necesaria. El uso de cebadores apropiados para Etiqueta 1 y Etiqueta 2 se puede utilizar para crear ácidos nucleicos de doble cadena y ciclos adicionales darían como resultado una amplificación exponencial. Los cebadores pueden tener secuencias adicionales en sus extremos 5'.

La Figura 28 muestra un ejemplo no limitante de un proceso DMI. Un proceso de este tipo es útil para etiquetar ácido nucleico fragmentado, tal como ácido nucleico libre de células (p. ej., de sangre, saliva u orina), por ejemplo, en un único recipiente. En el proceso DMI ilustrado, las mismas condiciones utilizadas para añadir la etiqueta de polinucleótido DMI también se pueden utilizar para añadir un Sitio de Cebado OligodT y, a través de él, una Secuencia de Cebador Universal. Con referencia a la Figura 28, se configura una reacción DMI en la Etapa 1 combinando en un solo tubo ADN purificado, tampón de Transferasa Terminal (incluido CoCl_2), una mezcla de dNTP que contiene los cuatro dNTPs (es decir, dATP, dTTP, dCTP y dGTP) y Transferasa Terminal (TdT). Esta mezcla se incuba luego durante 30 minutos a 37°C , y luego la TdT se inactiva calentando a 75°C durante 20 minutos. Este proceso en la Etapa 1 genera una gran diversidad de DMIs. Ahora que el ADN está etiquetado con DMI, puede utilizarse en una diversidad de aplicaciones posteriores. En el método que se muestra en la Figura 28, los dNTPs son inactivados por la Fosfatasa Alcalina de Camarón (SAP) en la Etapa 2. Después de que la SAP es a su vez inactivada por calor, se puede añadir una cola poliA en la etapa 3 añadiendo al mismo tubo dATP y TdT fresca seguido de incubación durante 30 minutos a 37°C . La TdT se inactiva luego calentando a 75°C durante 20 minutos. Esto crea una cola poliA en todas las moléculas de ADN etiquetadas con DMI. En la Etapa 4, la amplificación lineal se logra añadiendo un cebador OligodT que tiene en su extremo 5' una Secuencia Universal de Cebador, dNTPs, una mezcla de ajuste de tampón para modificar las concentraciones de cationes y una ADN polimerasa, en este ejemplo, ADN polimerasa Phusion HS II. Dado que cada uno de los eventos de amplificación lineales es un hecho independiente, se puede derivar la secuencia nativa correcta y original más adelante al comparar todas las secuencias para cada una de las DMI particulares. En la Etapa 5, la amplificación exponencial se emplea para amplificar preferentemente dianas seleccionados, p. ej., una región particular de un oncogén. Esta amplificación preferencial se realiza añadiendo un cebador específico para el locus (p. ej., aguas arriba del codón 600 del oncogén BRAF) y un Cebador Universal que incluye la Secuencia del Cebador Universal que se adjuntó en la Etapa 4. La amplificación puede tener lugar a una temperatura estricta compatible tanto con el cebador específico para el locus como el Cebador Universal. Téngase en cuenta que, dado que solo se utiliza un cebador específico para el locus, la "huella" del ensayo (es decir, la cantidad real de secuencia de ADN necesaria para que funcione el ensayo) puede ser muy pequeña, una gran ventaja cuando se utiliza ADN altamente fragmentado. Además, los cebadores de amplificación pueden tener en sus extremos 5' secuencias adicionales necesarias para una secuenciación altamente multiplexada (HMS) o secuenciación de nueva generación (NGS) en cualquier plataforma elegida. También se pueden incluir códigos de barras específicos para la muestra (índices específicos para la muestra). En la Etapa 6, los dNTPs se inactivan mediante tratamiento con SAP y los cebadores se eliminan mediante Exonucleasa I. Ya que no se utilizan adaptadores de doble cadena, la preparación para la NGS es simple y directa. Se puede realizar un procesamiento adicional si se desea. Por ejemplo, si se utilizaron códigos de barras/índices específicos de la muestra, entonces las muestras pueden agruparse y purificarse y concentrarse adicionalmente para su aplicación a una plataforma NGS de elección. Aunque el protocolo que se muestra en la Figura 28 hacía uso de ADN, se pueden diseñar protocolos similares para su uso con ARN, incluidos ARNm, ARNm_i y ARN_{Inc}. Variaciones del protocolo mostradas en la Figura 28 también se pueden adaptar fácilmente a la metodología ilustrada en la Figura 15 a la Figura 25 y descrita en esta memoria.

Ejemplo 7: Curso en el tiempo de adiciones de dATP y dNTP por transferasa terminal

La transferasa terminal es una enzima robusta, bien conocida. Sus condiciones de reacción están bien definidas para la adición de nucleótidos individuales, pero no están definidas para mezclas de nucleótidos. Por lo tanto, se generó un oligonucleótido sintético de 30 meros y se monitorizaron los ciclos de tiempo para determinar las condiciones para la adición de aproximadamente 35 a 70 nucleótidos utilizando dATP y 10 a 70 nucleótidos utilizando una mezcla de los cuatro dNTPs. Los resultados se analizaron utilizando un gel de Urea TBE de acrilamida Novex al 15% teñido con SYBR Gold. Al comparar los resultados con las escalas de marcadores de ADN de 10 pb incluidas en el gel, se determinó que en 25 minutos se añadieron aproximadamente 35 a 65 residuos dA, y se añadieron aproximadamente 10 a 60 residuos dN (véase, *p. ej.*, la FIG. 26). El intervalo de adiciones puede aumentarse fácilmente aumentando el tiempo de reacción.

10 ***Ejemplo 8: Adición aleatoria de dNTP por transferasa terminal***

Se testan condiciones para la adición aleatoria de residuos dN. Utilizando un oligonucleótido de 30 meros que contiene varios residuos inosina, se utilizó dATP con transferasa terminal o dNTP. Después de la extensión, los productos fueron digeridos con Endonucleasa V para destruir el oligonucleótido, permitiendo que los productos de extensión sean analizados por espectrometría de masas MALDI-TOF. El espectro en la parte superior de la Figura 27, por ejemplo, muestra resultados para la adición de dA. Se observan picos agudos con una separación apropiada para la adición secuencial de residuos dA. En contraste, el espectro en la parte inferior de la Figura 27, por ejemplo, muestra múltiples picos para cada uno de los residuos añadidos, lo que indica una adición aleatoria de nucleótidos y poca o ninguna preferencia de nucleótidos cuando los dNTPs se mezclan entre sí. Con cada uno de los nucleótidos adicionales, los picos se vuelven más amplios, como consecuencia del aumento de la diversidad de masa con la adición aleatoria de un nucleótido a cada una de las cadenas en crecimiento. El espectro inferior subestima en gran medida la diversidad de secuencias, ya que múltiples secuencias diferentes pueden tener la misma composición y, por lo tanto, la misma masa.

REIVINDICACIONES

1. Un método para modificar un ácido nucleico, que comprende:
- 5 (a) poner en contacto un ácido nucleico de cadena sencilla con una actividad de transferasa terminal y una mezcla de dos o más nucleótidos diferentes en condiciones en las que los nucleótidos en la mezcla se añaden secuencial y aleatoriamente al extremo 3' del ácido nucleico mediante la actividad de transferasa terminal, añadiendo con ello un heteropolinucleótido que comprende nucleótidos en la mezcla al extremo 3' del ácido nucleico y generando un primer ácido nucleico modificado;
- 10 (b) poner en contacto el primer ácido nucleico modificado con una actividad de transferasa terminal y una composición que comprende monómeros de un solo nucleótido en condiciones en las que los monómeros se añaden al extremo 3' del primer ácido nucleico modificado mediante la actividad de transferasa terminal, añadiendo con ello un homopolinucleótido que comprende los monómeros al extremo 3' del primer ácido nucleico modificado y generando un segundo ácido nucleico modificado que comprende, 5' a 3', el heteropolinucleótido y homopolinucleótido; y
- 15 (c) poner en contacto el segundo ácido nucleico modificado con un segundo ácido nucleico de cadena sencilla que comprende una secuencia que es complementaria al, o sustancialmente complementaria al homopolinucleótido del segundo ácido nucleico modificado y una actividad de ligasa, en condiciones en las que se ligan el extremo 3' del segundo ácido nucleico modificado y el extremo 5' del segundo ácido nucleico de cadena sencilla, generando con ello un ácido nucleico ligado, en donde el extremo 5' del segundo ácido nucleico de cadena sencilla está fosforilado o adenilado.
- 20 2. El método de la reivindicación 1, en el que la mezcla comprende dos nucleótidos diferentes.
3. El método de la reivindicación 1, en el que la mezcla comprende tres nucleótidos diferentes.
4. El método de la reivindicación 1, en el que la mezcla comprende cuatro nucleótidos diferentes.
5. El método de una cualquiera de las reivindicaciones 1 a 4, en el que los diferentes nucleótidos en la mezcla son diferentes desoxinucleótidos trifosfato.
- 25 6. El método de la reivindicación 5, en el que los diferentes nucleótidos en la mezcla se eligen entre adenosina trifosfato (ATP), guanosina trifosfato (GTP), citidina trifosfato (CTP) y timidina trifosfato (TTP).
7. El método de una cualquiera de las reivindicaciones 1 a 6, que comprende poner en contacto el ácido nucleico indexado (o ácido nucleico modificado) con una actividad de transferasa terminal y una composición que comprende monómeros de un solo nucleótido en condiciones en las que los monómeros se añaden al extremo 3' del ácido nucleico indexado (o ácido nucleico modificado) por la actividad de la transferasa terminal,
- 30 añadiendo con ello un homopolinucleótido indicador que comprende los monómeros al extremo 3' del ácido nucleico indexado (o ácido nucleico modificado) y generando un ácido nucleico con indicador que comprende, 5' a 3', el índice de heteropolinucleótido y el homopolinucleótido indicador.
8. El método de una cualquiera de las reivindicaciones 1 a 7, que comprende poner en contacto el ácido nucleico indexado (o ácido nucleico modificado) con una actividad de transferasa terminal y una composición que comprende monómeros de un solo nucleótido en condiciones en las que los monómeros se añaden al extremo 3' del ácido nucleico indexado (o ácido nucleico modificado) por la actividad de la transferasa terminal,
- 35 añadiendo con ello un homopolinucleótido que comprende los monómeros al extremo 3' del ácido nucleico indexado (o ácido nucleico modificado) y generando un ácido nucleico modificado adicional que comprende, 5' a 3', el heteropolinucleótido y el homopolinucleótido.
- 40 9. El método de una cualquiera de las reivindicaciones 1 a 8, en el que el heteropolinucleótido es de aproximadamente 5 bases de nucleótidos a aproximadamente 30 bases de nucleótidos o más de longitud.
10. El método de una cualquiera de las reivindicaciones 1 a 9, que comprende amplificar el ácido nucleico que comprende el heteropolinucleótido.
- 45 11. Uso de un método según se define en una cualquiera de las reivindicaciones 1 a 10 en la indexación molecular.
12. Un método para producir una molécula de ácido nucleico de cadena sencilla, comprendiendo el método:
- poner en contacto una primera molécula de ácido nucleico de cadena sencilla (ANcs) que comprende un extremo 5' fosforilado, una secuencia diana y un extremo 3' con una actividad de transferasa terminal y dNTPs en condiciones de reacción en las que se añade una secuencia aleatoria de residuos de heteropolinucleótidos al extremo 3' de la primera molécula de ANcs, produciendo con ello un ácido nucleico indexado;
- 50 añadir una secuencia de homopolímero cola 3' mediante el uso de una actividad de transferasa terminal y ribonucleótidos;

poner en contacto la primera molécula de ANcs con una actividad ligasa de ácido nucleico de cadena sencilla para ligar el extremo 3' de la primera molécula de ANcs con su resto de 5' fosfato para producir una única molécula de NAcS circular cerrada, y

5 amplificar una porción de la única molécula de ANcs circular cerrada de tal manera que se copien el índice molecular y porciones de interés de la molécula nativa.

13. Un método para modificar un ácido nucleico, que comprende:

10 (a) poner en contacto un primer ácido nucleico con una primera proteína que tiene una actividad de transferasa terminal y una mezcla de dos o más nucleótidos diferentes en condiciones en las que los nucleótidos en la mezcla se añaden secuencial y aleatoriamente a un extremo 3' del primer ácido nucleico mediante la primera proteína, añadiendo con ello un heteropolinucleótido que comprende nucleótidos en la mezcla al extremo 3' del primer ácido nucleico y generando un primer ácido nucleico modificado;

15 (b) poner en contacto el primer ácido nucleico modificado con una segunda proteína que tiene una actividad de transferasa terminal y una pluralidad de monómeros de un solo nucleótido en condiciones en las que la pluralidad de monómeros de un solo nucleótido se añade al extremo 3' del primer ácido nucleico modificado por la segunda proteína, añadiendo con ello un homopolinucleótido que comprende los monómeros al extremo 3' del primer ácido nucleico modificado y generando un segundo ácido nucleico modificado que comprende, 5' a 3', el heteropolinucleótido y el homopolinucleótido;

20 (c) poner en contacto el segundo ácido nucleico modificado con un segundo ácido nucleico de cadena sencilla que comprende una secuencia que es complementaria o sustancialmente complementaria al homopolinucleótido del segundo ácido nucleico modificado, en condiciones que permitan la hibridación o reasociación del homopolinucleótido al segundo ácido nucleico modificado, seguido de extender el segundo ácido nucleico modificado utilizando el segundo ácido nucleico de cadena sencilla como molde, generando con ello un tercer ácido nucleico modificado que tiene, 5' a 3', el primer ácido nucleico modificado y el segundo ácido nucleico modificado, en donde opcionalmente la extensión del segundo ácido nucleico modificado para generar un equivalente complementario del segundo ácido nucleico de cadena sencilla se realiza en condiciones de amplificación, 25 opcionalmente condiciones de reacción en cadena de la polimerasa (PCR).

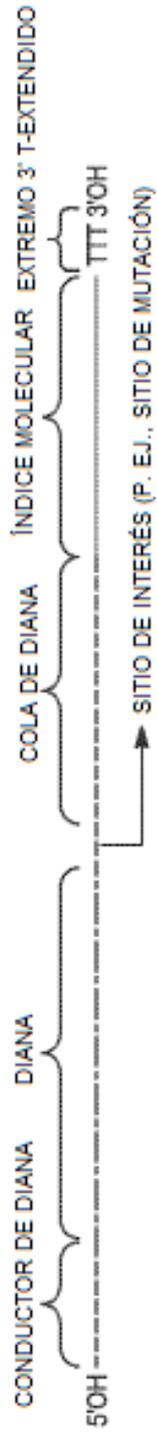


FIG. 1

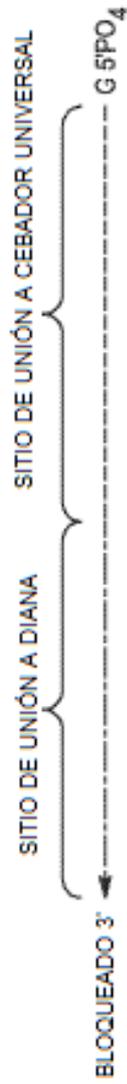


FIG. 2

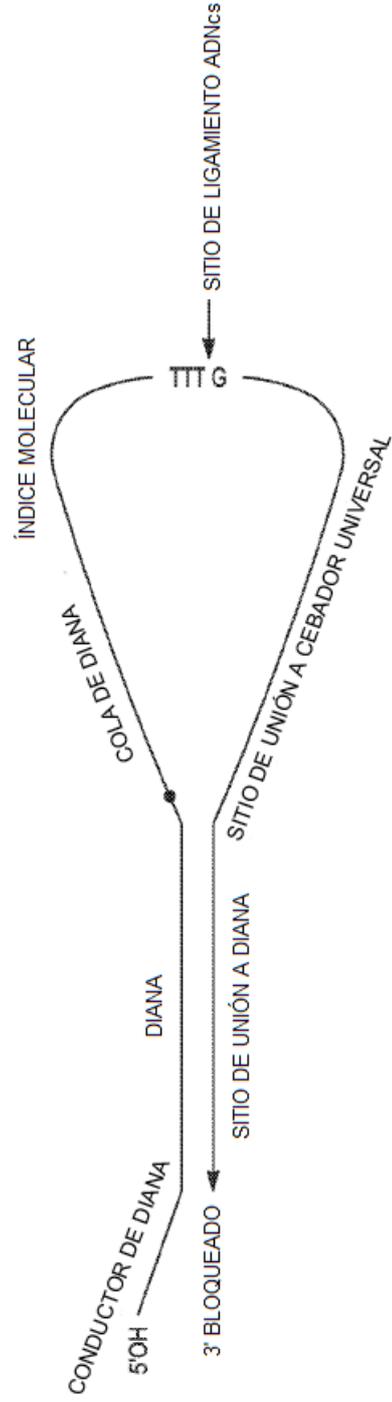


FIG. 3



FIG. 4

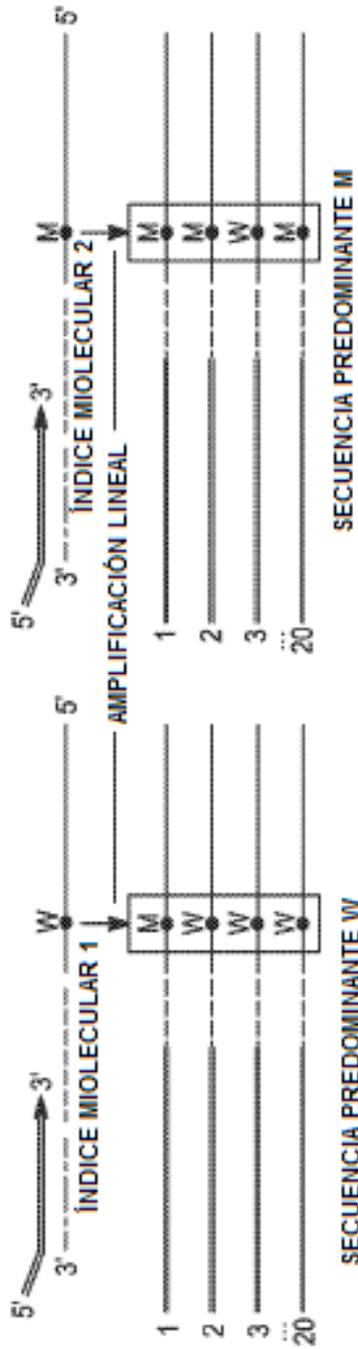


FIG. 5

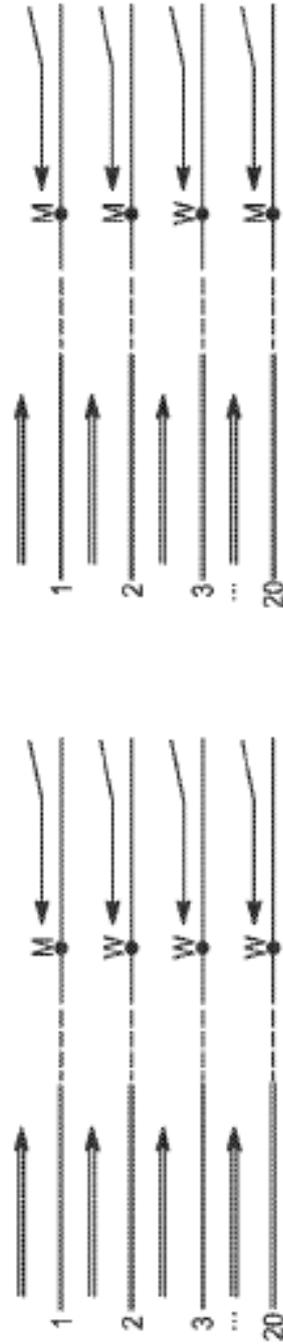


FIG. 6

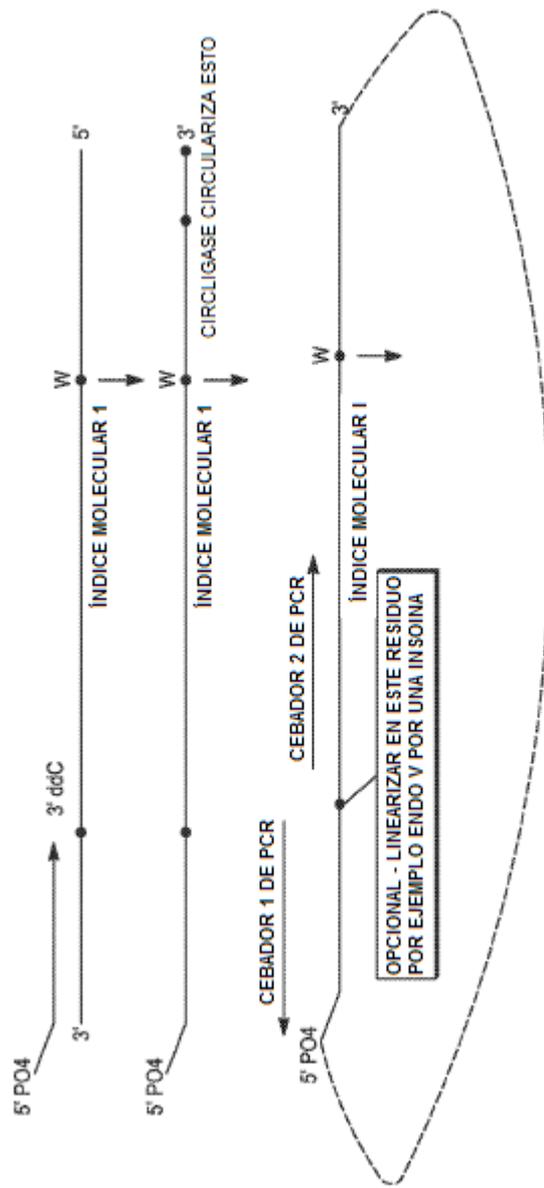


FIG. 7

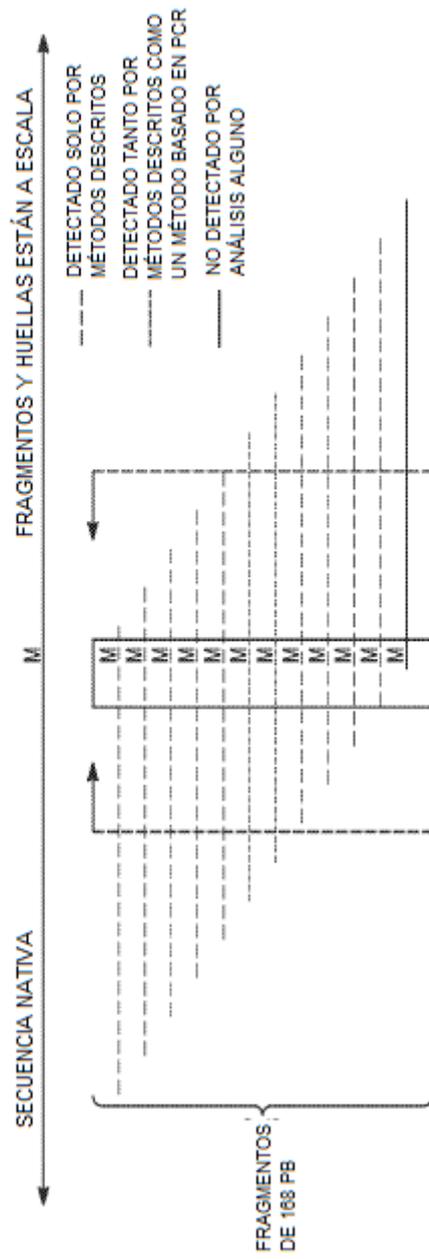


FIG. 8

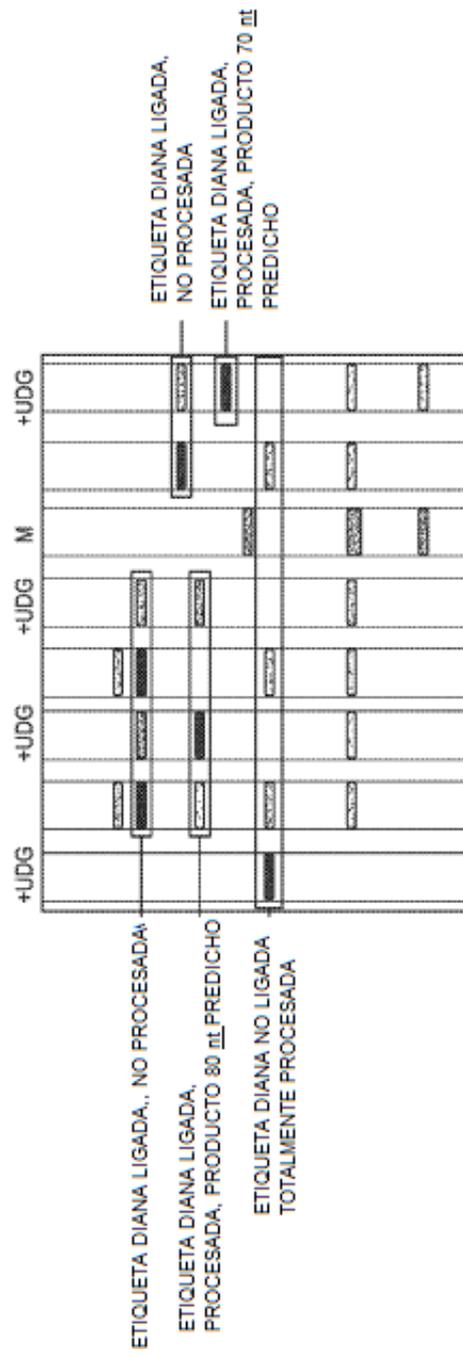


FIG. 9

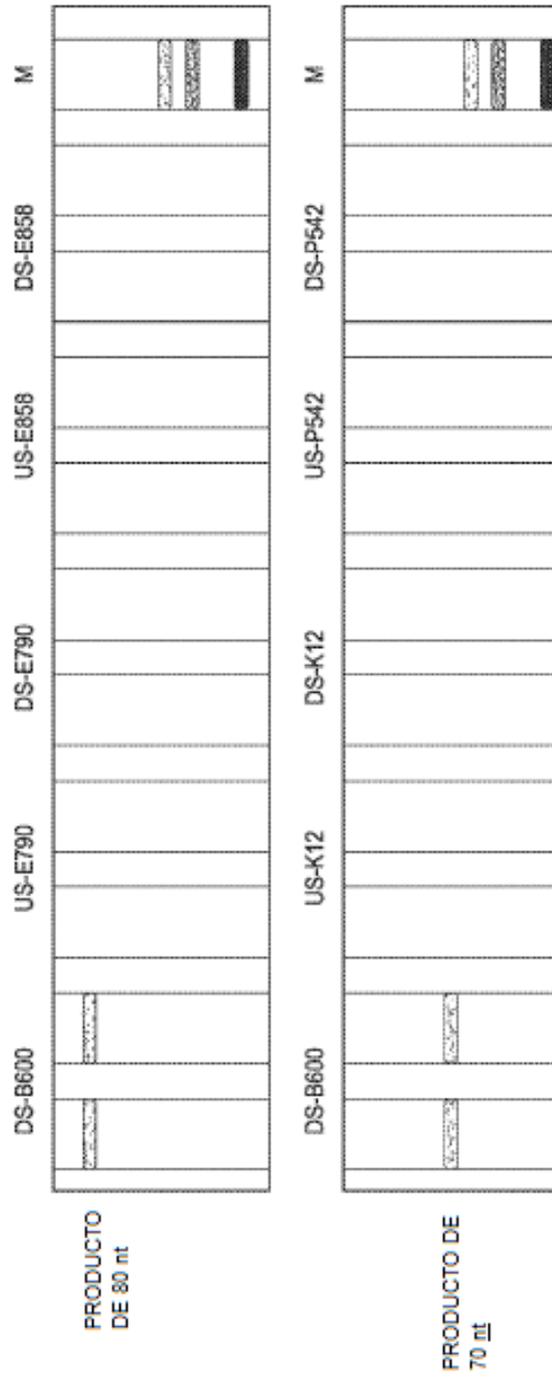
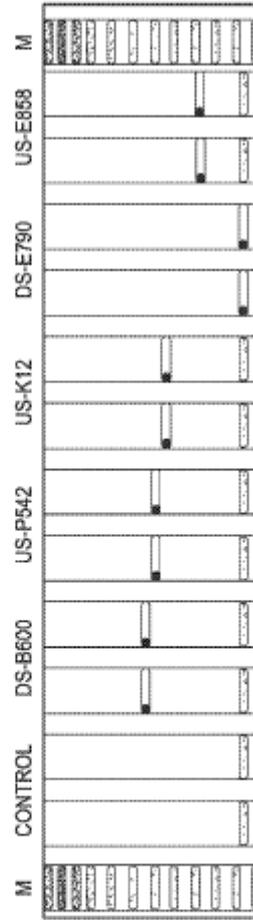
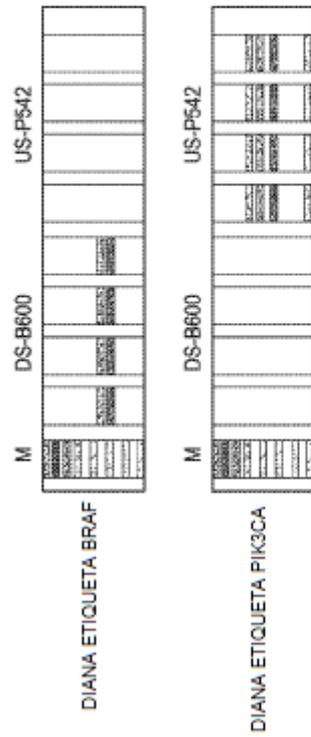


FIG. 10



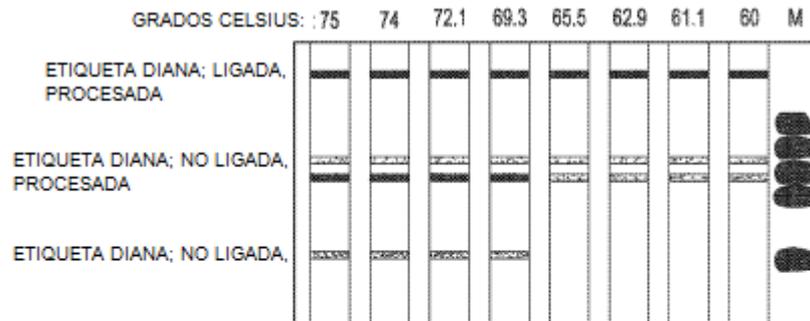


FIG. 13

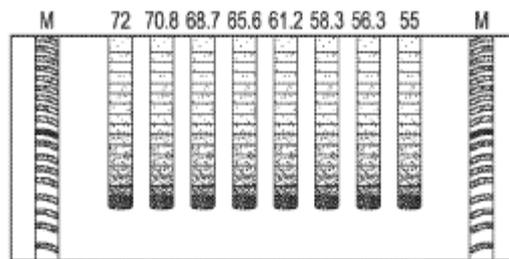


FIG. 14A

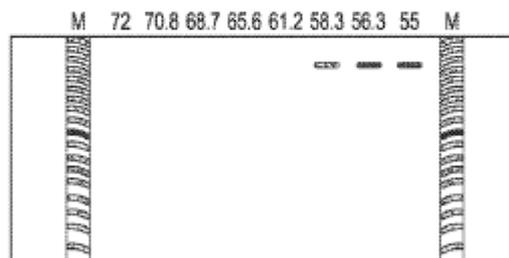


FIG. 14B

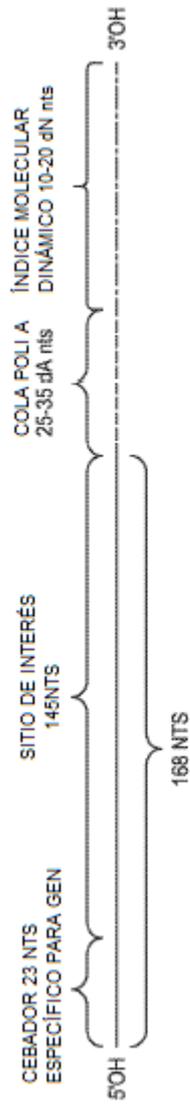


FIG. 15

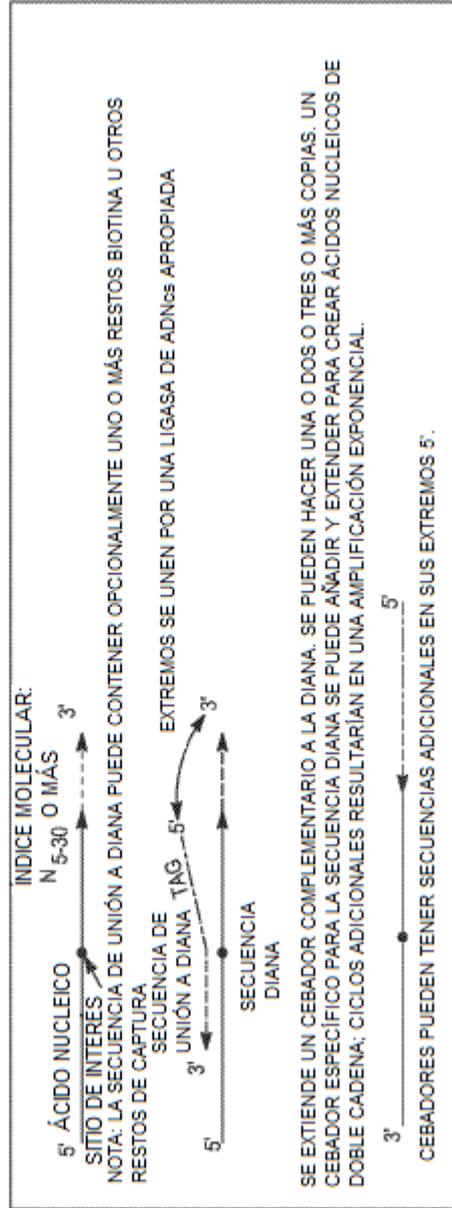
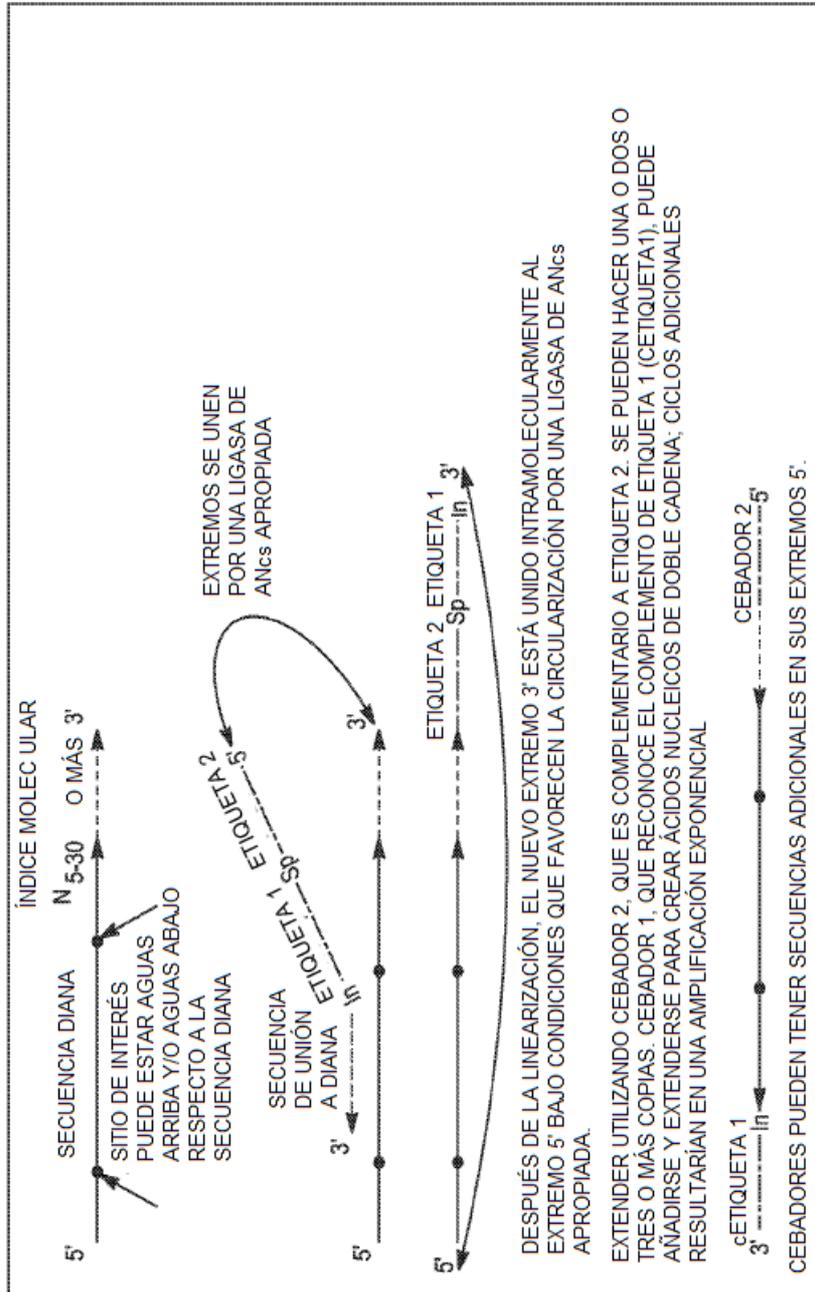


FIG. 16



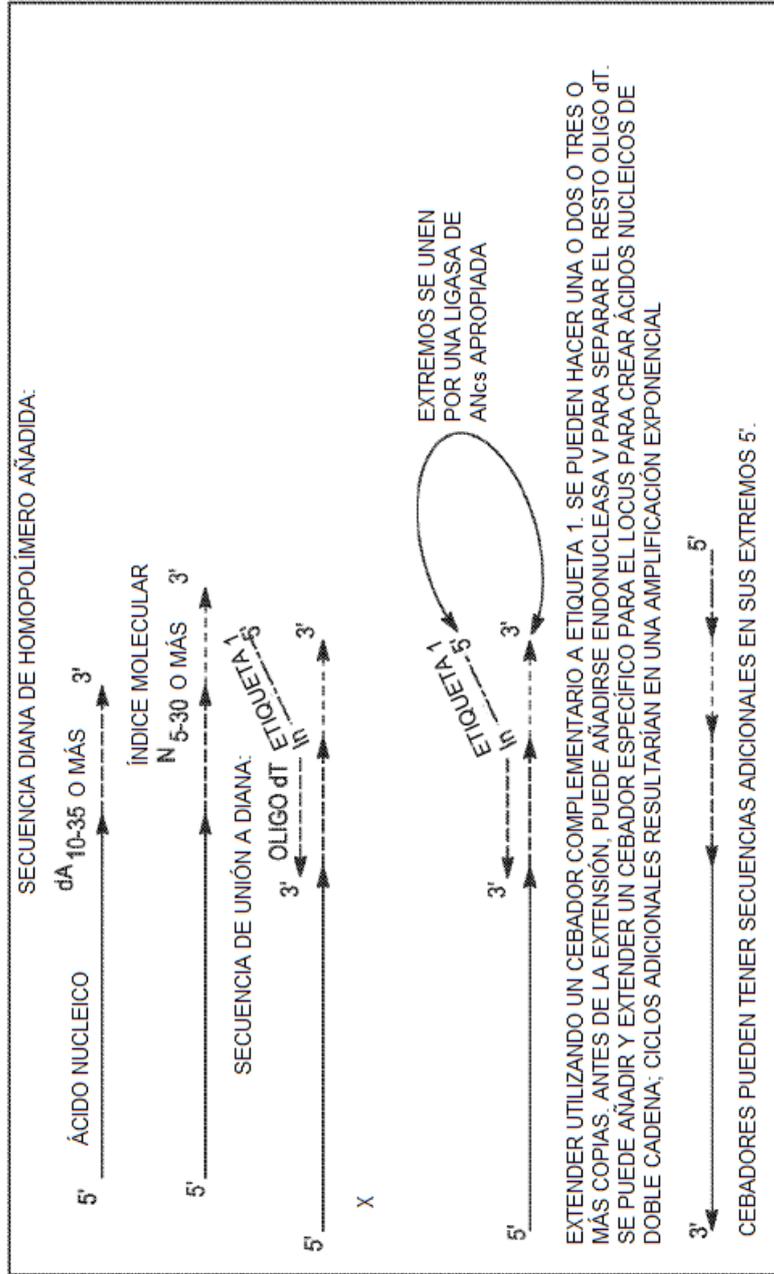


FIG. 18

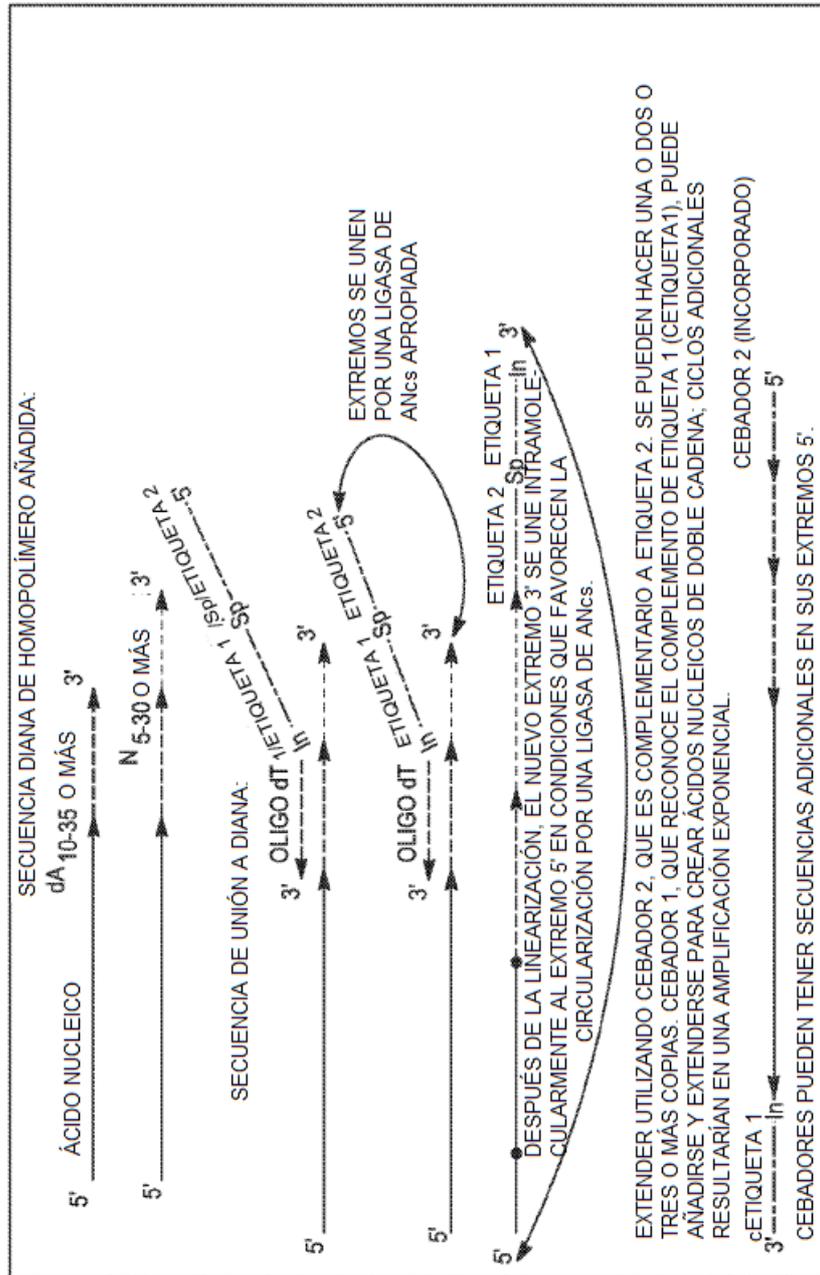


FIG. 19

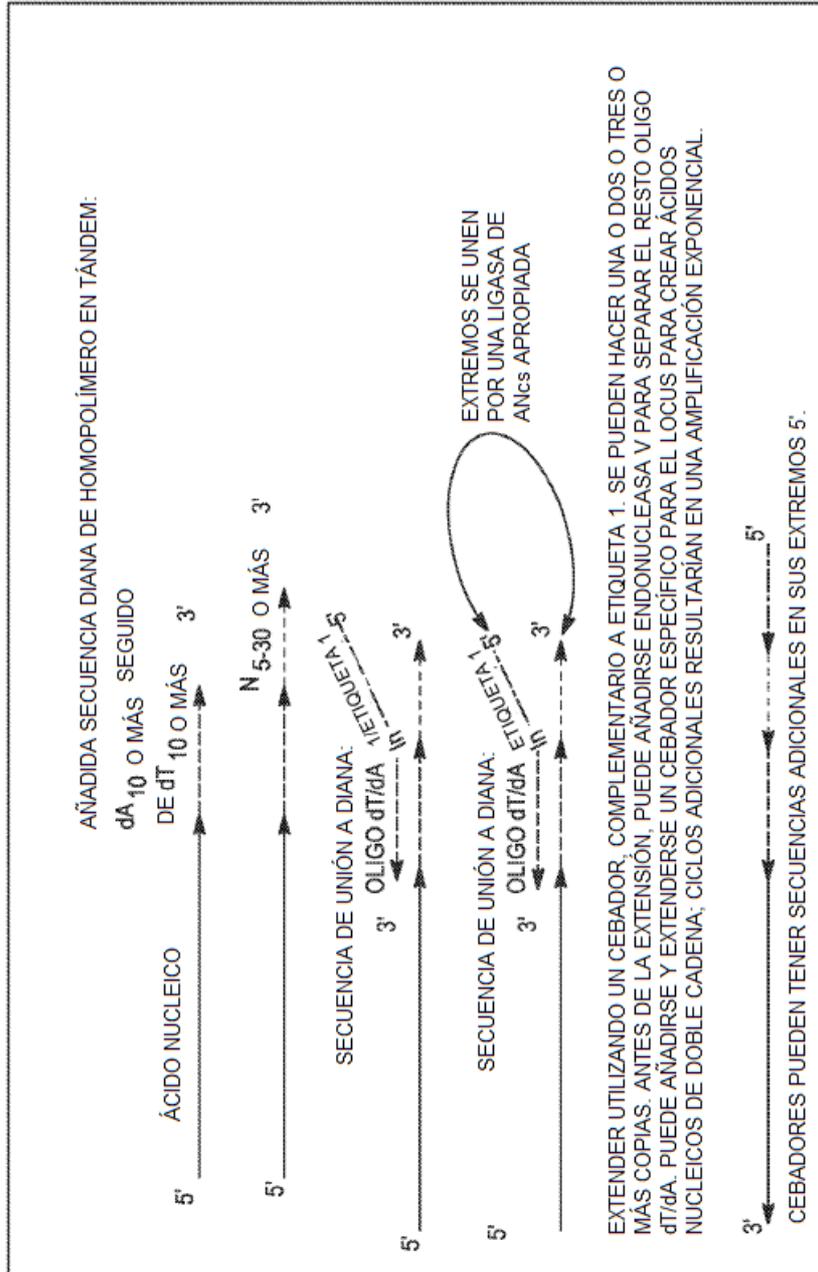


FIG. 20

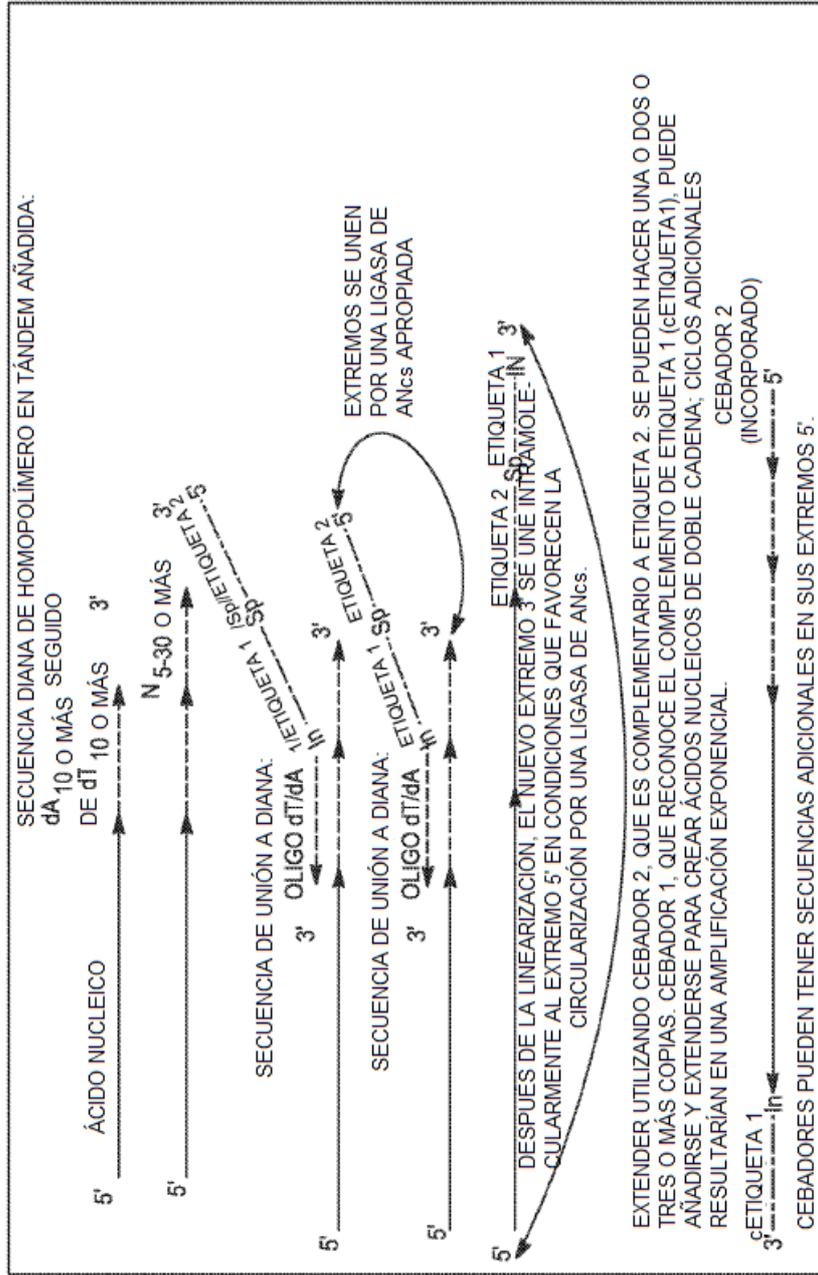


FIG. 21

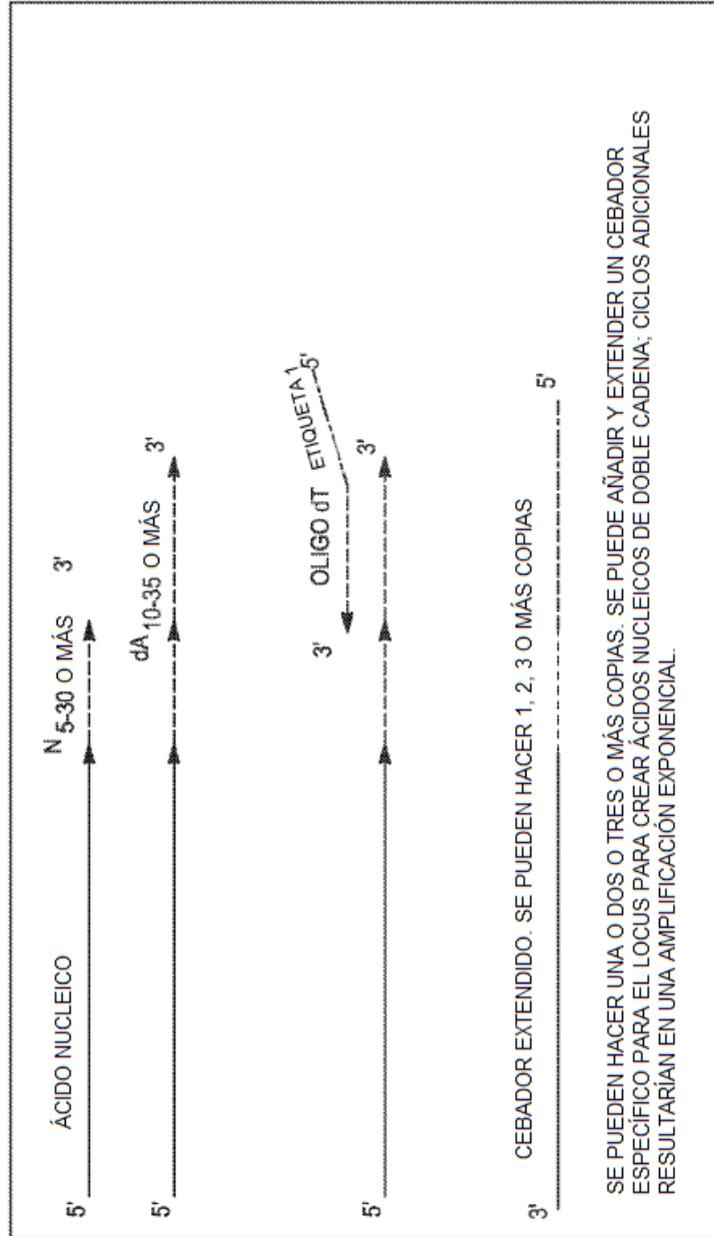


FIG. 22

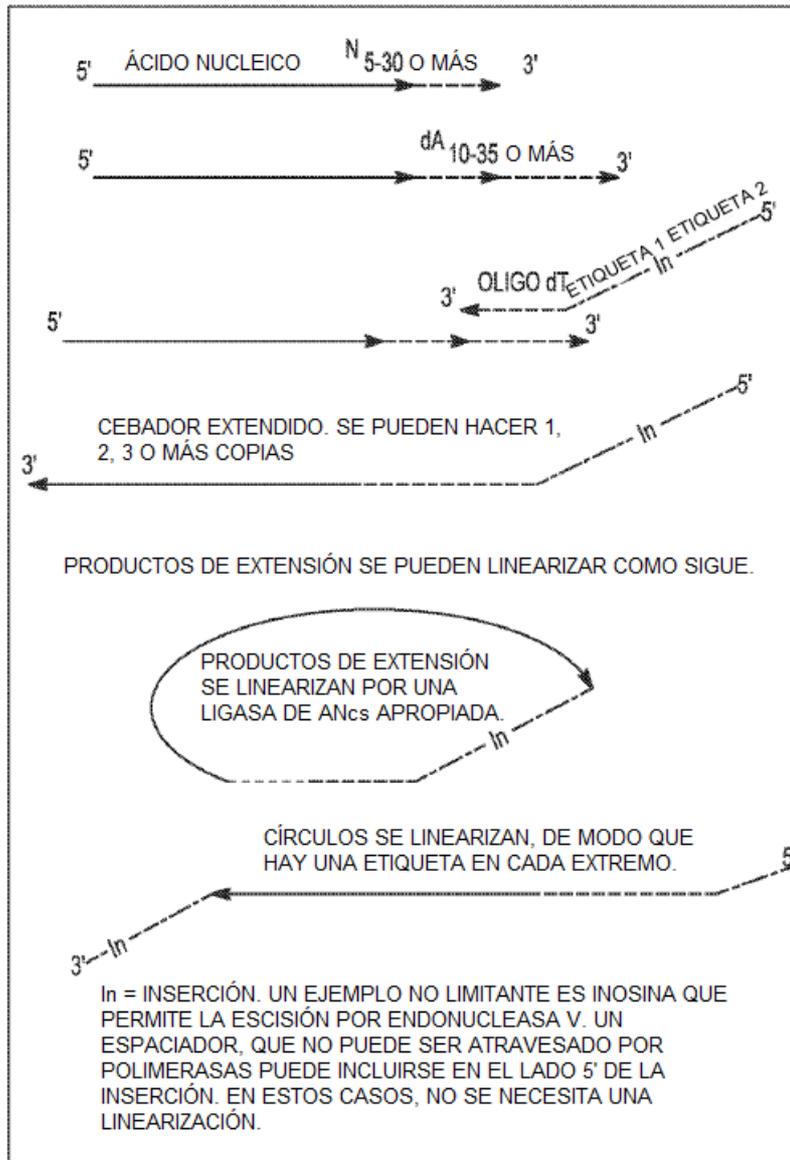


FIG. 23

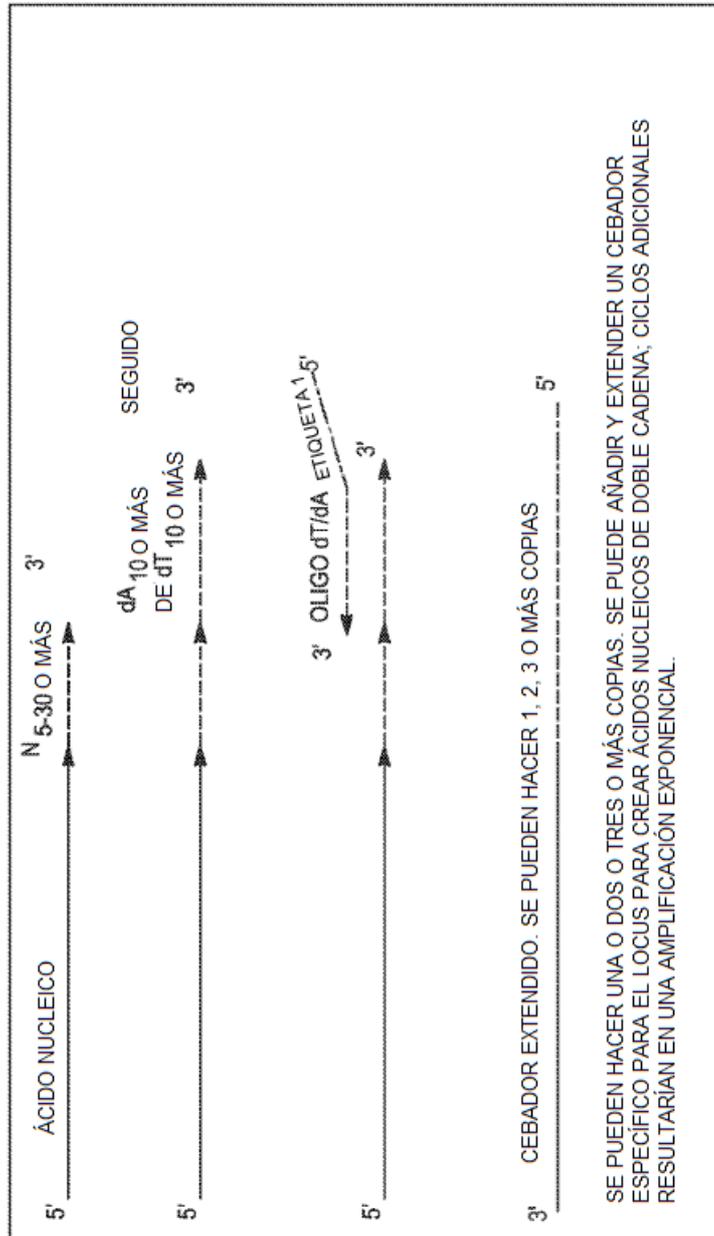


FIG. 24

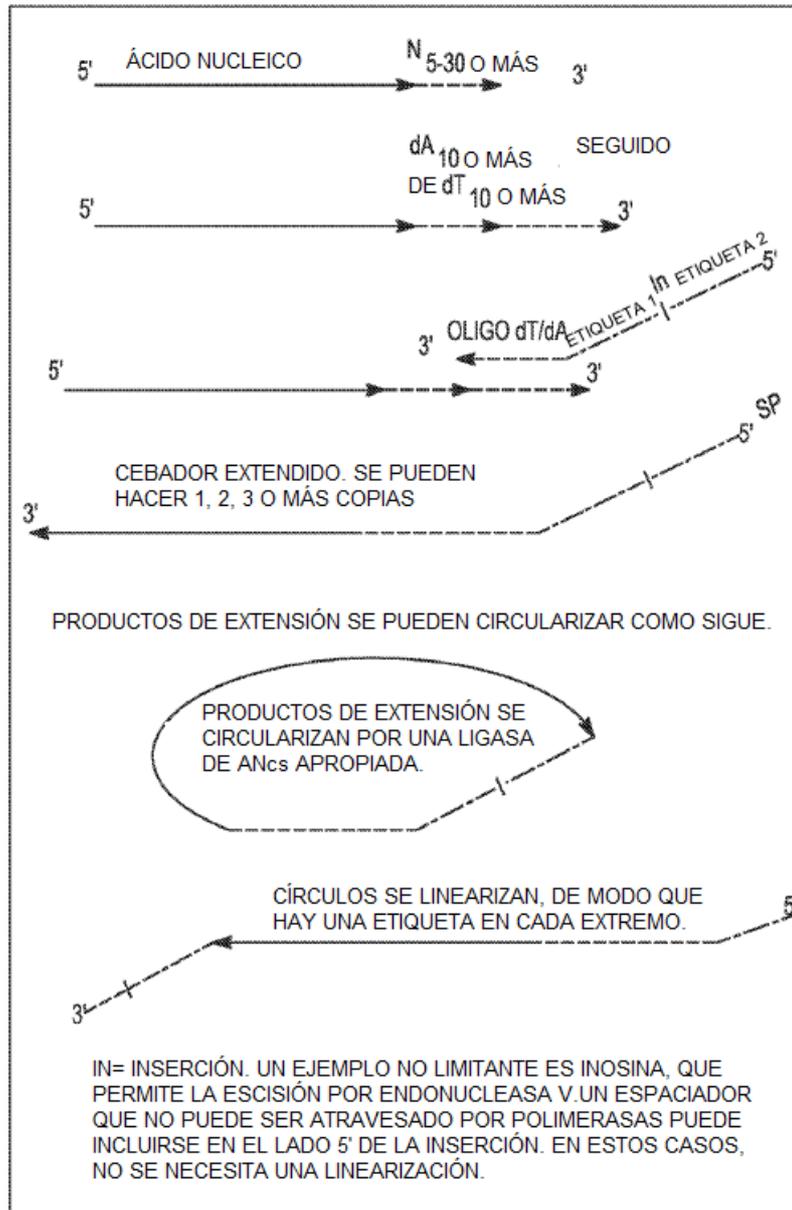


FIG. 25

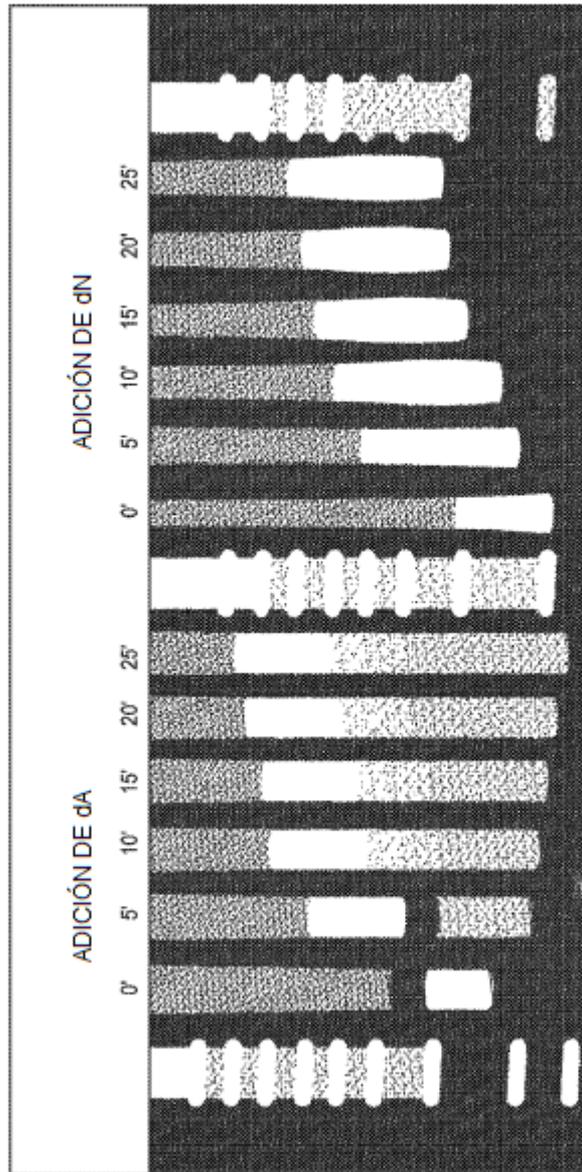


FIG. 26

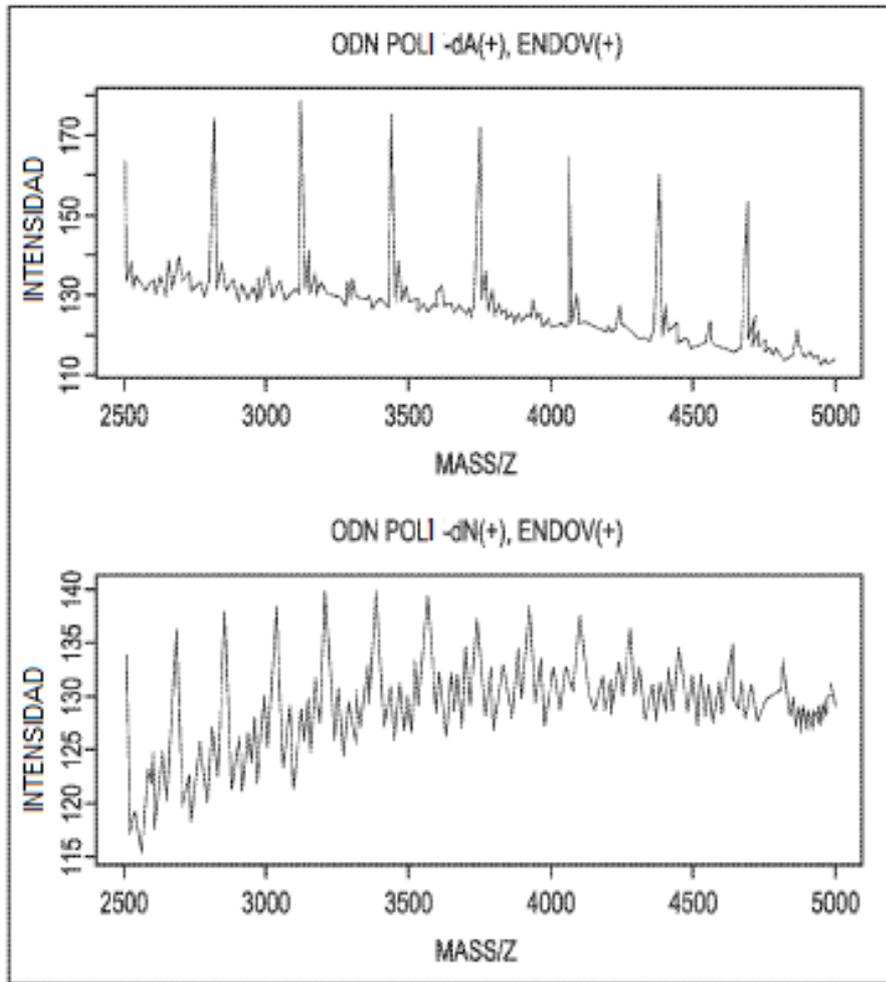


FIG. 27

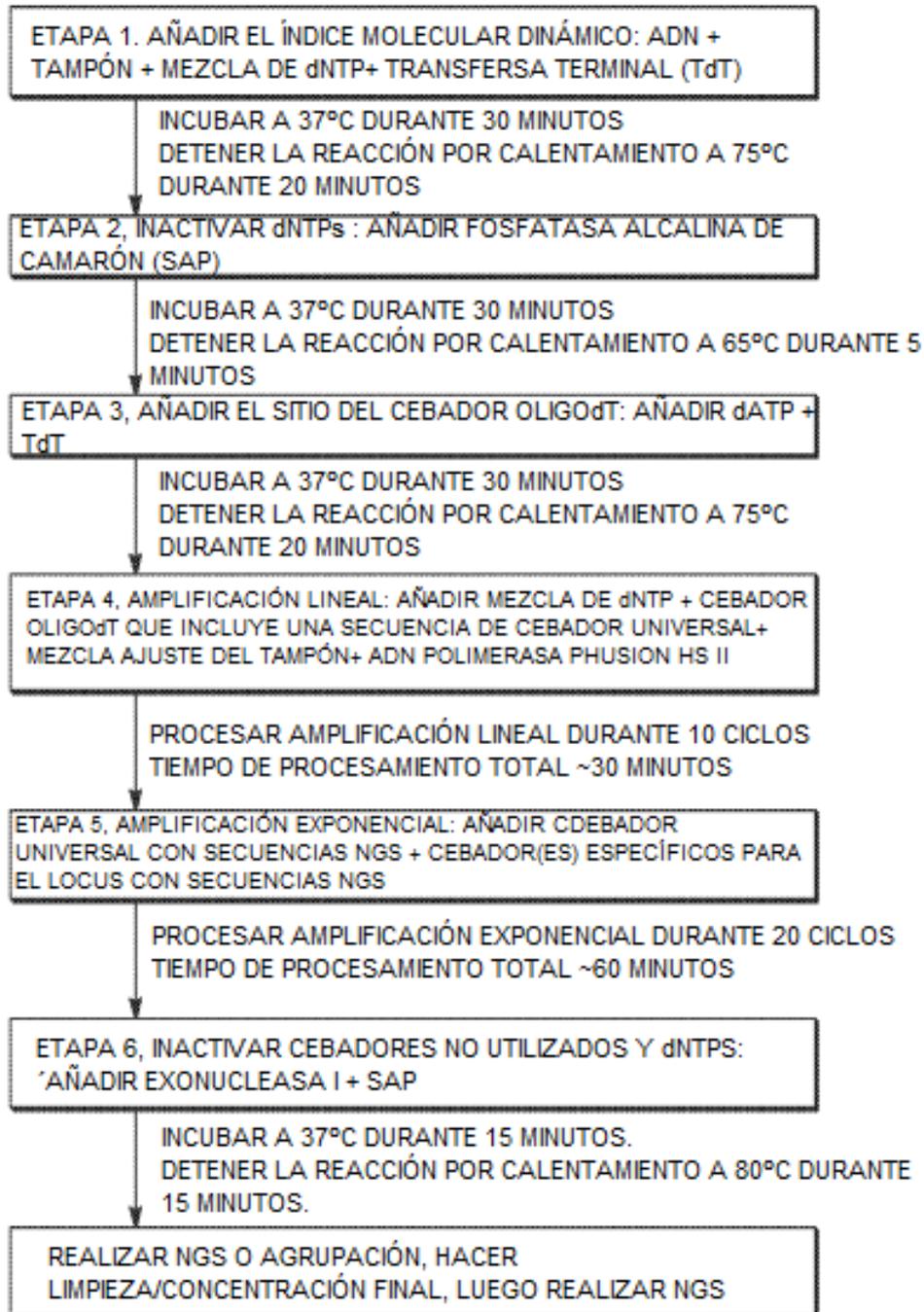


FIG. 28