

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **3 014 208**

51 Int. Cl.:

C12N 15/10 (2006.01)

C12Q 1/6806 (2008.01)

C12Q 1/6834 (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **01.05.2019 PCT/US2019/030175**

87 Fecha y número de publicación internacional: **07.11.2019 WO19213237**

96 Fecha de presentación y número de la solicitud europea: **01.05.2019 E 19723988 (2)**

97 Fecha y número de publicación de la concesión europea: **01.01.2025 EP 3788170**

54 Título: **Identificación con códigos de barras moleculares en extremos opuestos de transcrito**

30 Prioridad:

03.05.2018 US 201862666506 P

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

21.04.2025

73 Titular/es:

BECTON, DICKINSON AND COMPANY (100.00%)
1 Becton Drive
Franklin Lakes, NJ 07417, US

72 Inventor/es:

SHUM, ELEEN y
FAN, CHRISTINA

74 Agente/Representante:

IZQUIERDO BLANCO, María Alicia

ES 3 014 208 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

Identificación con códigos de barras moleculares en extremos opuestos de transcrito

5 **Campo**

La presente divulgación se refiere de manera general al campo de la biología molecular y, en particular, a análisis multiómicos que usan identificación con códigos de barras moleculares.

10 **Descripción de la técnica relacionada**

Los métodos y técnicas como la identificación con códigos de barras molecular son útiles para el análisis transcriptómico de células individuales, incluyendo el descifrado de perfiles de expresión génica para determinar los estados de las células usando, por ejemplo, transcripción inversa, amplificación de reacción en cadena de la polimerasa (PCR) y secuenciación de próxima generación (NGS). La identificación con códigos de barras moleculares también es útil para el análisis proteómico de células individuales.

El documento US2016017320 describe códigos de barras semialeatorios para el análisis de ácidos nucleicos: los oligonucleótidos que comprenden las secuencias de códigos de barras semialeatorios pueden incorporarse a cebadores de transcripción inversa, cebadores de PCR o porciones de adaptadores de secuenciación en la preparación de bibliotecas de secuenciación. El documento WO2016160965 describe métodos y composiciones para la reparación de extremos de ADN mediante múltiples actividades enzimáticas. La generación de ácidos nucleicos competentes para la ligación puede realizarse con composiciones que comprenden Exonucleasa III, ADN Polimerasa T4, Klenow, y/o polinucleótido quinasa T4. El documento WO2012103154 describe cebadores de adaptadores de ARN-ADN compuestos de estructura de horquilla, para la generación y amplificación de bibliotecas. El documento WO2016191272 describe métodos y composiciones para el paseo genómico de próxima generación: identificación de secuencias de ADN raras y/o desconocidas mediante enfoques de secuenciación de próxima generación. El documento WO2014071361 describe métodos de identificación con códigos de barras de ácidos nucleicos, como ADN genómico, en donde un fragmento de ADN genómico puede comprender un primer y un segundo código de barras.

30 **SUMARIO**

La invención se expone en el conjunto de reivindicaciones adjunto. Los términos "realización" y "realizaciones" deben interpretarse como realización o realizaciones de la invención únicamente en la medida en que entren dentro del alcance de las reivindicaciones adjuntas. De lo contrario, se refieren únicamente a realizaciones de la divulgación.

La invención proporciona métodos para unir códigos de barras de oligonucleótidos a una diana en una muestra. A modo de ejemplo, la diana puede comprender o consistir en una diana de ácido nucleico. El método comprende: identificar con códigos de barras copias de una diana de ácido nucleico en una muestra usando una pluralidad de códigos de barras de oligonucleótidos para generar una pluralidad de moléculas de ácido nucleico identificadas con códigos de barras cada una de las cuales comprendiendo secuencia de la diana de ácido nucleico, un marcador molecular y una región de unión a la diana, y en donde por lo menos 10 de la pluralidad de códigos de barras de oligonucleótidos comprenden diferentes secuencias de marcadores moleculares; unir un oligonucleótido que comprende un complemento de la región de unión a la diana a la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras para generar una pluralidad de moléculas de ácido nucleico identificadas con códigos de barras, cada una de las cuales comprendiendo la región de unión a la diana y el complemento de la región de unión a la diana; hibridar la región de unión a la diana y el complemento de la región de unión a la diana dentro de cada una de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras para formar una estructura de horquilla; y extender los extremos 3' de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras para extender la estructura de horquilla para generar una pluralidad de moléculas de ácido nucleico identificadas con códigos de barras extendidas que comprenden cada una el marcador molecular y un complemento del marcador molecular. Opcionalmente, el método comprende además amplificar la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras para generar una pluralidad de moléculas de ácido nucleico identificadas con códigos de barras amplificadas. La unión del oligonucleótido que comprende el complemento de la región de unión a la diana puede comprender unir el oligonucleótido a la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras amplificadas para generar la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras, cada una de las cuales comprendiendo la región de unión a la diana y un complemento de la región de unión a la diana. Opcionalmente, el método comprende además amplificar la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras extendidas. La cantidad de dianas de ácido nucleico en la muestra puede determinarse después de amplificar la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras extendidas. En algunas realizaciones, el método comprende: determinar la cantidad de diana de ácido nucleico en la muestra basándose en el número de marcadores moleculares con secuencias distintas, complementos de las mismas, o una combinación de las mismas, asociadas con la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras extendidas. En algunas realizaciones, el método comprende: determinar la cantidad de diana de

ácido nucleico en la muestra basándose en el número de marcadores moleculares con secuencias distintas asociadas con la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras extendidas.

La invención también proporciona métodos para determinar la cantidad de diana en una muestra. A modo de ejemplo, las dianas pueden comprender o consistir en un ácido nucleico. El método comprende: identificar con códigos de barras copias de una diana de ácido nucleico en una muestra usando una pluralidad de códigos de barras de oligonucleótidos para generar una pluralidad de moléculas de ácido nucleico identificadas con códigos de barras cada una de las cuales comprendiendo secuencia de la diana para ácido nucleico, un marcador molecular y una región de unión a la diana, y en donde por lo menos 10 de la pluralidad de códigos de barras de oligonucleótidos comprenden diferentes secuencias de marcadores moleculares; unir un oligonucleótido que comprende un complemento de la región de unión a la diana a la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras para generar una pluralidad de moléculas de ácido nucleico identificadas con códigos de barras, cada una de las cuales comprendiendo la región de unión a la diana y el complemento de la región de unión a la diana; hibridar la región de unión a la diana y el complemento de la región de unión a la diana dentro de cada una de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras para formar una estructura de horquilla; extender los extremos 3' de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras para extender la estructura de horquilla y generar una pluralidad de moléculas de ácido nucleico identificadas con códigos de barras extendidas, cada una de las cuales comprendiendo el marcador molecular y un complemento del marcador molecular; y determinar la cantidad de diana de ácido nucleico en la muestra basándose en el número de complementos de marcadores moleculares con secuencias distintas asociadas con la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras extendidas. Opcionalmente, el método comprende además amplificar la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras para generar una pluralidad de moléculas de ácido nucleico identificadas con códigos de barras amplificadas. La unión del oligonucleótido que comprende el complemento de la región de unión a la diana puede comprender unir el oligonucleótido a la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras amplificadas para generar la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras, cada una de las cuales comprendiendo la región de unión a la diana y un complemento de la región de unión a la diana. Opcionalmente, el método comprende además amplificar la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras extendidas. El número de dianas de ácido nucleico en la muestra puede determinarse después de amplificar la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras extendidas. En algunas realizaciones, el método comprende determinar la cantidad de la diana de ácido nucleico en la muestra basándose en el número de marcadores moleculares con secuencias distintas, complementos de las mismas, o una combinación de las mismas, asociadas con la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras extendidas.

En algunas realizaciones, cualquiera de los métodos descritos en la presente comprende identificar con códigos de barras las copias de la pluralidad de dianas comprende: poner en contacto copias de la diana de ácido nucleico con la pluralidad de códigos de barras de oligonucleótidos, en donde cada uno de la pluralidad de códigos de barras de oligonucleótidos comprende la región de unión a la diana capaz de hibridar con la diana de ácido nucleico; y extender las copias de la diana de ácido nucleico hibridadas con los códigos de barras de oligonucleótidos para generar la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras. En algunas realizaciones, cualquiera de los métodos descritos en la presente comprende: identificar con códigos de barras las copias de la pluralidad de dianas comprende: poner en contacto las copias de la diana de ácido nucleico con la pluralidad de códigos de barras de oligonucleótidos, en donde cada uno de la pluralidad de códigos de barras de oligonucleótidos comprende la región de unión a la diana. La región de unión a la diana puede hibridar con la diana de ácido nucleico. El método puede comprender además extender las copias de la diana de ácido nucleico hibridadas con los códigos de barras de oligonucleótidos para generar la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras.

En algunas realizaciones, cualquiera de los métodos descritos en la presente comprende: amplificar la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras para generar una pluralidad de moléculas de ácido nucleico identificadas con códigos de barras amplificadas, en donde la unión del oligonucleótido que comprende el complemento de la región de unión a la diana comprende: unir el oligonucleótido que comprende el complemento de la región de unión a la diana a la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras amplificadas para generar una pluralidad de moléculas de ácido nucleico identificadas con códigos de barras que comprenden cada una la región de unión a la diana y un complemento de la región de unión a la diana.

En algunas realizaciones, cualquiera de los métodos descritos en la presente comprende: amplificar la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras extendidas para generar una pluralidad de moléculas de ácido nucleico marcadas individualmente que comprenden cada una el complemento del marcador molecular, en donde determinar la cantidad de la diana de ácido nucleico en la muestra comprende: determinar la cantidad de la diana de ácido nucleico en la muestra basándose en el número de complementos de marcadores moleculares con secuencias distintas asociados con la pluralidad de moléculas de ácido nucleico marcadas individualmente.

En algunas realizaciones, el método comprende: amplificar la pluralidad de moléculas de ácido nucleico

identificadas con códigos de barras extendidas para generar copias de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras extendidas, en donde determinar la cantidad de la diana de ácido nucleico en la muestra comprende: determinar la cantidad de la diana de ácido nucleico en la muestra basándose en el número de complementos de marcadores moleculares con secuencias distintas asociados con las copias de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras extendidas.

La invención proporciona métodos para determinar las cantidades de una diana de ácido nucleico en una muestra. El método comprende: poner en contacto copias de una diana de ácido nucleico en una muestra con una pluralidad de códigos de barras de oligonucleótidos, en donde cada uno de la pluralidad de códigos de barras de oligonucleótidos comprende un marcador molecular y una región de unión a la diana capaz de hibridar con la diana de ácido nucleico, y en donde por lo menos 10 de la pluralidad de códigos de barras de oligonucleótidos comprenden diferentes secuencias de marcadores moleculares; extender las copias de la diana de ácido nucleico hibridadas con los códigos de barras de oligonucleótidos para generar una pluralidad de moléculas de ácido nucleico, cada una de las cuales comprendiendo una secuencia complementaria a por lo menos una parte de la diana de ácido nucleico; amplificar la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras para generar una pluralidad de moléculas de ácido nucleico identificadas con códigos de barras amplificadas; unir un oligonucleótido que comprenda el complemento de la región de unión a la diana a la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras amplificadas para generar una pluralidad de moléculas de ácido nucleico identificadas con códigos de barras que comprenda cada una la región de unión a la diana y un complemento de la región de unión a la diana; hibridar la región de unión a la diana y el complemento de la región de unión a la diana dentro de cada una de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras para formar una estructura de horquilla; extender los extremos 3' de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras para extender la estructura de horquilla y generar una pluralidad de moléculas de ácido nucleico identificadas con códigos de barras extendidas, cada una de las cuales comprendiendo el marcador molecular y un complemento del marcador molecular; amplificar la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras extendidas para generar una pluralidad de moléculas de ácido nucleico marcadas individualmente, cada una de las cuales comprendiendo el complemento del marcador molecular; y determinar la cantidad de la diana de ácido nucleico en la muestra basándose en el número de complementos de marcadores moleculares con secuencias distintas asociadas con la pluralidad de moléculas de ácido nucleico marcadas individualmente.

En algunas realizaciones, para cualquiera de los métodos descritos en la presente, el marcador molecular se hibrida con el complemento del marcador molecular después de extender los extremos 3' de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras, el método comprendiendo desnaturalizar la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras extendidas antes de amplificar la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras extendidas para generar la pluralidad de moléculas de ácido nucleico marcadas individualmente. Poner en contacto copias de la diana de ácido nucleico en la muestra puede comprender poner en contacto copias de una pluralidad de dianas de ácido nucleico con una pluralidad de códigos de barras de oligonucleótidos. Extender las copias de la diana de ácido nucleico puede comprender extender las copias de la pluralidad de dianas de ácido nucleico hibridadas con los códigos de barras de oligonucleótidos para generar una pluralidad de moléculas de ácido nucleico identificadas con códigos de barras, cada una de las cuales comprendiendo una secuencia complementaria a por lo menos una parte de una de la pluralidad de dianas de ácido nucleico. Determinar la cantidad de la diana de ácido nucleico puede comprender determinar la cantidad de cada una de la pluralidad de dianas de ácido nucleico en la muestra basándose en el número de los complementos de los marcadores moleculares con secuencias distintas asociados con moléculas de ácido nucleico marcadas individualmente de la pluralidad de moléculas de ácido nucleico marcadas individualmente que comprenden una secuencia de cada una de la pluralidad de dianas de ácido nucleico. La secuencia de cada una de la pluralidad de dianas de ácido nucleico puede comprender una subsecuencia de cada una de la pluralidad de dianas de ácido nucleico.

En algunas realizaciones, para cualquiera de los métodos descritos en la presente, la secuencia de la diana de ácido nucleico en la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras comprende una subsecuencia de la diana de ácido nucleico. La región de unión a la diana puede comprender una secuencia específica de gen. La unión del oligonucleótido que comprende el complemento de la región de unión a la diana puede comprender ligar el oligonucleótido que comprende el complemento de la región de unión a la diana a la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras.

En algunas realizaciones, para cualquiera de los métodos descritos en la presente, la región de unión a la diana puede comprender una secuencia poli(dT), en donde la unión del oligonucleótido que comprende el complemento de la región de unión a la diana comprende: añadir una pluralidad de monofosfatos de adenosina a la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras usando una desoxinucleotidil transferasa terminal.

En algunas realizaciones, para cualquiera de los métodos descritos en la presente, extender las copias de la diana de ácido nucleico hibridadas con los códigos de barras de oligonucleótidos puede comprender la transcripción inversa de las copias de la diana de ácido nucleico hibridadas con los códigos de barras de oligonucleótidos para

generar una pluralidad de moléculas de ácido desoxirribonucleico complementario (ADNc) identificadas con códigos de barras. Extender las copias de la diana de ácido nucleico hibridadas con los códigos de barras de oligonucleótidos puede comprender extender las copias de la diana de ácido nucleico hibridadas con los códigos de barras de oligonucleótidos usando una ADN polimerasa que carezca de por lo menos una de actividad de exonucleasa de 5' a 3' y actividad de exonucleasa de 3' a 5'. La ADN polimerasa puede comprender un fragmento de Klenow.

En algunas realizaciones, cualquiera de los métodos descritos en la presente comprende: obtener información de secuencia de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras extendidas. Obtener la información de secuencia puede comprender unir adaptadores de secuenciación a la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras extendidas.

En algunas realizaciones, para cualquiera de los métodos descritos en la presente, el complemento de la región de unión a la diana puede comprender la secuencia complementaria inversa de la región de unión a la diana. El complemento de la región de unión a la diana puede comprender la secuencia complementaria de la región de unión a la diana. El complemento del marcador molecular puede comprender una secuencia complementaria inversa del marcador molecular. El complemento del marcador molecular puede comprender una secuencia complementaria del marcador molecular.

En algunas realizaciones, para cualquiera de los métodos descritos en la presente, la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras puede comprender moléculas de ácido desoxirribonucleico (ADN) identificadas con códigos de barras. Las moléculas de ácido nucleico identificadas con códigos de barras pueden comprender moléculas de ácido ribonucleico (ARN) identificadas con códigos de barras. La diana de ácido nucleico puede comprender una molécula de ácido nucleico. La molécula de ácido nucleico puede comprender ácido ribonucleico (ARN), ARN mensajero (ARNm), microARN, ARN interferente pequeño (ARNip), producto de degradación de ARN, ARN que comprende una cola poli(A) o cualquier combinación de los mismos.

En algunas realizaciones, para cualquiera de los métodos descritos en la presente, la diana de ácido nucleico puede comprender un reactivo de unión a componente celular. La molécula de ácido nucleico puede asociarse con el reactivo de unión a componente celular. El método puede comprender: disociar la molécula de ácido nucleico y el reactivo de unión a componente celular.

En algunas realizaciones, para cualquiera de los métodos descritos en la presente, cada marcador molecular de la pluralidad de códigos de barras de oligonucleótidos comprende por lo menos 6 nucleótidos. El código de barras de oligonucleótidos puede comprender un marcador de muestra idéntico. Cada marcador de muestra de la pluralidad de códigos de barras de oligonucleótidos puede comprender por lo menos 6 nucleótidos. El código de barras de oligonucleótidos puede comprender un marcador celular idéntico. Cada marcador celular de la pluralidad de códigos de barras de oligonucleótidos puede comprender por lo menos 6 nucleótidos.

En algunas realizaciones, para cualquiera de los métodos descritos en la presente, por lo menos una de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras se asocia con un soporte sólido al hibridar la región de unión a la diana y el complemento de la región de unión a la diana dentro de cada una de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras para formar la estructura de horquilla. Por lo menos una de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras puede disociarse de un soporte sólido al hibridar la región de unión a la diana y el complemento de la región de unión a la diana dentro de cada una de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras para formar la estructura de horquilla. Por lo menos una de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras puede asociarse con un soporte sólido al hibridar la región de unión a la diana y el complemento de la región de unión a la diana dentro de cada una de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras para formar la estructura de horquilla.

En algunas realizaciones, para cualquiera de los métodos descritos en la presente, por lo menos una de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras está asociada con un soporte sólido cuando se extienden los extremos 3' de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras para extender la estructura de horquilla y generar la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras extendidas que comprenden cada una el marcador molecular y un complemento del marcador molecular. Por lo menos una de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras puede disociarse de un soporte sólido al extender los extremos 3' de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras para extender la estructura de horquilla y generar la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras extendidas que comprenden cada una el marcador molecular y un complemento del marcador molecular. Por lo menos una de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras puede asociarse con un soporte sólido al extender los extremos 3' de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras para extender la estructura de horquilla y generar la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras extendidas que comprenden cada una el marcador molecular y un complemento del marcador molecular. El soporte sólido puede comprender una partícula sintética. El soporte sólido puede comprender una superficie plana (por ejemplo, un portaobjetos, como un portaobjetos de

microscopio y un cubreobjetos). En algunas realizaciones, una solución como la descrita en la presente puede repartirse en una división que no comprenda más de una célula. La división puede comprender por lo menos una microgotita, un pocillo (por ejemplo, un micropocillo), como en un sustrato, o una cámara de un dispositivo fluido, como un dispositivo microfluídico. La microgotita puede comprender un hidrogel.

En algunas realizaciones, para cualquiera de los métodos descritos en la presente, por lo menos una de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras está en solución cuando hibridan la región de unión a la diana y el complemento de la región de unión a la diana dentro de cada una de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras para formar la estructura de horquilla. Por lo menos una de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras puede estar en solución al extender los extremos 3' de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras para extender la estructura de horquilla y generar la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras extendidas, cada una de las cuales comprendiendo el marcador molecular y un complemento del marcador molecular.

En algunas realizaciones, para cualquiera de los métodos descritos en la presente, la muestra comprende una célula individual, el método comprendiendo asociar una partícula sintética que comprende la pluralidad de los códigos de barras de oligonucleótidos con la célula individual de la muestra. El método puede comprender: lisis de la célula individual después de asociar la partícula sintética con la célula individual. La lisis de la célula individual puede comprender calentar la muestra, poner en contacto la muestra con un detergente, cambiar el pH de la muestra o cualquier combinación de los mismos. La partícula sintética y la célula individual pueden estar en el mismo pocillo. La partícula sintética y la célula individual pueden estar en la misma gotita.

En algunas realizaciones, para cualquiera de los métodos descritos en la presente, por lo menos uno de la pluralidad de códigos de barras de oligonucleótidos puede estar inmovilizado en la partícula sintética. Por lo menos uno de la pluralidad de códigos de barras de oligonucleótidos puede estar parcialmente inmovilizado en la partícula sintética. Por lo menos uno de la pluralidad de códigos de barras de oligonucleótidos puede estar encerrado en la partícula sintética. Por lo menos uno de la pluralidad de códigos de barras de oligonucleótidos puede estar parcialmente encerrado en la partícula sintética. La partícula sintética puede ser alterable. La partícula sintética puede comprender una perla. La perla puede comprender una perla de Sepharose, una perla de estreptavidina, una perla de agarosa, una perla magnética, una perla conjugada, una perla conjugada con proteína A, una perla conjugada con proteína G, una perla conjugada con proteína A/G, una perla conjugada con proteína L, una perla conjugada con oligo(dT), una perla de sílice, una perla similar a la sílice, una microesfera antibiotina, una microesfera antifluorocromo, o cualquier combinación de las mismas. La partícula sintética puede comprender un material seleccionado del grupo que consiste en polidimetilsiloxano (PDMS), poliestireno, vidrio, polipropileno, agarosa, gelatina, hidrogel, paramagnético, cerámica, plástico, vidrio, metilestireno, polímero acrílico, titanio, látex, Sepharose, celulosa, nylon, silicona, y cualquier combinación de los mismos. La partícula sintética puede comprender una partícula de hidrogel alterable. Cada uno de la pluralidad de códigos de barras de oligonucleótidos puede comprender un grupo funcional conector. La partícula sintética puede comprender un grupo funcional de soporte sólido. El grupo funcional de soporte y el grupo funcional conector pueden estar asociados entre sí. El grupo funcional conector y el grupo funcional de soporte pueden seleccionarse individualmente del grupo formado por C6, biotina, estreptavidina, aminas primarias, aldehídos, cetonas, y cualquier combinación de los mismos.

En algunas realizaciones, para cualquier método descrito en la presente, la región de unión a la diana comprende una secuencia específica de gen, una secuencia oligo(dT), un multímero aleatorio o cualquier combinación de los mismos. El código de barras de oligonucleótidos puede comprender un marcador de muestra idéntico y/o un marcador celular idéntico. Cada marcador de muestra y/o marcador celular de la pluralidad de códigos de barras de oligonucleótidos puede comprender por lo menos 6 nucleótidos. Cada marcador molecular de la pluralidad de códigos de barras de oligonucleótidos puede comprender por lo menos 6 nucleótidos.

En algunas realizaciones, para cualquier método descrito en la presente, por lo menos uno de la pluralidad de códigos de barras de oligonucleótidos está inmovilizado en la partícula sintética. Por lo menos uno de la pluralidad de códigos de barras de oligonucleótidos puede estar parcialmente inmovilizado en la partícula sintética. Por lo menos uno de la pluralidad de códigos de barras de oligonucleótidos puede estar encerrado en la partícula sintética. Por lo menos uno de la pluralidad de códigos de barras de oligonucleótidos puede estar parcialmente encerrado en la partícula sintética. La partícula sintética puede ser alterable. La partícula sintética puede comprender una perla. La perla puede comprender una perla de Sepharose, una perla de estreptavidina, una perla de agarosa, una perla magnética, una perla conjugada, una perla conjugada con proteína A, una perla conjugada con proteína G, una perla conjugada con proteína A/G, una perla conjugada con proteína L, una perla conjugada con oligo(dT), una perla de sílice, una perla similar a sílice, una microesfera antibiotina, una microesfera antifluorocromo, o cualquier combinación de las mismas. La partícula sintética puede comprender un material seleccionado del grupo que consiste en polidimetilsiloxano (PDMS), poliestireno, vidrio, polipropileno, agarosa, gelatina, hidrogel, paramagnético, cerámica, plástico, vidrio, metilestireno, polímero acrílico, titanio, látex, Sepharose, celulosa, nylon, silicona, y cualquier combinación de los mismos. La partícula sintética puede comprender una partícula de hidrogel alterable. Cada uno de la pluralidad de códigos de barras de oligonucleótidos puede comprender un grupo funcional conector. La partícula

sintética puede comprender un grupo funcional de soporte sólido. El grupo funcional de soporte y el grupo funcional conector pueden estar asociados entre sí. El grupo funcional conector y el grupo funcional soporte pueden seleccionarse individualmente del grupo formado por C6, biotina, estreptavidina, aminas primarias, aldehídos, cetonas, y cualquier combinación de los mismos.

BREVE DESCRIPCIÓN DE LOS DIBUJOS

La FIG. 1 ilustra un código de barras ejemplar no limitativo.

La FIG. 2 muestra un flujo de trabajo ejemplar no limitativo de identificación con códigos de barras y recuento digital.

La FIG. 3 es una ilustración esquemática que muestra un proceso ejemplar no limitativo para generar una biblioteca indexada de dianas identificadas con códigos de barras en los extremos 3' a partir de una pluralidad de dianas.

Las FIGS. 4A-4B muestran una ilustración esquemática de un método ejemplar no limitativo de marcado específico de genes de dianas de ácido nucleico en los extremos 5'.

Las FIGS. 5A-5B muestran una ilustración esquemática de un método ejemplar no limitativo de marcado de dianas de ácido nucleico en los extremos 5' para el análisis del transcriptoma completo.

DESCRIPCIÓN DETALLADA

En la siguiente descripción detallada, se hace referencia a los dibujos acompañantes, que forman parte de la misma. En los dibujos, símbolos similares identifican típicamente componentes similares, a menos que el contexto indique lo contrario. No se pretende que las realizaciones ilustrativas descritas en la descripción detallada, los dibujos y las reivindicaciones sean limitativas.

Cuantificar números pequeños de ácidos nucleicos, por ejemplo moléculas de ácido ribonucleótido mensajero (ARNm), es clínicamente importante para determinar, por ejemplo, los genes que se expresan en una célula en diferentes etapas de desarrollo o bajo diferentes condiciones ambientales. Sin embargo, también puede ser muy difícil determinar el número absoluto de moléculas de ácido nucleico (por ejemplo, moléculas de ARNm), especialmente cuando el número de moléculas es muy pequeño. Un método para determinar el número absoluto de moléculas en una muestra es la reacción en cadena de la polimerasa (PCR) digital. Idealmente, la PCR produce una copia idéntica de una molécula en cada ciclo. Sin embargo, la PCR puede tener desventajas tales como que cada molécula se replica con una probabilidad estocástica, y esta probabilidad varía según el ciclo de PCR y la secuencia del gen, lo que da como resultado un sesgo de la amplificación y mediciones inexactas de la expresión génica. Los códigos de barras estocásticos con marcadores moleculares únicos (también denominados índices moleculares (MI)) pueden usarse para contar el número de moléculas y corregir el sesgo de amplificación. Los códigos de barras estocásticos, como el ensayo Precise™ (Cellular Research, Inc. (Palo Alto, CA)) y el ensayo Rhapsody™ (Becton, Dickinson and Company (Franklin Lakes, NJ)), pueden corregir el sesgo inducido por la PCR y los pasos de preparación de bibliotecas usando marcadores moleculares (ML) para marcar los ARNm durante la transcripción inversa (RT). Hay una necesidad de métodos y técnicas para identificar con códigos de barras moleculares dianas de ácidos nucleicos en uno o ambos de los extremos 5' y los extremos 3'.

El ensayo Precise™ puede utilizar un grupo no agotable de códigos de barras estocásticos con un gran número, por ejemplo de 6561 a 65536, de secuencias de marcadores moleculares únicos en oligonucleótidos poli(T) para hibridar con todos los ARNm poli(A) de una muestra durante el paso de RT. Un código de barras estocástico puede comprender un sitio de cebado de PCR universal. Durante la RT, las moléculas del gen diana reaccionan aleatoriamente con los códigos de barras estocásticos. Cada molécula diana puede hibridar con un código de barras estocástico resultante para generar moléculas de ácido ribonucleótido complementario (ADNc) identificadas con códigos de barras estocásticos. Después del marcado, las moléculas de ADNc identificadas estocásticamente con códigos de barras a partir de los micropocillos de una placa de micropocillos pueden agruparse en un único tubo para la amplificación por PCR y la secuenciación. Los datos brutos de secuenciación pueden analizarse para producir el número de lecturas, el número de códigos de barras estocásticos con secuencias de marcadores moleculares únicos y las cantidades de moléculas de ARNm.

La invención proporciona un método para unir códigos de barras de oligonucleótidos a una diana en una muestra. El método comprende: identificar con códigos de barras copias de una diana de ácido nucleico en una muestra usando una pluralidad de códigos de barras de oligonucleótidos para generar una pluralidad de moléculas de ácido nucleico identificadas con códigos de barras cada una de las cuales comprendiendo una secuencia de la diana de ácido nucleico, un marcador molecular y una región de unión a la diana, y en donde por lo menos 10 de la pluralidad de códigos de barras de oligonucleótidos comprenden diferentes secuencias de marcadores moleculares; unir un oligonucleótido que comprenda un complemento de la región de unión a la diana a la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras para generar una pluralidad de moléculas de ácido nucleico identificadas con códigos de barras, cada una de las cuales comprendiendo la región de unión a la diana y el complemento de la región de unión a la diana; hibridar la región de unión a la diana y el complemento de la región de unión a la diana dentro de cada una de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras para formar una estructura de horquilla; y extender los extremos 3' de la pluralidad de moléculas de ácido nucleico identificadas

con códigos de barras para extender la estructura de horquilla y generar una pluralidad de moléculas de ácido nucleico identificadas con códigos de barras extendidas, cada una de las cuales comprendiendo el marcador molecular y un complemento del marcador molecular.

La invención también proporciona métodos para determinar la cantidad de dianas en una muestra. El método comprende: identificar con códigos de barras copias de una diana de ácido nucleico en una muestra usando una pluralidad de códigos de barras de oligonucleótidos para generar una pluralidad de moléculas de ácido nucleico identificadas con códigos de barras cada una de las cuales comprendiendo una secuencia de la diana de ácido nucleico, un marcador molecular y una región de unión a la diana, y en donde por lo menos 10 de la pluralidad de códigos de barras de oligonucleótidos comprenden diferentes secuencias de marcadores moleculares; unir un oligonucleótido que comprenda un complemento de la región de unión a la diana a la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras para generar una pluralidad de moléculas de ácido nucleico identificadas con códigos de barras, cada una de las cuales comprendiendo la región de unión a la diana y el complemento de la región de unión a la diana; hibridar la región de unión a la diana y el complemento de la región de unión a la diana dentro de cada una de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras para formar una estructura de horquilla; extender los extremos 3' de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras para extender la estructura de horquilla y generar una pluralidad de moléculas de ácido nucleico identificadas con códigos de barras extendidas, cada una de las cuales comprendiendo el marcador molecular y un complemento del marcador molecular; y determinar la cantidad de dianas de ácido nucleico en la muestra basándose en el número de complementos de marcadores moleculares con secuencias distintas asociados a la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras extendidas.

La invención también proporciona métodos para determinar las cantidades de una diana de ácido nucleico en una muestra. El método comprende: poner en contacto copias de una diana de ácido nucleico en una muestra con una pluralidad de códigos de barras de oligonucleótidos, en donde cada uno de la pluralidad de códigos de barras de oligonucleótidos comprende un marcador molecular y una región de unión a la diana capaz de hibridar con la diana de ácido nucleico, y en donde por lo menos 10 de la pluralidad de códigos de barras de oligonucleótidos comprenden diferentes secuencias de marcadores moleculares; extender las copias de la diana de ácido nucleico hibridadas con los códigos de barras de oligonucleótidos para generar una pluralidad de moléculas de ácido nucleico, cada una de las cuales comprendiendo una secuencia complementaria con por lo menos una parte de la diana de ácido nucleico; amplificar la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras para generar una pluralidad de moléculas de ácido nucleico identificadas con códigos de barras amplificadas; unir un oligonucleótido que comprende el complemento de la región de unión a la diana a la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras amplificadas para generar una pluralidad de moléculas de ácido nucleico identificadas con códigos de barras que comprenda cada una la región de unión a la diana y un complemento de la región de unión a la diana; hibridar la región de unión a la diana y el complemento de la región de unión a la diana dentro de cada una de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras para formar una estructura de horquilla; extender los extremos 3' de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras para extender la estructura de horquilla y generar una pluralidad de moléculas de ácido nucleico identificadas con códigos de barras extendidas, cada una de las cuales comprendiendo el marcador molecular y un complemento del marcador molecular; amplificar la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras extendidas para generar una pluralidad de moléculas de ácido nucleico marcadas individualmente, cada una de las cuales comprendiendo el complemento del marcador molecular; y determinar la cantidad de la diana de ácido nucleico en la muestra basándose en el número de complementos de marcadores moleculares con secuencias distintas asociados a la pluralidad de moléculas de ácido nucleico marcadas individualmente.

A menos que se definan de otro modo, los términos técnicos y científicos usados en la presente tienen el mismo significado que el entendido comúnmente por un experto en la técnica a la que pertenece la presente divulgación. Ejemplos de referencias disponibles para los expertos en la técnica incluyen, Singleton et al., Dictionary of Microbiology and Molecular Biology 2ª ed., J. Wiley & Sons (New York, NY 1994); Sambrook et al., Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press (Cold Spring Harbor, NY 1989).

Como se usa en la presente, el término "adaptador" tiene su significado habitual y ordinario en la técnica a la vista de esta memoria descriptiva. Puede referirse a una secuencia para facilitar la amplificación o secuenciación de ácidos nucleicos asociados. Los ácidos nucleicos asociados pueden comprender ácidos nucleicos diana. Los ácidos nucleicos asociados pueden comprender uno o más de los marcadores espaciales, marcadores diana, marcadores de muestra, marcador de indexación o secuencias de código de barras (por ejemplo, marcadores moleculares). Los adaptadores pueden ser lineales. Los adaptadores pueden ser adaptadores preadenilados. Los adaptadores pueden ser de cadena doble o sencilla. Uno o más adaptadores pueden estar situados en el extremo 5' o 3' de un ácido nucleico. Cuando los adaptadores comprenden secuencias conocidas en los extremos 5' y 3', las secuencias conocidas pueden ser secuencias iguales o diferentes. Un adaptador situado en los extremos 5' y/o 3' de un polinucleótido puede ser capaz de hibridar con uno o más oligonucleótidos inmovilizados en una superficie. En algunas realizaciones, un adaptador puede comprender una secuencia universal. Una secuencia universal puede ser una región de secuencia de nucleótidos que es común a dos o más moléculas de ácido nucleico. Las dos o más moléculas de ácido nucleico también pueden tener regiones de secuencia diferente. Así, por ejemplo, los adaptadores 5' pueden

comprender secuencias de ácido nucleico idénticas y/o universales y los adaptadores 3' pueden comprender secuencias idénticas y/o universales. Una secuencia universal que puede estar presente en diferentes miembros de una pluralidad de moléculas de ácido nucleico puede permitir la replicación o amplificación de múltiples secuencias diferentes usando un único cebador universal que sea complementario a la secuencia universal. De manera similar, por lo menos una, dos (por ejemplo, un par) o más secuencias universales que pueden estar presentes en diferentes miembros de una colección de moléculas de ácido nucleico pueden permitir la replicación o amplificación de múltiples secuencias diferentes usando por lo menos uno, dos (por ejemplo, un par) o más cebadores universales individuales que son complementarios a las secuencias universales. Así, un cebador universal incluye una secuencia que puede hibridar con dicha secuencia universal. Las moléculas portadoras de secuencias de ácido nucleico diana pueden modificarse para unir adaptadores universales (por ejemplo, secuencias de ácido nucleico no diana) a uno o ambos extremos de las diferentes secuencias de ácido nucleico diana. Los uno o más cebadores universales unidos al ácido nucleico diana pueden proporcionar sitios para la hibridación de cebadores universales. Los uno o más cebadores universales unidos al ácido nucleico diana pueden ser iguales o diferentes entre sí.

Como se usa en la presente, el término "asociado" o "asociado con" tiene su significado habitual y ordinario en la técnica a la vista de esta memoria descriptiva. Puede significar que dos o más especies son identificables como colocalizadas en un punto en el tiempo. Una asociación puede significar que dos o más especies están o estaban dentro de un contenedor similar. Una asociación puede ser una asociación informática. Por ejemplo, la información digital referente a dos o más especies puede almacenarse y puede usarse para determinar que una o más de las especies estaban colocalizadas en un momento dado. Una asociación también puede ser una asociación física. En algunas realizaciones, dos o más especies asociadas están "atadas", "unidas" o "inmovilizadas" entre sí o a una superficie sólida o semisólida común. Una asociación puede referirse a medios covalentes o no covalentes para unir marcadores a soportes sólidos o semisólidos como perlas. Una asociación puede ser una unión covalente entre una diana y un marcador. Una asociación puede comprender la hibridación entre dos moléculas (como una molécula diana y un marcador).

Como se usa en la presente, el término "complementario" tiene su significado habitual y ordinario en la técnica a la vista de esta memoria descriptiva. Puede referirse a la capacidad de emparejamiento preciso entre dos nucleótidos. Por ejemplo, si un nucleótido en una posición dada de un ácido nucleico es capaz de formar un enlace de hidrógeno con un nucleótido de otro ácido nucleico, entonces se considera que los dos ácidos nucleicos son complementarios entre sí en esa posición. La complementariedad entre dos moléculas de ácido nucleico de cadena sencilla puede ser "parcial", en la que sólo algunos de los nucleótidos se unen, o puede ser completa cuando existe complementariedad total entre las moléculas de cadena sencilla. Se puede decir que una primera secuencia de nucleótidos es el "complemento" de una segunda secuencia si la primera secuencia de nucleótidos es complementaria con la segunda secuencia de nucleótidos. Se puede decir que una primera secuencia de nucleótidos es el "complemento inverso" de una segunda secuencia, si la primera secuencia de nucleótidos es complementaria con una secuencia que es inversa (es decir, el orden de los nucleótidos está invertido) de la segunda secuencia. Como se usa en la presente, una secuencia "complementaria" puede referirse a un "complemento" o a un "complemento inverso" de una secuencia. De la divulgación se entiende que si una molécula puede hibridar con otra molécula puede ser complementaria, o parcialmente complementaria, con la molécula que está hibridando.

Como se usa en la presente, el término "recuento digital" puede referirse a un método para estimar una cantidad de moléculas diana en una muestra. El recuento digital puede incluir el paso de determinar un número de marcadores únicos que se han asociado con objetivos en una muestra. Esta metodología, que puede ser de naturaleza estocástica, transforma el problema del recuento de moléculas de uno de localización e identificación de moléculas idénticas a una serie de preguntas digitales de sí/no referentes a la detección de un conjunto de marcadores predefinidos.

Como se usa en la presente, el término "marcador" o "marcadores" tiene su significado habitual y ordinario en la técnica a la vista de esta memoria descriptiva. Puede referirse a códigos de ácido nucleico asociados a una diana dentro de una muestra. Un marcador puede ser, por ejemplo, un marcador de ácido nucleico. Un marcador puede ser total o parcialmente amplificable. Un marcador puede ser total o parcialmente secuenciable. Un marcador puede ser una porción de un ácido nucleico nativo identificable como distinto. Un marcador puede ser una secuencia conocida. Un marcador puede comprender una unión de secuencias de ácido nucleico, por ejemplo una unión de una secuencia nativa y no nativa. Como se usa en la presente, el término "marcador" puede usarse indistintamente con los términos "índice", "etiqueta" o "marcador-etiqueta". Los marcadores pueden transmitir información. Por ejemplo, en varias realizaciones, los marcadores pueden usarse para determinar la identidad de una muestra, la fuente de una muestra, la identidad de una célula y/o una diana.

Como se usa en la presente, el término "depósitos no agotables" puede referirse a un grupo de códigos de barras (por ejemplo, códigos de barras estocásticos) compuesto por muchos marcadores diferentes. Un depósito no agotable puede comprender un gran número de códigos de barras diferentes, de tal manera que cuando el depósito no agotable se asocia con un grupo de dianas, es probable que cada diana se asocie con un código de barras único. La unicidad de cada molécula diana marcada puede determinarse mediante la estadística de elección aleatoria, y depende del número de copias de moléculas diana idénticas en la colección en comparación con la diversidad de

marcadores. El tamaño del conjunto resultante de moléculas diana marcadas puede determinarse por la naturaleza estocástica del proceso de identificación con códigos de barras, y el análisis del número de códigos de barras detectados permite entonces calcular el número de moléculas diana presentes en la colección o muestra original. Cuando la relación entre el número de copias de una molécula diana presente y el número de códigos de barras únicos es baja, las moléculas diana marcadas son altamente únicas (es decir, hay una probabilidad muy baja de que más de una molécula diana haya sido marcada con un marcador dado).

Como se usa en la presente, el término "ácido nucleico" tiene su significado habitual y ordinario en la técnica a la vista de esta memoria descriptiva. Se refiere a una secuencia de polinucleótidos, o fragmento de la misma. Un ácido nucleico puede comprender nucleótidos. Un ácido nucleico puede ser exógeno o endógeno a una célula. Un ácido nucleico puede existir en un entorno libre de células. Un ácido nucleico puede ser un gen o un fragmento del mismo. Un ácido nucleico puede ser ADN. Un ácido nucleico puede ser ARN. Un ácido nucleico puede comprender uno o más análogos (por ejemplo, una estructura principal, azúcar o nucleobase alterados). Algunos ejemplos no limitativos de análogos incluyen: 5-bromouracilo, ácido nucleico peptídico, ácido xeno nucleico, morfolinos, ácidos nucleicos bloqueados, ácidos nucleicos glicólicos, ácidos nucleicos de treosa, dideoxynucleótidos, cordicepina, 7-deaza-GTP, fluoróforos (por ejemplo, rodamina o fluoresceína enlazada al azúcar), nucleótidos que contienen tioles, nucleótidos enlazados a biotina, análogos de bases fluorescentes, islas CpG, metil-7-guanosina, nucleótidos metilados, inosina, tiouridina, pseudouridina, dihidrouridina, queuosina y wiosina. "Ácido nucleico", "polinucleótido", "polinucleótido diana" y "ácido nucleico diana" pueden usarse indistintamente.

Un ácido nucleico puede comprender una o más modificaciones (por ejemplo, una modificación de la base, una modificación de la estructura principal), para proporcionar al ácido nucleico una característica nueva o mejorada (por ejemplo, estabilidad mejorada). Un ácido nucleico puede comprender una etiqueta de afinidad de ácido nucleico. Un nucleósido puede ser una combinación de base y azúcar. La porción de base del nucleósido puede ser una base heterocíclica. Las dos clases más comunes de tales bases heterocíclicas son las purinas y las pirimidinas. Los nucleótidos pueden ser nucleósidos que incluyen además un grupo fosfato enlazado covalentemente a la porción de azúcar del nucleósido. En el caso de los nucleósidos que incluyen un azúcar de pentofuranosilo, el grupo fosfato puede estar enlazado a la fracción hidroxilo 2', 3' o 5' del azúcar. Al formar ácidos nucleicos, los grupos fosfato pueden enlazar covalentemente nucleósidos adyacentes entre sí para formar un compuesto polimérico lineal. A su vez, los extremos respectivos de este compuesto polimérico lineal pueden unirse de nuevo para formar un compuesto circular; sin embargo, los compuestos lineales son generalmente adecuados. Además, los compuestos lineales pueden tener complementariedad interna de bases nucleotídicas y, por lo tanto, pueden plegarse de manera que produzcan un compuesto total o parcialmente de cadena doble. Dentro de los ácidos nucleicos, puede hacerse referencia comúnmente a los grupos fosfato como que forman la estructura principal internucleosídica del ácido nucleico. El enlace o estructura principal puede ser un enlace fosfodiéster de 3' a 5'.

Un ácido nucleico puede comprender una estructura principal modificada y/o enlaces internucleosídicos modificados. Las estructuras principales modificadas pueden incluir aquellas que retienen un átomo de fósforo en la estructura principal y aquellas que no tienen un átomo de fósforo en la estructura principal. Las estructuras principales de ácidos nucleicos modificadas adecuadas que contienen un átomo de fósforo pueden incluir, por ejemplo, fosforotioatos, fosforotioatos quirales, fosforoditioatos, fosfotriesteres, aminoalquilfosfotriesteres, fosfonatos de metilo y otros alquilo, como fosfonatos de 3'-alquileo, fosfonatos de 5'-alquileo, fosfonatos quirales, fosfinatos, fosforamidatos, incluyendo 3'-aminofosforamidato y aminoalquilfosforamidatos, fosforodiamidatos, tionofosforamidatos, tionoalquilfosfonatos, tionoalquilfosfotriesteres, selenofosfatos y boranofosfatos con enlaces normales 3'-5', análogos 2'-5' enlazados y aquellos que tienen polaridad invertida en los que uno o más enlaces internucleotídicos son un enlace de 3' a 3', de 5' a 5' o de 2' a 2'.

Un ácido nucleico puede comprender estructuras principales de polinucleótidos que están formadas por enlaces internucleosídicos de alquilo o cicloalquilo de cadena corta, enlaces internucleosídicos mixtos de heteroátomo y alquilo o cicloalquilo, o uno o más enlaces internucleosídicos heteroatómicos o heterocíclicos de cadena corta. Estos pueden incluir los que tienen enlaces morfolinos (formados en parte a partir de la porción de azúcar de un nucleósido); estructuras principales de siloxano; estructuras principales de sulfuro, sulfóxido y sulfona; estructuras principales de formacetilo y tioformacetilo; estructuras principales de metileno formacetilo y tioformacetilo; estructuras principales de riboacetilo; estructuras principales que contienen alqueno; estructuras principales de sulfamato; estructuras principales de metilenimino y metilenhidrazino; estructuras principales de sulfonato y sulfonamida; estructuras principales de amida; y otras que tienen partes componentes de N, O, S y CH₂ mixtas.

Un ácido nucleico puede comprender un mimético de ácido nucleico. Puede pretenderse que el término "mimético" incluya polinucleótidos en los que sólo el anillo de furanosa o tanto el anillo de furanosa como el enlace internucleotídico se sustituyen por grupos no de furanosa; la sustitución de sólo el anillo de furanosa también puede denominarse sustituto de azúcar. La fracción de base heterocíclica o una fracción de base heterocíclica modificada puede mantenerse para la hibridación con un ácido nucleico diana apropiado. Uno de tales ácidos nucleicos puede ser un ácido nucleico peptídico (PNA). En un PNA, la estructura principal de azúcar de un polinucleótido puede sustituirse por una estructura principal que contenga amida, en particular una estructura principal de aminoetilglicina. Los nucleótidos pueden conservarse y unirse directa o indirectamente a los átomos de nitrógeno aza de la porción

amida de la estructura principal. La estructura principal de los compuestos de PNA puede comprender dos o más unidades de aminoetilglicina enlazadas, lo que confiere al PNA una estructura principal que contiene amida. Las moléculas de base heterocíclica pueden unirse directa o indirectamente a los átomos de nitrógeno aza de la porción amida de la estructura principal.

Un ácido nucleico puede comprender una estructura principal de morfolino. Por ejemplo, un ácido nucleico puede comprender un anillo morfolino de 6 miembros en lugar de un anillo ribosa. En algunas de estas realizaciones, un fosforodiamidato u otro enlace internucleosídico no fosfodiéster puede sustituir a un enlace fosfodiéster.

Un ácido nucleico puede comprender unidades de morfolino enlazadas (por ejemplo, ácido nucleico morfolino) que tienen bases heterocíclicas unidas al anillo de morfolino. Los grupos de enlace pueden enlazar las unidades monoméricas de morfolino en un ácido nucleico de morfolino. Los compuestos oligoméricos no iónicos basados en morfolinos pueden tener menos interacciones no deseadas con las proteínas celulares. Los polinucleótidos basados en morfolinos pueden ser imitadores no iónicos de ácidos nucleicos. Una variedad de compuestos dentro de la clase morfolino pueden unirse usando diferentes grupos de enlace. Una clase adicional de polinucleótidos miméticos puede denominarse ácidos nucleicos de ciclohexenilo (CeNA). El anillo de furanosa presente normalmente en una molécula de ácido nucleico puede sustituirse por un anillo de ciclohexenilo. Pueden prepararse monómeros de fosforamidita protegidos con DMT de CeNA y usarse para la síntesis de compuestos oligoméricos mediante química de fosforamidita. La incorporación de monómeros de CeNA en una cadena de ácido nucleico puede aumentar la estabilidad de un híbrido ADN/ARN. Los oligoadenilatos de CeNA pueden formar complejos con complementos de ácido nucleico con una estabilidad similar a la de los complejos nativos. Una modificación adicional puede incluir los ácidos nucleicos bloqueados (LNA) en los que el grupo 2'-hidroxilo está enlazado al átomo de carbono 4' del anillo de azúcar formando de este modo un enlace 2'-C, 4'-C-oximetileno formando de este modo una fracción de azúcar bicíclico. El enlace puede ser un grupo metileno (-CH₂), que forma un puente entre el átomo de oxígeno 2' y el átomo de carbono 4' en donde n es 1 o 2. El LNA y los análogos del LNA pueden mostrar estabilidades térmicas dúplex muy altas con el ácido nucleico complementario (Tm=+3 a +10°C), estabilidad frente a la degradación 3'-exonucleolítica y buenas propiedades de solubilidad.

Un ácido nucleico también puede incluir modificaciones o sustituciones de nucleobases (a menudo denominadas simplemente "bases"). Como se usa en la presente, las nucleobases "no modificadas" o "naturales" pueden incluir las bases de purina, (por ejemplo, adenina (A) y guanina (G)), y las bases de pirimidina, (por ejemplo, timina (T), citosina (C) y uracilo (U)). Las nucleobases modificadas pueden incluir otras nucleobases sintéticas y naturales, como la 5-metilcitosina (5-me-C), la 5-hidroximetilcitosina, la xantina, la hipoxantina, la 2-aminoadenina, la 6-metil y otros derivados alquílicos de la adenina y la guanina, 2-propilo y otros derivados alquílicos de la adenina y la guanina, 2-tiouracilo, 2-tiotimina y 2-tiocitosina, 5-halouracilo y citosina, 5-propinil (-C≡C-CH₃) uracilo y citosina y otros derivados alquílicos de las bases de pirimidínicas, 6-azo uracilo, citosina y timina, 5-uracilo (pseudouracilo), 4-tiouracilo, 8-halo, 8-amino, 8-tiol, 8-tioalquilo, 8-hidroxilo y otras adeninas y guaninas 8-sustituidas, 5-halo particularmente 5-bromo, 5-trifluorometilo y otros uracilos y citosinas 5-sustituidos, 7-metilguanina y 7-metiladenina, 2-F-adenina, 2-aminoadenina, 8-azaguanina y 8-azaadenina, 7-deazaguanina y 7-deazaadenina y 3-deazaguanina y 3-deazaadenina. Las nucleobases modificadas pueden incluir pirimidinas tricíclicas como la fenoxazina citidina (1H-pirimido(5,4-b)(1,4)benzoxazin-2(3H)-ona), fenotiazina citidina (1H-pirimido(5,4-b)(1,4)benzotiazin-2(3H)-ona), abrazaderas G como una fenoxazina citidina sustituida (por ejemplo, 9-(2-aminoetoxi)-H-pirimido(5,4-b)(1,4)benzoxazin-2(3H)-ona), fenotiazina citidina (1H-pirimido(5,4-b)(1,4)benzotiazin-2(3H)-ona), abrazaderas G como una fenoxazina citidina sustituida (por ejemplo, 9-(2-aminoetoxi)-H-pirimido(5,4-b)(1,4)benzoxazin-2(3H)-ona), citidina carbazol (2H-pirimido(4,5-b)indol-2-ona), citidina piridoindol (H-pirido(3',2':4,5)pirrolo[2,3-d]pirimidin-2-ona).

Como se usa en la presente, el término "muestra" tiene su significado habitual y ordinario en la técnica a la vista de esta memoria descriptiva. Puede referirse a una composición que comprende dianas. Las muestras adecuadas para el análisis mediante los métodos, dispositivos y sistemas divulgados incluyen células, tejidos, órganos u organismos. En algunas realizaciones, la muestra comprende una célula individual. En algunas realizaciones, la muestra comprende, consiste esencialmente en, o consiste en por lo menos 100.000, 200.000, 300.000, 500.000, 800.000, o 1.000.000 células individuales.

Como se usa en la presente, el término "dispositivo de muestreo" o "dispositivo" puede referirse a un dispositivo que puede tomar una sección de una muestra y/o colocar la sección en un sustrato. Un dispositivo de muestreo puede referirse, por ejemplo, a una máquina de clasificación celular activada por fluorescencia (FACS), una máquina clasificadora de células, una aguja de biopsia, un dispositivo de biopsia, un dispositivo de seccionamiento de tejidos, un dispositivo microfluidico, una rejilla de cuchillas y/o un micrótopo.

Como se usa en la presente, el término "soporte sólido" tiene su significado habitual y ordinario en la técnica a la vista de esta memoria descriptiva. Puede referirse a superficies sólidas o semisólidas discretas a las que pueden unirse una pluralidad de códigos de barras (por ejemplo, códigos de barras estocásticos). Un soporte sólido puede abarcar cualquier tipo de esfera, bola, cojinete, cilindro u otra configuración similar sólidos, porosos o huecos compuestos de plástico, cerámica, metal o material polimérico (por ejemplo, hidrogel) sobre los que pueden inmovilizarse un ácido nucleico (por ejemplo, covalente o no covalentemente). Un soporte sólido puede comprender

una partícula discreta que puede ser esférica (por ejemplo, microesferas) o tener una forma no esférica o irregular, como cúbica, cuboide, piramidal, cilíndrica, cónica, oblonga o en forma de disco, y similares. Una perla puede tener una forma no esférica. Una pluralidad de soportes sólidos espaciados en una matriz puede no comprender un sustrato. Un soporte sólido puede usarse indistintamente con el término "perla". Se contempla que siempre que en la presente se describa un soporte sólido como una partícula o superficie (por ejemplo, si un código de barras está inmovilizado en un soporte sólido, partícula, perla o similar), otra opción es repartir las soluciones en divisiones, para asociar los marcadores celulares uno a uno con las divisiones (y por tanto asociar los marcadores celulares uno a uno con las células individuales en las divisiones). Las divisiones de ejemplo pueden incluir gotitas como microgotitas, pocillos (por ejemplo, micropocillos) que pueden estar en un sustrato como una placa de múltiples pocillos, y cámaras en un dispositivo fluídico (como un dispositivo microfluídico). En algunas realizaciones, una solución como la descrita en la presente puede repartirse en una división que comprende no más de una célula. La división puede comprender por lo menos una microgotita, un micropocillo o una cámara de un dispositivo fluídico, como un dispositivo microfluídico. La microgotita puede comprender un hidrogel. Los códigos de barras pueden estar inmovilizados en la división del sustrato, o pueden estar libres en solución en la división.

Como se usa en la presente, el término "código de barras estocástico" puede referirse a una secuencia de polinucleótidos que comprende marcadores de la presente divulgación. Un código de barras estocástico puede ser una secuencia de polinucleótidos que puede usarse para identificar con códigos de barras estocásticos. Los códigos de barras estocásticos pueden usarse para cuantificar dianas dentro de una muestra. Los códigos de barras estocásticos pueden usarse para controlar los errores que pueden producirse después de que un marcador se haya asociado a una diana. Por ejemplo, un código de barras estocástico puede usarse para evaluar errores de amplificación o secuenciación. Un código de barras estocástico asociado a una diana puede denominarse código de barras estocástico-diana o código de barras estocástico-etiqueta-diana.

Como se usa en la presente, el término "código de barras estocástico específico de gen" puede referirse a una secuencia de polinucleótidos que comprende marcadores y una región de unión a la diana que es específica de un gen. Un código de barras estocástico puede ser una secuencia de polinucleótidos que puede usarse para la identificación con códigos de barras estocásticos. Los códigos de barras estocásticos pueden usarse para cuantificar dianas dentro de una muestra. Los códigos de barras estocásticos pueden usarse para controlar los errores que pueden producirse después de asociar un marcador a una diana. Por ejemplo, un código de barras estocástico puede usarse para evaluar errores de amplificación o secuenciación. Un código de barras estocástico asociado a una diana puede denominarse código de barras estocástico-diana o código de barras estocástico-etiqueta-diana.

Como se usa en la presente, el término "identificación con códigos de barras estocásticos" puede referirse al marcado aleatorio (por ejemplo, identificación con códigos de barras) de ácidos nucleicos. La identificación con códigos de barras estocásticos puede utilizar una estrategia de Poisson recursiva para asociar y cuantificar marcadores asociados a dianas. Como se usa en la presente, el término "identificación con códigos de barras estocásticos" puede usarse indistintamente con "marcado estocástico".

Como se usa en la presente, el término "diana" tiene su significado habitual y ordinario en la técnica a la vista de esta memoria descriptiva. Puede referirse a una composición que puede asociarse con un código de barras (por ejemplo, un código de barras estocástico). Las dianas ejemplares adecuadas para el análisis mediante los métodos, dispositivos y sistemas divulgados incluyen oligonucleótidos, ADN, ARN, ARNm, microARN, ARNt y similares. Las dianas pueden ser de cadena simple o doble. En algunas realizaciones, las dianas pueden ser proteínas, péptidos o polipéptidos. En algunas realizaciones, las dianas son lípidos. En la presente, "diana" puede usarse indistintamente con "especie".

Como se usa en la presente, el término "transcriptasas inversas" tiene su significado habitual y ordinario en la técnica a la vista de esta memoria descriptiva. Puede referirse a un grupo de enzimas que tienen actividad de transcriptasa inversa (es decir, que catalizan la síntesis de ADN a partir de una plantilla de ARN). En general, dichas enzimas incluyen, entre otras, transcriptasa inversa retroviral, transcriptasa inversa retrotransposónica, transcriptasas inversas retroplásmidas, transcriptasas inversas retrónicas, transcriptasas inversas bacterianas, transcriptasas inversas derivadas de intrones del grupo II y mutantes, variantes o derivados de las mismas. Las transcriptasas inversas no retrovirales incluyen transcriptasas inversas de retrotransposones no LTR, transcriptasas inversas de retroplásmidos, transcriptasas inversas de retrones y transcriptasas inversas de intrones del grupo II. Ejemplos de transcriptasas inversas de intrones del grupo II incluyen la transcriptasa inversa de intrones LI.LtrB de *Lactococcus lactis*, la transcriptasa inversa de intrones Tel4c de *Thermosynechococcus elongatus* o la transcriptasa inversa de intrones Gsl-IIC de *Geobacillus stearothermophilus*. Otras clases de transcriptasas inversas pueden incluir muchas clases de transcriptasas inversas no retrovirales (es decir, retrones, intrones del grupo II y retroelementos generadores de diversidad, entre otros).

Los términos "cebador adaptador universal", "adaptador cebador universal" o "secuencia adaptadora universal" se usan indistintamente para referirse a una secuencia de nucleótidos que puede usarse para hibridar con códigos de barras (por ejemplo, códigos de barras estocásticos) para generar códigos de barras específicos de genes. Una secuencia adaptadora universal puede, por ejemplo, ser una secuencia conocida que es universal en todos los

códigos de barras usados en los métodos de la divulgación. Por ejemplo, cuando se están marcando múltiples dianas usando los métodos divulgados en la presente, cada una de las secuencias específicas de la diana puede enlazarse a la misma secuencia adaptadora universal. En algunas realizaciones, en los métodos divulgados en la presente puede usarse más de una secuencia adaptadora universal. Por ejemplo, cuando se están marcando múltiples dianas usando los métodos divulgados en la presente, por lo menos dos de las secuencias específicas de la diana se enlazan a diferentes secuencias adaptadoras universales. Un cebador adaptador universal y su complemento pueden estar incluidos en dos oligonucleótidos, uno de los cuales comprende una secuencia específica de diana y el otro comprende un código de barras. Por ejemplo, una secuencia adaptadora universal puede formar parte de un oligonucleótido que comprende una secuencia específica de la diana para generar una secuencia de nucleótidos que sea complementaria a un ácido nucleico diana. Un segundo oligonucleótido que comprende un código de barras y una secuencia complementaria de la secuencia adaptadora universal puede hibridar con la secuencia de nucleótidos y generar un código de barras específico de la diana (por ejemplo, un código de barras estocástico específico de la diana). En algunas realizaciones, un cebador adaptador universal tiene una secuencia que es diferente de un cebador de PCR universal usado en los métodos de la presente divulgación.

Códigos de barras

La identificación con códigos de barras, como la identificación con códigos de barras estocásticos, se ha descrito, por ejemplo, en los documentos US 2015/0299784, WO 2015/031691, y Fu et al, Proc Natl Acad Sci U.S.A. 31 de mayo de 2011;108(22):9026-31. En algunas realizaciones, el código de barras divulgado en la presente puede ser un código de barras estocástico que puede ser una secuencia de polinucleótidos que puede usarse para marcar estocásticamente (por ejemplo, código de barras, etiqueta) una diana. Los códigos de barras pueden denominarse códigos de barras estocásticos si la relación entre el número de secuencias de código de barras diferentes de los códigos de barras estocásticos y el número de apariciones de cualquiera de las dianas que se van a marcar puede ser, o ser de aproximadamente, 1:1, 2:1, 3:1, 4:1, 5:1, 6:1, 7:1, 8:1, 9:1, 10:1, 11:1, 12:1, 13:1, 14:1, 15:1, 16:1, 17:1, 18:1, 19:1, 20:1, 30:1, 40:1, 50:1, 60:1, 70:1, 80:1, 90:1, 100:1, o un número o un intervalo entre dos cualesquiera de estos valores. Una diana puede ser una especie de ARNm que comprende moléculas de ARNm con secuencias idénticas o casi idénticas. Los códigos de barras pueden denominarse códigos de barras estocásticos si la relación entre el número de secuencias de código de barras diferentes de los códigos de barras estocásticos y el número de apariciones de cualquiera de las dianas que se van a marcar es por lo menos, o como máximo, de 1:1, 2:1, 3:1, 4:1, 5:1, 6:1, 7:1, 8:1, 9:1, 10:1, 11:1, 12:1, 13:1, 14:1, 15:1, 16:1, 17:1, 18:1, 19:1, 20:1, 30:1, 40:1, 50:1, 60:1, 70:1, 80:1, 90:1, o 100:1. Las secuencias de códigos de barras de los códigos de barras estocásticos pueden denominarse marcadores moleculares.

Un código de barras, por ejemplo un código de barras estocástico, puede comprender uno o más marcadores. Los marcadores ejemplares pueden incluir un marcador universal, un marcador celular, una secuencia de código de barras (por ejemplo, un marcador molecular), un marcador de muestra, un marcador de placa, un marcador espacial y/o un marcador preespacial. La FIG. 1 ilustra un código de barras 104 ejemplar con un marcador espacial. El código de barras 104 puede comprender una amina 5' que puede enlazar el código de barras a un soporte sólido 105. El código de barras puede comprender un marcador universal, un marcador de dimensión, un marcador espacial, un marcador celular y/o un marcador molecular. El código de barras puede comprender un marcador universal, un marcador celular y un marcador molecular. El código de barras puede comprender un marcador universal, un marcador espacial, un marcador celular y un marcador molecular. El código de barras puede comprender un marcador universal, un marcador de dimensión, un marcador celular, y un marcador molecular. El orden de los diferentes marcadores (incluyendo pero no limitándose al marcador universal, el marcador de dimensión, el marcador espacial, el marcador celular, y/o el marcador molecular) en el código de barras puede variar. Por ejemplo, como se muestra en la FIG. 1, el marcador universal puede ser el marcador más 5', y el marcador molecular puede ser el marcador más 3'. El marcador espacial, el marcador de dimensión y el marcador celular pueden estar en cualquier orden. En algunas realizaciones, el marcador universal, el marcador espacial, el marcador de dimensión, el marcador celular y el marcador molecular están en cualquier orden. El código de barras puede incluir una región de unión a la diana. La región de unión a la diana puede interactuar con una diana (por ejemplo, ácido nucleico, ARN, ARNm, ADN diana) en una muestra. Por ejemplo, una región de unión a la diana puede comprender una secuencia oligo(dT) que puede interactuar con colas poli(A) de ARNm. En algunos casos, los marcadores del código de barras (por ejemplo, el marcador universal, el marcador de dimensión, el marcador espacial, el marcador celular y la secuencia del código de barras) pueden estar separados por 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 o 20 nucleótidos o más.

Un marcador, por ejemplo el marcador celular, puede comprender un conjunto único de subsecuencias de ácido nucleico de longitud definida, por ejemplo, de siete nucleótidos cada una (equivalente al número de bits usados en algunos códigos de corrección de errores Hamming), que pueden diseñarse para proporcionar capacidad de corrección de errores. El conjunto de subsecuencias de corrección de errores comprende siete secuencias de nucleótidos puede diseñarse de tal manera que cualquier combinación de secuencias por pares en el conjunto muestre una "distancia genética" definida (o número de bases mal apareadas); por ejemplo, un conjunto de subsecuencias de corrección de errores puede diseñarse para que muestre una distancia genética de tres nucleótidos. En este caso, la revisión de las secuencias de corrección de errores en el conjunto de datos de secuencia para moléculas de ácido nucleico diana marcadas (descritas más detalladamente a continuación) puede permitir detectar o corregir errores de

amplificación o secuenciación. En algunas realizaciones, la longitud de las subsecuencias de ácido nucleico usadas para crear códigos de corrección de errores puede variar, por ejemplo, pueden tener, o tener aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30, 31, 40, 50, o un número o un intervalo entre cualquiera de estos dos valores, nucleótidos de longitud. En algunas realizaciones, pueden usarse subsecuencias de ácido nucleico de otras longitudes para crear códigos de corrección de errores.

El código de barras puede comprender una región de unión a la diana. La región de unión a la diana puede interactuar con una diana en una muestra. La diana puede ser, o comprender, ácidos ribonucleicos (ARN), ARN mensajeros (ARNm), microARN, ARN interferentes pequeños (ARNip), productos de degradación de ARN, ARN que comprenden cada uno una cola poli(A), o cualquier combinación de los mismos. En algunas realizaciones, la pluralidad de dianas puede incluir ácidos desoxirribonucleicos (ADN).

En algunas realizaciones, una región de unión a la diana puede comprender una secuencia oligo(dT) que puede interactuar con colas poli(A) de ARNm. Uno o más de los marcadores del código de barras (por ejemplo, el marcador universal, el marcador de dimensión, el marcador espacial, el marcador celular y las secuencias del código de barras (por ejemplo, el marcador molecular)) pueden estar separadas por un espaciador de otro u otros dos de los marcadores restantes del código de barras. El espaciador puede tener, por ejemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, o 20, o más nucleótidos. En algunas realizaciones, ninguno de los marcadores del código de barras está separado por un espaciador.

Marcadores universales

Un código de barras puede comprender uno o más marcadores universales. En algunas realizaciones, los uno o más marcadores universales pueden ser los mismos para todos los códigos de barras en el conjunto de códigos de barras unidos a un soporte sólido dado. En algunas realizaciones, los uno o más marcadores universales pueden ser los mismos para todos los códigos de barras unidos a una pluralidad de perlas. En algunas realizaciones, un marcador universal puede comprender una secuencia de ácido nucleico capaz de hibridar con un cebador de secuenciación. Los cebadores de secuenciación pueden usarse para secuenciar códigos de barras que comprenden un marcador universal. Los cebadores de secuenciación (por ejemplo, cebadores de secuenciación universales) pueden comprender cebadores de secuenciación asociados con plataformas de secuenciación de alto rendimiento. En algunas realizaciones, un marcador universal puede comprender una secuencia de ácido nucleico capaz de hibridar con un cebador de PCR. En algunas realizaciones, el marcador universal puede comprender una secuencia de ácido nucleico capaz de hibridar con un cebador de secuenciación y un cebador de PCR. La secuencia de ácido nucleico del marcador universal capaz de hibridar con un cebador de secuenciación o de PCR puede denominarse sitio de unión a cebador. Un marcador universal puede comprender una secuencia que puede usarse para iniciar la transcripción del código de barras. Un marcador universal puede comprender una secuencia que puede usarse para la extensión del código de barras o de una región dentro del código de barras. Un marcador universal puede tener, o tener aproximadamente, 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, o un número o un intervalo entre dos cualesquiera de estos valores, nucleótidos de longitud. Por ejemplo, un marcador universal puede comprender por lo menos aproximadamente 10 nucleótidos. Un marcador universal puede tener por lo menos, o como máximo, 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 100, 200 o 300 nucleótidos de longitud. En algunas realizaciones, puede formar parte de la secuencia del marcador universal un conector escindible o nucleótido modificado para permitir que el código de barras se escinda del soporte.

Marcadores de dimensión

Un código de barras puede comprender uno o más marcadores de dimensión. En algunas realizaciones, un marcador de dimensión puede comprender una secuencia de ácido nucleico que proporciona información sobre una dimensión en la que se produjo el marcado (por ejemplo, marcado estocástico). Por ejemplo, un marcador de dimensión puede proporcionar información sobre el momento en el que se identificó con código de barras una diana. Un marcador de dimensión puede estar asociado a un momento de la identificación con códigos de barras (por ejemplo, identificación con códigos de barras estocásticos) en una muestra. Un marcador de dimensión puede activarse en el momento del marcado. Diferentes marcadores de dimensión pueden activarse en diferentes momentos. El marcador de dimensión proporciona información sobre el orden en el que se marcaron las dianas, grupos de dianas y/o muestras. Por ejemplo, una población de células puede ser identificada con códigos de barras en la fase G0 del ciclo celular. Las células pueden ser pulsadas de nuevo con códigos de barras (por ejemplo, códigos de barras estocásticos) en la fase G1 del ciclo celular. Las células pueden ser pulsadas de nuevo con códigos de barras en la fase S del ciclo celular, y así sucesivamente. Los códigos de barras en cada pulso (por ejemplo, cada fase del ciclo celular), pueden comprender diferentes marcadores de dimensión. De esta manera, el marcador de dimensión proporciona información sobre qué dianas se marcaron y en qué fase del ciclo celular. Los marcadores de dimensión pueden interrogar muchos momentos biológicos diferentes. Los momentos biológicos ejemplares pueden incluir, entre otros, el ciclo celular, la transcripción (por ejemplo, el inicio de la transcripción) y la degradación de la transcripción. En otro ejemplo, una muestra (por ejemplo, una célula, una población de células) puede marcarse antes y/o después del tratamiento con un fármaco y/o terapia. Los cambios en el número de copias de distintas dianas pueden ser indicativos de la respuesta de la muestra al fármaco y/o terapia.

Un marcador de dimensión puede ser activable. Un marcador de dimensión activable puede activarse en un punto temporal específico. El marcador activable puede, por ejemplo, activarse constitutivamente (por ejemplo, no apagarse). El marcador de dimensión activable puede, por ejemplo, activarse reversiblemente (por ejemplo, el marcador de dimensión activable puede encenderse y apagarse). El marcador de dimensión puede ser reversiblemente activable, por ejemplo, por lo menos 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, o más veces. El marcador de dimensión puede ser reversiblemente activable, por ejemplo, por lo menos 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más veces. En algunas realizaciones, el marcador de dimensión puede activarse con fluorescencia, luz, un evento químico (por ejemplo, escisión, ligadura de otra molécula, adición de modificaciones (por ejemplo, pegilado, sumoilado, acetilado, metilado, desacetilado, desmetilado), un evento fotoquímico (por ejemplo, fotoencapsulación) y la introducción de un nucleótido no natural.

El marcador de dimensión puede, en algunas realizaciones, ser idéntico para todos los códigos de barras (por ejemplo, códigos de barras estocásticos) unidos a un soporte sólido dado (por ejemplo, una perla), pero diferente para diferentes soportes sólidos (por ejemplo, perlas). En algunas realizaciones, por lo menos el 60%, 70%, 80%, 85%, 90%, 95%, 97%, 99% o 100% de los códigos de barras en el mismo soporte sólido pueden comprender el mismo marcador de dimensión. En algunas realizaciones, por lo menos el 60% de los códigos de barras en el mismo soporte sólido pueden incluir el mismo marcador de dimensión. En algunas realizaciones, por lo menos el 95% de los códigos de barras en el mismo soporte sólido pueden incluir el mismo marcador de dimensión.

En una pluralidad de soportes sólidos (por ejemplo, perlas) puede haber hasta 10^6 o más secuencias de marcadores de dimensión únicos representadas. Un marcador de dimensión puede tener, o tener aproximadamente 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, o un número o un intervalo entre dos cualesquiera de estos valores, nucleótidos de longitud. Un marcador de dimensión puede tener por lo menos, o como máximo, 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 100, 200 o 300 nucleótidos de longitud. Un marcador de dimensión puede comprender entre aproximadamente 5 y aproximadamente 200 nucleótidos. Un marcador de dimensión puede comprender entre aproximadamente 10 y aproximadamente 150 nucleótidos. Un marcador de dimensión puede comprender entre aproximadamente 20 y aproximadamente 125 nucleótidos de longitud.

Marcadores espaciales

Un código de barras puede comprender uno o más marcadores espaciales. En algunas realizaciones, un marcador espacial puede comprender una secuencia de ácido nucleico que proporciona información sobre la orientación espacial de una molécula diana que está asociada con el código de barras. Un marcador espacial puede asociarse a una coordenada en una muestra. La coordenada puede ser una coordenada fija. Por ejemplo, una coordenada puede ser fija con referencia a un sustrato. Un marcador espacial puede estar en referencia a una cuadrícula bidimensional o tridimensional. Una coordenada puede fijarse con referencia a un punto de referencia. El punto de referencia puede ser identificable en el espacio. Un punto de referencia puede ser una estructura de la que pueden obtenerse imágenes. Un punto de referencia puede ser una estructura biológica, por ejemplo un punto de referencia anatómico. Un punto de referencia puede ser un punto de referencia celular, por ejemplo un orgánulo. Un punto de referencia puede ser un punto de referencia no natural, como una estructura con un identificador identificable, como un código de color, un código de barras, una propiedad magnética, fluorescentes, radiactividad, o un tamaño o forma únicos. Un marcador espacial puede estar asociado con una división física (por ejemplo, un pocillo, un recipiente o una gotita). En algunas realizaciones, se usan múltiples marcadores espaciales juntos para codificar una o más posiciones en el espacio.

El marcador espacial puede ser idéntico para todos los códigos de barras unidos a un soporte sólido dado (por ejemplo, una perla), pero diferente para diferentes soportes sólidos (por ejemplo, perlas). En algunas realizaciones, el porcentaje de códigos de barras en el mismo soporte sólido que comprenden el mismo marcador espacial puede ser, o ser aproximadamente, del 60%, 70%, 80%, 85%, 90%, 95%, 97%, 99%, 100%, o un número o un intervalo entre dos cualesquiera de estos valores. En algunas realizaciones, el porcentaje de códigos de barras en el mismo soporte sólido que comprenden el mismo marcador espacial puede ser como mínimo, o como máximo, del 60%, 70%, 80%, 85%, 90%, 95%, 97%, 99% o 100%. En algunas realizaciones, por lo menos el 60% de los códigos de barras del mismo soporte sólido pueden incluir el mismo marcador espacial. En algunas realizaciones, por lo menos el 95% de los códigos de barras del mismo soporte sólido pueden incluir el mismo marcador espacial.

Puede haber hasta 10^6 o más secuencias de marcadores espaciales únicos representados en una pluralidad de soportes sólidos (por ejemplo, perlas). Un marcador espacial puede tener, o tener aproximadamente, 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, o un número o un intervalo entre dos cualesquiera de estos valores, nucleótidos de longitud. Un marcador espacial puede tener por lo menos o como máximo 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 100, 200 o 300 nucleótidos de longitud. Un marcador espacial puede comprender entre aproximadamente 5 y aproximadamente 200 nucleótidos. Un marcador espacial puede tener entre aproximadamente 10 y aproximadamente 150 nucleótidos. Un marcador espacial puede tener entre aproximadamente 20 y aproximadamente 125 nucleótidos de longitud.

Marcadores celulares

Un código de barras (por ejemplo, un código de barras estocástico) puede comprender uno o más marcadores celulares. En algunas realizaciones, un marcador celular puede comprender una secuencia de ácido nucleico que proporciona información para determinar qué ácido nucleico diana se originó a partir de qué célula. En algunas realizaciones, el marcador celular es idéntico para todos los códigos de barras unidos a un soporte sólido dado (por ejemplo, una perla), pero diferente para diferentes soportes sólidos (por ejemplo, perlas). En algunas realizaciones, el porcentaje de códigos de barras en el mismo soporte sólido que comprenden el mismo marcador celular puede ser, o ser aproximadamente, el 60%, 70%, 80%, 85%, 90%, 95%, 97%, 99%, 100%, o un número o un intervalo entre dos cualesquiera de estos valores. En algunas realizaciones, el porcentaje de códigos de barras en el mismo soporte sólido que comprenden el mismo marcador celular puede ser, o ser de aproximadamente, el 60%, 70%, 80%, 85%, 90%, 95%, 97%, 99%, o 100%. Por ejemplo, por lo menos el 60% de los códigos de barras en el mismo soporte sólido pueden comprender el mismo marcador celular. Como otro ejemplo, por lo menos el 95% de los códigos de barras en el mismo soporte sólido pueden comprender el mismo marcador celular.

Puede haber tantas como 10^6 o más secuencias de marcadores celulares únicos representadas en una pluralidad de soportes sólidos (por ejemplo, perlas). Un marcador celular puede tener, o tener aproximadamente, 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, o un número o un intervalo entre dos cualesquiera de estos valores, nucleótidos de longitud. Un marcador celular puede tener por lo menos, o como máximo, 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 100, 200 o 300 nucleótidos de longitud. Por ejemplo, un marcador celular puede comprender entre aproximadamente 5 y aproximadamente 200 nucleótidos. Como otro ejemplo, un marcador celular puede comprender entre aproximadamente 10 y aproximadamente 150 nucleótidos. Como otro ejemplo, un marcador celular puede comprender entre aproximadamente 20 y aproximadamente 125 nucleótidos de longitud.

Secuencias de códigos de barras

Un código de barras puede comprender una o más secuencias de código de barras. En algunas realizaciones, una secuencia de código de barras puede comprender una secuencia de ácido nucleico que proporciona información de identificación para el tipo específico de especie de ácido nucleico diana hibridada con el código de barras. Una secuencia de código de barras puede comprender una secuencia de ácido nucleico que proporciona un contador (por ejemplo, que proporciona una aproximación) para la aparición específica de la especie de ácido nucleico diana hibridada con el código de barras (por ejemplo, región de unión a la diana).

En algunas realizaciones, se unen un conjunto diverso de secuencias de código de barras a un soporte sólido dado (por ejemplo, una perla). En algunas realizaciones, puede haber, o haber aproximadamente, 10^2 , 10^3 , 10^4 , 10^5 , 10^6 , 10^7 , 10^8 , 10^9 , o un número o un intervalo entre dos cualesquiera de estos valores, secuencias de marcadores moleculares únicos. Por ejemplo, una pluralidad de códigos de barras puede comprender aproximadamente 6561 secuencias de códigos de barras con secuencias distintas. Como otro ejemplo, una pluralidad de códigos de barras puede comprender aproximadamente 65536 secuencias de códigos de barras con secuencias distintas. En algunas realizaciones, puede haber por lo menos, o haber como máximo, 10^2 , 10^3 , 10^4 , 10^5 , 10^6 , 10^7 , 10^8 , o 10^9 , secuencias de código de barras únicas. Las secuencias de marcadores moleculares únicos pueden unirse a un soporte sólido dado (por ejemplo, una perla). En algunas realizaciones, la secuencia de marcador molecular único está parcial o totalmente abarcada por una partícula (por ejemplo, una perla de hidrogel).

En diferentes implementaciones la longitud de un código de barras puede ser diferente. Por ejemplo, un código de barras puede tener, o tener aproximadamente, 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, o un número o un intervalo entre dos cualesquiera de estos valores, nucleótidos de longitud. Como otro ejemplo, un código de barras puede tener como mínimo, o como máximo, 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 100, 200 o 300 nucleótidos de longitud.

Marcadores moleculares

Un código de barras (por ejemplo, un código de barras estocástico) comprende uno o más marcadores moleculares. Los marcadores moleculares pueden incluir secuencias de códigos de barras. En algunas realizaciones, un marcador molecular puede comprender una secuencia de ácido nucleico que proporciona información de identificación para el tipo específico de especie de ácido nucleico diana hibridada con el código de barras. Un marcador molecular puede comprender una secuencia de ácido nucleico que proporciona un contador para la aparición específica de la especie de ácido nucleico diana hibridada con el código de barras (por ejemplo, región de unión a la diana).

En algunas realizaciones, se une un conjunto diverso de marcadores moleculares a un soporte sólido dado (por ejemplo, una perla). En algunas realizaciones, puede haber, o haber aproximadamente, 10^2 , 10^3 , 10^4 , 10^5 , 10^6 , 10^7 , 10^8 , 10^9 , o un número o un intervalo entre dos cualesquiera de estos valores, de secuencias de marcadores moleculares únicos. Por ejemplo, una pluralidad de códigos de barras puede comprender aproximadamente 6561 marcadores moleculares con secuencias distintas. Como otro ejemplo, una pluralidad de códigos de barras puede

comprender aproximadamente 65536 marcadores moleculares con secuencias distintas. En algunas realizaciones, puede haber por lo menos, o como máximo, 10^2 , 10^3 , 10^4 , 10^5 , 10^6 , 10^7 , 10^8 , o 10^9 , secuencias de marcadores moleculares únicos. Los códigos de barras con secuencias de marcadores moleculares únicos pueden unirse a un soporte sólido dado (por ejemplo, una perla).

Para la identificación con códigos de barras (por ejemplo, identificación con códigos de barras estocásticos) usando una pluralidad de códigos de barras estocásticos, la relación entre el número de secuencias de marcadores moleculares diferentes y la cantidad de apariciones de cualquiera de las dianas puede ser, o ser de aproximadamente, 1:1, 2:1, 3:1, 4:1, 5:1, 6:1, 7:1, 8:1, 9:1, 10:1, 11:1, 12:1, 13:1, 14:1, 15:1, 16:1, 17:1, 18:1, 19:1, 20:1, 30:1, 40:1, 50:1, 60:1, 70:1, 80:1, 90:1, 100:1, o un número o intervalo entre dos cualesquiera de estos valores. Una diana puede ser una especie de ARNm que comprende moléculas de ARNm con secuencias idénticas o casi idénticas. En algunas realizaciones, la relación entre el número de secuencias de marcadores moleculares diferentes y el número de apariciones de cualquiera de las dianas es por lo menos, o es como máximo, 1:1, 2:1, 3:1, 4:1, 5:1, 6:1, 7:1, 8:1, 9:1, 10:1, 11:1, 12:1, 13:1, 14:1, 15:1, 16:1, 17:1, 18:1, 19:1, 20:1, 30:1, 40:1, 50:1, 60:1, 70:1, 80:1, 90:1, o 100:1.

Un marcador molecular puede tener, o tener aproximadamente, 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, o un número o un intervalo entre dos cualesquiera de estos valores, nucleótidos de longitud. Un marcador molecular puede tener por lo menos, o como máximo, 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 100, 200 o 300 nucleótidos de longitud.

Región de unión a la diana

Un código de barras puede comprender una o más regiones de unión a la diana, como sondas de captura. En algunas realizaciones, una región de unión a la diana puede hibridar con una diana de interés. En algunas realizaciones, las regiones de unión a la diana pueden comprender una secuencia de ácido nucleico que hibrida específicamente con una diana (por ejemplo, ácido nucleico diana, molécula diana, por ejemplo, un ácido nucleico celular a analizar), por ejemplo con una secuencia génica específica. En algunas realizaciones, una región de unión a la diana puede comprender una secuencia de ácido nucleico que puede unirse (por ejemplo, hibridar) con una ubicación específica de un ácido nucleico diana específico. En algunas realizaciones, la región de unión a la diana puede comprender una secuencia de ácido nucleico que es capaz de hibridar específicamente con un saliente de sitio de enzima de restricción (por ejemplo, un saliente de extremo pegajoso de EcoRI). El código de barras puede entonces ligarse a cualquier molécula de ácido nucleico que comprenda una secuencia complementaria al saliente del sitio de restricción.

En algunas realizaciones, una región de unión a la diana puede comprender una secuencia de ácido nucleico diana no específica. Una secuencia de ácido nucleico diana no específica puede referirse a una secuencia que puede unirse a múltiples ácidos nucleicos diana, independientemente de la secuencia específica del ácido nucleico diana. Por ejemplo, la región de unión a la diana puede comprender una secuencia multimérica aleatoria, o una secuencia oligo(dT) que hibrida con la cola poli(A) de las moléculas de ARNm. Una secuencia multimérica aleatoria puede ser, por ejemplo, un dímero, trímero, cuatrímero, pentámero, hexámero, septámero, octámero, nonámero, decámero o una secuencia multimérica superior aleatoria de cualquier longitud. En algunas realizaciones, la región de unión diana es la misma para todos los códigos de barras unidos a una perla dada. En algunas realizaciones, las regiones de unión a la diana para la pluralidad de códigos de barras unidos a una perla dada pueden comprender dos o más secuencias de unión a la diana diferentes. Una región de unión a la diana puede tener, o tener aproximadamente, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, o un número o intervalo entre dos cualesquiera de estos valores, nucleótidos de longitud. Una región de unión a la diana puede tener como máximo aproximadamente 5, 10, 15, 20, 25, 30, 35, 40, 45, 50 o más nucleótidos de longitud.

En algunas realizaciones, una región de unión a la diana puede comprender un oligo(dT) que puede hibridar con ARNm que comprendan extremos poliadenilados. Una región de unión a la diana puede ser específica de un gen. Por ejemplo, una región de unión a la diana puede configurarse para que hibride con una región específica de una diana. En algunas realizaciones, una región de unión a la diana no comprende un oligo(dT). Una región de unión a la diana puede tener, o tener aproximadamente, 1, 2, 3, 4, 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, o un número o intervalo entre dos cualesquiera de estos valores, nucleótidos de longitud. Una región de unión a la diana puede tener por lo menos, o como máximo, 1, 2, 3, 4, 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, o 30, nucleótidos de longitud. Una región de unión a la diana puede tener aproximadamente 5-30 nucleótidos de longitud. Cuando un código de barras comprende una región de unión a la diana específica de un gen, en la presente el código de barras puede denominarse código de barras específico de un gen.

Propiedad de orientación

Un código de barras estocástico (por ejemplo, un código de barras estocástico) puede comprender una o más propiedades de orientación que pueden usarse para orientar (por ejemplo, alinear) los códigos de barras. Un código de barras puede comprender una fracción para enfoque isoeléctrico. Diferentes códigos de barras pueden comprender

diferentes puntos de enfoque isoeléctrico. Cuando estos códigos de barras se introducen en una muestra, la muestra puede experimentar un enfoque isoeléctrico para orientar los códigos de barras de una manera conocida. De esta manera, la propiedad de orientación puede usarse para desarrollar un mapa conocido de códigos de barras en una muestra. Las propiedades de orientación ejemplares pueden incluir, movilidad electroforética (por ejemplo, basada en el tamaño del código de barras), punto isoeléctrico, espín, conductividad y/o autoensamblaje. Por ejemplo, los códigos de barras con una propiedad de orientación de autoensamblaje pueden autoensamblarse en una orientación específica (por ejemplo, nanoestructura de ácido nucleico) tras la activación.

Propiedad de afinidad

Un código de barras (por ejemplo, un código de barras estocástico) puede comprender una o más propiedades de afinidad. Por ejemplo, un marcador espacial puede comprender una propiedad de afinidad. Una propiedad de afinidad puede incluir una fracción química y/o biológica que puede facilitar la unión del código de barras a otra entidad (por ejemplo, un receptor celular). Por ejemplo, una propiedad de afinidad puede comprender un anticuerpo, por ejemplo, un anticuerpo específico para una fracción específica (por ejemplo, receptor) en una muestra. En algunas realizaciones, el anticuerpo puede guiar el código de barras a un tipo de célula o molécula específica. Las dianas en y/o cerca del tipo de célula o molécula específica pueden estar marcadas (por ejemplo, marcadas estocásticamente). En algunas realizaciones, la propiedad de afinidad puede proporcionar información espacial además de la secuencia de nucleótidos del marcador espacial, ya que el anticuerpo puede guiar el código de barras a una ubicación específica. El anticuerpo puede ser un anticuerpo terapéutico, por ejemplo un anticuerpo monoclonal o un anticuerpo policlonal. El anticuerpo puede ser humanizado o quimérico. El anticuerpo puede ser un anticuerpo desnudo o un anticuerpo de fusión.

El anticuerpo puede ser una molécula de inmunoglobulina de longitud completa (es decir, de origen natural o formada mediante procesos recombinatorios de fragmentos de genes de inmunoglobulina normales) (por ejemplo, un anticuerpo IgG) o una porción inmunológicamente activa (es decir, de unión específica) de una molécula de inmunoglobulina, como un fragmento de anticuerpo.

El fragmento de anticuerpo puede ser, por ejemplo, una porción de un anticuerpo como F(ab')₂, Fab', Fab, Fv, sFv y similares. En algunas realizaciones, el fragmento de anticuerpo puede unirse al mismo antígeno reconocido por el anticuerpo de longitud completa. El fragmento de anticuerpo puede incluir fragmentos aislados que consisten en las regiones variables de los anticuerpos, como los fragmentos "Fv" que consisten en las regiones variables de las cadenas pesada y ligera y moléculas polipeptídicas recombinantes de cadena sencilla en las que las regiones variables ligera y pesada están conectadas por un conector peptídico ("proteínas scFv"). Los anticuerpos ejemplares pueden incluir, entre otros, anticuerpos contra células cancerosas, anticuerpos contra virus, anticuerpos que se unen a receptores de la superficie celular (CD8, CD34, CD45) y anticuerpos terapéuticos.

Cebador adaptador universal

Un código de barras puede comprender uno o más cebadores adaptadores universales. Por ejemplo, un código de barras específico de un gen, como un código de barras estocástico específico de un gen, puede comprender un cebador adaptador universal. Un cebador adaptador universal puede referirse a una secuencia de nucleótidos que es universal para todos los códigos de barras. Un cebador adaptador universal puede usarse para construir códigos de barras específicos de un gen. Un cebador adaptador universal puede tener, o tener aproximadamente, 1, 2, 3, 4, 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, o un número o intervalo entre dos cualesquiera de estos nucleótidos de longitud. Un cebador adaptador universal puede tener por lo menos, o como máximo, 1, 2, 3, 4, 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 o 30 nucleótidos de longitud. Un cebador adaptador universal puede tener una longitud de 5-30 nucleótidos.

Conector

Cuando un código de barras comprende más de un tipo de marcador (por ejemplo, más de un marcador celular o más de una secuencia de código de barras, como un marcador molecular), los marcadores pueden intercalarse con una secuencia de marcador conector. Una secuencia de marcador conector puede tener por lo menos aproximadamente 5, 10, 15, 20, 25, 30, 35, 40, 45, 50 o más nucleótidos de longitud. Una secuencia de marcador conector puede tener como máximo aproximadamente 5, 10, 15, 20, 25, 30, 35, 40, 45, 50 o más nucleótidos de longitud. En algunos casos, una secuencia de marcador conector tiene 12 nucleótidos de longitud. Una secuencia de marcador conector puede usarse para facilitar la síntesis del código de barras. El marcador conector puede incluir un código de corrección de errores (por ejemplo, Hamming).

Soportes sólidos

En algunas realizaciones, los códigos de barras, como los códigos de barras estocásticos, divulgados en la presente pueden estar asociados con un soporte sólido. El soporte sólido puede ser, por ejemplo, una partícula sintética. En algunas realizaciones, algunas o todas las secuencias de códigos de barras, como marcadores

moleculares para códigos de barras estocásticos (por ejemplo, las primeras secuencias de códigos de barras) de una pluralidad de códigos de barras (por ejemplo, la primera pluralidad de códigos de barras) sobre un soporte sólido difieren en por lo menos un nucleótido. Los marcadores celulares de los códigos de barras en el mismo soporte sólido pueden ser iguales. Los marcadores celulares de los códigos de barras en diferentes soportes sólidos pueden diferir en por lo menos un nucleótido. Por ejemplo, los primeros marcadores celulares de una primera pluralidad de códigos de barras en un primer soporte sólido pueden tener la misma secuencia, y los segundos marcadores celulares de una segunda pluralidad de códigos de barras en un segundo soporte sólido pueden tener la misma secuencia. Los primeros marcadores celulares de la primera pluralidad de códigos de barras en el primer soporte sólido y los segundos marcadores celulares de la segunda pluralidad de códigos de barras en el segundo soporte sólido pueden diferir en por lo menos un nucleótido. Un marcador celular puede tener, por ejemplo, aproximadamente entre 5 y 20 nucleótidos de longitud. Una secuencia de código de barras puede tener, por ejemplo, aproximadamente entre 5 y 20 nucleótidos de longitud. La partícula sintética puede ser, por ejemplo, una perla.

La perla puede ser, por ejemplo, una perla de gel de sílice, una perla de vidrio de poro controlado, una perla magnética, una Dynabead, una perla Sephadex/Sepharose, una perla de celulosa, una perla de poliestireno, o cualquier combinación de las mismas. La perla puede comprender un material como polidimetilsiloxano (PDMS), poliestireno, vidrio, polipropileno, agarosa, gelatina, hidrogel, paramagnético, cerámica, plástico, vidrio, metilestireno, polímero acrílico, titanio, látex, Sepharose, celulosa, nailon, silicona, o cualquier combinación de los mismos.

En algunas realizaciones, la perla puede ser una perla polimérica, por ejemplo una perla deformable o una perla de gel, funcionalizada con códigos de barras o códigos de barras estocásticos (como las perlas de gel de 10X Genomics (San Francisco, CA). En alguna implementación, una perla de gel puede comprender un gel a base de polímero. Las perlas de gel pueden generarse, por ejemplo, encapsulando uno o más precursores poliméricos en gotitas. Tras la exposición de los precursores poliméricos a un acelerador (por ejemplo, tetrametiletilendiamina (TEMED)), puede generarse una perla de gel.

En algunas realizaciones, la partícula puede ser alterable (por ejemplo, soluble, degradable). Por ejemplo, la perla polimérica puede disolverse, fundirse o degradarse, por ejemplo, bajo una condición deseada. La condición deseada puede incluir una condición ambiental. La condición deseada puede dar como resultado que la perla polimérica se disuelva, funda o degrade de manera controlada. Una perla de gel puede disolverse, fundirse o degradarse debido a un estímulo químico, un estímulo físico, un estímulo biológico, un estímulo térmico, un estímulo magnético, un estímulo eléctrico, un estímulo luminoso o cualquier combinación de los mismos.

Los analitos y/o reactivos, como los códigos de barras de oligonucleótidos, por ejemplo, pueden acoplarse/inmovilizarse a la superficie interior de una perla de gel (por ejemplo, el interior accesible mediante difusión de un código de barras de oligonucleótidos y/o materiales usados para generar un código de barras de oligonucleótidos) y/o la superficie exterior de una perla de gel o cualquier otra microcápsula descrita en la presente. El acoplamiento/inmovilización puede realizarse mediante cualquier forma de enlace químico (por ejemplo, enlace covalente, enlace iónico) o fenómeno físico (por ejemplo, fuerzas de Van der Waals, interacciones dipolo-dipolo, etc.). En algunas realizaciones, el acoplamiento/inmovilización de un reactivo a una perla de gel o a cualquier otra microcápsula descrita en la presente puede ser reversible como, por ejemplo, a través de una fracción lábil (por ejemplo, a través de un reticulante químico, incluyendo los reticulantes químicos descritos en la presente). Tras la aplicación de un estímulo, puede escindirse la fracción lábil y puede liberarse el reactivo inmovilizado. En algunas realizaciones, la fracción lábil es un enlace disulfuro. Por ejemplo, en el caso de que un código de barras de oligonucleótidos esté inmovilizado en una perla de gel mediante un enlace disulfuro, la exposición del enlace disulfuro a un agente reductor puede escindir el enlace disulfuro y liberar el código de barras de oligonucleótidos de la perla. La fracción lábil puede incluirse como parte de una perla o microcápsula de gel, como parte de un conector químico que enlaza un reactivo o analito a una perla o microcápsula de gel, y/o como parte de un reactivo o analito. En algunas realizaciones, por lo menos un código de barras de la pluralidad de códigos de barras puede estar inmovilizado en la partícula, parcialmente inmovilizado en la partícula, encerrado en la partícula, parcialmente encerrado en la partícula, o cualquier combinación de los mismos.

En algunas realizaciones, una perla de gel puede comprender una amplia variedad de polímeros diferentes, incluyendo, entre otros: polímeros, polímeros sensibles al calor, polímeros fotosensibles, polímeros magnéticos, polímeros sensibles al pH, polímeros sensibles a sales, polímeros químicamente sensibles, polielectrolitos, polisacáridos, péptidos, proteínas y/o plásticos. Los polímeros pueden incluir, entre otros, materiales como poli(N-isopropilacrilamida) (PNIPAAm), poli(sulfonato de estireno) (PSS), poli(alil amina) (PAAm), poli(ácido acrílico) (PAA), poli(etileno imina) (PEI), poli(cloruro de dialildimetil-amonio) (PDADMAC), poli(pirrol) (PPy), poli(vinilpirrolidona) (PVPON), poli(vinilpiridina) (PVP), poli(ácido metacrílico) (PMAA), poli(metacrilato de metilo) (PMMA), poliestireno (PS), poli(tetrahidrofurano) (PTHF), poli(ftaladehído) (PTHF), poli(hexil viologeno) (PHV), poli(L-lisina) (PLL), poli(L-arginina) (PARG), poli(ácido láctico-co-glicólico) (PLGA).

Para desencadenar la alteración, disolución o degradación de las perlas pueden usarse numerosos estímulos químicos. Los ejemplos de estos cambios químicos pueden incluir, pero no se limitan a, cambios en la pared de la perla mediados por el pH, disgregación de la pared de la perla a través de la escisión química de los enlaces cruzados,

despolimerización desencadenada de la pared de la perla y reacciones de cambio de la pared de la perla. Para desencadenar la disgregación de las perlas también pueden usarse cambios de volumen.

Los cambios físicos o de volumen en la microcápsula a través de varios estímulos también ofrecen muchas ventajas en el diseño de cápsulas para liberar reactivos. Los cambios físicos o de volumen se producen a escala macroscópica, en la que la ruptura de la perla es el resultado de fuerzas mecanofísicas inducidas por un estímulo. Estos procesos pueden incluir, entre otros, la ruptura inducida por presión, la fusión de la pared de la perla o cambios en la porosidad de la pared de la perla.

Para desencadenar la alteración, disolución o degradación de las perlas también pueden usarse estímulos biológicos. Generalmente, los desencadenantes biológicos se asemejan a los desencadenantes químicos, pero muchos ejemplos usan biomoléculas, o moléculas que se encuentran comúnmente en los sistemas vivos, como enzimas, péptidos, sacáridos, ácidos grasos, ácidos nucleicos y similares. Por ejemplo, las perlas pueden comprender polímeros con enlaces cruzados peptídicos sensibles a la escisión por proteasas específicas. Más concretamente, un ejemplo puede comprender una microcápsula que comprende enlaces cruzados de péptidos GFLGK. Tras la adición de un desencadenante biológico, como la proteasa Catepsina B, los enlaces cruzados peptídicos del pocillo de la cubierta se escinden y se libera el contenido de las perlas. En otros casos, las proteasas pueden activarse por calor. En otro ejemplo, las perlas comprenden una pared de cubierta que comprende celulosa. La adición de la enzima hidrolítica quitosano sirve como desencadenante biológico para la escisión de los enlaces celulósicos, la despolimerización de la pared de la cubierta y la liberación de su contenido interno.

También puede inducirse a las perlas para que liberen su contenido mediante la aplicación de un estímulo térmico. Un cambio de temperatura puede provocar varios cambios en las perlas. Un cambio de calor puede provocar la fusión de una perla de tal manera que la pared de la perla se disgregue. En otros casos, el calor puede aumentar la presión interna de los componentes internos de la perla de tal manera que la perla se rompa o explote. En otros casos más, el calor puede transformar la perla en un estado deshidratado y encogido. El calor también puede actuar sobre los polímeros sensibles al calor de la pared de la perla para provocar la alteración de la perla.

La inclusión de nanopartículas magnéticas en la pared de la perla de las microcápsulas puede permitir la ruptura desencadenada de las perlas, así como la guía de las perlas en una matriz. Un dispositivo de la presente divulgación puede comprender perlas magnéticas para ambos propósitos. En un ejemplo, la incorporación de nanopartículas de Fe_3O_4 en perlas que contienen polielectrolitos desencadena la ruptura en presencia de un estímulo de campo magnético oscilante.

Una perla también puede alterarse, disolverse o degradarse como resultado de la estimulación eléctrica. De manera similar a las partículas magnéticas descritas en la sección anterior, las perlas eléctricamente sensibles pueden permitir tanto la ruptura desencadenada de las perlas como otras funciones como la alineación en un campo eléctrico, la conductividad eléctrica o las reacciones redox. En un ejemplo, las perlas que contienen material eléctricamente sensible se alinean en un campo eléctrico de tal manera que pueda controlarse la liberación de reactivos internos. En otros ejemplos, los campos eléctricos pueden inducir reacciones redox dentro de la propia pared de la perla que pueden aumentar la porosidad.

Para alterar las perlas también puede usarse un estímulo luminoso. Son posibles numerosos activadores luminosos y pueden incluir sistemas que usen varias moléculas, como nanopartículas y cromóforos, capaces de absorber fotones de intervalos específicos de longitudes de onda. Por ejemplo, como desencadenantes de cápsulas pueden usarse recubrimientos de óxido metálico. La irradiación UV de cápsulas de polielectrolito recubiertas con SiO_2 puede provocar la disgregación de la pared de la perla. En otro ejemplo más, pueden incorporarse a la pared de la perla materiales fotoactivables, como grupos azobenceno. Tras la aplicación de luz ultravioleta o visible, estas sustancias químicas experimentan una isomerización reversible de cis a trans por absorción de fotones. En este aspecto, la incorporación de interruptores de fotones da como resultado una pared de la perla que puede disgregarse o volverse más porosa tras la aplicación de un desencadenante luminoso.

Por ejemplo, en un ejemplo no limitativo de identificación con códigos de barras (por ejemplo, identificación con códigos de barras estocásticos) ilustrado en la FIG. 2, después de introducir células como células individuales en una pluralidad de micropocillos de una matriz de micropocillos en el bloque 208, pueden introducirse perlas en la pluralidad de micropocillos de la matriz de micropocillos en el bloque 212. Cada micropocillo puede comprender una perla. Las perlas pueden comprender una pluralidad de códigos de barras. Un código de barras puede comprender una región de amina 5' unida a una perla. El código de barras puede comprender un marcador universal, una secuencia de código de barras (por ejemplo, un marcador molecular), una región de unión a la diana o cualquier combinación de las mismas.

Los códigos de barras divulgados en la presente pueden asociarse con (por ejemplo, unirse a) un soporte sólido (por ejemplo, una perla). Los códigos de barras asociados con un soporte sólido pueden comprender cada uno una secuencia de código de barras seleccionada de un grupo que comprende por lo menos 100 o 1000 secuencias de código de barras con secuencias únicas. En algunas realizaciones, diferentes códigos de barras asociados con un

soporte sólido pueden comprender códigos de barras con secuencias diferentes. En algunas realizaciones, un porcentaje de los códigos de barras asociados a un soporte sólido comprende el mismo marcador celular. Por ejemplo, el porcentaje puede ser, o ser de aproximadamente el 60%, 70%, 80%, 85%, 90%, 95%, 97%, 99%, 100%, o un número o un intervalo entre dos cualesquiera de estos valores. Como otro ejemplo, el porcentaje puede ser por lo menos, o ser como máximo del 60%, 70%, 80%, 85%, 90%, 95%, 97%, 99%, o 100%. En algunas realizaciones, los códigos de barras asociados a un soporte sólido pueden tener el mismo marcador celular. Los códigos de barras asociados con diferentes soportes sólidos pueden tener diferentes marcadores celulares seleccionadas de un grupo que comprende por lo menos 100 o 1000 marcadores celulares con secuencias únicas.

Los códigos de barras divulgados en la presente pueden asociarse a (por ejemplo, unirse a) un soporte sólido (por ejemplo, una perla). En algunas realizaciones, la identificación con códigos de barras de la pluralidad de dianas en la muestra puede realizarse con un soporte sólido que incluya una pluralidad de partículas sintéticas asociadas con la pluralidad de códigos de barras. En algunas realizaciones, el soporte sólido puede incluir una pluralidad de partículas sintéticas asociadas a la pluralidad de códigos de barras. Los marcadores espaciales de la pluralidad de códigos de barras en diferentes soportes sólidos pueden diferir en por lo menos un nucleótido. El soporte sólido puede, por ejemplo, incluir la pluralidad de códigos de barras en dos dimensiones o en tres dimensiones. Las partículas sintéticas pueden ser perlas. Las perlas pueden ser perlas de gel de sílice, perlas de vidrio de poro controlado, perlas magnéticas, Dynabeads, perlas Sephadex/Sepharose, perlas de celulosa, perlas de poliestireno, o cualquier combinación de las mismas. El soporte sólido puede incluir un polímero, una matriz, un hidrogel, un dispositivo de matriz de agujas, un anticuerpo o cualquier combinación de los mismos. En algunas realizaciones, los soportes sólidos pueden flotar libremente. En algunas realizaciones, los soportes sólidos pueden estar incrustados en una matriz semisólida o sólida. Los códigos de barras pueden no estar asociados con soportes sólidos. Los códigos de barras pueden ser nucleótidos individuales. Los códigos de barras pueden estar asociados con un sustrato.

Como se usa en la presente, los términos "atado", "unido" e "inmovilizado" se usan indistintamente y pueden referirse a medios covalentes o no covalentes para unir códigos de barras a un soporte sólido. Como soporte sólido para unir códigos de barras presintetizados o para la síntesis *in situ* en fase sólida de códigos de barras puede usarse cualquiera de una variedad de soportes sólidos.

En algunas realizaciones, el soporte sólido es una perla. La perla puede comprender uno o más tipos de esfera, bola, cojinete, cilindro u otra configuración similar sólida, porosa o hueca que pueda inmovilizar un ácido nucleico (por ejemplo, covalentemente o no covalentemente). La perla puede estar compuesta, por ejemplo, de plástico, cerámica, metal, material polimérico o cualquier combinación de los mismos. Una perla puede ser, o comprender, una partícula discreta que es esférica (por ejemplo, microesferas) o tener una forma no esférica o irregular, como cúbica, cuboide, piramidal, cilíndrica, cónica, oblonga o en forma de disco, y similares. En algunas realizaciones, una perla puede tener forma no esférica.

Las perlas pueden comprender una variedad de materiales que incluyen, entre otros, materiales paramagnéticos (por ejemplo, magnesio, molibdeno, litio y tántalo), materiales superparamagnéticos (por ejemplo, nanopartículas de ferrita (Fe_3O_4 ; magnetita)), materiales ferromagnéticos (por ejemplo, hierro, níquel, cobalto, algunas aleaciones de los mismos y algunos compuestos metálicos de tierras raras), cerámica, plástico, vidrio, poliestireno, sílice, metilmetacrilato, polímeros acrílicos, titanio, látex, Sepharose, agarosa, hidrogel, polímero, celulosa, nailon o cualquier combinación de los mismos.

En algunas realizaciones, la perla (por ejemplo, la perla a la que se unen los marcadores) es una perla de hidrogel. En algunas realizaciones, la perla comprende hidrogel.

Algunas realizaciones divulgadas en la presente incluyen una o más partículas (por ejemplo, perlas). Cada una de las partículas puede comprender una pluralidad de oligonucleótidos (por ejemplo, códigos de barras). Cada uno de la pluralidad de oligonucleótidos puede comprender una secuencia de código de barras (por ejemplo, una secuencia de marcador molecular), un marcador celular y una región de unión a la diana (por ejemplo, una secuencia oligo(dT), una secuencia específica de gen, un multímero aleatorio o una combinación de los mismos). La secuencia del marcador celular de cada uno de la pluralidad de oligonucleótidos puede ser la misma. Las secuencias de los marcadores celulares de los oligonucleótidos en partículas diferentes pueden ser diferentes, de tal manera que puedan identificarse los oligonucleótidos en partículas diferentes. En diferentes implementaciones el número de secuencias de marcadores celulares diferentes puede ser diferente. En algunas realizaciones, el número de secuencias de marcadores celulares puede ser, o ser aproximadamente de 10, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 20000, 30000, 40000, 50000, 60000, 70000, 80000, 90000, 100000, 10^6 , 10^7 , 10^8 , 10^9 , un número o un intervalo entre dos cualesquiera de estos valores, o más. En algunas realizaciones, el número de secuencias de marcadores celulares puede ser como mínimo o como máximo de 10, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 20000, 30000, 40000, 50000, 60000, 70000, 80000, 90000, 100000, 10^6 , 10^7 , 10^8 o 10^9 . En algunas realizaciones, no más de 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, o más de la pluralidad de las partículas incluyen oligonucleótidos con la misma secuencia celular. En algunas realizaciones, la pluralidad de partículas que incluyen oligonucleótidos con la misma secuencia celular puede ser como máximo del

0,1%, 0,2%, 0,3%, 0,4%, 0,5%, 0,6%, 0,7%, 0,8%, 0,9%, 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, o más. En algunas realizaciones, ninguna de la pluralidad de partículas tiene la misma secuencia de marcador celular.

La pluralidad de oligonucleótidos en cada partícula puede comprender diferentes secuencias de código de barras (por ejemplo, marcadores moleculares). En algunas realizaciones, el número de secuencias de código de barras puede ser, o ser aproximadamente de 10, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 20000, 30000, 40000, 50000, 60000, 70000, 80000, 90000, 100000, 10^6 , 10^7 , 10^8 , 10^9 , o un número o un intervalo entre dos cualesquiera de estos valores. En algunas realizaciones, el número de secuencias de códigos de barras puede ser como mínimo o como máximo de 10, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 20000, 30000, 40000, 50000, 60000, 70000, 80000, 90000, 100000, 10^6 , 10^7 , 10^8 o 10^9 . Por ejemplo, por lo menos 100 de la pluralidad de oligonucleótidos comprenden secuencias de código de barras diferentes. Como otro ejemplo, en una partícula individual, por lo menos 100, 500, 1000, 5000, 10000, 15000, 20000, 50000, un número o un intervalo entre dos cualesquiera de estos valores, o más de la pluralidad de oligonucleótidos comprenden secuencias de código de barras diferentes. Algunas realizaciones proporcionan una pluralidad de partículas que comprenden códigos de barras. En algunas realizaciones, la relación de una aparición (o una copia o un número) de una diana a marcar y las diferentes secuencias de códigos de barras puede ser de por lo menos de 1:1, 1:2, 1:3, 1:4, 1:5, 1:6, 1:7, 1:8, 1:9, 1:10, 1:11, 1:12, 1:13, 1:14, 1:15, 1:16, 1:17, 1:18, 1:19, 1:20, 1:30, 1:40, 1:50, 1:60, 1:70, 1:80, 1:90, o más. En algunas realizaciones, cada uno de la pluralidad de oligonucleótidos comprende además un marcador de muestra, un marcador universal, o ambos. La partícula puede ser, por ejemplo, una nanopartícula o una micropartícula.

El tamaño de las perlas puede variar. Por ejemplo, el diámetro de la perla puede variar de 0,1 micrómetros a 50 micrómetros. En algunas realizaciones, el diámetro de la perla puede ser, o ser de aproximadamente, 0,1, 0,5, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50 micrómetros, o un número o un intervalo entre dos de cualesquiera estos valores.

El diámetro de la perla puede estar relacionado con el diámetro de los pocillos del sustrato. En algunas realizaciones, el diámetro de la microesfera puede ser, o ser aproximadamente, un 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, o un número o un intervalo entre dos cualesquiera de estos valores, más largo o más corto que el diámetro del pocillo. El diámetro de las perlas puede estar relacionado con el diámetro de una célula (por ejemplo, una célula individual atrapada por un pocillo del sustrato). En algunas realizaciones, el diámetro de la perla puede ser por lo menos, o como máximo, un 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% o 100% más largo o más corto que el diámetro del pocillo. El diámetro de las microesferas puede estar relacionado con el diámetro de una célula (por ejemplo, una célula individual atrapada por un pocillo del sustrato). En algunas realizaciones, el diámetro de la perla puede ser, o ser aproximadamente, un 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, 150%, 200%, 250%, 300%, o un número o un intervalo entre dos cualesquiera de estos valores, más largo o más corto que el diámetro de la célula. En algunas realizaciones, el diámetro de las perlas puede ser como mínimo, o como máximo, un 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, 150%, 200%, 250% o 300% más largo o más corto que el diámetro de la célula.

Una perla puede unirse y/o incrustarse en un sustrato. Una perla puede unirse y/o incrustarse en un gel, hidrogel, polímero y/o matriz. La posición espacial de una perla dentro de un sustrato (por ejemplo, gel, matriz, andamiaje o polímero) puede identificarse usando el marcador espacial presente en el código de barras en la perla, que puede servir como dirección de localización.

Ejemplos de perlas pueden incluir, pero no se limitan a, perlas de estreptavidina, perlas de agarosa, perlas magnéticas, perlas Dynabeads®, microperlas MACS®, perlas conjugadas con anticuerpos (por ejemplo, perlas antiinmunoglobulina), perlas conjugadas con proteína A, perlas conjugadas con proteína G, perlas conjugadas con proteína A/G, perlas conjugadas con proteína L, perlas conjugadas con oligo(dT), perlas de sílice, perlas similares a la sílice, microperlas antibiotina, microperlas antifluorocromo, y perlas magnéticas terminadas en carboxilo BcMag™.

Una perla puede estar asociada con (por ejemplo, impregnada con) puntos cuánticos o colorantes fluorescentes para hacerla fluorescente en un canal óptico de fluorescencia o en múltiples canales ópticos. Una perla puede asociarse con óxido de hierro u óxido de cromo para hacerla paramagnética o ferromagnética. Las perlas pueden ser identificables. Por ejemplo, pueden obtenerse imágenes de una perla con una cámara. Una perla puede tener un código detectable asociado con la perla. Por ejemplo, una perla puede comprender un código de barras. Una perla puede cambiar de tamaño, por ejemplo, debido al hinchamiento en una solución orgánica o inorgánica. Una perla puede ser hidrófoba. Una perla puede ser hidrófila. Una perla puede ser biocompatible.

Puede visualizarse un soporte sólido (por ejemplo, una perla). El soporte sólido puede comprender una etiqueta de visualización (por ejemplo, un colorante fluorescente). Un soporte sólido (por ejemplo, una perla) puede grabarse con un identificador (por ejemplo, un número). El identificador puede visualizarse a través de obtención de imágenes de las perlas.

Un soporte sólido puede comprender un material insoluble, semisoluble o soluble. Un soporte sólido puede denominarse "funcionalizado" cuando incluye un conector, un andamiaje, una unidad básica u otra fracción reactiva

unida al mismo, mientras que un soporte sólido puede ser "no funcionalizado" cuando carece de dicha fracción reactiva unida al mismo. El soporte sólido puede emplearse libre en solución, como en un formato de pocillo de microtitulación; en un formato de flujo continuo, como en una columna; o en una tira reactiva.

El soporte sólido puede comprender una membrana, papel, plástico, superficie recubierta, superficie plana, vidrio, portaobjetos, chip, o cualquier combinación de los mismos. Un soporte sólido puede adoptar la forma de resinas, geles, microesferas u otras configuraciones geométricas. Un soporte sólido puede comprender chips de sílice, micropartículas, nanopartículas, placas, matrices, capilares, soportes planos como filtros de fibra de vidrio, superficies de vidrio, superficies metálicas (acero, plata dorada, aluminio, silicio y cobre), soportes de vidrio, soportes de plástico, soportes de silicio, chips, filtros, membranas, placas de micropocillos, portaobjetos, materiales plásticos, incluyendo placas o membranas multipocillos (por ejemplo, formadas de polietileno, polipropileno, poliamida, polivinilidendifluoruro), y/u obleas, peines, alfileres o agujas (por ejemplo, matrices de alfileres adecuadas para la síntesis o el análisis combinatorios) o perlas en una matriz de fosas o pocillos de nanolitros de superficies planas como obleas (por ejemplo, obleas de silicio), obleas con fosas con o sin fondos filtrantes.

El soporte sólido puede comprender una matriz polimérica (por ejemplo, gel, hidrogel). La matriz polimérica puede ser capaz de penetrar el espacio intracelular (por ejemplo, alrededor de orgánulos). La matriz polimérica puede ser capaz de ser bombeada a través del sistema circulatorio.

Sustratos y matriz de micropocillos

Como se usa en la presente, un sustrato puede referirse a un tipo de soporte sólido. Un sustrato puede referirse a un soporte sólido que puede comprender códigos de barras o códigos de barras estocásticos de la divulgación. Un sustrato puede, por ejemplo, comprender una pluralidad de micropocillos. Por ejemplo, un sustrato puede ser una matriz de pocillos que comprende dos o más micropocillos. En algunas realizaciones, un micropocillo puede comprender una cámara de reacción pequeña de volumen definido. En algunas realizaciones, un micropocillo puede atrapar una o más células. En algunas realizaciones, un micropocillo puede atrapar sólo una célula. En algunas realizaciones, un micropocillo puede atrapar uno o más soportes sólidos. En algunas realizaciones, un micropocillo puede atrapar sólo un soporte sólido. En algunas realizaciones, un micropocillo atrapa una única célula y un único soporte sólido (por ejemplo, una perla). Un micropocillo puede comprender reactivos de código de barras de la divulgación.

Métodos de identificación con códigos de barras

La divulgación proporciona métodos para estimar el número de dianas distintas en ubicaciones distintas en una muestra física (por ejemplo, tejido, órgano, tumor, célula). Los métodos pueden comprender colocar códigos de barras (por ejemplo, códigos de barras estocásticos) en estrecha proximidad con la muestra, lisar la muestra, asociar objetivos distintos con los códigos de barras, amplificar los objetivos y/o contar digitalmente las dianas. El método puede comprender además analizar y/o visualizar la información obtenida a partir de los marcadores espaciales en los códigos de barras. En algunas realizaciones, un método comprende visualizar la pluralidad de dianas en la muestra. Mapear la pluralidad de dianas en el mapa de la muestra puede comprender generar un mapa bidimensional o un mapa tridimensional de la muestra. El mapa bidimensional y el mapa tridimensional pueden generarse antes o después de identificar con códigos de barras (por ejemplo, identificar estocásticamente con códigos de barras) la pluralidad de objetivos de la muestra. Visualizar la pluralidad de dianas en la muestra puede incluir mapear la pluralidad de dianas en un mapa de la muestra. Mapear la pluralidad de dianas en el mapa de la muestra puede incluir generar un mapa bidimensional o tridimensional de la muestra. En algunas realizaciones, el mapa bidimensional y el mapa tridimensional pueden generarse antes o después de lisar la muestra. La lisis de la muestra antes o después de generar el mapa bidimensional o el mapa tridimensional puede incluir calentar la muestra, poner en contacto la muestra con un detergente, cambiar el pH de la muestra, o cualquier combinación de los mismos.

En algunas realizaciones, la identificación con códigos de barras de la pluralidad de dianas comprende la hibridación de una pluralidad de códigos de barras con una pluralidad de dianas para crear dianas identificadas con códigos de barras (por ejemplo, dianas identificadas estocásticamente con códigos de barras). La identificación con códigos de barras de la pluralidad de dianas puede comprender la generación de una biblioteca indexada de las dianas identificadas con códigos de barras. La generación de una biblioteca indexada de las dianas identificadas con códigos de barras puede realizarse con un soporte sólido que comprende la pluralidad de códigos de barras (por ejemplo, códigos de barras estocásticos).

Contacto de una muestra y un código de barras

La divulgación proporciona métodos para poner en contacto una muestra (por ejemplo, células) con un sustrato de la divulgación. Una muestra que comprende, por ejemplo, una sección delgada de célula, órgano o tejido, puede ponerse en contacto con códigos de barras (por ejemplo, códigos de barras estocásticos). Las células pueden ponerse en contacto, por ejemplo, por flujo de gravedad en donde las células pueden asentarse y crear una monocapa. La muestra puede ser una sección delgada de tejido. La sección delgada puede colocarse sobre el sustrato. La

muestra puede ser unidimensional (por ejemplo, formar una superficie plana). La muestra (por ejemplo, células) puede extenderse por el sustrato, por ejemplo, haciendo crecer/cultivando las células en el sustrato.

Cuando los códigos de barras están muy cerca de las dianas, las dianas pueden hibridar con el código de barras. Los códigos de barras pueden ponerse en contacto en una proporción no agotable, de tal manera que cada diana distinta pueda asociarse con un código de barras distinto de la divulgación. Para garantizar una asociación eficaz entre la diana y el código de barras, las dianas pueden reticularse con el código de barras.

Lisis celular

Después de la distribución de las células y códigos de barras, las células pueden lisarse para liberar las moléculas diana. La lisis celular puede lograrse mediante cualquiera de una variedad de medios, por ejemplo, mediante medios químicos o bioquímicos, mediante choque osmótico, o por medio de lisis térmica, lisis mecánica o lisis óptica. Las células pueden lisarse mediante la adición de un tampón de lisis celular que comprenda un detergente (por ejemplo, SDS, dodecil sulfato de Li, Triton X-100, Tween-20 o NP-40), un solvente orgánico (por ejemplo, metanol o acetona) o enzimas digestivas (por ejemplo, proteinasa K, pepsina o tripsina), o cualquier combinación de los mismos. Para aumentar la asociación de una diana y un código de barras, la velocidad de difusión de las moléculas diana puede alterarse, por ejemplo, reduciendo la temperatura y/o aumentando la viscosidad del lisado.

En algunas realizaciones, la muestra puede lisarse usando un papel de filtro. El papel de filtro puede empaparse con un tampón de lisis sobre la parte superior del papel de filtro. El papel de filtro puede aplicarse a la muestra con presión, lo que puede facilitar la lisis de la muestra y la hibridación de las dianas de la muestra al sustrato.

En algunas realizaciones, la lisis puede realizarse mediante lisis mecánica, lisis por calor, lisis óptica y/o lisis química. La lisis química puede incluir el uso de enzimas digestivas como proteinasa K, pepsina y tripsina. La lisis puede realizarse mediante la adición de un tampón de lisis al sustrato. Un tampón de lisis puede comprender Tris HCl. Un tampón de lisis puede comprender por lo menos aproximadamente 0,01, 0,05, 0,1, 0,5 o 1 M o más de Tris HCl. Un tampón de lisis puede contener como máximo aproximadamente 0,01, 0,05, 0,1, 0,5 o 1 M o más de Tris HCl. Un tampón de lisis puede comprender aproximadamente 0,1 M Tris HCl. El pH del tampón de lisis puede ser por lo menos de aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más. El pH del tampón de lisis puede ser como máximo de aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, o más. En algunas realizaciones, el pH del tampón de lisis es de aproximadamente 7,5. El tampón de lisis puede contener una sal (por ejemplo, LiCl). La concentración de sal en el tampón de lisis puede ser de por lo menos aproximadamente 0,1, 0,5 o 1 M o más. La concentración de sal en el tampón de lisis puede ser como máximo de aproximadamente 0,1, 0,5 o 1 M o más. En algunas realizaciones, la concentración de sal en el tampón de lisis es de aproximadamente 0,5 M. El tampón de lisis puede comprender un detergente (por ejemplo, SDS, dodecil sulfato de Li, Triton X, tween, NP-40). La concentración del detergente en el tampón de lisis puede ser de por lo menos aproximadamente 0,0001%, 0,0005%, 0,001%, 0,005%, 0,01%, 0,05%, 0,1%, 0,5%, 1%, 2%, 3%, 4%, 5%, 6% o 7%, o más. La concentración del detergente en el tampón de lisis puede ser como máximo de aproximadamente el 0,0001%, 0,0005%, 0,001%, 0,005%, 0,01%, 0,05%, 0,1%, 0,5%, 1%, 2%, 3%, 4%, 5%, 6% o 7%, o más. En algunas realizaciones, la concentración del detergente en el tampón de lisis es de aproximadamente un 1% de dodecil sulfato de Li. El tiempo usado en el método para la lisis puede depender de la cantidad de detergente usado. En algunas realizaciones, cuanto más detergente se use, menos tiempo se necesitará para la lisis. El tampón de lisis puede comprender un agente quelante (por ejemplo, EDTA, EGTA). La concentración de un agente quelante en el tampón de lisis puede ser de por lo menos de aproximadamente 1, 5, 10, 15, 20, 25 o 30 mM o más. La concentración de un agente quelante en el tampón de lisis puede ser como máximo de aproximadamente 1, 5, 10, 15, 20, 25 o 30 mM o más. En algunas realizaciones, la concentración de agente quelante en el tampón de lisis es de aproximadamente 10 mM. El tampón de lisis puede comprender un reactivo reductor (por ejemplo, beta-mercaptoetanol, DTT). La concentración del reactivo reductor en el tampón de lisis puede ser de por lo menos de aproximadamente 1, 5, 10, 15 o 20 mM o más. La concentración del reactivo reductor en el tampón de lisis puede ser como máximo de aproximadamente 1, 5, 10, 15 o 20 mM o más. En algunas realizaciones, la concentración de reactivo reductor en el tampón de lisis es de aproximadamente 5 mM. En algunas realizaciones, un tampón de lisis puede comprender aproximadamente 0,1M TrisHCl, aproximadamente pH 7,5, aproximadamente 0,5M LiCl, aproximadamente un 1% de dodecil sulfato de litio, aproximadamente 10mM EDTA, y aproximadamente 5mM DTT.

La lisis puede realizarse a una temperatura de aproximadamente 4, 10, 15, 20, 25 o 30° C. La lisis puede realizarse durante aproximadamente 1, 5, 10, 15 o 20 minutos o más. Una célula lisada puede comprender por lo menos aproximadamente 100000, 200000, 300000, 400000, 500000, 600000 o 700000 o más moléculas de ácido nucleico diana. Una célula lisada puede comprender como máximo aproximadamente 100.000, 200.000, 300.000, 400.000, 500.000, 600.000 o 700.000 moléculas de ácido nucleico diana.

Unión de códigos de barras a moléculas de ácido nucleico diana

Después de la lisis de las células y la liberación de moléculas de ácido nucleico de las mismas, las moléculas de ácido nucleico pueden asociarse aleatoriamente con los códigos de barras del soporte sólido colocalizado. La asociación puede comprender la hibridación de la región de reconocimiento de la diana de un código de barras con

una porción complementaria de la molécula de ácido nucleico diana (por ejemplo, el oligo(dT) del código de barras puede interactuar con una cola de poli(A) de una diana). Las condiciones de ensayo utilizadas para la hibridación (por ejemplo, pH del tampón, fuerza iónica, temperatura, etc.) pueden elegirse para promover la formación de híbridos específicos y estables. En algunas realizaciones, las moléculas de ácido nucleico liberadas de las células lisadas pueden asociarse con la pluralidad de sondas del sustrato (por ejemplo, hibridar con las sondas en el sustrato). Cuando las sondas comprenden oligo(dT), las moléculas de ARNm pueden hibridar con las sondas y transcribirse inversamente. La porción de oligo(dT) del oligonucleótido puede actuar como cebador para la síntesis de la primera cadena de la molécula de ADNc. Por ejemplo, en un ejemplo no limitativo de identificación con códigos de barras ilustrado en la FIG. 2, en el bloque 216, las moléculas de ARNm pueden hibridar con códigos de barras en perlas. Por ejemplo, los fragmentos de nucleótidos de cadena sencilla pueden hibridar con las regiones de unión a la diana de los códigos de barras.

La unión puede comprender además la ligación de una región de reconocimiento de la diana de un código de barras y una porción de la molécula de ácido nucleico diana. Por ejemplo, la región de unión a la diana puede comprender una secuencia de ácido nucleico que puede ser capaz de hibridación específica a un saliente de sitio de restricción (por ejemplo, un saliente de extremo pegajoso EcoRI). El procedimiento de ensayo puede comprender además tratar los ácidos nucleicos diana con una enzima de restricción (por ejemplo, EcoRI) para crear un saliente de sitio de restricción. A continuación, el código de barras puede ligarse a cualquier molécula de ácido nucleico que comprenda una secuencia complementaria al saliente del sitio de restricción. Para unir los dos fragmentos puede usarse una ligasa (por ejemplo, T4 ADN ligasa).

Por ejemplo, en un ejemplo no limitativo de código de barras ilustrado en la FIG. 2, en el bloque 220, las dianas marcadas de una pluralidad de células (o una pluralidad de muestras) (por ejemplo, moléculas de diana-código de barras) pueden agruparse posteriormente, por ejemplo, en un tubo. Las dianas marcadas pueden agruparse, por ejemplo, recuperando los códigos de barras y/o las perlas a las que están unidas las moléculas de diana-código de barras.

La recuperación de colecciones basadas en soporte sólido de moléculas de diana-código de barras unidas puede implementarse mediante el uso de perlas magnéticas y un campo magnético aplicado externamente. Una vez agrupadas las moléculas de diana-código de barras, todo el procesamiento posterior puede realizarse en un único recipiente de reacción. El procesamiento posterior puede incluir, por ejemplo, reacciones de transcripción inversa, reacciones de amplificación, reacciones de escisión, reacciones de disociación y/o reacciones de extensión de ácidos nucleicos. Las reacciones de procesamiento posterior pueden realizarse dentro de los micropocillos, es decir, sin agrupar primero las moléculas de ácido nucleico diana marcadas de una pluralidad de células.

Transcripción inversa

La divulgación proporciona un método para crear un conjugado diana-código de barras usando transcripción inversa (por ejemplo, en el bloque 224 de la FIG. 2). El conjugado diana-código de barras puede comprender el código de barras y una secuencia complementaria de todo o parte del ácido nucleico diana (es decir, una molécula de ADNc identificada con código de barras, como una molécula de ADNc identificada estocásticamente con código de barras). La transcripción inversa de la molécula de ARN asociada puede producirse mediante la adición de un cebador de transcripción inversa junto con la transcriptasa inversa. El cebador de transcripción inversa puede ser un cebador oligo(dT), un cebador de hexanucleótidos aleatorio o un cebador de oligonucleótidos específico de la diana. Los cebadores oligo(dT) pueden tener, o tener aproximadamente, entre 12 y 18 nucleótidos de longitud y se unen a la cola de poli(A) endógena en el extremo 3' del ARNm de mamífero. Los cebadores de hexanucleótidos aleatorios pueden unirse al ARNm en una variedad de sitios complementarios. Los cebadores de hexanucleótidos específicos de la diana típicamente ceban selectivamente el ARNm de interés.

En algunas realizaciones, la transcripción inversa de la molécula de ARN marcado puede producirse mediante la adición de un cebador de transcripción inversa. En algunas realizaciones, el cebador de transcripción inversa es un cebador oligo(dT), un cebador de hexanucleótidos aleatorio o un cebador de oligonucleótidos específico de la diana. Generalmente, los cebadores oligo(dT) tienen entre 12 y 18 nucleótidos de longitud y se unen a la cola de poli(A) endógena en el extremo 3' del ARNm de mamíferos. Los cebadores de hexanucleótidos aleatorios pueden unirse al ARNm en una variedad de sitios complementarios. Los cebadores de oligonucleótidos específicos de diana típicamente ceban selectivamente el ARNm de interés.

La transcripción inversa puede producirse repetidamente para producir múltiples moléculas de ADNc marcado. Los métodos divulgados en la presente pueden comprender realizar por lo menos aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 o 20 reacciones de transcripción inversa. El método puede comprender realización por lo menos aproximadamente 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95 o 100 reacciones de transcripción inversa.

Amplificación

Pueden realizarse una o más reacciones de amplificación de ácido nucleico (por ejemplo, en el bloque 228 de la FIG. 2) para crear múltiples copias de las moléculas de ácido nucleico diana marcadas. La amplificación puede realizarse de manera multiplexada, en donde múltiples secuencias de ácido nucleico diana se amplifican simultáneamente. La reacción de amplificación puede usarse para añadir adaptadores de secuenciación a las moléculas de ácido nucleico. Las reacciones de amplificación pueden comprender la amplificación de por lo menos una parte de un marcador de muestra, si lo hay. Las reacciones de amplificación pueden comprender la amplificación de por lo menos una parte del marcador celular y/o la secuencia del código de barras (por ejemplo, un marcador molecular). Las reacciones de amplificación pueden comprender la amplificación de por lo menos una parte de una etiqueta de muestra, un marcador celular, un marcador espacial, una secuencia de código de barras (por ejemplo, un marcador molecular), un ácido nucleico diana, o una combinación de los mismos. Las reacciones de amplificación pueden comprender la amplificación del 0,5%, 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 100%, o un intervalo o un número entre dos cualesquiera de estos valores, de la pluralidad de ácidos nucleicos. El método puede comprender además llevar a cabo una o más reacciones de síntesis de ADNc para producir una o más copias de ADNc de moléculas de código de barras diana que comprenden un marcador de muestra, un marcador celular, un marcador espacial y/o una secuencia de código de barras (por ejemplo, un marcador molecular).

En algunas realizaciones, la amplificación puede realizarse usando una reacción en cadena de la polimerasa (PCR). Como se usa en la presente, PCR puede referirse a una reacción para la amplificación in vitro de secuencias específicas de ADN mediante la extensión simultánea de cebadores de cadenas complementarias de ADN. Como se usa en la presente, la PCR puede abarcar formas derivadas de la reacción, incluyendo pero no limitado a, RT-PCR, PCR en tiempo real, PCR anidada, PCR cuantitativa, PCR digital PCR multiplexada y PCR de ensamblaje.

La amplificación de los ácidos nucleicos marcados puede comprender métodos no basados en RCP. Ejemplos de métodos no basados en PCR incluyen, pero no se limitan a, amplificación por desplazamiento múltiple (MDA), amplificación mediada por transcripción (TMA), amplificación basada en secuencias de ácido nucleico (NASBA), amplificación por desplazamiento de cadena (SDA), SDA en tiempo real, amplificación de círculo rodante o amplificación de círculo a círculo. Otros métodos de amplificación no basados en la PCR incluyen ciclos múltiples de amplificación por transcripción de ARN dependiente de ADN impulsada por ARN polimerasa o síntesis y transcripción de ADN dirigida por ARN para amplificar dianas de ADN o ARN, una reacción en cadena de ligasa (LCR), y un método de Q β replicasa (Q β), uso de sondas palindrómicas, amplificación por desplazamiento de cadena, amplificación impulsada por oligonucleótidos usando una endonucleasa de restricción, un método de amplificación en el que un cebador se hibrida con una secuencia de ácido nucleico y el dúplex resultante se escinde antes de la reacción de extensión y amplificación, amplificación por desplazamiento de cadena usando una polimerasa de ácido nucleico que carece de actividad exonucleasa 5', amplificación por círculo rodante y amplificación por extensión ramificada (RAM). En algunas realizaciones, la amplificación no produce transcritos circularizados.

En algunas realizaciones, los métodos divulgados en la presente comprenden además llevar a cabo una reacción en cadena de la polimerasa en el ácido nucleico marcado (por ejemplo, ARN marcado, ADN marcado, ADNc marcado) para producir un amplicón marcado (por ejemplo, un amplicón marcado estocásticamente). El amplicón marcado puede ser una molécula de cadena doble. La molécula de cadena doble puede comprender una molécula de cadena doble de ARN, una molécula cadena doble de ADN o una molécula de ARN hibridada con una molécula de ADN. Una o ambas cadenas de la molécula cadena doble pueden incluir un marcador de muestra, un marcador espacial, un marcador celular y/o una secuencia de código de barras (por ejemplo, un marcador molecular). El amplicón marcado puede ser una molécula de cadena sencilla. La molécula cadena sencilla puede comprender ADN, ARN o una combinación de los mismos. Los ácidos nucleicos de la divulgación pueden comprender ácidos nucleicos sintéticos o alterados.

La amplificación puede comprender el uso de uno o más nucleótidos no naturales. Los nucleótidos no naturales pueden comprender nucleótidos fotolábiles o activables. Ejemplos de nucleótidos no naturales pueden incluir, pero no se limitan a, ácido nucleico peptídico (PNA), morfolino y ácido nucleico bloqueado (LNA), así como ácido nucleico glicólico (GNA) y ácido nucleico de treosa (TNA). Los nucleótidos no naturales pueden añadirse a uno o más ciclos de una reacción de amplificación. La adición de los nucleótidos no naturales puede usarse para identificar productos como ciclos específicos o puntos temporales en la reacción de amplificación.

La realización de una o más reacciones de amplificación puede comprender el uso de uno o más cebadores. Los uno o más cebadores pueden comprender, por ejemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, o 15 o más nucleótidos. Los uno o más cebadores pueden comprender por lo menos 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, o 15 o más nucleótidos. Los uno o más cebadores pueden comprender menos de 12-15 nucleótidos. Los uno o más cebadores pueden aparearse en por lo menos una parte de la pluralidad de dianas marcadas (por ejemplo, dianas marcadas estocásticamente). Los uno o más cebadores pueden aparearse al extremo 3' o al extremo 5' de la pluralidad de dianas marcadas. Los uno o más cebadores pueden aparearse en una región interna de la pluralidad de dianas marcadas. La región interna puede estar por lo menos a aproximadamente 50, 100, 150, 200, 220, 230, 240, 250, 260, 270, 280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 380, 390, 400, 410, 420, 430, 440, 450, 460, 470, 480, 490, 500, 510, 520, 530, 540, 550, 560, 570, 580, 590, 600, 650, 700, 750, 800, 850, 900 o 1000 nucleótidos de los extremos

3' de la pluralidad de dianas marcadas. Los uno o más cebadores puede comprender un panel fijo de cebadores. Los uno o más cebadores pueden comprender por lo menos uno o más cebadores personalizados. Los uno o más cebadores puede comprender por lo menos uno o más cebadores de control. Los uno o más cebadores puede comprender por lo menos uno o más cebadores específicos de gen.

Los uno o más cebadores puede comprender un cebador universal. El cebador universal puede aparearse a un sitio de unión de cebador universal. Los uno o más cebadores personalizados pueden aparearse a un primer marcador de muestra, un segundo marcador de muestra, un marcador espacial, un marcador celular, una secuencia de código de barras (por ejemplo, un marcador molecular), una diana o cualquier combinación de los mismos. Los uno o más cebadores puede comprender un cebador universal y un cebador personalizado. El cebador personalizado puede diseñarse para amplificar una o más dianas. Las dianas pueden comprender un subconjunto de los ácidos nucleicos totales en una o más muestras. Las dianas pueden comprender un subconjunto del total de dianas marcadas en una o más muestras. Los uno o más cebadores pueden comprender por lo menos 96 o más cebadores personalizados. Los uno o más cebadores pueden comprender por lo menos 960 o más cebadores personalizados. Los uno o más cebadores pueden comprender por lo menos 9600 o más cebadores personalizados. Los uno o más cebadores personalizados pueden aparearse a dos o más ácidos nucleicos marcados diferentes. Los dos o más ácidos nucleicos marcados diferentes pueden corresponder a uno o más genes.

En los métodos de la presente divulgación puede usarse cualquier esquema de amplificación. Por ejemplo, en un esquema, la primera ronda de PCR puede amplificar moléculas unidas a la perla usando un cebador específico de gen y un cebador contra la secuencia del cebador 1 de secuenciación universal de Illumina. La segunda ronda de PCR puede amplificar los productos de la primera PCR usando un cebador específico de gen anidado flanqueado por la secuencia del cebador 2 de secuenciación de Illumina, y un cebador contra la secuencia del cebador 1 de secuenciación de Illumina universal. La tercera ronda de PCR añade P5 y P7 e índice de muestra para convertir los productos de PCR en una biblioteca de secuenciación de Illumina. La secuenciación usando secuenciación de 150 pb x 2 puede revelar el marcador celular y la secuencia del código de barras (por ejemplo, el marcador molecular) en la lectura 1, el gen en la lectura 2 y el índice de la muestra en la lectura del índice 1.

En algunas realizaciones, los ácidos nucleicos pueden eliminarse del sustrato usando escisión química. Por ejemplo, puede usarse un grupo químico o una base modificada presente en un ácido nucleico para facilitar su eliminación de un soporte sólido. Por ejemplo, puede usarse una enzima para eliminar un ácido nucleico de un sustrato. Por ejemplo, un ácido nucleico puede eliminarse de un sustrato mediante una digestión con endonucleasas de restricción. Por ejemplo, puede usarse tratamiento de un ácido nucleico que contiene un dUTP o ddUTP con uracilo-d-glicosilasa (UDG) para eliminar un ácido nucleico de un sustrato. Por ejemplo, un ácido nucleico puede eliminarse de un sustrato usando una enzima que realiza la escisión de nucleótidos, como una enzima de reparación por escisión de bases, como una endonucleasa apurínica/apirimidinica (AP). En algunas realizaciones, un ácido nucleico puede eliminarse de un sustrato usando un grupo fotoescindible y luz. En algunas realizaciones, para eliminar un ácido nucleico del sustrato puede usarse un conector escindible. Por ejemplo, el conector escindible puede comprender por lo menos uno de biotina/avidina, biotina/estreptavidina, biotina/neutravidina, Ig-proteína A, un conector foto-lábil, un grupo conector lábil a ácido o base, o un aptámero.

Cuando las sondas son específicas de un gen, las moléculas pueden hibridar con las sondas y ser transcritas inversamente y/o amplificadas. En algunas realizaciones, después de que el ácido nucleico ha sido sintetizado (por ejemplo, transcrito inversamente), puede amplificarse. La amplificación puede realizarse de forma multiplex, en donde múltiples secuencias de ácido nucleico diana se amplifican simultáneamente. La amplificación puede añadir adaptadores de secuenciación al ácido nucleico.

En algunas realizaciones, la amplificación puede realizarse en el sustrato, por ejemplo, con amplificación en puente. A los ADNc se les pueden añadir colas de homopolímero con el fin de generar un extremo compatible para la amplificación con puente usando sondas oligo(dT) en el sustrato. En la amplificación con puente, el cebador complementario al extremo 3' del ácido nucleico plantilla puede ser el primer cebador de cada par que está unido covalentemente a la partícula sólida. Cuando se pone en contacto una muestra que contiene el ácido nucleico plantilla con la partícula y se realiza un único ciclo térmico, la molécula plantilla aparearse con el primer cebador y el primer cebador se alarga en la dirección directa mediante la adición de nucleótidos para formar una molécula dúplex que consiste en la molécula plantilla y una cadena de ADN recién formada que es complementaria a la plantilla. En el paso de calentamiento del siguiente ciclo, la molécula dúplex puede desnaturalizarse, liberando la molécula plantilla de la partícula y dejando la cadena de ADN complementaria unida a la partícula a través del primer cebador. En la etapa de apareamiento el paso de apareamiento y elongación que sigue, la cadena complementaria puede hibridar con el segundo cebador, que es complementario a un segmento de la cadena complementaria en una ubicación alejada del primer cebador. Esta hibridación puede hacer que la cadena complementaria forme un puente entre el primer y el segundo cebadores, fijado al primer cebador por un enlace covalente y al segundo cebador por hibridación. En la etapa de elongación, el segundo cebador puede alargarse en la dirección inversa mediante la adición de nucleótidos en la misma mezcla de reacción, convirtiendo de este modo el puente en un puente de cadena doble. A continuación comienza el siguiente ciclo, y el puente de cadena doble puede desnaturalizarse para proporcionar dos moléculas de ácido nucleico de cadena sencilla, cada una de las cuales tiene un extremo unido a la superficie de la partícula a través

del primer y del segundo cebadores, respectivamente, y el otro extremo de cada una de ellas no unido. En el paso de apareamiento y elongación de este segundo ciclo, cada cadena puede hibrida con un cebador complementario adicional, no usado previamente, en la misma partícula, para formar nuevos puentes de cadena sencilla. Los dos cebadores no usados previamente que ahora están hibridados se alargan para convertir los dos nuevos puentes en puentes de cadena doble.

Las reacciones de amplificación pueden comprender la amplificación de por lo menos el 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97% o 100% de la pluralidad de ácidos nucleicos.

La amplificación de los ácidos nucleicos marcados puede comprender métodos basados en PCR o métodos no basados en PCR. La amplificación de los ácidos nucleicos marcados puede comprender la amplificación exponencial de los ácidos nucleicos marcados. La amplificación de los ácidos nucleicos marcados puede comprender la amplificación lineal de los ácidos nucleicos marcados. La amplificación puede realizarse mediante reacción en cadena de la polimerasa (PCR). La PCR puede referirse a una reacción para la amplificación in vitro de secuencias específicas de ADN mediante la extensión simultánea de cebadores de cadenas complementarias de ADN. La PCR puede abarcar formas derivadas de la reacción, incluyendo, entre otras, RT-PCR, PCR en tiempo real, PCR anidada, PCR cuantitativa, PCR multiplexada, PCR digital, PCR de supresión, PCR semisupresiva y PCR de ensamblaje.

En algunas realizaciones, la amplificación de los ácidos nucleicos marcados comprende métodos no basados en RCP. Ejemplos de métodos no basados en PCR incluyen, pero no se limitan a, amplificación por desplazamiento múltiple (MDA), amplificación mediada por transcripción (TMA), amplificación basada en secuencias de ácido nucleico (NASBA), amplificación por desplazamiento de cadena (SDA), SDA en tiempo real, amplificación de círculo rodante o amplificación de círculo a círculo. Otros métodos de amplificación no basados en la PCR incluyen ciclos múltiples de amplificación por transcripción de ARN dependiente de ADN impulsada por ARN polimerasa o síntesis y transcripción de ADN dirigida por ARN para amplificar dianas de ADN o ARN, una reacción en cadena de ligasa (LCR), una replicasa Q β (Q β), uso de sondas palindrómicas, amplificación por desplazamiento de cadena, amplificación impulsada por oligonucleótidos usando una endonucleasa de restricción, un método de amplificación en el que un cebador hibrida con una secuencia de ácido nucleico y el dúplex resultante se escinde antes de la reacción de extensión y amplificación, amplificación por desplazamiento de cadena usando una polimerasa de ácido nucleico que carece de actividad de exonucleasa 5', amplificación por círculo rodante, y/o amplificación por extensión ramificada (RAM).

En algunas realizaciones, los métodos divulgados en la presente comprenden además llevar a cabo una reacción en cadena de la polimerasa anidada en el amplicón amplificado (por ejemplo, diana). El amplicón puede ser una molécula de cadena doble. La molécula de cadena doble puede comprender una molécula de ARN de cadena doble, una molécula de ADN de cadena doble o una molécula de ARN hibridada con una molécula de ADN. Una o ambas cadenas de la molécula de cadena doble pueden contener un marcador de muestra o un marcador de identificador molecular. Alternativamente, el amplicón puede ser una molécula de cadena sencilla. La molécula de cadena sencilla puede comprender ADN, ARN o una combinación de los mismos. Los ácidos nucleicos descritos en la presente pueden comprender ácidos nucleicos sintéticos o alterados.

En algunas realizaciones, el método comprende amplificar repetidamente el ácido nucleico marcado para producir múltiples amplicones. Los métodos divulgados en la presente pueden comprender llevar a cabo por lo menos aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, o 20 reacciones de amplificación. Alternativamente, el método comprende realizar por lo menos aproximadamente 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95 o 100 reacciones de amplificación.

La amplificación puede comprender además añadir uno o más ácidos nucleicos de control a una o más muestras que comprenden una pluralidad de ácidos nucleicos. La amplificación puede comprender además añadir uno o más ácidos nucleicos de control a una pluralidad de ácidos nucleicos. Los ácidos nucleicos de control pueden comprender un marcador de control.

La amplificación puede comprender el uso de uno o más nucleótidos no naturales. Los nucleótidos no naturales pueden comprender nucleótidos fotolábiles y/o activables. Ejemplos de nucleótidos no naturales incluyen, pero no se limitan a, ácido nucleico peptídico (PNA), morfolino y ácido nucleico bloqueado (LNA), así como ácido nucleico glicólico (GNA) y ácido nucleico de treosa (TNA). Los nucleótidos no naturales pueden añadirse a uno o más ciclos de una reacción de amplificación. La adición de los nucleótidos no naturales puede usarse para identificar productos como ciclos específicos o puntos temporales en la reacción de amplificación.

La realización de una o más reacciones de amplificación puede comprender el uso de uno o más cebadores. Los uno o más cebadores pueden comprender uno o más oligonucleótidos. Los uno o más oligonucleótidos pueden comprender por lo menos aproximadamente 7-9 nucleótidos. Los uno o más oligonucleótidos pueden comprender menos de 12-15 nucleótidos. Los uno o más cebadores pueden apararse con por lo menos una parte de la pluralidad de ácidos nucleicos marcados. Los uno o más cebadores pueden apararse al extremo 3' y/o al extremo 5' de la pluralidad de ácidos nucleicos marcados. Los uno o más cebadores pueden apararse en una región interna de la

pluralidad de ácidos nucleicos marcados. La región interna puede estar por lo menos a aproximadamente 50, 100, 150, 200, 220, 230, 240, 250, 260, 270, 280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 380, 390, 400, 410, 420, 430, 440, 450, 460, 470, 480, 490, 500, 510, 520, 530, 540, 550, 560, 570, 580, 590, 600, 650, 700, 750, 800, 850, 900 o 1000 nucleótidos de los extremos 3' de la pluralidad de ácidos nucleicos marcados. Los uno o más cebadores puede comprender un panel fijo de cebadores. Los uno o más cebadores pueden comprender por lo menos uno o más cebadores personalizados. Los uno o más cebadores pueden comprender por lo menos uno o más cebadores de control. Los uno o más cebadores puede comprender por lo menos uno o más cebadores de genes constitutivos. Los uno o más cebadores puede comprender un cebador universal. El cebador universal puede aparearse a un sitio de unión del cebador universal. Los uno o más cebadores personalizados pueden aparearse a la primera etiqueta de muestra, la segunda etiqueta de muestra, el marcador identificador molecular, el ácido nucleico o un producto de los mismos. Los uno o más cebadores pueden comprender un cebador universal y un cebador personalizado. El cebador personalizado puede estar diseñado para amplificar uno o más ácidos nucleicos diana. Los ácidos nucleicos diana pueden comprender un subconjunto de los ácidos nucleicos totales en una o más muestras. En algunas realizaciones, los cebadores son las sondas unidas a la matriz de la divulgación.

En algunas realizaciones, la identificación con códigos de barras (por ejemplo, identificar estocásticamente con códigos de barras) de la pluralidad de dianas en la muestra comprende además generar una biblioteca indexada de las dianas identificadas con códigos de barras (por ejemplo, dianas identificadas estocásticamente con códigos de barras) o fragmentos identificados con códigos de barras de las dianas. Las secuencias de códigos de barras de diferentes códigos de barras (por ejemplo, los marcadores moleculares de diferentes códigos de barras estocásticos) pueden ser diferentes entre sí. La generación de una biblioteca indexada de dianas identificadas con códigos de barras incluye la generación de una pluralidad de polinucleótidos indexados a partir de la pluralidad de dianas de la muestra. Por ejemplo, para una biblioteca indexada de dianas identificadas con códigos de barras que comprende una primera diana indexada y una segunda diana indexada, la región del marcador del primer polinucleótido indexado puede diferir de la región del marcador del segundo polinucleótido indexado en, aproximadamente, por lo menos, o como máximo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, o un número o un intervalo entre dos cualesquiera de estos valores, nucleótidos. En algunas realizaciones, la generación de una biblioteca indexada de las dianas identificadas con códigos de barras incluye poner en contacto una pluralidad de dianas, por ejemplo moléculas de ARNm, con una pluralidad de oligonucleótidos que incluyen una región poli(T) y una región de marcador; y llevar a cabo una síntesis de primera cadena usando una transcriptasa inversa para producir moléculas de ADNc marcadas de cadena sencilla cada una de las cuales comprendiendo una región de ADNc y una región de marcador, en donde la pluralidad de dianas incluye por lo menos dos moléculas de ARNm de secuencias diferentes y la pluralidad de oligonucleótidos incluye por lo menos dos oligonucleótidos de secuencias diferentes. La generación de una biblioteca indexada de las dianas identificadas con códigos de barras puede comprender además la amplificación de las moléculas de ADNc marcadas de cadena sencilla para producir moléculas de ADNc marcadas de cadena doble; y la realización de PCR anidada en las moléculas de ADNc marcadas de cadena doble para producir amplicones marcados. En algunas realizaciones, el método puede incluir la generación de un amplicón marcado con adaptador.

La identificación con códigos de barras (por ejemplo, identificación con códigos de barras estocásticos) puede incluir el uso de códigos de barras o etiquetas de ácido nucleico para marcar moléculas individuales de ácido nucleico (por ejemplo, ADN o ARN). En algunas realizaciones, implica añadir códigos de barras o etiquetas de ADN a moléculas de ADNc a medida que se generan a partir de ARNm. Puede realizarse una PCR anidada para minimizar el sesgo de amplificación de la PCR. Pueden añadirse adaptadores para la secuenciación usando, por ejemplo, la secuenciación de próxima generación (NGS). Los resultados de la secuenciación pueden usarse para determinar marcadores celulares, marcadores moleculares y secuencias de fragmentos de nucleótidos de una o más copias de las dianas, por ejemplo en el bloque 232 de la FIG. 2.

La FIG. 3 es una ilustración esquemática que muestra un proceso ejemplar no limitativo de generación de una biblioteca indexada de dianas identificadas con códigos de barras (por ejemplo, dianas identificadas con códigos de barras estocásticos), como ARNm identificados con códigos de barras o fragmentos de los mismos. Como se muestra en el paso 1, el proceso de transcripción inversa puede codificar cada molécula de ARNm con una secuencia de marcador molecular única, una secuencia de marcador celular y un sitio de PCR universal. En particular, las moléculas de ARN 302 pueden transcribirse inversamente para producir moléculas de ADNc marcadas 304, incluyendo una región de ADNc 306, mediante hibridación (por ejemplo, hibridación estocástica) de un conjunto de códigos de barras (por ejemplo, códigos de barras estocásticos) 310 a la región de cola poli(A) 308 de las moléculas de ARN 302. Cada uno de los códigos de barras 310 puede comprender una región de unión a la diana, por ejemplo una región poli(dT) 312, una región de marcador 314 (por ejemplo, una secuencia de código de barras o una molécula), y una región de PCR universal 316.

En algunas realizaciones, la secuencia de marcador celular puede incluir de 3 a 20 nucleótidos. En algunas realizaciones, la secuencia de marcador molecular puede incluir de 3 a 20 nucleótidos. En algunas realizaciones, cada uno de la pluralidad de códigos de barras estocásticos comprende además uno o más de un marcador universal y un marcador celular, en donde los marcadores universales son los mismos para la pluralidad de códigos de barras estocásticos en el soporte sólido y los marcadores celulares son los mismos para la pluralidad de códigos de barras estocásticos en el soporte sólido. En algunas realizaciones, el marcador universal puede incluir de 3 a 20 nucleótidos.

En algunas realizaciones, el marcador celular comprende de 3 a 20 nucleótidos.

En algunas realizaciones, la región de marcador 314 puede incluir una secuencia de código de barras o un marcador molecular 318 y un marcador celular 320. En algunas realizaciones, la región de marcador 314 puede incluir uno o más de un marcador universal, un marcador de dimensión y un marcador celular. La secuencia de código de barras o marcador molecular 318 puede tener, puede tener aproximadamente, puede tener por lo menos, o puede tener como máximo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, o un número o un intervalo entre cualquiera de estos valores, nucleótidos de longitud. El marcador celular 320 puede tener, puede tener aproximadamente, puede tener por lo menos, o puede tener como máximo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, o un número o un intervalo entre cualquiera de estos valores, nucleótidos de longitud. El marcador universal puede tener, puede tener aproximadamente, puede tener por lo menos, o puede tener como máximo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, o un número o un intervalo entre cualquiera de estos valores, de nucleótidos de longitud. Los marcadores universales pueden ser los mismos para la pluralidad de códigos de barras estocásticos en el soporte sólido y los marcadores celulares son los mismos para la pluralidad de códigos de barras estocásticos en el soporte sólido. El marcador de dimensión puede tener, puede tener aproximadamente, puede tener por lo menos, o puede tener como máximo 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, o un número o un intervalo entre cualquiera de estos valores, de nucleótidos de longitud.

En algunas realizaciones, la región de marcador 314 puede comprender, comprender aproximadamente, comprender por lo menos, o comprender como máximo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, o un número o un intervalo entre cualquiera de estos valores, marcadores diferentes, como una secuencia de código de barras o un marcador molecular 318 y un marcador celular 320. Cada marcador puede tener, puede tener aproximadamente, puede tener por lo menos, o puede tener como máximo 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, o un número o un intervalo entre cualquiera de estos valores, nucleótidos de longitud. Un conjunto de códigos de barras o códigos de barras estocásticos 310 puede contener, contener aproximadamente, contener por lo menos, o puede ser como máximo de, 10, 20, 40, 50, 70, 80, 90, 10², 10³, 10⁴, 10⁵, 10⁶, 10⁷, 10⁸, 10⁹, 10¹⁰, 10¹¹, 10¹², 10¹³, 10¹⁴, 10¹⁵, 10²⁰, o un número o un intervalo entre cualquiera de estos valores, códigos de barras o códigos de barras estocásticos 310. Y el conjunto de códigos de barras o códigos de barras estocásticos 310 puede, por ejemplo, contener cada uno una región de marcador única 314. Las moléculas de ADNc marcadas 304 pueden purificarse para eliminar el exceso de códigos de barras o códigos de barras estocásticos 310. La purificación puede comprender la purificación con perlas Ampure.

Como se muestra en el paso 2, los productos del proceso de transcripción inversa del paso 1 pueden agruparse en 1 tubo y amplificarse por PCR con un 1º grupo de cebadores de PCR y un 1º cebador de PCR universal. La agrupación es posible gracias a la región de marcador única 314. En particular, las moléculas de ADNc marcadas 304 pueden amplificarse para producir amplicones marcados por PCR anidada 322. La amplificación puede comprender una amplificación por PCR multiplex. La amplificación puede comprender una amplificación por PCR multiplex con 96 cebadores multiplex en un único volumen de reacción. En algunas realizaciones, la amplificación por PCR multiplex puede utilizar, utilizar aproximadamente, utilizar por lo menos, o utilizar como máximo, 10, 20, 40, 50, 70, 80, 90, 10², 10³, 10⁴, 10⁵, 10⁶, 10⁷, 10⁸, 10⁹, 10¹⁰, 10¹¹, 10¹², 10¹³, 10¹⁴, 10¹⁵, 10²⁰, o un número o un intervalo entre cualquiera de estos valores, cebadores multiplex en un único volumen de reacción. La amplificación puede comprender usar un 1º grupo de cebadores de PCR 324 que comprende cebadores personalizados 326A-C dirigidos a genes específicos y un cebador universal 328. Los cebadores personalizados 326 pueden hibridar con una región dentro de la porción de ADNc 306' de la molécula de ADNc marcada 304. El cebador universal 328 puede hibridar con la región de PCR universal 316 de la molécula de ADNc marcada 304.

Como se muestra en el paso 3 de la FIG. 3, los productos de la amplificación PCR en el paso 2 pueden amplificarse con un grupo de cebadores de PCR anidados y un 2º cebador de PCR universal. La PCR anidada puede minimizar el sesgo de amplificación por PCR. En particular, los amplicones marcados por PCR anidada 322 pueden amplificarse adicionalmente por PCR anidada. La PCR anidada puede comprender PCR multiplex con un grupo de cebadores de PCR anidada 330 de cebadores de PCR anidada 332a-c y un 2º cebador de PCR universal 328' en un único volumen de reacción. El grupo de cebadores de PCR anidada 328 puede contener, contener aproximadamente, contener por lo menos, o contener como máximo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, o un número o un intervalo entre cualquiera de estos valores, cebadores de PCR anidada 330 diferentes. Los cebadores de PCR anidada 332 pueden contener un adaptador 334 e hibridar con una región dentro de la porción de ADNc 306" del amplicón marcado 322. El cebador universal 328' puede contener un adaptador 336 e hibridar con la región de PCR universal 316 del amplicón marcado 322. Por tanto, el paso 3 produce el amplicón marcado con adaptador 338. En algunas realizaciones, los cebadores de PCR anidada 332 y el 2º cebador de PCR universal 328' pueden no contener los adaptadores 334 y 336. En su lugar, los adaptadores 334 y 336 pueden ligarse a los productos de la PCR anidada para producir el amplicón marcado con el adaptador 338.

Como se muestra en el paso 4, los productos de PCR del paso 3 pueden amplificarse por PCR para secuenciación usando cebadores de amplificación de biblioteca. En particular, los adaptadores 334 y 336 pueden usarse para realizar uno o más ensayos adicionales en el amplicón marcado con adaptador 338. Los adaptadores 334 y 336 pueden hibridar con los cebadores 340 y 342. Los uno o más cebadores 340 y 342 pueden ser cebadores de

amplificación por PCR. Los uno o más cebadores 340 y 342 pueden ser cebadores de secuenciación. Los uno o más adaptadores 334 y 336 pueden usarse para la amplificación adicional de los amplicones marcados con adaptador 338. Los uno o más adaptadores 334 y 336 pueden usarse para secuenciar el amplicón marcado con adaptador 338. El cebador 342 puede contener un índice de placa 344 para que los amplicones generados usando el mismo conjunto de códigos de barras o códigos de barras estocásticos 310 puedan secuenciarse en una reacción de secuenciación usando secuenciación de próxima generación (NGS).

Identificación con códigos de barras en los extremos 5' de las dianas de ácidos nucleicos

En la presente se divulgan métodos para la unión de códigos de barras (por ejemplo, códigos de barras estocásticos) con marcadores moleculares (o índices moleculares) a los extremos 5' de dianas de ácido nucleico que se estén identificando con códigos de barras o codificando (por ejemplo, moléculas de ácido desoxirribonucleico y moléculas de ácido ribonucleico). Los métodos de recuento de transcritos basados en 5' divulgados en la presente pueden complementar o suplementar, por ejemplo, los métodos de recuento de transcritos basados en 3' (por ejemplo, el ensayo Rhapsody™ (Becton, Dickinson and Company (Franklin Lakes, NJ)), Solución 3' de células individuales Chromium™ (10X Genomics (San Francisco, CA)). Las dianas de ácido nucleico identificadas con código de barras pueden usarse para la identificación de secuencias, el recuento de transcritos, el análisis de corte y empalme alternativo, el cribado de mutaciones y/o la secuenciación de longitud completa de alto rendimiento. El recuento de transcritos en el extremo 5' (5' con respecto a las dianas de ácido nucleico que se están marcando) puede revelar isoformas de corte y empalme alternativas y variantes (incluyendo, pero no limitadas a, variantes de corte y empalme, polimorfismos de un único nucleótido (SNP), inserciones, deleciones, sustituciones) en, o más cerca de, los extremos 5' de las moléculas de ácido nucleico. En algunas realizaciones, el método puede incluir hibridación intramolecular.

Las FIGS. 4A-4B muestran una ilustración esquemática de un método ejemplar no limitativo 400 de marcado específico de genes diana de ácido nucleico en los extremos 5'. Un código de barras 420 (por ejemplo, un código de barras estocástico) con una región de unión a la diana (por ejemplo, una cola de poli(dT) 422) puede unirse a transcritos de ARN poliadenilados 424 a través de la cola de poli(dA) 426, u otras dianas de ácido nucleico, para el marcado o identificación con código de barras (por ejemplo, marcado único). Los códigos de barras 420 pueden incluir marcadores moleculares (ML) 428 y marcadores de muestra (SL) 430 para marcar los transcritos 424 y rastrear los orígenes de muestra de los transcritos de ARN 424, respectivamente, junto con una o más secuencias adicionales (por ejemplo, secuencias consenso, como una secuencia adaptadora 432), flanqueando la región de marcador molecular 428/marcador de muestra 430 de cada código de barras 420 para reacciones posteriores. El repertorio de secuencias de los marcadores moleculares en los códigos de barras por muestra puede ser suficientemente amplio para el marcado estocástico de los transcritos de ARN.

Después de la síntesis de ADNc en el bloque 402 para generar moléculas de ADNc identificadas con códigos de barras 434 que comprenden los transcritos de ARN 424 (o una porción de los mismos), puede usarse un método específico de gen para la identificación con códigos de barras molecular 5'. Después de la amplificación específica del gen en el bloque 404, que puede ser opcional, puede añadirse una transferasa terminal y desoxiadenosina trifosfatos (dATP) en el bloque 406 para facilitar la adición de cola 3' de poli(dA) para generar amplicones 436 con una cola 438 de poli(A). Un breve paso de desnaturalización en el bloque 408 permite la separación de las cadenas directa 436m e inversa 436c (por ejemplo, moléculas de ADNc codificadas con códigos de barras con colas de poli(dA)) del amplicón 436. La cadena inversa 436c del amplicón 436 se separa de la cadena directa 436m. La cadena inversa 436c del amplicón 436 puede hibridar intramolecularmente a través de su cola de poli(dA) 438 en el extremo 3' y la región de poli(dT) 422 final de la cadena para formar una horquilla o estructura de horquilla 440 en el bloque 410. A continuación, puede usarse una polimerasa (por ejemplo, un fragmento de Klenow) para que se extienda desde la cola de poli(dA) 438 para duplicar el código de barras y formar la cadena inversa identificada con código de barras extendida 442 en el bloque 412. Luego puede realizarse la amplificación específica de genes en el bloque 414 (por ejemplo, opcionalmente) para amplificar genes de interés para producir amplicones 444 con códigos de barras en el extremo 5' (con respecto a los transcritos de ARN 424) para secuenciación en el bloque 416. En algunas realizaciones, el método 400 incluye una o ambas amplificaciones específicas de genes de la molécula de ADNc identificada con código de barras 434 en el bloque 404 y la amplificación específica de genes de la cadena inversa identificada con código de barras extendida 442 en el bloque 414.

Las FIGS. 5A-5B muestran una ilustración esquemática de un método 500 ejemplar no limitativo de marcado de dianas de ácido nucleico en los extremos 5' para el análisis del transcriptoma completo. Un código de barras 420 (por ejemplo, un código de barras estocástico) con una región de unión a la diana (por ejemplo, una cola de poli(dT) 422) puede unirse a transcritos de ARN poliadenilados 424 a través de la cola de poli(dA) 426, u otras dianas de ácido nucleico, para el marcado o identificación con códigos de barras (por ejemplo, marcado único). Por ejemplo, un código de barras 420 con una región de unión a la diana puede unirse a una diana de ácido nucleico para el marcado o la identificación con código de barras. Un código de barras 420 puede incluir un marcador molecular (ML) 428 y un marcador de muestra (SL) 430. Los marcadores moleculares 428 y los marcadores de muestra 430 pueden usarse para marcar los transcritos 424, o las dianas de ácido nucleico (por ejemplo, oligonucleótidos de anticuerpos, ya estén asociados con anticuerpos o se hayan disociado de anticuerpos) y rastrear los orígenes de muestra de los transcritos 424, respectivamente, junto con una o más secuencias adicionales (por ejemplo, secuencias consenso, como una

secuencia adaptadora 432), flanqueando la región de marcador molecular 428/marcador de muestra 430 de cada código de barras 420 para reacciones posteriores. El repertorio de secuencias de los marcadores moleculares 428 en los códigos de barras por muestra puede ser suficientemente amplio para el marcado estocástico de transcritos de ARN 424, o dianas de ácidos nucleicos.

Después de la síntesis de ADNc para generar moléculas de ADNc identificadas con códigos de barras 434 en el bloque 402, puede usarse una enzima transferasa terminal para la adición de cola A del extremo 3' de las moléculas de ADNc identificadas con códigos de barras 434 (equivalente al extremo 5' de los transcritos de ARN marcados) para generar moléculas de ADNc 436c cada una con una cola poli(dA) 3' 438 en el bloque 406. La hibridación intramolecular de las moléculas de ADNc 436c con colas de poli(dA) 3' 438 puede iniciarse (por ejemplo, con un ciclo de calentamiento y enfriamiento, o diluyendo las moléculas de ADNc identificadas con códigos de barras 436c con colas de poli(dA) 438) de tal manera que la nueva cola de poli(dA) 3' 438 se aparee con la cola de poli(dT) (aunque se contempla que para secuencias de región de unión a la diana distintas de poli(dA), los complementos correspondientes de la secuencia relevante de unión a la diana pueda aparearse con la región de unión a la diana) 422 de la misma molécula de ADNc marcada para generar una molécula de ADNc identificada con código de barras con una estructura de horquilla o tallo lazo 440 en el bloque 410. Puede añadirse una polimerasa (por ejemplo, enzima Klenow) con dNTP para facilitar una extensión 3' más allá de la nueva cola 3' de poli(dA) 438 para duplicar los códigos de barras (por ejemplo, marcadores moleculares 428 que están en los extremos 5' de las moléculas de ADNc marcadas con estructuras de horquilla 440 en el bloque 412. Puede realizarse una amplificación del transcriptoma completo (WTA) en el bloque 414 usando adaptadores reflejados 432, 432rc o cebadores que contengan secuencias (o subsecuencias) de los adaptadores 432, 432rc. Pueden usarse métodos, como la tagmentación o el cebado aleatorio, para generar fragmentos más pequeños de amplicones 444 con adaptadores de secuenciación (por ejemplo, la secuencia P5 446 y P7 448) para la secuenciación en el bloque 418 (por ejemplo, usando un secuenciador Illumina (San Diego, CA, Estados Unidos). En algunas realizaciones, los adaptadores de secuenciación para otros métodos de secuenciación o secuenciadores (por ejemplo, secuenciadores de Pacific Biosciences of California, Inc. (Menlo Park, CA, Estados Unidos) u Oxford Nanopore Technologies Limited (Oxford, Reino Unido)) pueden ligarse directamente para generar amplicones para secuenciación.

Lo divulgado en la presente incluye métodos para determinar los números de una diana de ácido nucleico en una muestra. En algunas realizaciones, el método comprende: poner en contacto copias de una diana de ácido nucleico 424 en una muestra con una pluralidad de códigos de barras de oligonucleótidos 420, en donde cada uno de la pluralidad de códigos de barras de oligonucleótidos 420 comprende una secuencia de marcador molecular 428 y una región de unión a la diana (por ejemplo, una secuencia de poli(dT) 422) capaz de hibridar con la diana de ácido nucleico 424, y en donde por lo menos 10 de la pluralidad de códigos de barras de oligonucleótidos 420 comprenden secuencias de marcador molecular diferentes 428; extender las copias de la diana de ácido nucleico 424 hibridadas con los códigos de barras de oligonucleótidos 420 para generar una pluralidad de moléculas de ácido nucleico 434 cada una de las cuales comprende una secuencia complementaria 450c a por lo menos una porción de la diana de ácido nucleico 424 en el bloque 402; amplificar la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras 434 en el bloque 404 para generar una pluralidad de moléculas de ácido nucleico identificadas con códigos de barras amplificadas 436; unir un oligonucleótido que comprende el complemento 438 de la región de unión a la diana 422 a la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras amplificadas 436 para generar una pluralidad de moléculas de ácido nucleico identificadas con códigos de barras 436c que comprenden cada una la región de unión a la diana 422 y un complemento 438 de la región de unión a la diana en el bloque 406; hibridar la región de unión a la diana 422 y el complemento 438 de la región de unión a la diana 422 dentro de cada una de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras 436c para formar una estructura de horquilla 440 en el bloque 410; extender los extremos 3' de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras cada una con la estructura de horquilla 440 en el bloque 412 para extender la estructura de horquilla 440 para generar una pluralidad de moléculas de ácido nucleico identificadas con códigos de barras extendidas 442 que comprenden cada una el marcador molecular 428 y un complemento 428rc del marcador molecular; amplificar la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras extendidas 442 en el bloque 414 para generar una pluralidad de moléculas de ácido nucleico marcadas individualmente 444c que comprenden cada una el complemento 428rc del marcador molecular; y determinar la cantidad de la diana de ácido nucleico en la muestra basándose en el número de complementos 428rc de marcadores moleculares con secuencias distintas asociadas con la pluralidad de moléculas de ácido nucleico marcadas individualmente.

En algunas realizaciones, el marcador molecular 428 se hibrida con el complemento 428rc del marcador molecular después de extender los extremos 3' de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras con las estructuras de horquilla 440. El método puede comprender desnaturalizar la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras extendidas 442 antes de amplificar la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras extendidas 442 para generar la pluralidad de moléculas de ácido nucleico marcadas individualmente 444c (que pueden formar parte de los amplicones 444c). Poner en contacto copias de la diana de ácido nucleico 424 en la muestra puede comprender poner en contacto copias de una pluralidad de dianas de ácido nucleico 424 con una pluralidad de códigos de barras de oligonucleótidos 420. Extender las copias de la diana de ácido nucleico 424 puede comprender extender las copias de la pluralidad de dianas de ácido nucleico 424 hibridadas con los códigos de barras de oligonucleótidos 420 para generar una pluralidad

de moléculas de ácido nucleico identificadas con códigos de barras 436c cada una de las cuales comprende una secuencia complementaria 450c a por lo menos una parte de una de la pluralidad de dianas de ácido nucleico 424. Determinar la cantidad de la diana de ácido nucleico 424 puede comprender determinar la cantidad de cada una de la pluralidad de dianas de ácido nucleico 424 en la muestra basándose en el número de los complementos 428rc de los marcadores moleculares con secuencias distintas asociadas con moléculas de ácido nucleico marcadas individualmente de la pluralidad de moléculas de ácido nucleico marcadas individualmente 444c que comprenden una secuencia 452c de cada una de la pluralidad de dianas de ácido nucleico 424. La secuencia 452c de cada una de la pluralidad de dianas de ácido nucleico puede comprender una subsecuencia (incluyendo un complemento o un complemento inverso) de cada una de la pluralidad de dianas de ácido nucleico 424.

Lo divulgado en la presente incluye métodos para determinar la cantidad de dianas en una muestra. En algunas realizaciones, el método comprende: identificar con códigos de barras 402 copias de una diana de ácido nucleico 424 en una muestra usando una pluralidad de códigos de barras de oligonucleótidos 420 para generar una pluralidad de moléculas de ácido nucleico identificadas con códigos de barras 434 cada una de las cuales comprendiendo secuencia 450c (por ejemplo, una secuencia complementaria, una secuencia complementaria inversa, o una combinación de las mismas) de la diana de ácido nucleico 424, un marcador molecular 428, y una región de unión a la diana (por ejemplo, una región poli(dT) 422), y en donde por lo menos 10 de la pluralidad de códigos de barras de oligonucleótidos 420 comprenden secuencias diferentes de marcadores moleculares 428; unir 406 un oligonucleótido que comprende un complemento 438 de la región de unión a la diana 422 a la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras 434 para generar una pluralidad de moléculas de ácido nucleico identificadas con códigos de barras 436 que comprenden cada una la región de unión a la diana 422 y el complemento 438 de la región de unión a la diana 422; hibridar 410 la región de unión a la diana 422 y el complemento 438 de la región de unión a la diana dentro de cada una de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras 436c para formar una estructura de horquilla 440; extender 412 los extremos 3' de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras para extender la estructura de horquilla 440 para generar una pluralidad de moléculas de ácido nucleico identificadas con códigos de barras extendidas 442 que comprenden cada una el marcador molecular 428 y un complemento 428rc del marcador molecular; y determinar la cantidad de la diana de ácido nucleico 424 en la muestra basándose en el número de complementos 428rc de marcadores moleculares con secuencias distintas asociadas con la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras extendidas 442.

Lo divulgado en la presente incluye métodos para unir códigos de barras de oligonucleótidos a una diana en una muestra. En algunas realizaciones, el método comprende: identificar con códigos de barras 402 copias de una diana de ácido nucleico 424 en una muestra usando una pluralidad de códigos de barras de oligonucleótidos 420 para generar una pluralidad de moléculas de ácido nucleico identificadas con códigos de barras 434 cada una de las cuales comprendiendo secuencia 450c de la diana de ácido nucleico 424, un marcador molecular 428, y una región de unión a la diana 422, y en donde por lo menos 10 de la pluralidad de códigos de barras de oligonucleótidos 420 comprenden diferentes secuencias de marcadores moleculares 428; unir un oligonucleótido que comprende un complemento 438 de la región de unión a la diana 422 a la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras 434 para generar una pluralidad de moléculas de ácido nucleico identificadas con códigos de barras 436c que comprenden cada una la región de unión a la diana 422 y el complemento 438 de la región de unión a la diana 422; hibridar 410 la región de unión a la diana 422 y el complemento 438 de la región de unión a la diana 422 dentro de cada una de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras 436c para formar una estructura de horquilla 440; y extender 412 los extremos 3' de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras para extender la estructura de horquilla 440 para generar una pluralidad de moléculas de ácido nucleico identificadas con códigos de barras extendidas 442 que comprenden cada una el marcador molecular 428 y un complemento 428rc del marcador molecular 428. En algunas realizaciones, el método comprende: determinar la cantidad de la diana de ácido nucleico 424 en la muestra basándose en el número de marcadores moleculares 428 con secuencias distintas, complementos 428rc de las mismas, o una combinación de las mismas, asociadas con la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras extendidas 442. Por ejemplo, la cantidad de la diana de ácido nucleico 424 puede determinarse basándose en uno o ambos de los marcadores moleculares 428 con secuencias distintas, complementos 428rc de las mismas.

En algunas realizaciones, el método comprende: identificar con códigos de barras 402 las copias de la pluralidad de dianas 424 comprende: poner en contacto las copias de la diana de ácido nucleico 424 con la pluralidad de códigos de barras de oligonucleótidos 420, en donde cada uno de la pluralidad de códigos de barras de oligonucleótidos 420 comprende la región de unión a la diana 422 capaz de hibridar con la diana de ácido nucleico 424; y extender 402 las copias de la diana de ácido nucleico 424 hibridadas con los códigos de barras de oligonucleótidos 420 para generar la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras 434.

En algunas realizaciones, el método comprende: amplificar 404 la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras 434 para generar una pluralidad de moléculas de ácido nucleico identificadas con códigos de barras amplificadas 436c, en donde unir el oligonucleótido que comprende el complemento 438 de la región de unión a la diana 422 comprende: unir el oligonucleótido que comprende el complemento 438 de la región de unión

a la diana a la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras amplificadas para generar una pluralidad de moléculas de ácido nucleico identificadas con códigos de barras 436r que comprenden cada una la región de unión a la diana 422 y un complemento 438 de la región de unión a la diana.

5 Análisis específico de genes. En algunas realizaciones, el método (por ejemplo, el método 400) comprende: amplificar 414 la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras extendidas 442 para generar una pluralidad de moléculas de ácido nucleico marcadas individualmente 444c, cada una de las cuales comprende el complemento 428rc del marcador molecular 428. Las moléculas de ácido nucleico marcadas individualmente 444c pueden generarse cuando los amplicones 444 que las contienen se desnaturalizan. Determinar la cantidad de la diana de ácido nucleico 424 en la muestra puede comprender: determinar la cantidad de la diana de ácido nucleico 424 en la muestra basándose en el número de complementos 428rc de marcadores moleculares 428 con secuencias distintas asociadas con la pluralidad de moléculas de ácido nucleico marcadas individualmente 444c.

15 Análisis del Transcriptoma Completo. En algunas realizaciones, el método (por ejemplo, el método 500) comprende: amplificar 414 la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras extendidas 442 para generar copias 444c de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras extendidas. Determinar la cantidad de la diana de ácido nucleico 424 en la muestra comprende: determinar la cantidad de la diana de ácido nucleico 424 en la muestra basándose en el número de complementos 428rc de marcadores moleculares 428 con secuencias distintas asociadas con las copias 444c de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras extendidas. Las copias 444c de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras extendidas pueden formarse cuando los amplicones 444 que las contienen se desnaturalizan.

25 En algunas realizaciones, la secuencia de la diana de ácido nucleico en la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras comprende una subsecuencia 452c de la diana de ácido nucleico. La región de unión a la diana puede comprender una secuencia específica de gen. La unión 406 del oligonucleótido que comprende el complemento 438 de la región de unión a la diana 422 puede comprender ligar el oligonucleótido que comprende el complemento 438 de la región de unión a la diana 422 a la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras 434.

30 En algunas realizaciones, la región de unión a la diana puede comprender una secuencia poli(dT) 422 (que también puede denominarse en la presente secuencia oligo(dT)). Unir el oligonucleótido que comprende el complemento 438 de la región de unión a la diana 422 comprende: añadir una pluralidad de monofosfatos de adenosina a la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras 434 usando una desoxinucleotidil transferasa terminal. En algunas realizaciones, la región de unión a la diana no comprende una secuencia poli(dT).

40 En algunas realizaciones, extender las copias de la diana de ácido nucleico 424 hibridadas con los códigos de barras de oligonucleótidos 420 puede comprender transcribir inversamente las copias de la diana de ácido nucleico 424 hibridadas con los códigos de barras de oligonucleótidos 420 para generar una pluralidad de moléculas de ácido desoxirribonucleico complementario (ADNc) identificadas con códigos de barras 434. Extender las copias de la diana de ácido nucleico 424 hibridadas con los códigos de barras de oligonucleótidos 420 puede comprender extender 402 las copias de la diana de ácido nucleico 424 hibridadas con los códigos de barras de oligonucleótidos 420 usando una ADN polimerasa que carece de por lo menos una de las actividades exonucleasa de 5' a 3' y exonucleasa de 3' a 5'. La ADN polimerasa puede comprender un fragmento de Klenow.

50 En algunas realizaciones, el método comprende: obtener información de secuencia de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras extendidas 442. Obtener la información de secuencia puede comprender unir adaptadores de secuenciación (por ejemplo, el adaptador P5 446 y el adaptador P7 448) a la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras extendidas 442.

55 En algunas realizaciones, el complemento 438 de la región de unión a la diana puede comprender la secuencia complementaria inversa de la región de unión a la diana. El complemento 438 de la región de unión a la diana puede comprender la secuencia complementaria de la región de unión a la diana. El complemento 428rc del marcador molecular puede comprender una secuencia complementaria inversa del marcador molecular. El complemento del marcador molecular puede comprender una secuencia complementaria de el marcador molecular.

60 En algunas realizaciones, la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras 434 puede comprender moléculas de ácido desoxirribonucleico (ADN) identificadas con códigos de barras. Las moléculas de ácido nucleico identificadas con códigos de barras 434 pueden comprender moléculas de ácido ribonucleico (ARN) identificadas con códigos de barras. La diana de ácido nucleico 424 puede comprender una molécula de ácido nucleico. La molécula de ácido nucleico puede comprender ácido ribonucleico (ARN), ARN mensajero (ARNm), microARN, ARN interferente pequeño (ARNip), producto de degradación de ARN, ARN que comprende una cola poli(A), o cualquier combinación de los mismos.

Oligonucleótidos de anticuerpo. En algunas realizaciones, la diana de ácido nucleico puede comprender un reactivo de unión a componentes celulares. Los reactivos de unión celular asociados a dianas de ácido nucleico (por ejemplo, oligonucleótidos de anticuerpos, como oligonucleótidos de indexación de muestras) se han descrito en el documento US2018/0088112; y la Solicitud de Estados Unidos N° 15/937.713, presentada el 27 de marzo de 2018.

En algunas realizaciones, la información multiómica, como la genómica, la accesibilidad de la cromatina, la metilómica, la transcriptómica y la proteómica, de células individuales puede obtenerse usando métodos de identificación con códigos de barras 5' de la divulgación. La molécula de ácido nucleico puede asociarse con el reactivo de unión a componentes celulares. El método puede comprender: disociar la molécula de ácido nucleico y el reactivo de unión a componente celular.

En algunas realizaciones, cada marcador molecular 428 de la pluralidad de códigos de barras de oligonucleótidos 420 comprende por lo menos 6 nucleótidos. El código de barras de oligonucleótidos 420 puede comprender un marcador de muestra idéntico 430. Cada marcador de muestra 430 de la pluralidad de códigos de barras de oligonucleótidos 420 puede comprender por lo menos 6 nucleótidos. El código de barras de oligonucleótidos 420 puede comprender un marcador celular de la pluralidad de códigos de barras de oligonucleótidos 420 puede comprender por lo menos 6 nucleótidos.

En algunas realizaciones, por lo menos una de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras 436c se asocia con un soporte sólido cuando hibridan 410 la región de unión a la diana y el complemento de la región de unión a la diana dentro de cada una de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras para formar la estructura de horquilla. Por lo menos una de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras 436c puede disociarse de un soporte sólido cuando hibridan 410 la región de unión a la diana 422 y el complemento 438 de la región de unión a la diana 422 dentro de cada una de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras 436c para formar la estructura de horquilla 440. Por lo menos una de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras 436c puede asociarse con un soporte sólido cuando hibridan 410 la región de unión a la diana 422 y el complemento 438 de la región de unión a la diana dentro de cada una de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras 436c para formar la estructura de horquilla 440.

En algunas realizaciones, por lo menos una de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras está asociada con un soporte sólido cuando extienden 412 los extremos 3' de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras para extender la estructura de horquilla 440 para generar la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras extendidas 442 que comprenden cada una el marcador molecular 428 y un complemento 428rc del marcador molecular. Por lo menos una de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras puede disociarse de un soporte sólido cuando se extienden 412 los extremos 3' de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras para extender la estructura de horquilla 440 para generar la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras extendidas 442 que comprenden cada una el marcador molecular 428 y un complemento 428rc del marcador molecular. Por lo menos una de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras 436c puede asociarse con un soporte sólido cuando se extienden 412 los extremos 3' de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras para extender la estructura de horquilla 440 para generar la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras extendidas 442 que comprenden cada una el marcador molecular 428 y un complemento 428rc del marcador molecular. El soporte sólido puede comprender una partícula sintética 454. El soporte sólido puede comprender una superficie plana o una superficie sustancialmente plana (por ejemplo, un portaobjetos, como un portaobjetos de microscopio o un cubreobjetos).

En algunas realizaciones, por lo menos una de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras 436c está en solución cuando hibridan 410 la región de unión a la diana 422 y el complemento 438 de la región de unión a la diana 422 dentro de cada una de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras 436c para formar la estructura de horquilla 440. Por ejemplo, cuando la concentración de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras 436c en solución es suficientemente baja, puede producirse dicha hibridación intramolecular. Por lo menos una de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras puede estar en solución cuando se extienden 412 los extremos 3' de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras para extender la estructura de horquilla 440 y generar la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras extendidas 442, cada una de las cuales comprende el marcador molecular 428 y un complemento 428rc del marcador molecular.

En algunas realizaciones, la muestra comprende una célula individual, el método comprende asociar una partícula sintética 454 que comprende la pluralidad de los códigos de barras de oligonucleótidos 420 con la célula individual de la muestra. El método puede comprender: lisis de la célula individual después de asociar la partícula sintética 454 con la célula individual. La lisis de la célula individual puede comprender calentar la muestra, poner en contacto la muestra con un detergente, cambiar el pH de la muestra, o cualquier combinación de los mismos. La partícula sintética y la célula individual pueden estar en el mismo pocillo. La partícula sintética y la célula individual pueden estar en la misma gotita.

En algunas realizaciones, por lo menos uno de la pluralidad de códigos de barras de oligonucleótidos 420 puede estar inmovilizado en la partícula sintética 454. Por lo menos uno de la pluralidad de códigos de barras de oligonucleótidos 420 puede estar parcialmente inmovilizado en la partícula sintética 454. Por lo menos uno de la pluralidad de códigos de barras de oligonucleótidos 420 puede estar encerrado en la partícula sintética 454. Por lo menos uno de la pluralidad de códigos de barras de oligonucleótidos 420 puede estar parcialmente encerrado en la partícula sintética 454. La partícula sintética 454 puede ser alterable. La partícula sintética 454 puede comprender una perla. La perla puede comprender una perla de Sepharose, una perla de estreptavidina, una perla de agarosa, una perla magnética, una perla conjugada, una perla conjugada con proteína A, una perla conjugada con proteína G, una perla conjugada con proteína A/G, una perla conjugada con proteína L, una perla conjugada con oligo(dT), una perla de sílice, una perla similar a la sílice, una perla antibiotina, una perla antifluorocromo, o cualquier combinación de las mismas. La partícula sintética 454 puede comprender un material seleccionado del grupo que consiste en polidimetilsiloxano (PDMS), poliestireno, vidrio, polipropileno, agarosa, gelatina, hidrogel, paramagnético, cerámica, plástico, vidrio, metilestireno, polímero acrílico, titanio, látex, Sepharose, celulosa, nylon, silicona, y cualquier combinación de los mismos. La partícula sintética 454 puede comprender una partícula de hidrogel alterable. Cada uno de la pluralidad de códigos de barras de oligonucleótidos 420 puede comprender un grupo funcional conector. La partícula sintética 454 puede comprender un grupo funcional de soporte sólido. El grupo funcional de soporte y el grupo funcional conector pueden estar asociados entre sí. El grupo funcional conector y el grupo funcional de soporte pueden seleccionarse individualmente del grupo que consiste en C6, biotina, estreptavidina, aminos primarios, aldehídos, cetonas, y cualquier combinación de los mismos.

En algunas realizaciones, la región de unión a la diana comprende una secuencia específica de gen, una secuencia oligo(dT), un multímero aleatorio o cualquier combinación de los mismos. El código de barras de oligonucleótidos puede comprender un marcador de muestra idéntico y/o un marcador celular idéntico. Cada marcador de muestra y/o marcador celular de la pluralidad de códigos de barras de oligonucleótidos puede comprender por lo menos 6 nucleótidos. Cada marcador molecular de la pluralidad de códigos de barras de oligonucleótidos puede comprender por lo menos 6 nucleótidos.

En algunas realizaciones, por lo menos uno de la pluralidad de códigos de barras de oligonucleótidos 420 está inmovilizado en la partícula sintética 454. Por lo menos uno de la pluralidad de códigos de barras de oligonucleótidos 420 puede estar parcialmente inmovilizado en la partícula sintética 454. Por lo menos uno de la pluralidad de códigos de barras de oligonucleótidos 420 puede estar encerrado en la partícula sintética 454. Por lo menos uno de la pluralidad de códigos de barras de oligonucleótidos 420 puede estar parcialmente encerrado en la partícula sintética 454. La partícula sintética 454 puede ser alterable. La partícula sintética 454 puede comprender una perla. La microesfera puede comprender una perla de Sepharose, una perla de estreptavidina, una perla de agarosa, una perla magnética, una perla conjugada, una perla conjugada con proteína A, una perla conjugada con proteína G, una perla conjugada con proteína A/G, una perla conjugada con proteína L, una perla conjugada con oligo(dT), una perla de sílice, una perla similar a la sílice, una microperla antibiotina, una microperla antifluorocromo, o cualquier combinación de las mismas. La partícula sintética puede comprender un material seleccionado del grupo que consiste en polidimetilsiloxano (PDMS), poliestireno, vidrio, polipropileno, agarosa, gelatina, hidrogel, paramagnético, cerámica, plástico, vidrio, metilestireno, polímero acrílico, titanio, látex, Sepharose, celulosa, nylon, silicona, y cualquier combinación de los mismos. La partícula sintética 454 puede comprender una partícula de hidrogel alterable. Cada uno de la pluralidad de códigos de barras de oligonucleótidos puede comprender un grupo funcional conector. La partícula sintética 454 puede comprender un grupo funcional de soporte sólido. El grupo funcional de soporte y el grupo funcional conector pueden estar asociados entre sí. El grupo funcional conector y el grupo funcional de soporte pueden seleccionarse individualmente del grupo que consiste en C6, biotina, estreptavidina, aminos primarios, aldehídos, cetonas, y cualquier combinación de los mismos.

REIVINDICACIONES

1. Un método para unir códigos de barras de oligonucleótidos a una diana de ácido nucleico en una muestra, que comprende:

identificar con códigos de barras copias de la diana de ácido nucleico en una muestra usando una pluralidad de códigos de barras de oligonucleótidos para generar una pluralidad de moléculas de ácido nucleico identificadas con códigos de barras, cada una de las cuales comprendiendo una secuencia de la diana de ácido nucleico, un marcador molecular y una región de unión a la diana, y en donde por lo menos 10 de la pluralidad de códigos de barras de oligonucleótidos comprenden secuencias de marcadores moleculares diferentes; unir un oligonucleótido que comprende un complemento de la región de unión a la diana a la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras para generar una pluralidad de moléculas de ácido nucleico identificadas con códigos de barras, cada una de las cuales comprendiendo la región de unión a la diana y el complemento de la región de unión a la diana; hibridar la región de unión a la diana y el complemento de la región de unión a la diana dentro de cada una de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras para formar una estructura de horquilla; y extender los extremos 3' de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras para extender la estructura de horquilla para generar una pluralidad de moléculas de ácido nucleico identificadas con códigos de barras extendidas, cada una de las cuales comprendiendo el marcador molecular y un complemento del marcador molecular.

2. El método de la reivindicación 1, que comprende además determinar la cantidad de la diana de ácido nucleico en la muestra basándose en:

(a) el número de marcadores moleculares con secuencias distintas, complementos de las mismas, o una combinación de las mismas, asociados con la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras extendidas; o
(b) el número de complementos de marcadores moleculares con secuencias distintas asociadas a la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras extendidas.

3. El método de la reivindicación 1 o de la reivindicación 2, en donde:

(a) la identificación con códigos de barras de las copias de la pluralidad de dianas comprende:

(i) poner en contacto copias de la diana de ácido nucleico con la pluralidad de códigos de barras de oligonucleótidos, en donde cada uno de los códigos de barras de oligonucleótidos comprende la región de unión a la diana capaz de hibridar con la diana de ácido nucleico; y extender las copias de la diana de ácido nucleico hibridadas con los códigos de barras de oligonucleótidos para generar la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras; o
(ii) poner en contacto copias de la diana de ácido nucleico con la pluralidad de códigos de barras de oligonucleótidos, en donde cada uno de la pluralidad de códigos de barras de oligonucleótidos comprende la región de unión a la diana, en donde la región de unión a la diana hibrida con la diana de ácido nucleico; y extender las copias de la diana de ácido nucleico hibridadas con los códigos de barras de oligonucleótidos para generar la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras; y/o

(b) que comprende amplificar la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras para generar una pluralidad de moléculas de ácido nucleico identificadas con códigos de barras amplificadas, en donde unir el oligonucleótido que comprende el complemento de la región de unión a la diana comprende: unir el oligonucleótido que comprende el complemento de la región de unión a la diana a la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras amplificadas para generar la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras, cada una de las cuales comprendiendo la región de unión a la diana y un complemento de la región de unión a la diana.

4. El método de la reivindicación 2 o 3:

(a) que comprende amplificar la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras extendidas para generar una pluralidad de moléculas de ácido nucleico marcadas individualmente, cada una de las cuales comprendiendo el complemento del marcador molecular, en donde determinar la cantidad de la diana de ácido nucleico en la muestra comprende: determinar la cantidad de la diana de ácido nucleico en la muestra basándose en el número de complementos de marcadores moleculares con secuencias distintas asociadas a la pluralidad de moléculas de ácido nucleico marcadas individualmente; o
(b) que comprende amplificar la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras extendidas para generar copias de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras extendidas,

donde determinar la cantidad de la diana de ácido nucleico en la muestra comprende: determinar la cantidad de la diana de ácido nucleico en la muestra basándose en el número de complementos de marcadores moleculares con secuencias distintas asociadas con las copias de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras extendidas.

5. Un método para determinar la cantidad de una diana de ácido nucleico en una muestra, que comprende:

poner en contacto copias de una diana de ácido nucleico en una muestra con una pluralidad de códigos de barras de oligonucleótidos, en donde cada uno de la pluralidad de códigos de barras de oligonucleótidos comprende un marcador molecular y una región de unión a la diana capaz de hibridar con la diana de ácido nucleico, y en donde por lo menos 10 de los códigos de barras de oligonucleótidos comprenden secuencias de marcadores moleculares diferentes;

extender las copias de la diana de ácido nucleico hibridadas con los códigos de barras de oligonucleótidos para generar una pluralidad de moléculas de ácido nucleico, cada una de las cuales comprendiendo una secuencia complementaria a por lo menos una parte de la diana de ácido nucleico;

amplificar la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras para generar una pluralidad de moléculas de ácido nucleico identificadas con códigos de barras amplificadas;

unir un oligonucleótido que comprende el complemento de la región de unión a la diana a la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras amplificadas para generar una pluralidad de moléculas de ácido nucleico identificadas con códigos de barras, cada una de las cuales comprendiendo la región de unión a la diana y un complemento de la región de unión a la diana;

hibridar la región de unión a la diana y el complemento de la región de unión a la diana dentro de cada una de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras para formar una estructura de horquilla;

extender los extremos 3' de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras para extender la estructura de horquilla y generar una pluralidad de moléculas de ácido nucleico identificadas con códigos de barras extendidas, cada una de las cuales comprendiendo el marcador molecular y un complemento del marcador molecular;

amplificar la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras extendidas para generar una pluralidad de moléculas de ácido nucleico marcadas individualmente, cada una de las cuales comprendiendo el complemento del marcador molecular; y

determinar la cantidad de la diana de ácido nucleico en la muestra basándose en el número de complementos de marcadores moleculares con secuencias distintas asociadas con la pluralidad de moléculas de ácido nucleico marcadas individualmente.

6. El método de la reivindicación 4 o de la reivindicación 5, en donde el marcador molecular hibrida con el complemento del marcador molecular después de extender los extremos 3' de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras, el método comprendiendo desnaturalizar la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras extendidas antes de amplificar la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras extendidas para generar la pluralidad de moléculas de ácido nucleico marcadas individualmente.

7. El método de la reivindicación 5 o de la reivindicación 6,

en donde poner en contacto copias de la diana de ácido nucleico en la muestra comprende poner en contacto copias de una pluralidad de dianas de ácido nucleico con una pluralidad de códigos de barras de oligonucleótidos, en donde extender las copias de la diana de ácido nucleico comprende extender las copias de la pluralidad de dianas de ácido nucleico hibridadas con los códigos de barras de oligonucleótidos para generar una pluralidad de moléculas de ácido nucleico identificadas con códigos de barras, cada una de las cuales comprendiendo una secuencia complementaria a por lo menos una parte de una de la pluralidad de dianas de ácido nucleico, y

en donde determinar el número de dianas de ácido nucleico comprende determinar el número de cada una de la pluralidad de dianas de ácido nucleico en la muestra basándose en el número de los complementos de los marcadores moleculares con secuencias distintas asociadas con moléculas de ácido nucleico marcadas individualmente de la pluralidad de moléculas de ácido nucleico marcadas individualmente que comprenden una secuencia de cada una de la pluralidad de dianas de ácido nucleico,

opcionalmente, en donde la secuencia de cada una de la pluralidad de dianas de ácido nucleico comprende una subsecuencia de cada una de la pluralidad de dianas de ácido nucleico.

8. El método de cualquiera de las reivindicaciones 1-7:

(a) en donde:

(i) la secuencia de la diana de ácido nucleico en la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras comprende una subsecuencia de la diana de ácido nucleico;

(ii) la región de unión a la diana comprende una secuencia específica de un gen; y/o

(iii) unir el oligonucleótido que comprende el complemento de la región de unión a la diana comprende: ligar el oligonucleótido que comprende el complemento de la región de unión diana a la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras; o

5 (b) la región de unión a la diana comprende una secuencia poli(dT), y en donde unir el oligonucleótido que comprende el complemento de la región de unión a la diana comprende: añadir una pluralidad de monofosfatos de adenosina a la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras usando una desoxinucleotidil transferasa terminal.

10 9. El método de cualquiera de las reivindicaciones 3-8, en donde extender las copias de la diana de ácido nucleico hibridadas con los códigos de barras de oligonucleótidos comprende:

(a) transcribir inversamente las copias de la diana de ácido nucleico hibridadas con los códigos de barras de oligonucleótidos para generar una pluralidad de moléculas de ácido desoxirribonucleico complementario (ADNc) identificadas con códigos de barras; o

15 (b) extender las copias de la diana de ácido nucleico hibridadas con los códigos de barras de oligonucleótidos usando una ADN polimerasa que carezca de por lo menos una de la actividad de exonucleasa de 5' a 3' y de la actividad de exonucleasa de 3' a 5', opcionalmente en donde la ADN polimerasa comprende un fragmento de Klenow.

20 10. El método de cualquiera de las reivindicaciones 1-9:

(a) que comprende obtener información de secuencia de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras extendidas,

25 opcionalmente, en donde obtener la información de secuencia comprende unir adaptadores de secuenciación a la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras extendidas;

(b) en donde el complemento de la región de unión a la diana comprende:

30 (i) la secuencia complementaria inversa de la región de unión a la diana; o

(ii) la secuencia complementaria de la región de unión a la diana;

(c) en donde el complemento del marcador molecular comprende:

35 (i) una secuencia complementaria inversa del marcador molecular; o

(ii) una secuencia complementaria del marcador molecular;

(d) en donde:

40 (i) la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras comprende moléculas de ácido desoxirribonucleico (ADN) identificadas con códigos de barras; o

(ii) las moléculas de ácido nucleico identificadas con códigos de barras comprenden moléculas de ácido ribonucleico (ARN) identificadas con códigos de barras;

45 (e) en donde la diana de ácido nucleico comprende o consiste en una molécula de ácido nucleico, opcionalmente en donde:

(i) la molécula de ácido nucleico comprende ácido ribonucleico (ARN), ARN mensajero (ARNm), microARN, ARN interferente pequeño (ARNip), producto de degradación del ARN, ARN que comprende una cola poli(A), o cualquier combinación de los mismos; y/o

50 (ii) la diana de ácido nucleico comprende un reactivo de unión a componente celular, opcionalmente en donde la molécula de ácido nucleico está asociada con el reactivo de unión a componente celular, que comprende opcionalmente la disociación de la molécula de ácido nucleico y el reactivo de unión al componente celular;

55 (f) en donde cada marcador molecular de la pluralidad de códigos de barras de oligonucleótidos comprende por lo menos 6 nucleótidos;

(g) en donde el código de barras de oligonucleótidos comprende un marcador de muestra idéntico, opcionalmente, en donde cada marcador de muestra de la pluralidad de códigos de barras de oligonucleótidos comprende por lo menos 6 nucleótidos;

60 (h) en donde el código de barras de oligonucleótidos comprende un marcador celular idéntico, opcionalmente, en donde cada marcador celular de la pluralidad de códigos de barras de oligonucleótidos comprende por lo menos 6 nucleótidos; y/o

(j) en donde por lo menos una de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras:

65 (i) se asocia con un soporte sólido al hibridar la región de unión a la diana y el complemento de la región de

unión a la diana dentro de cada una de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras para formar la estructura de horquilla;

(ii) se disocia de un soporte sólido al hibridar la región de unión a la diana y el complemento de la región de unión a la diana dentro de cada una de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras para formar la estructura de horquilla;

(iii) se asocia con un soporte sólido al hibridar la región de unión a la diana y el complemento de la región de unión a la diana dentro de cada una de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras para formar la estructura de horquilla;

(iv) se asocia con un soporte sólido al extender los extremos 3' de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras para extender la estructura de horquilla y generar la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras extendidas, cada una de las cuales comprendiendo el marcador molecular y un complemento del marcador molecular;

(v) se disocia de un soporte sólido al hibridar la región de unión a la diana y el complemento de la región de unión a la diana dentro de cada una de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras para formar la estructura de horquilla; o

(vi) no se asocia con un soporte sólido al extender los extremos 3' de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras para extender la estructura de horquilla para generar la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras extendidas, cada una de las cuales comprendiendo el marcador molecular y un complemento del marcador molecular.

11. El método de la reivindicación 10(j), en donde el soporte sólido comprende:

- (a) una partícula sintética; o
- (b) una superficie plana.

12. El método de cualquiera de las reivindicaciones 1-10(h), en donde:

(a) por lo menos una de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras está en solución:

(i) al hibridar la región de unión a la diana y el complemento de la región de unión a la diana dentro de cada una de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras para formar la estructura de horquilla; o

(ii) al extender los extremos 3' de la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras para extender la estructura de horquilla y generar la pluralidad de moléculas de ácido nucleico identificadas con códigos de barras extendidas, cada una de las cuales comprendiendo el marcador molecular y un complemento del marcador molecular;

y opcionalmente

(b) en donde la solución se reparte en una división que no comprende más de una célula, opcionalmente en donde la división comprende por lo menos una de: una microgota, un micropocillo o una cámara de un dispositivo fluídico.

13. El método de cualquiera de las reivindicaciones 1-12, en donde la muestra comprende una célula individual, el método comprendiendo asociar una partícula sintética que comprende la pluralidad de los códigos de barras de oligonucleótidos con la célula individual en la muestra,

que comprende opcionalmente lisar la célula individual después de asociar la partícula sintética con la célula individual,

opcionalmente, en donde la lisis de la célula individual comprende calentar la muestra, poner en contacto la muestra con un detergente, cambiar el pH de la muestra o cualquier combinación de los mismos.

14. El método de la reivindicación 13, en donde:

- (a) la partícula sintética y la célula individual se encuentran en el mismo pocillo;
- (b) la partícula sintética y la célula individual se encuentran en la misma gotita;
- (c) por lo menos uno de la pluralidad de códigos de barras de oligonucleótidos:

(i) está inmovilizado en la partícula sintética;

(ii) está parcialmente inmovilizado en la partícula sintética;

(iii) está encerrado en la partícula sintética; o

(iv) está parcialmente encerrado en la partícula sintética; y/o

(d) la partícula sintética es alterable;

(e) la partícula sintética comprende una perla, opcionalmente en donde:

(i) la perla comprende una perla de Sepharose, una perla de estreptavidina, una perla de agarosa, una perla magnética, una perla conjugada, una perla conjugada con proteína A, una perla conjugada con proteína G, una perla conjugada con proteína A/G una perla conjugada con proteína L, una perla conjugada con oligo(dT), una perla de sílice, una perla similar a la sílice, una microperla antibiotina, una microperla antilfluorocromo, o cualquier combinación de las mismas; y/o

(ii) la partícula sintética comprende un material seleccionado del grupo que consiste en polidimetilsiloxano (PDMS), poliestireno, vidrio, polipropileno, agarosa, gelatina, hidrogel, paramagnético, cerámica, plástico, vidrio, metilestireno, polímero acrílico, titanio, látex, Sepharose, celulosa, nylon, silicona, y cualquier combinación de los mismos; y/o

(f) la partícula sintética comprende una partícula de hidrogel alterable.

15. El método de cualquiera de las reivindicaciones 13 o 14,

en donde cada uno de la pluralidad de códigos de barras de oligonucleótidos comprende un grupo funcional conector,

en donde la partícula sintética comprende un grupo funcional de soporte sólido, y

en donde el grupo funcional de soporte y el grupo funcional conector están asociados entre sí,

opcionalmente, en donde el grupo funcional conector y el grupo funcional de soporte se seleccionan individualmente del grupo que consiste en C6, biotina, estreptavidina, aminas primarias, aldehídos, cetonas y cualquier combinación de los mismos.

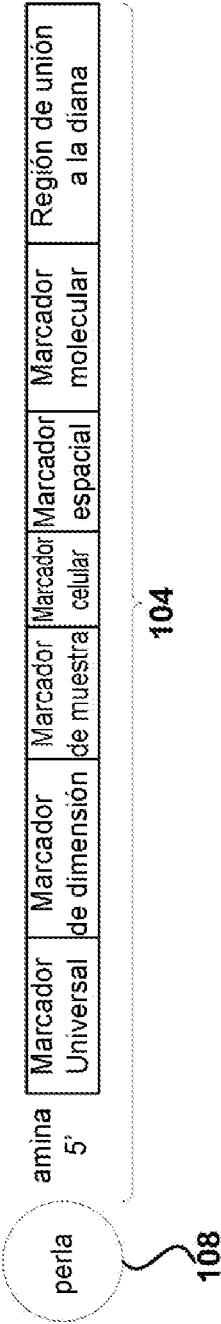


FIG. 1

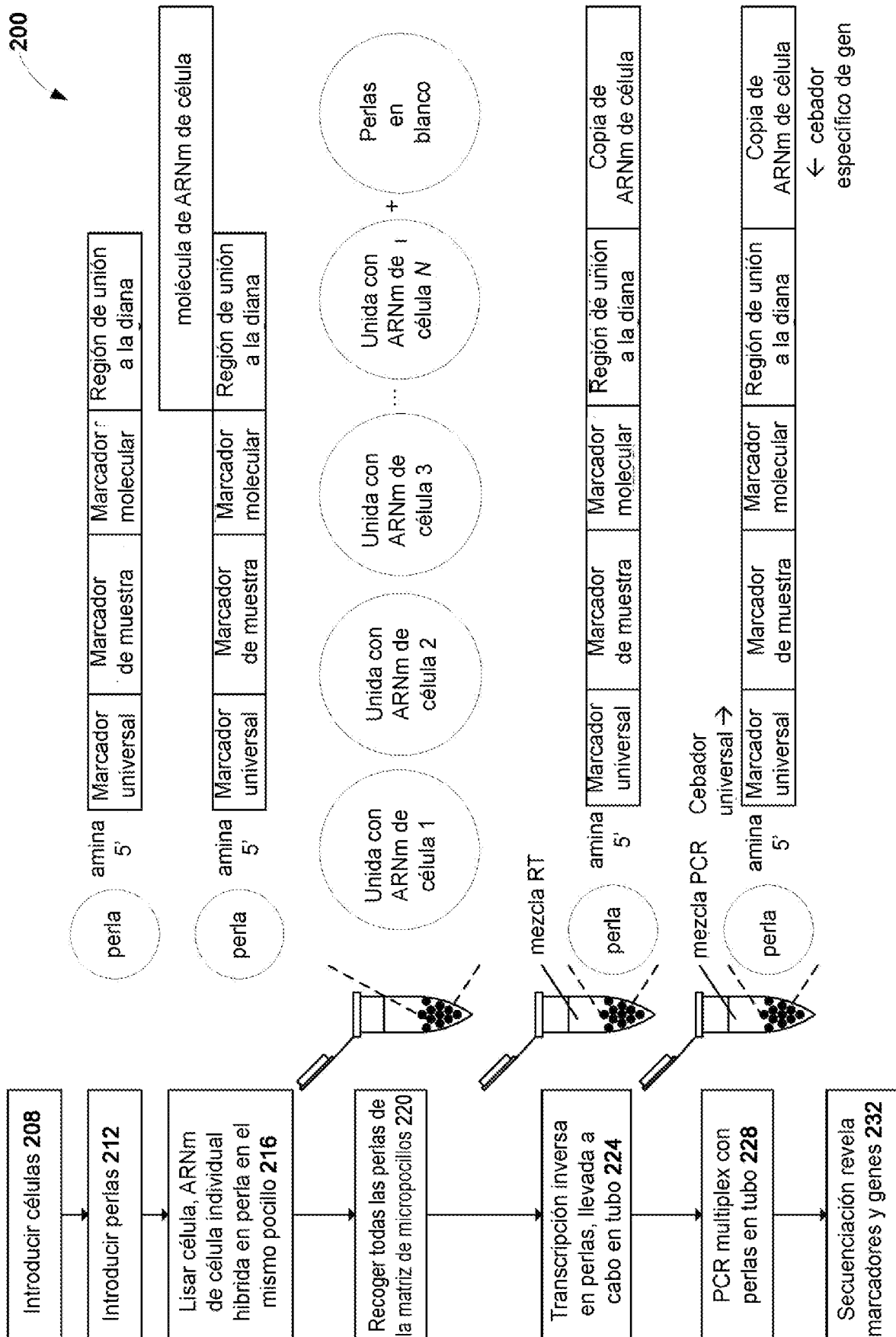
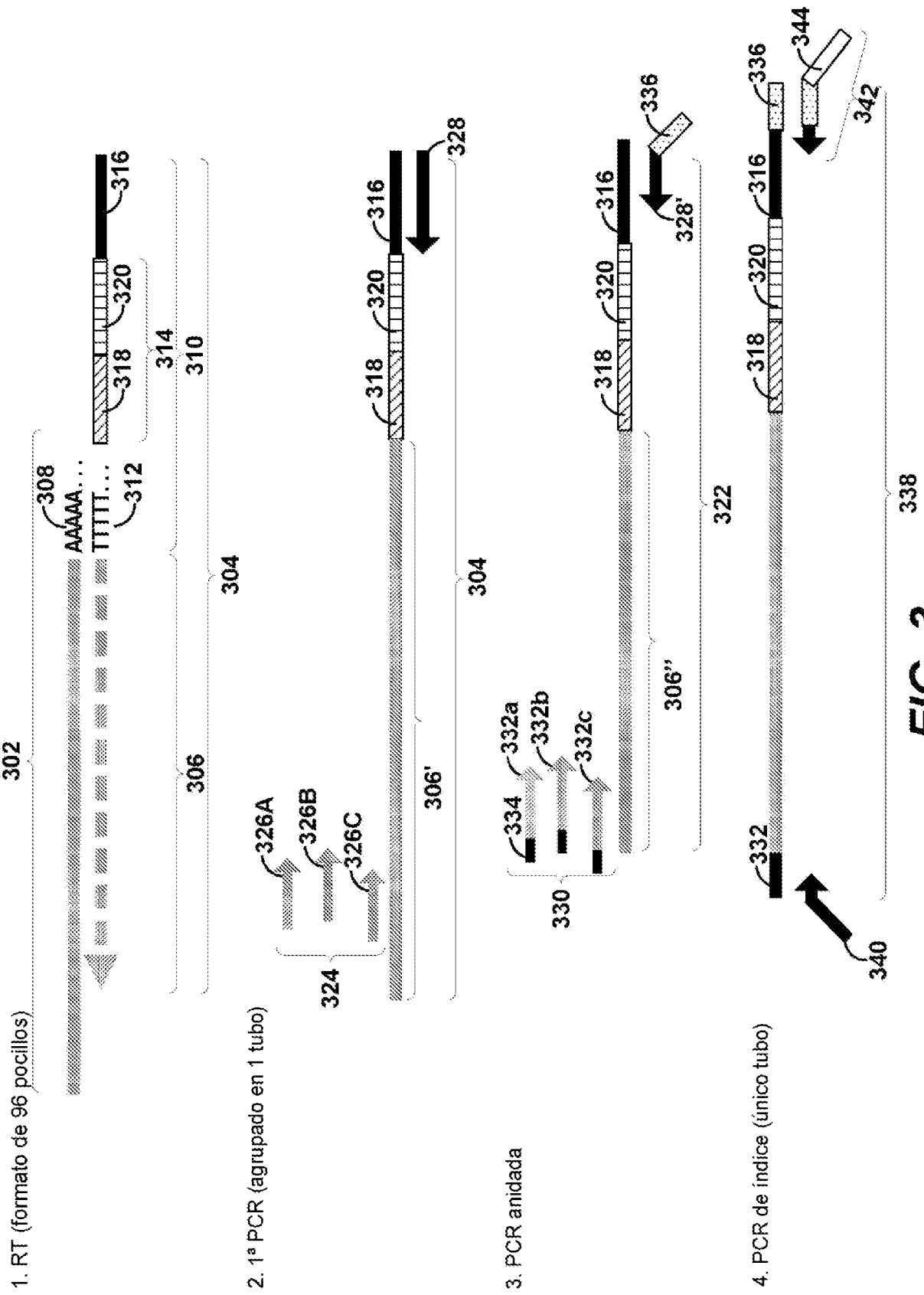
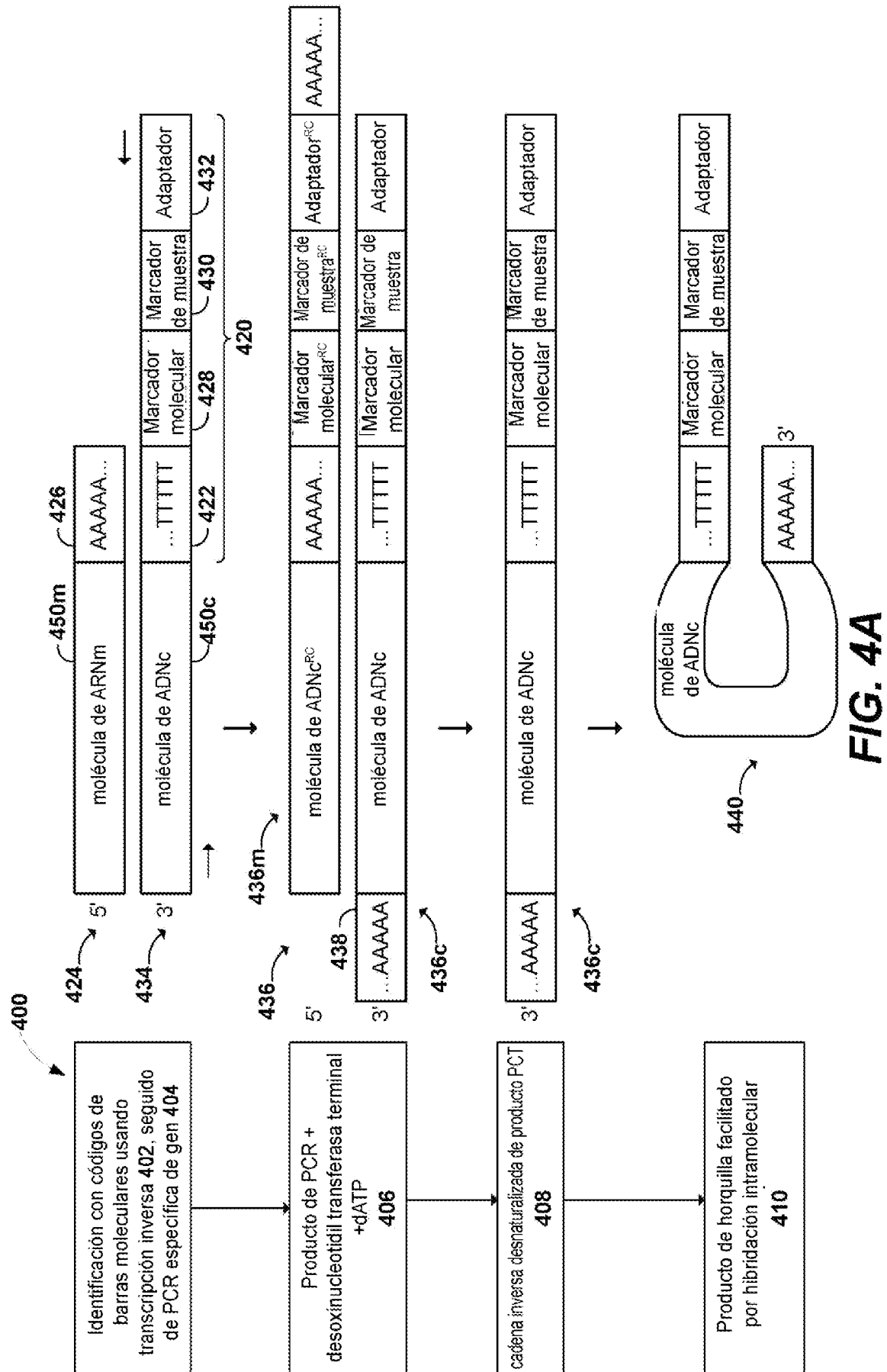


FIG. 2





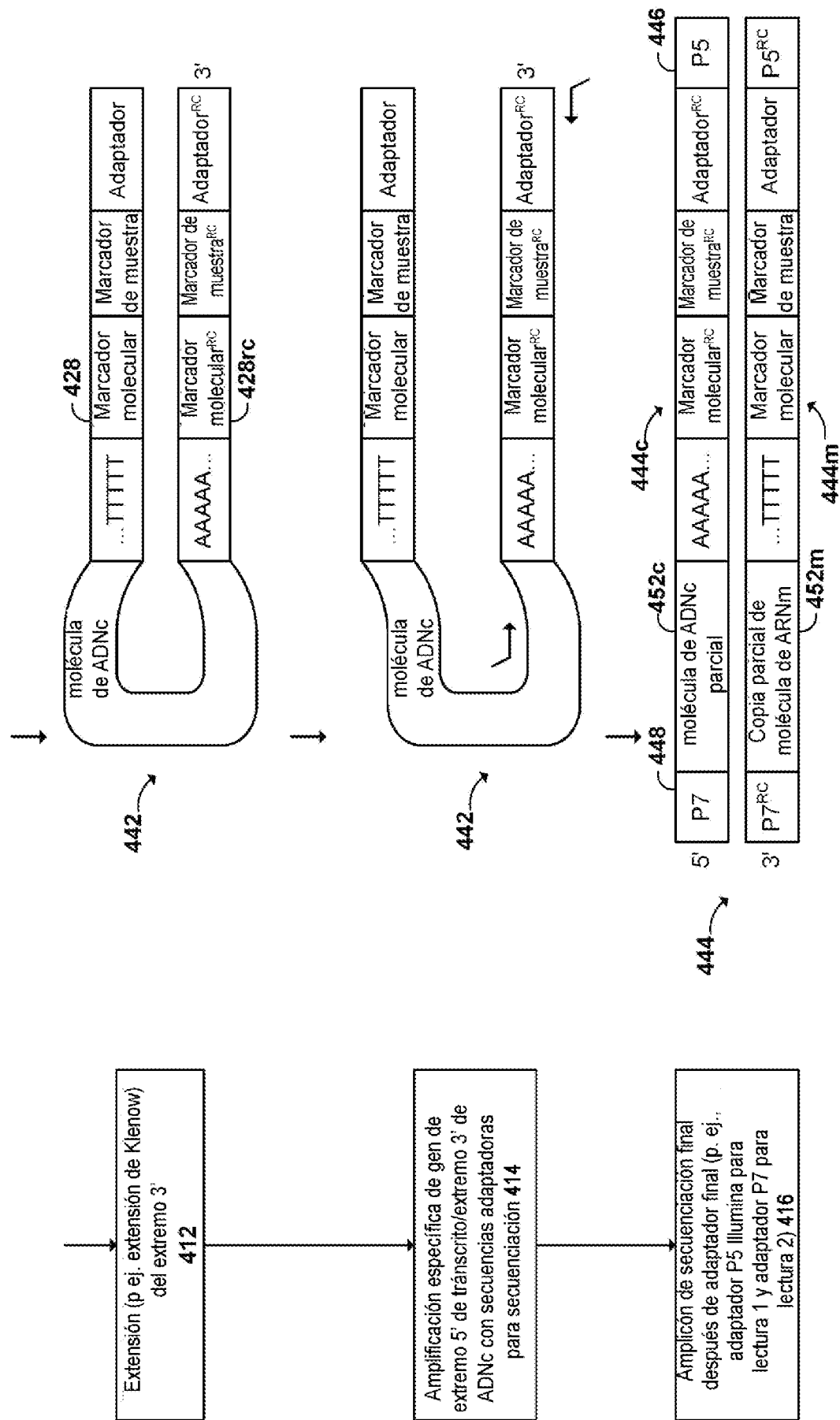


FIG. 4B

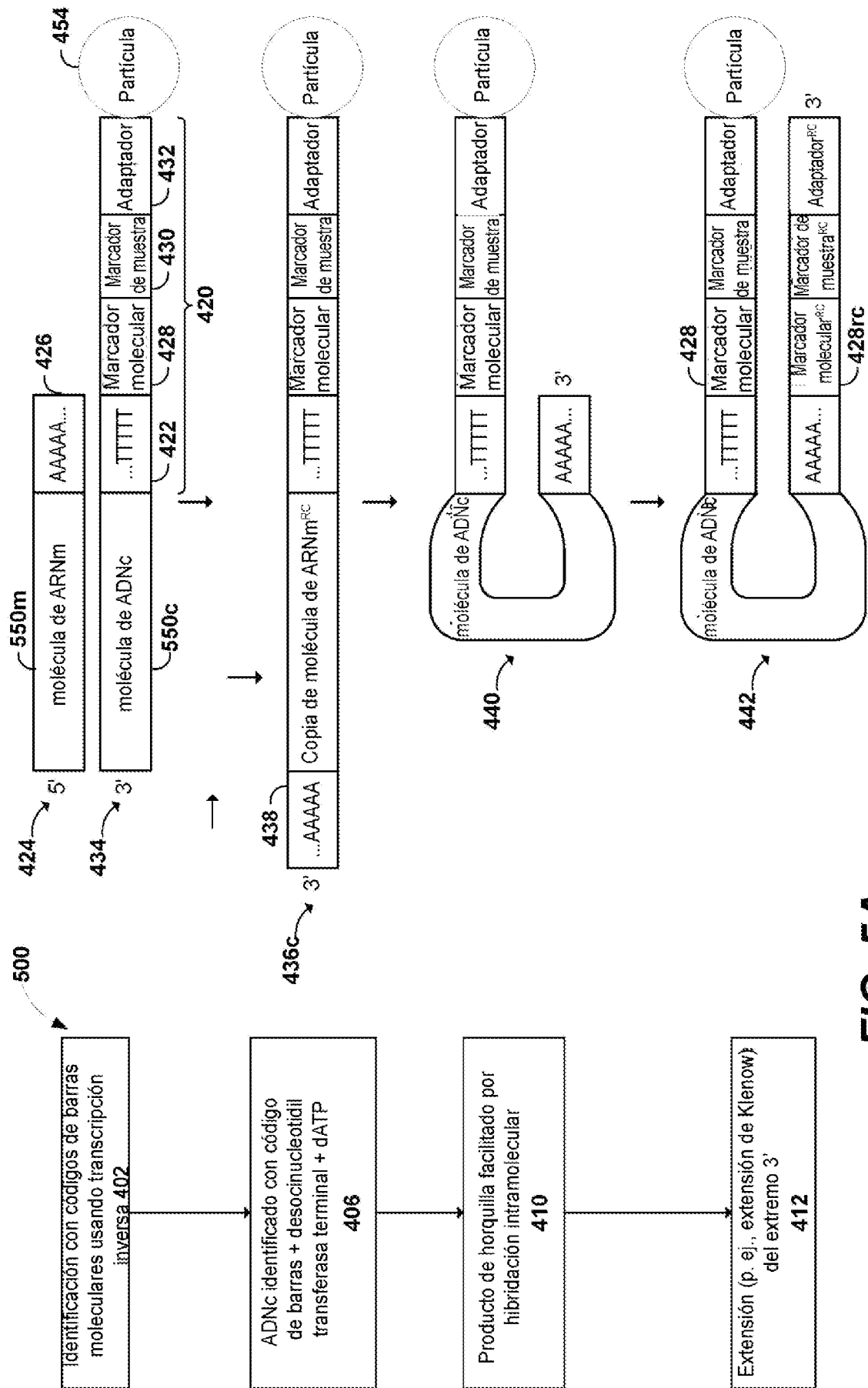


FIG. 5A

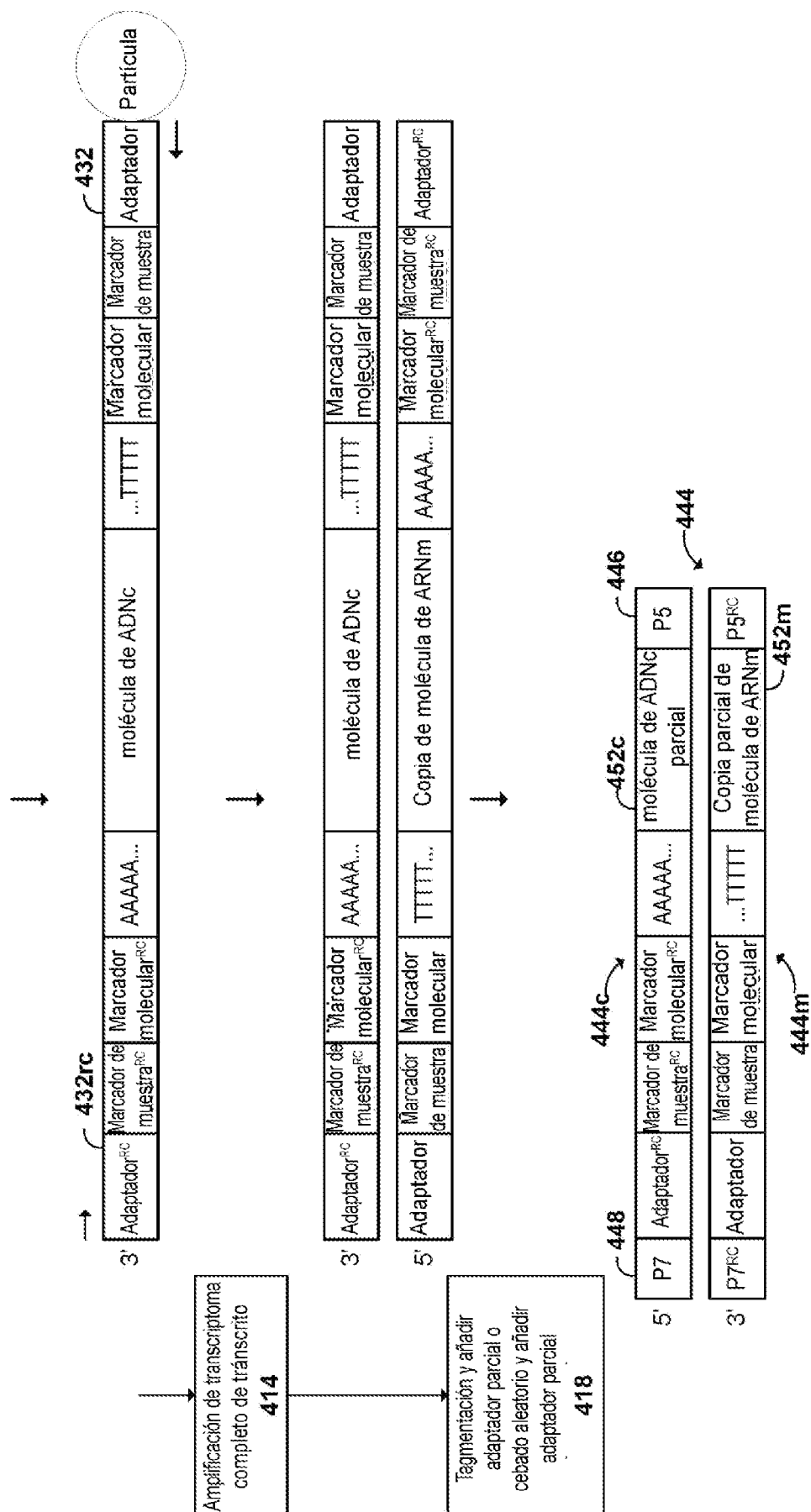


FIG. 5B