

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 745 694**

51 Int. Cl.:

**C12N 15/10** (2006.01)

**C12Q 1/68** (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **08.09.2016 PCT/US2016/050694**

87 Fecha y número de publicación internacional: **16.03.2017 WO17044574**

96 Fecha de presentación y número de la solicitud europea: **08.09.2016 E 16770134 (1)**

97 Fecha y número de publicación de la concesión europea: **26.06.2019 EP 3347465**

54 Título: **Métodos y composiciones para normalización de biblioteca de ácido nucleico**

30 Prioridad:

**11.09.2015 US 201562217220 P**

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

**03.03.2020**

73 Titular/es:

**CELLULAR RESEARCH, INC. (100.0%)  
4040 Campbell Avenue, Suite 110  
Menlo Park, CA 94025, US**

72 Inventor/es:

**BETTS, CRAIG y  
FU, GLENN**

74 Agente/Representante:

**IZQUIERDO BLANCO, María Alicia**

ES 2 745 694 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

Métodos y composiciones para normalización de biblioteca de ácido nucleico

## 5 ANTECEDENTES

[0001] En muchas aplicaciones, en particular la secuencia y la selección de la biblioteca, los investigadores quieren evaluar el catálogo completo de ácidos nucleicos en una muestra o biblioteca. Sin embargo, los ácidos nucleicos en una muestra (por ejemplo, una biblioteca de ADNc) a menudo se encuentran en un amplio rango de concentraciones que oscilan entre varios órdenes y una magnitud. Esto aumenta considerablemente la cantidad de secuenciación o selección necesaria para evaluar completamente la muestra. Para superar esto, los investigadores a menudo desean trabajar con una biblioteca normalizada donde las concentraciones de diferentes ácidos nucleicos se encuentran en una concentración similar, de modo que se requiere menos esfuerzo redundante para interrogar nuevas moléculas. El documento WO2013/191775 analiza métodos, composiciones y kits para la generación de bibliotecas de secuenciación de próxima generación en las que se han agotado o reducido sustancialmente las secuencias de ácidos nucleicos no deseadas. El documento WO2014/015098 describe protocolos de normalización para muestras biológicas, particularmente muestras que comprenden ácidos nucleicos para secuenciarse. El documento WO2012/108864 describe el enriquecimiento selectivo de ácidos nucleicos, mientras que el documento WO1999/015702 describe métodos para producir bibliotecas de ácido nucleico normalizadas en las que cada miembro de la biblioteca puede aislarse con una probabilidad aproximadamente equivalente. Patanjali et al (1991) PNAS, 88, 5, 1943-1947 discuten el uso de un enfoque cinético para construir bibliotecas de ADNc que contienen representaciones aproximadamente iguales de todas las secuencias en una preparación de ARN poli(A)+. El documento WO2013/117595 describe métodos para enriquecimiento dirigido y amplificación de ácidos nucleicos en un soporte y el documento WO2002/070684 se refiere a métodos para generar una biblioteca de genes caracterizada porque contiene sustancialmente todos los genes de una fuente deseada en cantidades sustancialmente equimolares, en donde cada gen está representado por una cantidad suficiente de fragmentos diferentes. Bogdanova et al. (2008) Molecular biosystems, 4, 3, 205 analiza la normalización de ADNc enriquecido de longitud completa utilizando métodos como la normalización por DSN, que se basa en las propiedades únicas de la nucleasa específica a dúplex (DSN) del cangrejo kamchatka y analiza la generación de bibliotecas de ADNc normalizadas con una alta tasa de descubrimiento de genes.

## SUMARIO

[0002] Algunas realizaciones descritas en este documento proporcionan métodos para la eliminación de especies de alta abundancia a partir de una pluralidad de moléculas de ácido nucleico, que comprende: hibridizar una pluralidad de primeros oligonucleótidos que comprende un resto de carácter definitivo con una primera pluralidad de moléculas de ácido nucleico, en el que la primera pluralidad de moléculas de ácido nucleico comprende al menos una especie de alta abundancia; extender la pluralidad de primeros oligonucleótidos para generar una pluralidad de cadenas complementarias de la primera pluralidad de moléculas de ácido nucleico que comprenden el resto de unión; desnaturalizando una pluralidad de moléculas de ácido nucleico de doble cadena que comprende la pluralidad de cadenas complementarias de la primera pluralidad de moléculas de ácido nucleico; la reanudación parcial de la pluralidad de cadenas complementarias de la primera pluralidad de moléculas de ácido nucleico; y la eliminación de las cadenas complementarias reanudadas de la primera pluralidad de moléculas de ácido nucleico mediante una molécula de captura inmovilizada en uno o más soportes sólidos para generar una segunda pluralidad de moléculas de ácido nucleico, en donde las moléculas de captura se unen específicamente al resto de unión, por lo que el contenido de al menos una especie de alta abundancia en la segunda pluralidad de moléculas de ácido nucleico se reduce en comparación con el contenido de la al menos una especie de alta abundancia en la primera pluralidad de moléculas de ácido nucleico.

[0003] En algunas realizaciones, el resto de unión es un grupo funcional seleccionado del grupo que consiste en biotina, estreptavidina, heparina, un aptámero, un resto de click-química, digoxigenina, amina(s) primaria(s), carboxilo(s), hidroxilo(s), aldehído(s), cetona(s) y cualquier combinación de los mismos. En algunas realizaciones, el resto de unión es biotina. En algunas realizaciones, la molécula de captura es estreptavidina. En algunas realizaciones, los métodos comprenden además sintetizar una segunda cadena para al menos una de la pluralidad de cadenas complementarias de la primera pluralidad de moléculas de ácido nucleico para generar una o más de la pluralidad de moléculas de ácido nucleico bicatenarias que comprenden la pluralidad de cadenas complementarias de la primera pluralidad de moléculas de ácido nucleico. En algunas realizaciones, la síntesis comprende hibridar una pluralidad de segundos oligonucleótidos con la pluralidad de cadenas complementarias de la primera pluralidad de moléculas de ácido nucleico y extender la pluralidad de segundo oligonucleótido. En algunas realizaciones, la pluralidad de primeros oligonucleótidos o la pluralidad de segundos oligonucleótidos comprende un sitio de unión a cebador universal. En algunas realizaciones, los métodos comprenden además amplificar la pluralidad de moléculas de ácido nucleico de doble cadena. En algunas realizaciones, la primera pluralidad de moléculas de ácido nucleico comprende una pluralidad de especies de alta abundancia. En algunas realizaciones, la al menos una especie de alta abundancia representa al menos el 50% de la primera pluralidad de moléculas de ácido nucleico. En algunas realizaciones, la al menos una especie de alta abundancia representa al menos el 60% de la primera pluralidad de moléculas de ácido nucleico. En algunas realizaciones, la al menos una especie de alta abundancia representa al menos el 70% de la

primera pluralidad de moléculas de ácido nucleico. En algunas realizaciones, la reducción del contenido de al menos una especie de alta abundancia es al menos del 80%. En algunas realizaciones, la reducción del contenido de al menos una especie de alta abundancia es de al menos el 90%. En algunas realizaciones, la reducción del contenido de al menos una especie de alta abundancia es de al menos el 95%. En algunas realizaciones, la reducción del contenido de al menos una especie de alta abundancia es de al menos el 99%. En algunas realizaciones, la segunda pluralidad de moléculas de ácido nucleico comprende la pluralidad de especies de alta abundancia. En algunas realizaciones, la pluralidad de especies de alta abundancia en la segunda pluralidad de moléculas de ácido nucleico representan menos del 50% de la segunda pluralidad de moléculas de ácido nucleico. En algunas realizaciones, la pluralidad de especies de alta abundancia en la segunda pluralidad de moléculas de ácido nucleico representa menos del 40% de la segunda pluralidad de moléculas de ácido nucleico. En algunas realizaciones, la pluralidad de especies de alta abundancia en la segunda pluralidad de moléculas de ácido nucleico representa menos del 30% de la segunda pluralidad de moléculas de ácido nucleico. En algunas realizaciones, la primera pluralidad de moléculas de ácido nucleico comprende una pluralidad de especies de baja abundancia. En algunas realizaciones, la pluralidad de especies de baja abundancia representa menos del 10% de la primera pluralidad de moléculas de ácido nucleico. En algunas realizaciones, la pluralidad de especies de baja abundancia representa menos del 5% de la primera pluralidad de moléculas de ácido nucleico. En algunas realizaciones, la pluralidad de especies de baja abundancia representa menos del 1% de la primera pluralidad de moléculas de ácido nucleico. En algunas realizaciones, la segunda pluralidad de moléculas de ácido nucleico comprende la pluralidad de especies de baja abundancia. En algunas realizaciones, la pluralidad de especies de baja abundancia en la segunda pluralidad de moléculas de ácido nucleico representa al menos el 5% de la segunda pluralidad de moléculas de ácido nucleico. En algunas realizaciones, la pluralidad de especies de baja abundancia en la segunda pluralidad de moléculas de ácido nucleico representa al menos el 10% de la segunda pluralidad de moléculas de ácido nucleico. En algunas realizaciones, la pluralidad de especies de baja abundancia en la segunda pluralidad de moléculas de ácido nucleico representa al menos el 20% de la segunda pluralidad de moléculas de ácido nucleico. En algunas realizaciones, cada una de la primera pluralidad de moléculas de ácido nucleico o cada una de la segunda pluralidad de moléculas de ácido nucleico comprende un código de barras estocástico. En algunas realizaciones, los métodos comprenden además la secuenciación de la segunda pluralidad de moléculas de ácido nucleico para generar una pluralidad de lecturas de secuenciación. En algunas realizaciones, las lecturas de secuencia para la pluralidad de especies de alta abundancia es menos del 50% del total de lecturas de secuencia. En algunas realizaciones, las lecturas de secuencia para la pluralidad de especies de alta abundancia es menos del 40% del total de lecturas de secuencia. En algunas realizaciones, las lecturas de secuencia para la pluralidad de especies de alta abundancia son menos del 30% de las lecturas de secuenciación totales. En algunas realizaciones, las lecturas de secuencia para la pluralidad de especies de baja abundancia son al menos el 5% de las lecturas de secuenciación totales. En algunas realizaciones, las lecturas de secuenciación para la pluralidad de especies de baja abundancia son al menos el 10% de las lecturas de secuenciación totales. En algunas realizaciones, las lecturas de secuencia para la pluralidad de especies de baja abundancia son al menos el 20% del total de lecturas de secuenciación. En algunas realizaciones, los métodos comprenden además la adición de una pluralidad de bloqueadores durante la etapa de recocido parcial. En algunas realizaciones, la pluralidad de híbridos bloqueadores al sitio de unión del cebador universal del primer oligonucleótido o el sitio de unión del cebador universal del segundo oligonucleótido. En algunas realizaciones, la pluralidad de bloqueadores impide la hibridación entre el sitio de unión del cebador universal del primer oligonucleótido o el sitio de unión del cebador universal del segundo oligonucleótido y su secuencia complementaria.

**[0004]** Algunas realizaciones descritas en este documento proporcionar métodos de generación de una biblioteca de ácido nucleico normalizada, que comprende: hibridación de una pluralidad de primeros oligonucleótidos que comprenden un resto de unión con una pluralidad de dianas de ácido nucleico; extender la pluralidad de primeros oligonucleótidos para generar una pluralidad de cadenas complementarias de la pluralidad de dianas de ácido nucleico que comprenden el resto de unión; desnaturalizar una pluralidad de moléculas de ácido nucleico de doble cadena que comprenden la pluralidad de cadenas complementarias de la pluralidad de dianas de ácido nucleico; volver a cerrar parcialmente la pluralidad de cadenas complementarias de la pluralidad de dianas de ácido nucleico; y la eliminación de las cadenas complementarias recocidas de la pluralidad de dianas de ácido nucleico por una molécula de captura inmovilizada en uno o más soportes sólidos, en donde las moléculas de captura se unen específicamente al resto de unión, donde se genera una biblioteca de ácido nucleico normalizada de la pluralidad de objetivos de ácido nucleico.

**[0005]** En algunas realizaciones, el resto de unión es un grupo funcional seleccionado del grupo que consiste en biotina, estreptavidina, heparina, un aptámero, un resto de click-química, digoxigenina, amina(s) primaria(s), carboxilo(s), hidroxilo(s), aldehído(s), cetona(s) y cualquier combinación de los mismos. En algunas realizaciones, el resto de unión es biotina. En algunas realizaciones, la molécula de captura es estreptavidina. En algunas realizaciones, los métodos comprenden además sintetizar una segunda cadena para una o más de la pluralidad de cadenas complementarias de la pluralidad de dianas de ácido nucleico para generar una o más de la pluralidad de moléculas de ácido nucleico bicatenarias que comprenden la pluralidad de cadenas complementarias de la pluralidad de dianas de ácido nucleico. En algunas realizaciones, la síntesis comprende hibridar una pluralidad de segundos oligonucleótidos con la pluralidad de cadenas complementarias de la pluralidad de dianas de ácido nucleico y extender la pluralidad de segundo oligonucleótido. En algunas realizaciones, la pluralidad de primeros oligonucleótidos o la pluralidad de segundos oligonucleótidos comprende un sitio de unión a cebador universal. En algunas realizaciones, los métodos comprenden además amplificar la pluralidad de moléculas de ácido nucleico de doble cadena. En algunas realizaciones, la pluralidad de dianas de ácido nucleico comprende una pluralidad de dianas de ácido nucleico de baja

abundancia. En algunas realizaciones, la pluralidad de objetivos de ácido nucleico de baja abundancia representa menos del 10% de la pluralidad de dianas de ácido nucleico. En algunas realizaciones, la pluralidad de dianas de ácido nucleico de baja abundancia representa menos del 5% de la pluralidad de dianas de ácido nucleico. En algunas realizaciones, la pluralidad de dianas de ácido nucleico de baja abundancia representa menos del 1% de la pluralidad de dianas de ácido nucleico. En algunas realizaciones, la biblioteca de ácidos nucleicos normalizada de la pluralidad de dianas de ácidos nucleicos comprende la pluralidad de dianas de ácidos nucleicos de baja abundancia. En algunas realizaciones, la pluralidad de dianas de ácido nucleico de baja abundancia en la biblioteca de ácidos nucleicos normalizada representa al menos el 5% de la pluralidad de dianas de ácido nucleico en la biblioteca de ácidos nucleicos normalizada. En algunas realizaciones, la pluralidad de dianas de ácidos nucleicos de baja abundancia en la biblioteca de ácidos nucleicos normalizados representa al menos el 10% de la pluralidad de dianas de ácidos nucleicos en la biblioteca de ácidos nucleicos normalizados. En algunas realizaciones, la pluralidad de dianas de ácido nucleico de baja abundancia en la biblioteca de ácidos nucleicos normalizada representa al menos el 20% de la pluralidad de dianas de ácido nucleico en la biblioteca de ácidos nucleicos normalizada. En algunas realizaciones, la pluralidad de dianas de ácido nucleico comprende una pluralidad de dianas de ácido nucleico de alta abundancia. En algunas realizaciones, la pluralidad de dianas de ácido nucleico de alta abundancia representa al menos el 50% de la pluralidad de dianas de ácido nucleico. En algunas realizaciones, la pluralidad de dianas de ácido nucleico de alta abundancia representa al menos el 60% de la pluralidad de dianas de ácido nucleico. En algunas realizaciones, la pluralidad de dianas de ácido nucleico de alta abundancia representa al menos el 70% de la pluralidad de dianas de ácido nucleico. En algunas realizaciones, el contenido de la pluralidad de especies de alta abundancia en la biblioteca de ácidos nucleicos normalizados se reduce en al menos un 80%. En algunas realizaciones, el contenido de la pluralidad de especies de alta abundancia en la biblioteca de ácidos nucleicos normalizados se reduce en al menos un 90%. En algunas realizaciones, el contenido de la pluralidad de especies de alta abundancia en la biblioteca de ácidos nucleicos normalizados se reduce en al menos el 95%. En algunas realizaciones, el contenido de la pluralidad de especies de alta abundancia en la biblioteca de ácidos nucleicos normalizados se reduce en al menos el 99%. En algunas realizaciones, la biblioteca de ácido nucleico normalizada de la pluralidad de dianas de ácido nucleico comprende la pluralidad de dianas de ácido nucleico de alta abundancia. En algunas realizaciones, la pluralidad de dianas de ácido nucleico de alta abundancia en la biblioteca de ácido nucleico normalizada representa menos del 50% de la pluralidad de dianas de ácido nucleico en la biblioteca de ácido nucleico normalizada. En algunas realizaciones, la pluralidad de objetivos de ácidos nucleicos de alta abundancia en la biblioteca de ácidos nucleicos normalizados representa menos del 40% de la pluralidad de dianas de ácidos nucleicos en la biblioteca de ácidos nucleicos normalizados. En algunas realizaciones, la pluralidad de dianas de ácidos nucleicos de alta abundancia en la biblioteca de ácidos nucleicos normalizada representa menos del 30% de la pluralidad de dianas de ácidos nucleicos en la biblioteca de ácidos nucleicos normalizada. En algunas realizaciones, cada uno de la pluralidad de primeros oligonucleótidos o cada uno de la pluralidad de segundos oligonucleótidos comprende un código de barras estocástico. En algunas realizaciones, los métodos comprenden además la secuenciación de la biblioteca de ácidos nucleicos normalizada para generar una pluralidad de lecturas de secuenciación. En algunas realizaciones, las lecturas de secuenciación para la pluralidad de dianas de ácido nucleico de alta abundancia son menos del 50% de las lecturas de secuenciación totales. En algunas realizaciones, las lecturas de secuencia para la pluralidad de dianas de ácido nucleico de alta abundancia es menos del 40% de las lecturas de secuenciación total. En algunas realizaciones, las lecturas de secuencia para la pluralidad de dianas de ácidos nucleicos de alta abundancia son menos del 30% de las lecturas de secuenciación totales. En algunas realizaciones, las lecturas de secuenciación para la pluralidad de dianas de ácido nucleico de baja abundancia son al menos el 5% de las lecturas de secuenciación total. En algunas realizaciones, las lecturas de secuencia para la pluralidad de dianas de ácido nucleico de baja abundancia son al menos el 10% de las lecturas de secuenciación total. En algunas realizaciones, las lecturas de secuencia para la pluralidad de dianas de ácido nucleico de baja abundancia son al menos el 20% de las lecturas de secuenciación totales. En algunas realizaciones, los métodos comprenden además la adición de una pluralidad de bloqueadores durante la etapa de reclasificación parcial. En algunas realizaciones, la pluralidad de híbridos bloqueadores al sitio de unión del cebador universal del primer oligonucleótido o el sitio de unión del cebador universal del segundo oligonucleótido. En algunas realizaciones, la pluralidad de bloqueadores previene la hibridación entre el sitio de unión del cebador universal del primer oligonucleótido o el sitio de unión del cebador universal del segundo oligonucleótido y su secuencia complementaria. En algunas realizaciones, la pluralidad de dianas de ácido nucleico comprende ARNm. En algunas realizaciones, la pluralidad de dianas de ácido nucleico comprende ARNm mitocondrial. En algunas realizaciones, la pluralidad de dianas de ácido nucleico comprende ARNm de proteína ribosomal. En algunas realizaciones, las dianas de ácido nucleico de baja abundancia comprenden 7.000 genes con el número más bajo de transcripciones. En algunas realizaciones, las dianas de ácido nucleico de baja abundancia comprenden 4.000 genes con el número más bajo de transcripciones. En algunas realizaciones, las dianas de ácido nucleico de baja abundancia comprenden 2.000 genes con el número más bajo de transcripciones. En algunas realizaciones, la pluralidad de primeros oligo-nucleótidos comprende cebadores específicos de diana. En algunas realizaciones, la pluralidad de primeros oligonucleótidos comprende cebadores no específicos de diana. En algunas realizaciones, la pluralidad de dianas de ácido nucleico comprende ADNc. En algunas realizaciones, la pluralidad de dianas de ácido nucleico comprende ADN genómico. En algunas realizaciones, las dianas de ácido nucleico de alta abundancia comprenden secuencias cortas de repetición en tándem. En algunas realizaciones, las dianas de ácidos nucleicos de alta abundancia comprenden secuencias teloméricas. En algunas realizaciones, las dianas de ácidos nucleicos de alta abundancia comprenden secuencias centroméricas. En algunas realizaciones, la pluralidad de dianas de ácido nucleico es de una sola célula.

65

**[0006]** Algunas realizaciones descritas en este documento proporcionan métodos de generación de una biblioteca de ácido nucleico normalizada, que comprende: hibridación de una pluralidad de primeros oligonucleótidos que comprenden un resto de unión con una pluralidad de dianas de ácido nucleico en una biblioteca de ácidos nucleicos no normalizada; extendiendo la pluralidad de primeros oligonucleótidos para generar una pluralidad de cadenas complementarias de la pluralidad de dianas de ácido nucleico que comprenden el resto de unión; definir una pluralidad de moléculas de ácido nucleico de doble cadena que comprenden la pluralidad de cadenas complementarias de la pluralidad de dianas de ácido nucleico; reanudar parcialmente la pluralidad de cadenas complementarias de la pluralidad de dianas de ácido nucleico; y eliminar las cadenas complementarias recocidas de la pluralidad de dianas de ácido nucleico, por lo que se genera una biblioteca de ácidos nucleicos normalizada de la pluralidad de dianas de ácido nucleico.

**[0007]** En algunas realizaciones, la biblioteca de ácido nucleico no normalizado comprende una o más dianas de ácido nucleico de alta abundancia y una o más dianas de ácido nucleico de baja abundancia. En algunas realizaciones, la una o más dianas de ácido nucleico de alta abundancia representan al menos el 50% de la biblioteca de ácido nucleico no normalizado. En algunas realizaciones, la una o más dianas de ácido nucleico de alta abundancia representan al menos el 60% de la biblioteca de ácido nucleico no normalizado. En algunas realizaciones, la una o más dianas de ácido nucleico de alta abundancia representan al menos el 70% de la biblioteca de ácido nucleico no normalizado. En algunas realizaciones, el contenido de una o más dianas de ácidos nucleicos de alta abundancia en la biblioteca de ácidos nucleicos normalizados se reduce en al menos un 80%. En algunas realizaciones, el contenido de una o más dianas de ácidos nucleicos de alta abundancia en la biblioteca de ácidos nucleicos normalizados se reduce en al menos un 90%. En algunas realizaciones, el contenido de una o más dianas de ácido nucleico de alta abundancia en la biblioteca de ácido nucleico normalizada se reduce en al menos el 95%. En algunas realizaciones, el contenido de una o más dianas de ácidos nucleicos de gran abundancia en la biblioteca de ácidos nucleicos normalizados se reduce en al menos un 99%. En algunas realizaciones, una o más dianas de ácido nucleico de baja abundancia representan menos del 10% de la biblioteca de ácido nucleico no normalizado. En algunas realizaciones, el objetivo de ácido nucleico de baja abundancia o más representa menos del 5% de la biblioteca de ácido nucleico no normalizado. En algunas realizaciones, la una o más dianas de ácido nucleico de baja abundancia representan menos del 1% de la biblioteca de ácido nucleico no normalizado. En algunas realizaciones, el objetivo de ácido nucleico de baja abundancia o más representa al menos el 5% de la biblioteca de ácido nucleico normalizada. En algunas realizaciones, la una o más dianas de ácido nucleico de baja abundancia representan al menos el 10% de la biblioteca de ácido nucleico normalizada. En algunas realizaciones, el objetivo de ácido nucleico de baja abundancia o más representa al menos el 20% de la biblioteca de ácido nucleico normalizada. En algunas realizaciones, la una o más dianas de ácido nucleico de alta abundancia representan menos del 50% de la biblioteca de ácido nucleico normalizada. En algunas realizaciones, el objetivo de ácido nucleico de alta abundancia o más representa menos del 40% de la biblioteca de ácido nucleico normalizada. En algunas realizaciones, la una o más dianas de ácido nucleico de alta abundancia representan menos del 30% de la biblioteca de ácido nucleico normalizada. En algunas realizaciones, la biblioteca de ácidos nucleicos no normalizada es una biblioteca de ADNc. En algunas realizaciones, la biblioteca de ácido nucleico no normalizado es una biblioteca genómica. En algunas realizaciones, la biblioteca de ácido nucleico no normalizado es una biblioteca de ácido nucleico unicelular.

**[0008]** En un aspecto, la descripción proporciona un método de normalización de biblioteca de ácidos nucleicos que comprende: generar una biblioteca de ADNc de doble hebra asimétricamente marcado, en el que una hebra de doble hebra de ADNc de la biblioteca de ADNc comprenden un resto de unión; cadenas de los ADNc de doble hebra en la biblioteca, lo que genera una mezcla de moléculas de ADNc que se re-hibrida que comprenden el resto de unión, moléculas de una sola hebra que comprenden el resto de unión, y las moléculas monocatenarias que carecen del resto de unión; y eliminar las moléculas que comprenden el resto de unión, mientras que deja atrás las moléculas de hebra simple que carecen del resto de unión, produciendo así una biblioteca normalizada. En algunas realizaciones, la generación comprende la transcripción inversa de un ARNm en una primera cadena de ADNc utilizando un cebador que comprende el resto de unión. En algunas realizaciones, el cebador comprende un código de barras estocástico. En algunas realizaciones, el método comprende además generar una segunda cadena de ADNc complementaria a la primera cadena de ADNc, generando de este modo el ADNc de doble cadena. En algunas realizaciones, la generación comprende realizar la extensión del cebador en una molécula de ADN usando un cebador que comprende el resto de unión, generando de este modo el ADNc de doble cadena. En algunas realizaciones, el resto de unión se selecciona del grupo que consiste en: biotina y estreptavidina. En algunas realizaciones, el resto de unión está unido a la cadena con sentido del ADNc de doble cadena. En algunas realizaciones, el enlace de unión está unido a la cadena antisentido del ADNc de doble cadena. En algunas realizaciones, la desnaturalización comprende calentar los ADNc de doble cadena. En algunas realizaciones, la desnaturalización comprende desnaturalizar al menos el 50% de los ADNc de doble cadena. En algunas realizaciones, la desnaturalización comprende desnaturalizar al menos el 90% de los ADNc de doble cadena. En algunas realizaciones, los ADNc recocidos son de una especie de ácido nucleico muy abundante. En algunas realizaciones, las cadenas de los ADNc re-recocidos recocinan al menos el doble de rápido que las cadenas de la biblioteca de ADNc de doble cadena de una especie de ácido nucleico menos abundante. En algunas realizaciones, las cadenas de los ADNc re-recocidos recocinan al menos cinco veces más rápido que las cadenas de la cadena de ADNc de doble cadena de una especie de ácido nucleico menos abundante. En algunas realizaciones, las hebras de los ADNc recocidos se vuelven a formar en una mayor abundancia que las hebras de la biblioteca de ADNc de doble cadena de una especie de ácido nucleico menos abundante. En algunas realizaciones, la eliminación comprende poner en contacto las moléculas con un soporte sólido. En algunas realizaciones, el soporte sólido

comprende una porción de captura que se une al resto de unión. En algunas realizaciones, el resto de captura se selecciona del grupo que consiste en: biotina y estreptavidina. En algunas realizaciones, el soporte sólido es una perla. En algunas realizaciones, el soporte sólido es magnético. En algunas realizaciones, los ADNc recocidos son de doble cadena. En algunas realizaciones, las moléculas monocatenarias que carecen del resto de unión son el complemento de las hebras individuales eliminadas durante la eliminación. En algunas realizaciones, el método comprende además amplificar la biblioteca normalizada. En algunas realizaciones, la amplificación comprende la generación de un ácido nucleico bicatenario a partir de ácidos nucleicos monocatenarios en la biblioteca normalizada. En algunas realizaciones, el método comprende además añadir adaptadores al ácido nucleico de doble cadena. En algunas realizaciones, los adaptadores de adición comprenden la unión de los adaptadores al ácido nucleico de doble cadena. En algunas realizaciones, la adición de adaptadores comprende la introducción de los adaptadores para el ácido nucleico de doble cadena mediante la amplificación. En algunas realizaciones, los adaptadores comprenden secuencias de células de flujo de secuenciación. En algunas realizaciones, el método comprende además la secuenciación de los ácidos nucleicos de doble cadena. En algunas realizaciones, una especie más abundante en la biblioteca normalizada es a lo sumo 20 veces más abundante que una especie menos abundante en la biblioteca normalizada. En algunas realizaciones, una especie más abundante en la biblioteca normalizada es a lo sumo 10 veces más abundante que una especie menos abundante en la biblioteca normalizada. En algunas realizaciones, una especie más abundante en la biblioteca normalizada es a lo sumo 5 veces más abundante que una especie menos abundante en la biblioteca normalizada.

## 20 BREVE DESCRIPCIÓN DE LOS DIBUJOS

**[0009]** Las características novedosas de la invención se exponen con particularidad en las reivindicaciones adjuntas. Se obtendrá una mejor comprensión de las características y ventajas de la presente invención haciendo referencia a la siguiente descripción detallada que presenta realizaciones ilustrativas, en las cuales se utilizan los principios de la invención, y los dibujos adjuntos de los cuales:

**La Figura 1** ilustra una realización ejemplar del método de normalización de biblioteca de la divulgación.

**La Figura 2** ilustra una realización ejemplar del método de codificación de barras estocástica de la divulgación.

**La Figura 3** ilustra una realización ejemplar del método de amplificación de la descripción para marcar asimétricamente un amplicón con un resto de unión.

**La Figura 4** ilustra una realización ejemplar del uso de bloqueadores en los métodos de la divulgación.

**Las Figuras 5A y 5B** representan una realización ejemplar de un método para agregar un código de barras estocástico a un objetivo utilizando un cebador adaptador universal.

## 40 DESCRIPCIÓN DETALLADA

**[0010]** La mayoría de las estrategias de normalización de la biblioteca se basan en uno de dos principios. El primero es hibridar la biblioteca con otro conjunto de ácidos nucleicos donde las secuencias están representadas uniformemente, como el ADN genómico del organismo de origen, y retener la fracción hibridada. El otro enfoque se basa en la dependencia de la concentración de la hibridación de la solución. Cuando un conjunto de moléculas de dsDNA se desnaturalizan, se rehibridarán a una tasa proporcional al cuadrado de sus concentraciones originales. Los métodos, composiciones y kits descritos en el presente documento, en algunas realizaciones, explotan esta propiedad para la normalización de bibliotecas desnaturalizando una mezcla y solo permitiendo que se reanude parcialmente; proporcionalmente, mucho más de las especies de alta concentración se habrán rehibridado a dsDNA, mientras que las especies menos abundantes seguirán siendo predominantemente monocatenarias.

**[0011]** Los métodos, composiciones y kits descritos en este documento, en algunas realizaciones, pueden evitar el uso de la separación física y enzimática de fracciones ssDNA y dsDNA durante la normalización de la biblioteca. Durante la preparación inicial de la biblioteca, la biblioteca está marcada asimétricamente en un extremo con un resto de unión (por ejemplo, mediante PCR con un cebador biotinilado 5' y otro cebador no marcado). Después de la desnaturalización y el recocido parcial, todas las cadenas marcadas se capturan en una matriz de soporte, como las perlas de estreptavidina paramagnéticas, y se separan las fracciones unidas y no unidas. Las secuencias muy abundantes serán predominantemente rehibridadas, y ambas hebras se eliminarán en la fracción unida. Sin embargo, será menos probable que las secuencias de baja abundancia se rehibridicen, por lo que el complemento de la cadena marcada estará presente en la fracción no unida. La fracción no unida representaría una biblioteca normalizada, y podría usarse directamente o amplificarse aún más para aplicaciones posteriores.

### Definiciones

**[0012]** A menos que se defina lo contrario, todos los términos técnicos utilizados en el presente documento tienen el mismo significado que comúnmente entiende un experto en la técnica en el campo al que pertenece esta descripción.

Tal como se utiliza en esta especificación y en las reivindicaciones adjuntas, las formas singulares "un", "una", "el", y "la" incluyen referencias plurales a menos que el contexto indique claramente lo contrario. Cualquier referencia a "o" aquí está destinada a abarcar "y/o" a menos que se indique lo contrario.

5 **[0013]** Tal como se utiliza aquí, el término "asociado" o "asociado con" puede significar que dos o más especies son identificables como co-situadas 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 de formato, donde, por ejemplo, se almacena información digital sobre dos o más especies y se puede utilizar para determinar que una o más de las especies se ubicaron en un punto en el tiempo. Una asociación también puede ser una asociación física. En algunos casos, dos o más especies asociadas están "atadas", "unidas" o "inmovilizadas" entre sí o con una superficie sólida o semisólida común. Una asociación puede referirse a medios covalentes o no covalentes para unir etiquetas a soportes sólidos o semisólidos, tales como perlas. Una asociación puede comprender la hibridación entre una diana y una etiqueta.

15 **[0014]** Tal como se utiliza aquí, el término "complementario" 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 enlaces 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 monocatenario puede ser "parcial", en que solo algunos de los nucleótidos se unen, o puede estar completo cuando existe una complementariedad total entre las moléculas de cadena simple. Se puede decir que una primera secuencia de nucleótidos es el "complemento" de una segunda secuencia si la primera secuencia de nucleótidos es complementaria de 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 de una secuencia que es inversa (es decir, se invierte el orden de los nucleótidos) de la segunda secuencia. Como se usa en este documento, los términos "complemento", "complementario" y "complemento inverso" se pueden usar indistintamente. Se entiende por la descripción que si una molécula puede hibridar con otra molécula, puede ser el complemento de la molécula que se está hibridando.

30 **[0015]** Tal como se utiliza aquí, el término "recuento digital" puede referirse a un método para estimar un número de moléculas diana en una muestra. El recuento digital puede incluir el paso de determinar una cantidad de etiquetas únicas que se han asociado con dianas en una muestra. Esta metodología estocástica transforma el problema de contar moléculas a partir de la localización e identificación de moléculas idénticas a una serie de preguntas digitales de sí/no relacionadas con la detección de un conjunto de etiquetas predefinidas.

35 **[0016]** Tal como se utiliza aquí, el término "etiqueta" o "etiquetas" puede referirse a los códigos de ácido nucleico asociados con una diana dentro de una muestra. Una etiqueta puede ser, por ejemplo, una etiqueta de ácido nucleico. Una etiqueta puede ser una etiqueta total o parcialmente amplificable. Una etiqueta puede ser total o parcialmente secuenciable. Una etiqueta puede ser una porción de un ácido nucleico nativo que es identificable como distinto. Una etiqueta puede ser una secuencia conocida. Una etiqueta puede comprender una unión de secuencias de ácido nucleico, por ejemplo, una unión de una secuencia nativa y no nativa. Tal como se usa en el presente documento, el término "etiqueta" se puede usar indistintamente con los términos "índice", "etiqueta" o "marca-etiqueta". Las etiquetas pueden transmitir información. Por ejemplo, en varias realizaciones, las etiquetas se pueden usar para determinar una identidad de una muestra, una fuente de una muestra, una identidad de una célula y/o una diana.

45 **[0017]** Tal como se utiliza aquí, el término "reservorios que no se agotan" puede referirse a un grupo de códigos de barras estocásticos compuesto de muchas etiquetas diferentes. Un depósito no agotador puede comprender grandes números de códigos de barras estocásticos diferentes, de manera que cuando el depósito no agotador está asociado con un conjunto de dianas, es probable que cada diana esté asociada con un código de barras estocástico único. La singularidad de cada molécula diana marcada puede determinarse por las estadísticas de elección aleatoria, y depende del número de copias de moléculas diana idénticas en la colección en comparación con la diversidad de etiquetas. El tamaño del conjunto resultante de moléculas diana marcadas puede determinarse por la naturaleza estocástica del proceso de código de barras, y el análisis del número de códigos de barras estocásticos detectados permite el cálculo del 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 estocásticos únicos es bajo, las moléculas diana marcadas son muy únicas (es decir, hay una probabilidad muy baja de que más de una molécula diana se haya etiquetado con una etiqueta dada).

60 **[0018]** Como se usa en este documento, un "ácido nucleico" puede referirse en general a una secuencia de polinucleótidos, o fragmento del mismo. Un ácido nucleico puede comprender nucleótidos. Un ácido nucleico puede ser exógeno o endógeno a una célula. Un ácido nucleico puede existir en un ambiente libre de células. Un ácido nucleico puede ser un gen o 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, espina alterada, azúcar o nucleobase). Algunos ejemplos no limitativos de análogos incluyen: 5-bromouracilo, ácido nucleico peptídico, ácido xeno nucleico, morfolinós, ácidos nucleicos bloqueados, ácidos nucleicos de glicol, ácidos nucleicos de treosa, didesoxinucleótidos, cordicepina, 7-deaza-GTP, fluoróforos (p. ej. rodamina o fluoresceína unida al azúcar), nucleótidos que contienen tior, nucleótidos unidos a biotina, análogos de bases fluorescentes, islas CpG, metil-7-guanosina, nucleótidos metilados,

inosina, tiouridina, pseudouridina, dihidrouridina, queosina y wiosina. El "ácido nucleico", el "polinucleótido", el "polinucleótido diana" y el "ácido nucleico diana" se pueden usar indistintamente.

5 **[0019]** Un ácido nucleico puede comprender una o más modificaciones (por ejemplo, una modificación de base, una modificación del esqueleto), para proporcionar el ácido nucleico con una función nueva o mejorada (por ejemplo, la mejora de la estabilidad). Un ácido nucleico puede comprender una etiqueta de afinidad de ácido nucleico. Un nucleósido puede ser una combinación de base-azúcar. La porción de base del nucleósido puede ser una base heterocíclica. Las dos clases más comunes de tales bases heterocíclicas son las proteínas y las pirimidinas. Los nucleótidos pueden ser nucleósidos que incluyen además un grupo fosfato unido covalentemente a la porción de  
10 azúcar del nucleósido. Para aquellos nucleósidos que incluyen un azúcar pentofuranosilo, el grupo fosfato se puede unir a los restos hidroxilo 2', 3' o 5' del azúcar. Al formar ácidos nucleicos, los grupos fosfato pueden unirse covalentemente a nucleósidos adyacentes entre sí para formar un compuesto polimérico lineal. A su vez, los extremos respectivos de este compuesto polimérico lineal pueden unirse aún más para formar un compuesto circular; sin embargo, los compuestos lineales son generalmente adecuados. Además, los compuestos lineales pueden tener una  
15 complementariedad interna con la base de nucleótidos y, por lo tanto, pueden plegarse de manera que produzcan un compuesto de cadena completa o parcialmente de dos cadenas. Dentro de los ácidos nucleicos, los grupos fosfato se pueden denominar comúnmente como formadores de la columna vertebral de nucleótidos del ácido nucleico. El enlace o esqueleto del ácido nucleico puede ser un enlace fosfodiéster de 3' a 5'.

20 **[0020]** Un ácido nucleico puede comprender una espina dorsal modificada y/o enlaces internucleósidos modificados. Los esqueletos modificados pueden incluir aquellos que retienen un átomo de fósforo en el esqueleto y aquellos que no tienen un átomo de fósforo en el esqueleto. Los esqueletos de ácidos nucleicos modificados adecuados que contienen un átomo de fósforo pueden incluir, por ejemplo, fosforotioatos, fosforotioatos quirales, fosforoditioatos, fosfotriésteres, aminoalquilfosfotriésteres, metilo y otros fosfonatos de alquilo, tales como fosfonatos de alquileo 3',  
25 fosfonatos de alquileo 5', fosfonatos quirales, fosfinatos, fosforamidatos incluyendo fosforamidato y aminoalquilfosforamidatos, fosforodiamidatos, tionoalquilfosfonatos, tionoalquilfosfotriésteres, selenofosfatos, y boranofosfatos que tienen enlaces 3'-5' normales, análogos enlazados 2'-5', y aquellos que tienen polaridad invertida en donde uno o más enlaces internucleótidos es un enlace 3' a 3', 5' a 5' o 2' a 2'.

30 **[0021]** Un ácido nucleico puede comprender cadenas principales de polinucleótidos que se forman por alquilo de cadena corta o enlaces internucleósidos de cicloalquilo, heteroátomo mixto y los enlaces internucleósidos de alquilo o cicloalquilo, o uno o más heterocíclicos de cadena más corta o enlaces internucleósidos heterocíclicos. Estos pueden incluir aquellos que tienen enlaces morfolinos (formados en parte por la porción de azúcar de un nucleósido);  
35 esqueletos de siloxano; esqueletos de sulfuro, sulfóxido y sulfona; esqueletos de formacetilo y tioformacetilo; esqueletos de metileno formacetilo y tioformacetilo; esqueletos de riboacetilo; esqueletos que contienen alquenos; esqueletos de sulfamato; esqueletos de metilnimino y metilhidracino; esqueletos de sulfonato y sulfonamida; esqueletos de amida; y otros que tienen partes componentes de N, O, S y CH<sub>2</sub> mixtas.

40 **[0022]** Un ácido nucleico puede comprender un mimético de ácido nucleico. El término "mimético" puede incluir polinucleótidos en los que solo el anillo de furanosa o tanto el anillo de furanosa como el enlace internucleótido se reemplazan con grupos que no contienen furanosa; el reemplazo de solo el anillo de furanosa también se puede denominar sustituto del azúcar. El resto de base heterocíclica o un resto de base heterocíclica modificada se puede mantener para la hibridación con un ácido nucleico diana apropiado. Uno de tales ácidos nucleicos puede ser un ácido  
45 nucleico peptídico (PNA). En un PNA, el esqueleto de azúcar de un polinucleótido puede reemplazarse con un esqueleto que contiene amida, en particular un esqueleto de aminoetilglicina. Los nucleótidos se pueden retener y se unen directa o indirectamente a los átomos de nitrógeno de la parte amida de la columna vertebral. El esqueleto en los compuestos de PNA puede comprender dos o más unidades de aminoetilglicina unidas que dan a PNA un esqueleto que contiene amida. Los restos de bases heterocíclicas se pueden unir directa o indirectamente a átomos  
50 de nitrógeno aza de la porción amida del esqueleto.

**[0023]** Un ácido nucleico puede comprender una estructura de espina dorsal morfolina. Por ejemplo, un ácido nucleico puede comprender un anillo morfolino de 6 miembros en lugar de un anillo de ribosa. En algunas de estas realizaciones, un fosforodiamidato u otro enlace internucleosídico no fosfodiéster puede reemplazar un enlace fosfodiéster.  
55

**[0024]** Un ácido nucleico puede comprender unidades de morfolino enlazadas (es decir, ácido nucleico morfolino) que tienen bases heterocíclicas unidas al anillo morfolino. Los grupos de enlace pueden enlazar las unidades monoméricas de morfolino en un ácido nucleico morfolino. Los compuestos oligoméricos a base de morfolino no iónicos pueden tener menos interacciones no deseadas con las proteínas celulares. Los polinucleótidos a base de morfolino pueden ser imitadores no iónicos de los ácidos nucleicos. Se pueden unir una variedad de compuestos dentro de la clase de morfolino utilizando diferentes grupos de enlace. Una clase adicional de polinucleótido mimético puede denominarse ácidos nucleicos de ciclohexenilo (CeNA). El anillo de furanosa normalmente presente en una molécula de ácido nucleico puede reemplazarse con un anillo de ciclohexenilo. Los monómeros de fosforamidita protegidos con CeNA DMT pueden prepararse y usarse para la síntesis de compuestos oligoméricos utilizando la química de la fosforamidita.  
60 La incorporación de monómeros de CeNA en una cadena de ácido nucleico puede aumentar la estabilidad de un híbrido de ADN/ARN. Los oligoadenilatos de CeNA pueden formar complejos con complementos de ácido nucleico  
65



con una estabilidad similar a los complejos nativos. Una modificación adicional puede incluir ácidos nucleicos bloqueados (LNA) en los cuales el grupo 2'-hidroxilo está unido al átomo de carbono 4' del anillo de azúcar formando así un enlace 2'-C, 4'-C-oximetileno formando así una porción de azúcar bicíclica. El enlace puede ser un grupo metileno (-CH<sub>2</sub>-) que une el átomo de oxígeno 2' y el átomo de carbono 4' en donde n es 1 o 2. Los análogos de LNA y LNA pueden mostrar estabildades térmicas dúplex muy altas con ácido nucleico complementario (T<sub>m</sub> = +3 a +10°C), estabilidad hacia degradación 3'-exonucleolítica y buenas propiedades de solubilidad.

**[0025]** Un ácido nucleico también puede incluir modificaciones de nucleobase (frecuentemente denominadas simplemente como "base") o sustituciones. Como se usa en este documento, 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 tales como 5-metilcitosina (5-me-C), 5-hidroxiometil-citosina, xantina, hipoxantina, 2-aminoadenina, 6-metilo y otros derivados alquílicos de adenina y guanina, 2-propil y otros derivados alquílicos de adenina y guanina, 2-tiouracilo, 2-tiotimina y 2-tiocitosina, 5-halouracilo y citosina, 5-propinilo (-C = C-CH<sub>3</sub>) uracilo y citosina y otros derivados alquílicos de bases de pirimidina, 6-azouracilo, citosina y timina, 5-uracilo (pseuourouracilo), 4-tiouracilo, 8-halo, 8-amino, 8-tiol, 8-tio-alquilo, 8-hidroxilo y otras adeninas y guaninas 8-sustituidas, 5-halo particularmente 5-bromo, 5-trifluorometilo y otros uracilos y citosinas 5-sustituidas, 7-metil-guanina 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 tales como citidina de fenoxazina (1H-pirimido(5,4-b)(1,4)benzoxazin-2(3H)-ona), citidina de fenotiazina (1H-pirimido (5,4-b)(1,4)benzotiazin-2(3H)-ona), pinzas G como una citidina de fenoxidina sustituida (por ejemplo, 9-(2-aminoetoxi)-H-pirimidina(5,4-b)(1,4)benzoxazin-2(3H)-ona), citidina de carbazol (2H-pirimido(4,5-b)indol-2-ona), citidina de piridoindol (Hpirido(3',:4,5)pirrolo[2,3-d]pirimidin-2-ona).

**[0026]** Tal como se utiliza aquí, el término "ácido nucleico cuasi-simétrico estocásticamente con código de barras" puede referirse a una molécula que comprende un código de barras estocástico de la revelación y los extremos son lo suficientemente simétricos como para hibridarse juntos para formar una estructura de franja (por ejemplo, para supresión de PCR), pero pueden no ser idénticos. Un ácido nucleico de código de barras estocásticamente cuasi simétrico puede comportarse como un ácido nucleico simétrico, pero tiene una secuencia asimétrica.

**[0027]** Como se usa en el presente documento, el término "muestra" puede referirse a una composición que comprende dianas. Las muestras adecuadas para el análisis mediante los métodos, dispositivos y sistemas descritos incluyen células, células individuales, tejidos, órganos u organismos.

**[0028]** Tal como se utiliza aquí, 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 sobre un sustrato. Un dispositivo de muestra puede referirse, por ejemplo, a una máquina de clasificación de células activadas por fluorescencia (FACS), una máquina clasificadora de células, una aguja de biopsia, un dispositivo de biopsia, un dispositivo de corte de tejido, un dispositivo de microfluidos, una rejilla de cuchillas, y/o un microtomo.

**[0029]** Tal como se utiliza aquí, el término "soporte sólido" se puede referir superficies sólidas o semi-sólidas a las que pueden unirse una serie de códigos de barras estocásticos. Un soporte sólido puede abarcar cualquier tipo de esfera sólida, porosa o hueca, bola, cojinete, cilindro u otra configuración similar compuesta de plástico, cerámica, metal o material polimérico (p. ej., hidrogel) sobre el cual puede estar un ácido nucleico. estar inmovilizado (por ejemplo, covalentemente o no covalentemente). Un soporte sólido puede comprender una partícula discreta que puede ser esférica (p. ej., microsferas) o tener una forma no esférica o irregular, como cúbica, cuboide, piramidal, cilíndrica, cónica, oblonga o en forma de disco, y similares. Una pluralidad de soportes sólidos espaciados en una matriz puede no comprender un sustrato. Un soporte sólido se puede usar de manera intercambiable con el término "perla". Como se usa en este documento, "soporte sólido" y "sustrato" se pueden usar indistintamente.

**[0030]** Tal como se utiliza aquí, el término "código de barras estocástico" puede referirse a una secuencia de polinucleótido que comprende etiquetas de la descripción. Un código de barras estocástico puede ser una secuencia polinucleotídica que puede usarse para la codificación de barras estocástica. Los códigos de barras estocásticos se pueden usar para cuantificar los objetivos dentro de una muestra. Los códigos de barras estocásticos se pueden usar para controlar los errores que pueden ocurrir después de que una etiqueta se asocie con una diana. Por ejemplo, un código de barras estocástico se puede utilizar para evaluar errores de amplificación o secuenciación. Un código de barras estocástico asociado con una diana puede llamarse una diana de código de barras estocástico o una diana de código de barras estocástico.

**[0031]** Tal como se utiliza aquí, el término "código de barras estocástico" puede referirse al etiquetado aleatorio (por ejemplo, códigos de barras) de ácidos nucleicos. Los códigos de barras estocásticos pueden utilizar una estrategia recursiva de Poisson para asociar y cuantificar etiquetas asociadas con dianas. Tal como se usa en el presente documento, el término "código de barras estocástico" se puede usar de manera intercambiable con "etiquetado estocástico".

[0032] Tal como se utiliza aquí, el término "diana" puede referirse a una composición que puede estar asociado con un código de barras estocástico. Las dianas adecuadas ejemplares para el análisis mediante los métodos, dispositivos y sistemas descritos 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. En algunas realizaciones los objetivos son los lípidos. Como se usa en este documento, "diana" se puede usar indistintamente con "especies".

[0033] El término "transcriptasas inversas" 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, tales enzimas incluyen, pero no se limitan a, transcriptasa inversa retroviral, transcriptasa inversa retrotransposón, transcriptasa inversa retroplasmídica, transcriptasa inversa retron, transcriptasa inversa bacteriana, transcriptasa inversa derivada intrón del grupo II y mutantes, variantes o derivados de los mismos. Las transcriptasas inversas no retro-virales incluyen las transcriptasas inversas retrotransposibles no LTR, las transcriptasas inversas retroplasmídicas, las transcriptasas inversas retron y las transcriptasas inversas internas del grupo II. Entre los ejemplos de transcriptasas inversas del intrón del grupo II se incluyen la transcriptasa inversa *Lactococcus lactis* L.LtrB, la transcriptasa inversa del intrón Tel4c de *Thermosynechococcus elongatus*, o la transcriptasa de la retranversa del intrón de *Geobacillus stearothermophilus* Gsl-IIC. Otras clases de transcriptasas inversas pueden incluir muchas clases de transcriptasas inversas no retrovirales (es decir, retron, intrones del grupo II y retroelementos generadores de diversidad, entre otros).

#### Métodos de eliminación de especies de alta abundancia

[0034] Algunas realizaciones descritas en este documento proporcionan métodos para la eliminación de especies de alta abundancia a partir de una pluralidad de moléculas de ácido nucleico. En alguna realización, los métodos descritos aquí pueden reducir el contenido de especies de alta abundancia de una pluralidad de moléculas de ácido nucleico sin eliminar significativamente la especie de baja abundancia o las especies de abundancia intermedia de la pluralidad de moléculas de ácido nucleico. Como se usa aquí, "remover significativamente" se refiere a remover al menos el 10%, al menos el 20%, al menos el 30%, al menos el 40%, al menos el 50% o más de una especie de baja abundancia o especies de abundancia intermedia de la pluralidad de moléculas de ácido nucleico. En algunas realizaciones, los métodos descritos en este documento pueden eliminar especies de alta abundancia y las especies de abundancia intermedia de una pluralidad de moléculas de ácido nucleico sin eliminar significativamente las especies de baja abundancia de la pluralidad de moléculas de ácido nucleico.

[0035] Como se usa en el presente documento, una "especie" se refiere a los polinucleótidos (por ejemplo, polinucleótidos de cadena simple) en la pluralidad de moléculas de ácido nucleico que son los mismos o el complemento de uno del otro, o son capaces de hibridarse entre sí, o son transcripciones del mismo locus genético, o codifican la misma proteína o fragmento de la misma, etc. En algunas realizaciones, los miembros de una especie son al menos el 80%, al menos el 90%, al menos el 95%, al menos 98%, al menos 99%, o 100% homólogos entre sí, o complemento de los mismos. En algunas realizaciones, los miembros de una especie pueden hibridar entre sí en condiciones de hibridación altamente estrictas. En algunas realizaciones, los miembros de una especie pueden hibridar con otras condiciones de hibridación rigurosas moderadas. En algunas realizaciones, los miembros de una especie pueden hibridarse entre sí en condiciones de hibridación bajo rigurosas. En algunas realizaciones, los miembros de una especie son transcripciones del mismo locus genético y las transcripciones pueden ser de la misma o diferente longitud. La especie es, en algunas realizaciones, ADNc o ARNm.

[0036] Como se usa en el presente documento, una "especie de abundancia alta" se refiere a una especie que está presente en gran cantidad en la pluralidad de ácidos nucleicos, por ejemplo la especie puede representar al menos el 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 20%, 30%, 40%, 50% o más de la pluralidad de moléculas de ácido nucleico. En algunas realizaciones, la pluralidad de moléculas de ácido nucleico puede comprender al menos 1, al menos 2, al menos 3, al menos 4, al menos 5, al menos 10, al menos 20, al menos 50, al menos 100, al menos 200, al menos 500, al menos 1,000, o más, especies de alta abundancia. En algunas realizaciones, el total de todas las especies de alta abundancia representa al menos el 10%, al menos el 20%, al menos el 30%, al menos el 40%, al menos el 50%, al menos el 60%, al menos el 70%, al menos 80%, o más de la pluralidad de moléculas de ácido nucleico. En algunas realizaciones, las especies de alta abundancia pueden comprender polinucleótidos que codifican una o más proteínas ribosómicas. En algunas realizaciones, las especies de alta abundancia pueden comprender polinucleótidos que codifican una o más proteínas mitocondriales. En algunas realizaciones, las especies de alta abundancia pueden comprender polinucleótidos que codifican una o más proteínas de mantenimiento.

[0037] Como se usa en el presente documento, una "especie en abundancia intermedia" se refiere a una especie que está presente en una cantidad en la pluralidad de ácido nucleico que es más baja que al menos una especie en la pluralidad de ácido nucleico y es más alta que al menos una especie en la pluralidad de ácidos nucleicos. En algunas realizaciones, una especie de abundancia intermedia puede representar aproximadamente el 10%, 5%, 4%, 3%, 2%, 1%, 0,1%, 0,01%, o un rango entre cualquiera de los dos valores anteriores, de la pluralidad de moléculas de ácido nucleico. En algunas realizaciones, la pluralidad de moléculas de ácido nucleico puede comprender al menos 1, al menos 2, al menos 3, al menos 4, al menos 5, al menos 10, al menos 20, al menos 50, al menos 100, al menos 200, al menos 500, al menos 1.000 o más especies de abundancia intermedia. En algunas realizaciones, el total de todas las especies de abundancia intermedia representa aproximadamente el 1%, aproximadamente el 2%, aproximadamente el 3%, aproximadamente el 4%, aproximadamente el 5%, aproximadamente el 10%,

aproximadamente el 20%, aproximadamente el 30%, o un intervalo entre cualesquiera dos de los valores anteriores, de la pluralidad de moléculas de ácido nucleico.

5 **[0038]** Como se usa en el presente documento, una "especie de abundancia baja" se refiere a una especie que está presente en baja cantidad en la pluralidad de ácidos nucleicos, por ejemplo la especie puede representar menos del 1%, 0,1%, 0,01%, 0,001%, 0,0001%, o menos de la pluralidad de moléculas de ácido nucleico. En algunas realizaciones, la pluralidad de moléculas de ácido nucleico puede comprender al menos 1, al menos 2, al menos 3, al menos 4, al menos 5, al menos 10, al menos 20, al menos 50, al menos 100, al menos 200, al menos 500, al menos 1.000 o más especies de baja abundancia. En algunas realizaciones, el total de todas las especies de baja abundancia representan menos del 20%, menos del 10%, menos del 5%, menos del 4%, menos del 3%, menos del 2%, menos del 1%, menos del 0,1% o menos de la pluralidad de moléculas de ácido nucleico. En algunas realizaciones, las especies de baja abundancia pueden comprender polinucleótidos que codifican uno o más factores de transcripción. En algunas realizaciones, las especies de alta abundancia pueden comprender polinucleótidos que codifican uno o más receptores de células T. En algunas realizaciones, las especies de alta abundancia pueden comprender polinucleótidos que codifican uno o más anticuerpos.

20 **[0039]** En algunas realizaciones, los métodos y las composiciones descritos en este documento pueden reducir el contenido de una o más especies alta abundancia a partir de la pluralidad de moléculas de ácido nucleico. Por ejemplo, los métodos y composiciones descritos en este documento pueden reducir el contenido de al menos 1, al menos 2, al menos 3, al menos 4, al menos 5, al menos 10, al menos 20, al menos 50, al menos 100, al menos 200, al menos 500, al menos 1.000 o más, especies de alta abundancia. En algunas realizaciones, los métodos y composiciones descritos aquí pueden reducir el contenido en al menos un 10%, al menos un 20%, al menos un 30%, al menos un 40%, al menos un 50%, al menos un 60%, al menos un 70%, al menos un 80%, al menos un 90%, al menos un 95% o 100% de cada una de las una o más especies de alta abundancia de la pluralidad de moléculas de ácido nucleico. En algunas realizaciones, los métodos y composiciones descritos en este documento pueden reducir el contenido en al menos un 10%, al menos un 20%, al menos un 30%, al menos un 40%, al menos un 50%, al menos un 60%, al menos un 70%, al menos un 80%, al menos un 90%, al menos un 95%, o 100% de al menos una de las una o más especies de alta abundancia de la pluralidad de moléculas de ácido nucleico. En algunas realizaciones, los métodos y composiciones descritos en este documento pueden reducir el contenido en al menos un 10%, al menos un 20%, al menos un 30%, al menos un 40%, al menos un 50%, al menos un 60%, al menos un 70%, al menos un 80%, al menos un 90%, al menos un 95% o el 100% del total de especies de alta abundancia de la pluralidad de moléculas de ácido nucleico.

35 **[0040]** En algunas realizaciones, los métodos y las composiciones descritos en este documento pueden reducir el contenido de una o más especies de alta abundancia a partir de la pluralidad de moléculas de ácido nucleico sin eliminar significativamente la especie de baja abundancia o la especie de abundancia intermedia a partir de la pluralidad de moléculas de ácido nucleico. En algunas realizaciones, los métodos y composiciones descritos en este documento pueden reducir el contenido en al menos un 10%, al menos un 20%, al menos un 30%, al menos un 40%, al menos un 50%, al menos un 60%, al menos un 70%, al menos un 80%, al menos un 90%, al menos un 95%, o 100% de cada una de las una o más especies de alta abundancia de la pluralidad de moléculas de ácido nucleico sin eliminar significativamente las especies de baja abundancia o la especie de abundancia intermedia de la pluralidad de moléculas de ácido nucleico. En algunas realizaciones, los métodos y composiciones descritos en el presente documento pueden reducir el contenido en al menos un 10%, al menos un 20%, al menos un 30%, al menos un 40%, al menos un 50%, al menos un 60%, al menos un 70%, al menos un 80%, al menos un 90%, al menos un 95%, o 100% del total de especies de alta abundancia de la pluralidad de moléculas de ácido nucleico sin eliminar significativamente las especies de baja abundancia o las especies de abundancia intermedia de la pluralidad de moléculas de ácido nucleico. En algunas realizaciones, los métodos y composiciones descritos en este documento pueden reducir el contenido de una o más especies de gran abundancia de la pluralidad de moléculas de ácido nucleico, manteniendo al menos el 50%, al menos el 60%, al menos el 70%, al menos el 80%, al menos el 90%, al menos el 95%, o el 100% de cada una de las una o más especies de baja abundancia. En algunas realizaciones, los métodos y composiciones descritos en este documento pueden reducir el contenido de una o más especies de alta abundancia de la pluralidad de moléculas de ácido nucleico, manteniendo al menos el 50%, al menos el 60%, al menos el 70%, al menos el 80%, al menos el 90%, al menos el 95%, o el 100% del total de especies de baja abundancia. En algunas realizaciones, los métodos y composiciones descritos aquí pueden reducir el contenido de una o más especies de gran abundancia a partir de la pluralidad de moléculas de ácido nucleico, manteniendo al menos el 50%, al menos el 60%, al menos el 70%, al menos el 80%, al menos el 90%, al menos el 95%, o el 100% del total de especies de baja abundancia. En algunas realizaciones, los métodos y composiciones descritos en este documento pueden reducir el contenido de una o más especies de alta abundancia de la pluralidad de moléculas de ácido nucleico, manteniendo al menos el 50%, al menos el 60%, al menos el 70%, al menos el 80%, al menos el 90%, al menos el 95%, o el 100% del total de especies de abundancia intermedia. En algunas realizaciones, los métodos y composiciones descritos en este documento pueden reducir el contenido de una o más especies de alta abundancia de la pluralidad de moléculas de ácido nucleico, manteniendo al menos el 50%, al menos el 60%, al menos el 70%, al menos el 80%, al menos el 90%, al menos el 95%, o el 100% de cada una de las especies de abundancia intermedia de la pluralidad de moléculas de ácido nucleico.

Pluralidad de moléculas de ácido nucleico

5 **[0041]** La pluralidad de moléculas de ácido nucleico descritas en el presente documento puede comprender una  
 variedad de moléculas de ácido nucleico. En algunas realizaciones, la pluralidad de moléculas de ácido nucleico  
 pueden comprender moléculas de ADN, moléculas de ARN, moléculas de ADN genómico, moléculas de ADNc,  
 moléculas de ARNm, moléculas de ARNr, moléculas de ARNip o una combinación de ellas, y pueden ser de hebra  
 10 doble o de una sola hebra. En algunas realizaciones, la pluralidad de moléculas de ácido nucleico comprenden al  
 menos 100, al menos 1.000, al menos 10.000, al menos 20.000, al menos 30.000, al menos 40.000, al menos 50.000,  
 al menos 100.000, al menos 1.000.000, o mas especies. En algunas realizaciones, la pluralidad de moléculas de ácido  
 nucleico puede ser de una muestra, tal como una célula única, o una pluralidad de células. En algunas realizaciones,  
 la pluralidad de moléculas de ácido nucleico puede ser agrupada de una pluralidad de muestras, como una pluralidad  
 de células individuales.

15 **[0042]** En algunas realizaciones, la pluralidad de moléculas de ácido nucleico comprende una biblioteca no  
 normalizada de ácido nucleico, una biblioteca de ácido nucleico parcialmente normalizada, o una biblioteca de ácido  
 nucleico que ha sido normalizada por otros métodos, tales como una biblioteca de ADNc, una biblioteca de ADN  
 genómico, o similar. En algunas realizaciones, la pluralidad de moléculas de ácido nucleico puede comprender una  
 20 biblioteca de ácido nucleico no normalizada agrupada, tal como una biblioteca de ácido nucleico no normalizada  
 combinada construida a partir de una pluralidad de bibliotecas de ácido nucleico no normalizadas, cada una  
 representando una célula única. En algunas realizaciones, la biblioteca de ácido nucleico no normalizada es una  
 biblioteca de ADNc. En algunas realizaciones, la biblioteca de ácidos nucleicos no normalizada es una biblioteca  
 genómica. En algunas realizaciones, la biblioteca de ácido nucleico no normalizada es una biblioteca de ácido nucleico  
 25 de una sola célula. Como se usa en este documento, una "biblioteca de ácido nucleico de una sola célula" significa  
 una colección de moléculas de ácido nucleico, como las moléculas de ADN o ARNm genómico, que se originan en  
 una sola célula. En algunas realizaciones, una biblioteca de ácidos nucleicos unicelular puede referirse a colecciones  
 de moléculas de ácido nucleico originadas a partir de una pluralidad de células individuales, en donde las moléculas  
 de ácido nucleico comprenden una etiqueta celular para identificar la célula única de la cual se originan las moléculas  
 de ácido nucleico.

30 **[0043]** En algunas realizaciones, la pluralidad de moléculas de ácido nucleico se puede someter a amplificación antes  
 de retirar las especies de alto abundancia. Por ejemplo, la pluralidad de moléculas de ácido nucleico puede  
 comprender una biblioteca de ácido nucleico amplificada. En algunas realizaciones, la pluralidad de moléculas de  
 ácido nucleico puede comprender al menos 2, al menos 4, al menos 8, al menos 16, al menos 100, al menos 1.000 o  
 35 más copias de cada molécula de ácido nucleico.

Resto vinculante

40 **[0044]** En algunas realizaciones, los procedimientos divulgados en el presente documento comprenden la hibridación  
 de una pluralidad de primeros oligonucleótidos que comprenden un resto de unión con la pluralidad de moléculas de  
 ácido nucleico. Se puede usar una variedad de restos de unión para los métodos y composiciones descritos en el  
 presente documento. Por ejemplo, un resto de unión puede ser parte de un par de unión. En algunas realizaciones, el  
 resto de unión puede ser un grupo funcional añadido a los oligonucleótidos. En algunas formas de realización, el resto  
 45 de unión puede ser biotina, estreptavidina, heparina, aptámero, una química de clic, digoxigenina, amina(s) primaria(s),  
 carboxilo(s), hidroxilo(s), aldehído(s), cetona(s), o cualquier combinación de los mismos.

**[0045]** Los restos de unión como se describe en el presente documento son capaces de enlazar para capturar restos  
 tales como moléculas de captura. En algunas realizaciones, el resto de unión y la molécula de captura pueden ser  
 50 miembros de un par de unión, por ejemplo, biotina/estreptavidina. La molécula de captura puede inmovilizarse sobre  
 un soporte sólido, como una perla, una micropartícula o una nanopartícula.

**[0046]** En algunas realizaciones, los primeros oligonucleótidos se pueden ampliar para generar una pluralidad de  
 hebras complementarias de la pluralidad de dianas de ácido nucleico que comprenden el resto de unión. En algunas  
 55 realizaciones, se puede sintetizar una segunda cadena utilizando un cebador que se une a un sitio de unión en las  
 cadenas complementarias para producir moléculas de ácido nucleico de doble cadena.

Reducir el contenido de especies de alta abundancia por desnaturalización/remodelación parcial

60 **[0047]** En algunas realizaciones, la reducción del contenido de especies de alta abundancia pueden comprender  
 desnaturalización seguido de reasociación parcial de las moléculas de ácido nucleico de doble cadena, seguido de la  
 eliminación de las cadenas complementarias recocidas de la pluralidad de dianas de ácido nucleico por una molécula  
 de captura inmovilizada en uno o más soportes sólidos, en donde las moléculas de captura se unen específicamente  
 al resto de unión.

65 **[0048]** La desnaturalización se puede realizar mediante una variedad de métodos que incluyen el calentamiento de  
 las moléculas de ácido nucleico de doble cadena, el tratamiento de las moléculas de ácido nucleico de doble cadena

con disolventes orgánicos (por ejemplo, DMS o formamida), cambiando la concentración de sal de las moléculas de núcleo nucleico de doble cadena, y/o cambio del pH de las moléculas de ácido nucleico de doble cadena.

5 **[0049]** Después de la desnaturalización, las moléculas de ácido nucleico de cadena simple se pueden recocer parcialmente. La reanimación parcial se puede realizar mediante cualquier método, por ejemplo, enfriamiento rápido en hielo, cambiando la concentración de sal (por ejemplo, invirtiendo la concentración de sal de la cantidad utilizada en la desnaturalización) y/o cambiando el pH (por ejemplo, invirtiendo el pH del nivel utilizado en la desnaturalización), y similares.

10 **[0050]** Se agradecería que la extensión de recocado puede ajustarse de acuerdo a varios factores, incluyendo pero no limitado a, el tipo de la especie a ser removido (especies por ejemplo, especies de alta abundancia y/o abundancia intermedia), el porcentaje deseado de especies de alta abundancia que se eliminarán, y/o el porcentaje de especies de abundancia intermedia o baja que se retendrán. Sin estar limitado por ninguna teoría en particular, se cree que las especies más abundantes (p. ej., especies de alta abundancia) se anidan más rápido que las especies con menor abundancia (p. ej., especies de abundancia intermedia y baja) en las mismas condiciones de recocado. Por ejemplo, al cambiar la temperatura, la concentración de sal, el pH y/o la duración de la etapa de resurtido, se puede ajustar el porcentaje de especies de gran abundancia que se eliminarán y/o el porcentaje de especies de abundancia intermedia o baja a conservar. En algunas realizaciones, la temperatura, la concentración de sal, el pH y/o la duración de la etapa de resurtido se pueden ajustar de manera que al menos el 10%, al menos el 20%, al menos el 30%, al menos el 40%, al menos el 50%, al menos el 60%, al menos el 70%, al menos el 80%, al menos el 90%, al menos el 95%, o al menos el 99% de una especie recocida de gran abundancia. En algunas realizaciones, la temperatura, la concentración de sal, el pH y/o la duración de la etapa de resurtido se pueden ajustar de manera que al menos el 10%, al menos el 20%, al menos el 30%, al menos el 40%, al menos el 50%, al menos el 60%, al menos el 70%, al menos el 80%, al menos el 90%, al menos el 95%, o el 100% de las especies de abundancia intermedia o baja permanecen monocatenarias.

30 **[0051]** La **Figura 1** representa una realización ejemplar de los métodos de la divulgación. Una muestra puede comprender una pluralidad de ácidos nucleicos. Algunos de los ácidos nucleicos pueden ser muy abundantes **106** y algunos de los ácidos nucleicos pueden ser menos abundantes **105**. Los ácidos nucleicos pueden transformarse en una biblioteca de ADNc de doble cadena **125** que está marcada asimétricamente con un resto de unión **120**. Por ejemplo, los ácidos nucleicos **105/106** pueden ser ARNm que se transcriben de manera inversa. La síntesis de la segunda cadena se puede realizar utilizando un cebador (por ejemplo, un cebador específico del gen o un cebador multimer aleatorio) que comprende el resto de unión **120**, generando así una biblioteca de ácido nucleico de doble cadena **125** marcada de forma asimétrica. En otro ejemplo, el nucleic Los ácidos **105/106** pueden ser ADN. El ADN puede extenderse usando un cebador que comprende el resto de unión **120**, generando de este modo una biblioteca de ADNc de doble cadena marcada asimétricamente **125**. La biblioteca de ADNc de doble cadena puede comprender una especie de ADNc de doble cadena altamente abundante **116** y poco abundantes especies de ADNc de doble cadena **115**. La biblioteca de ADNc de doble cadena **125** puede ligarse a los adaptadores **130/135**, generando así una biblioteca no normalizada **125**.

40 **[0052]** La biblioteca de ADNc de doble cadena **125** puede ser desnaturalizada por calor, separando de este modo las hebras del ADNc de doble cadena. La biblioteca desnaturalizada por calor se puede recocer (por ejemplo, parcialmente recocer) **140**. Los ácidos nucleicos más abundantes **106** puede hibridar más rápido que los ácidos nucleicos menos abundantes **105**. La desnaturalización y el recocado parcial puede resultar en una mezcla de especies **141** comprendiendo ADNc bicatenarios recocidos que comprenden dicho resto de unión **142**, moléculas monocatenarias que comprenden dicho resto de unión **143** y moléculas monocatenarias que carecen de dicho resto de unión **144**.

50 **[0053]** La biblioteca desnaturalizada y recocida puede contactarse a un soporte sólido **145**. El soporte sólido puede comprender un resto de captura que puede unirse a la porción de unión **120** de los ADNc asimétricamente marcados de doble hebra **115/116**. Los ácidos nucleicos que comprenden el resto de unión **120** pueden estar unidos por el soporte sólido. Estos ácidos nucleicos pueden incluir ADNcs de doble cadena recocidos que comprenden dicho resto de unión **142**, moléculas de cadena simple que comprenden dicho resto de unión **143**. Se puede usar un imán para eliminar el soporte sólido **145** unido a los ácidos nucleicos.

55 **[0054]** Los ácidos nucleicos restantes **150** representan el complemento de las especies que no lo hicieron recocado (por ejemplo, este es el complemento de una de una sola cadena que compone el resto de unión y se retiró por el soporte sólido, pero era a una concentración baja suficiente como para que no se hibridara con esa molécula monocatenaria). Los ácidos nucleicos **150** restantes pueden no contener el resto de unión **120**. Los ácidos nucleicos **150** dejados atrás pueden amplificarse y/o ligarse con un adaptador para la secuenciación. Estos ácidos nucleicos representan una biblioteca normalizada. En algunos casos, la normalización de la biblioteca puede no requerir el uso de enzimología para normalizar la biblioteca (p. ej., nucleasas).

#### Métodos de normalización de bibliotecas

65 **[0055]** La descripción proporciona métodos para la normalización de la biblioteca. Los métodos de la divulgación se pueden realizar en una muestra de ácido nucleico. La muestra de ácido nucleico puede comprender ácidos nucleicos.

La muestra puede ser de una muestra de la divulgación. La muestra puede ser una sola célula (por ejemplo, la muestra de ácido nucleico puede ser ácidos nucleicos de una sola célula). La muestra de ácido nucleico puede comprender ARN, ADN o tanto ARN como ADN. Los ácidos nucleicos de la muestra pueden ser monocatenarios, bicatenarios o una mezcla de tanto monocatenarios como bicatenarios. En algunos casos, todos o la mayoría de los ácidos nucleicos de la muestra son monocatenarios.

**[0056]** La muestra de ácido nucleico puede comprender dianas de ácido nucleico (es decir, especies de ácido nucleico, usadas intercambiabilmente en este documento con dianas de ácido nucleico) de diversas dianas. Por ejemplo, la muestra de ácido nucleico puede comprender dianas de gran abundancia (p. ej., actina, GapDH, globinas, genes de mantenimiento). La muestra de ácido nucleico puede comprender dianas de baja abundancia (p. ej., dianas poco comunes de células madre o células tumorales circulantes (CTC) o genes expresados con poca frecuencia). La muestra de ácido nucleico puede comprender una mezcla de dianas de alta abundancia y baja abundancia.

**[0057]** Los ácidos nucleicos de la muestra pueden ser contactados con un cebador que comprende un resto de unión para generar un ADNc de doble cadena. En algunas realizaciones, el ácido nucleico puede ser un ARN y el cebador puede ser un cebador de transcripción inversa. El cebador de transcripción inversa puede transcribir de manera inversa el ARN, generando de este modo un híbrido de ARN-ADNc (por ejemplo, la primera síntesis de la cadena). Se puede generar una segunda hebra utilizando técnicas estándar de síntesis de la segunda hebra. La primera hebra de ADNc puede comprender el resto de unión. La segunda hebra puede ser el complemento de la primera hebra. La segunda hebra puede no comprender el resto de unión. Este ADNc se puede referir como un ADNc marcado asimétricamente (por ejemplo, una cadena del ADNc está marcada con el resto de unión).

**[0058]** Cuando los ácidos nucleicos de la muestra son de ADN, un cebador que comprende un resto de unión puede contactarse con los ácidos nucleicos para generar una primera cadena (por ejemplo, hebra complementaria a la plantilla de ADN). La primera hebra puede comprender el resto de unión. En algunos casos, el cebador puede generar una segunda hebra que es complementaria de la primera hebra. La segunda hebra puede no comprender el resto de unión.

**[0059]** Con cualquiera de un ARN o de ADN molde de partida, el resultado puede ser una molécula de ADNc de doble cadena de forma asimétrica marcada donde una de las hebras comprende un resto de unión, y la otra hebra es el complemento. Un grupo de moléculas de ADNc de doble cadena marcadas asimétricamente se puede denominar una biblioteca no normalizada.

**[0060]** En algunas realizaciones, se añade el resto de unión a través de reacciones de amplificación que se producen después de la generación de la biblioteca (p. ej., primera hebra, segunda hebra, ligadura del adaptador). Como se muestra en la **Figura 3**, un ácido nucleico (por ejemplo, ARNm, ADN) **305** se puede transcribir o extender de manera inversa con un cebador que comprende un marcador molecular (es decir, índice molecular) **310**, un marcador de muestra (es decir, marcador celular, índice de muestra) **315**, y un marcador universal (es decir, la secuencia de unión del cebador universal) **320**. El producto puede experimentar una primera reacción de amplificación por PCR **325** anidada en la que se utiliza un cebador específico **330** específico y un cebador universal **335** que se une al marcador universal **320** para amplificar el producto, generando amplicones. En algunos casos, los amplicones pueden experimentar una segunda ronda de amplificación utilizando un cebador de PCR anidado **345** y un cebador universal **350** que comprende un resto de unión **355** que se une al marcador universal **320**. La reacción de amplificación puede producir amplicones (por ejemplo, amplicones de PCR anidados) etiquetados asimétricamente con un resto de unión. La **Figura 3** representa un esquema de amplificación representativo para añadir un resto de unión. El resto de unión se puede agregar en cualquier paso, como en la transcripción inversa, en la síntesis de la segunda hebra, antes de la ligadura del adaptador, después de la ligadura del adaptador y/o en cualquier etapa de amplificación por PCR. El cebador que comprende el resto de unión puede usarse en múltiples etapas de un esquema de preparación de biblioteca.

**[0061]** El resto de unión puede ser cualquier molécula pequeña que tiene una pareja de unión. Los restos de unión ejemplares incluyen, biotina, estreptavidina, heparina, un aptámero, un resto de química de clic, un segmento o estructura de unión a proteínas, y un segmento o estructura de unión a ácido nucleico, y similares.

**[0062]** Los métodos y composiciones descritos en este documento abordan los retos de la separación física y enzimática de las fracciones de ADNss y ADNds durante la normalización de la biblioteca. Como se muestra en las Figuras **5A** y **5B**, una biblioteca no normalizada **500** contiene una especie **505** de alta abundancia y una especie **510** de baja abundancia. Durante la preparación inicial de la biblioteca, la biblioteca, o una fracción de la biblioteca, está etiquetada asimétricamente en un extremo con un enlace de unión **515** (como biotina, etc.). Las moléculas **520** de ácido nucleico de doble cadena marcadas se desnaturalizan **530** para generar moléculas de ácido nucleico monocatenario que incluyen una especie de alta abundancia **535** y una especie de baja abundancia **540**. Las moléculas de ácido nucleico monocatenario se vuelven a reconocer parcialmente **550** para formar moléculas de doble cadena de las especies de alta abundancia **555** mientras que las especies de baja abundancia siguen siendo de cadena única **560**. Después de la desnaturalización y el recocido parcial, todas las cadenas marcadas se capturan en una matriz de soporte **565**, como perlas de estreptavidina paramagnéticas, y las fracciones unidas y no unidas se separan **570**. Las secuencias muy abundantes se rehibridarán predominantemente, y ambas cadenas se eliminarán

en la fracción unida. Sin embargo, las secuencias de baja abundancia no se volverán a recocer, por lo que el complemento de la cadena marcada estará presente en la fracción no unida. La fracción no unida que incluye la especie **575** de alta abundancia de hebra simple y la especie **580** de hebra simple de una sola hebra representaría una biblioteca **585** normalizada, y podría usarse directamente o **590** más amplificado para aplicaciones posteriores.

#### Desnaturalización y recocido parcial

**[0063]** La biblioteca no normalizada (por ejemplo, que comprende diana de ácido nucleico ADNc) puede ser desnaturalizada. La desnaturalización se puede realizar mediante una variedad de métodos que incluyen calentar la muestra, tratar la muestra con solventes orgánicos (p. ej., DMS o formamida), cambiar la concentración de sal de la muestra y/o cambiar el pH de la muestra.

**[0064]** En algunos casos, la desnaturalización se lleva a cabo por calentamiento de la muestra. La desnaturalización se puede realizar a una temperatura de al menos 50°C, 60°C, 70°C, 80°C, 90°C, o 95°C o más. La desnaturalización se puede realizar a una temperatura de 50°C, 60°C, 70°C, 80°C, 90°C, o 95°C o más. La desnaturalización se puede realizar por lo menos en 1, 2, 3, 4, 5, 6, 7, 8, 9, o 10 o más minutos. La desnaturalización se puede realizar durante un máximo de 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10 o más minutos. La desnaturalización puede dar como resultado la desnaturalización de al menos 50, 60, 70, 80, 90 o 100% de los ADNc. La desnaturalización puede dar como resultado una desnaturalización de como máximo 50, 60, 70, 80, 90 o 100% de los ADNc. La desnaturalización puede dar como resultado que al menos 50, 60, 70, 80, 90 o 100% de los ácidos nucleicos se encuentren en forma monocatenaria.

**[0065]** La muestra desnaturalizada se puede volver a recocer. La muestra desnaturalizada puede ser recocida parcialmente. El recocido parcial se puede realizar por cualquier método, por ejemplo, enfriamiento rápido en hielo, cambiando la concentración de sal (por ejemplo, invirtiendo la concentración de sal de la cantidad utilizada en la desnaturalización), y/o cambiando el pH (por ejemplo, invirtiendo el pH del nivel usado en la desnaturalización), y similares. En algunos casos, el recocido parcial comprende enfriar la muestra desnaturalizada (por ejemplo, en hielo). El recocido parcial puede comprender el recocido de al menos 10, 20, 30, 40, 50, 60, 70, 80, 90 o 100% de las hebras de la muestra desnaturalizada. El recocido parcial puede comprender un recocido de como máximo 10, 20, 30, 40, 50, 60, 70, 80, 90 o 100% de las hebras de la muestra desnaturalizada. Al menos 10, 20, 30, 40, 50, 60, 70, 80, 90 o 100% de las cadenas de ácidos nucleicos altamente abundantes pueden recocer de nuevo durante la etapa de recocido parcial. Como máximo, 10, 20, 30, 40, 50, 60, 70, 80, 90 o 100% de las cadenas de ácidos nucleicos altamente abundantes se pueden volver a recocer durante la etapa de recocido parcial. Al menos 10, 20, 30, 40, 50, 60, 70, 80, 90 o 100% de las cadenas de ácidos nucleicos de menor abundancia pueden recocer de nuevo durante la etapa de recocido parcial. Como máximo, 10, 20, 30, 40, 50, 60, 70, 80, 90 o 100% de las cadenas de ácidos nucleicos de menor abundancia pueden recocer durante la etapa de recocido parcial.

**[0066]** Las hebras de especies más abundantes pueden recocer al menos 50, 100, 150, 200, 250, 300, 350, 400, 450, o 500% o más rápidamente que las hebras de especies más abundantes. Las hebras de especies más abundantes pueden volver a recocer a lo sumo 50, 100, 150, 200, 250, 300, 350, 400, 450 o 500% o más rápidamente que las hebras de especies más abundantes. Las hebras de especies más abundantes pueden recocer al menos 1, 2, 3, 4, 5, 6, 7, 8, 9, o 10 veces o más que las hebras de especies de menor abundancia. Las hebras de especies más abundantes pueden re-recocer 1,2,3,4,5,5,6,7,8,9, o 10 veces más que las hebras de especies más abundantes.

**[0067]** Después de volver a recocer la muestra puede comprender una pluralidad de moléculas de doble cadena, una pluralidad de moléculas de cadena sencilla que comprenden un resto de unión, y una pluralidad de moléculas monocatenarias que pueden no comprender el resto de unión. Las moléculas de doble cadena pueden comprender una hebra que puede comprender el resto de unión y una hebra que no comprende el resto de unión.

**[0068]** La muestra de re-recocido puede contactarse con un soporte sólido. El soporte sólido puede comprender un resto de captura. Un resto de captura puede unirse al resto de unión. Por ejemplo, si el resto de unión es biotina, el resto de captura puede ser estreptavidina. El soporte sólido puede ser, por ejemplo, una resina, una suspensión, una perla, una resina, un hidrogel, un soporte semisólido, un soporte insoluble y/o un soporte semisólido. En algunos casos, el soporte sólido es una resina.

**[0069]** El soporte sólido puede tener una propiedad física. Por ejemplo, el soporte sólido puede ser soluble en ciertas condiciones de pH, condiciones de sal y/o condiciones de temperatura. El soporte sólido puede ser magnético, ferromagnético y/o paramagnético.

**[0070]** El soporte sólido se puede unir a las moléculas de las muestras de re-recocido que componen el resto vinculante. El soporte sólido se puede unir a ADNc de doble cadena (por ejemplo, que comprende el resto de unión). El soporte sólido puede unirse a hebras no recocidas de una sola hebra (por ejemplo, que comprenden el resto de unión).

**[0071]** El soporte sólido se puede utilizar para purificar la muestra. El soporte sólido se puede separar de la muestra (es decir, sobrenadante) (por ejemplo, por centrifugación, magnetismo). La muestra sobrante (es decir, el sobrenadante) se puede denominar una biblioteca normalizada. La biblioteca normalizada puede comprender

moléculas de ácido nucleico monocatenario que pueden no comprender un resto de unión. Las moléculas de ácido nucleico monocatenarias pueden ser moléculas que no se recoció durante la etapa de recocido parcial del método.

[0072] La biblioteca normalizada puede comprender relativamente más la misma cantidad de especies inferiores y superiores abundantes en comparación con una biblioteca no normalizada. Por ejemplo, las especies más abundantes pueden ser al menos 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, o 20 o más veces mayor que las especies de menor abundancia en una biblioteca normalizada en comparación con una biblioteca no normalizada. Las especies más abundantes pueden ser como máximo 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, o 20 o más veces mayor que especies menos abundantes en una biblioteca normalizada en comparación con una biblioteca no normalizada. Las especies de menor abundancia pueden ser al menos 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, o 20 o más veces mayor que especies más abundantes en una biblioteca normalizada en comparación con una biblioteca no normalizada. Las especies de menor abundancia pueden ser como máximo 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, o 20 o más veces que las especies más abundantes en una biblioteca normalizada en comparación con una biblioteca no normalizada.

#### Uso de bloqueadores

[0073] En algunas realizaciones, los métodos de la descripción proporcionan para el uso de bloqueadores durante la desnaturalización y re-recocido parcial. Como se usa en el presente documento, los bloqueadores se refieren a secuencias de oligonucleótidos que pueden hibridar con las secuencias universales de una biblioteca (por ejemplo, secuencias de cebadores universales, secuencias de flujo de secuenciación de células universales). Los bloqueadores se pueden usar para evitar que los blancos/amplificadores de diferentes genes se recocien juntos durante el recocido parcial a través de sus regiones universales sin tener en cuenta la secuencia del gen.

[0074] Un ejemplo de uso de los bloqueadores se muestra en **Figura 4**. Una primera diana **405** y una segunda diana **410** pueden cada una comprender secuencias universales **415/420** y un resto de carácter definitivo **421**. En algunos casos, durante la desnaturalización y recocido parcial, dianas con secuencias de diferencia **405/410** pueden hibridarse juntas a través de las secuencias universales que cada una de ellas comprenden **415/420**. Se pueden usar bloqueadores para evitar que esto suceda. La muestra puede ponerse en contacto con **425** con los bloqueadores **430**. Los bloqueadores **430** pueden hibridar con una o más secuencias universales de las dianas. Se pueden usar uno o más tipos diferentes de bloqueadores **430**. Los bloqueadores pueden ayudar al recocido parcial forzando a las cadenas a asociarse (p. ej., hibridar) a través de sus secuencias genéticas. De esta manera, los bloqueadores pueden usarse para ayudar a los métodos de normalización de bibliotecas de la divulgación.

[0075] Un bloqueador puede ser al menos 1, 2, 3, 4, 5, 6, 7, 8, 9, o 10 o más nucleótidos de longitud. Un bloqueador puede tener a lo sumo 1, 2, 3, 4, 5, 6, 7, 8, 9, o 10 o más nucleótidos de longitud. Un bloqueador puede hibridar con su diana con al menos 10, 20, 30, 40, 50, 60, 70, 80, 90 o 100% de complementariedad. Un bloqueador puede hibridar a su diana con un máximo de 10, 20, 30, 40, 50, 60, 70, 80, 90 o 100% de complementariedad.

[0076] En algunas realizaciones, los métodos de normalización de la biblioteca de la divulgación se pueden realizar en un soporte sólido. Por ejemplo, se puede generar una biblioteca en la que los amplicones de la biblioteca están marcados asimétricamente con una de las moléculas involucradas en la química de clics (por ejemplo, azida, alquino, para la cicloadición de azida-alquino). El soporte sólido puede comprender la otra molécula en la química del clic. Por ejemplo, el amplicón puede comprender un alquino y el soporte sólido puede comprender una azida. Los amplicones se pueden unir al soporte sólido (por ejemplo, mediante química de clic). El soporte sólido puede calentarse induciendo así la desnaturalización de los amplicones unidos. Durante el recocido parcial, los amplicones que son más abundantes pueden recocer de nuevo a las moléculas unidas al soporte sólido. Los amplicones que son menos abundantes pueden dejarse en solución (por ejemplo, por centrifugación, magnetismo, cromatografía). Los soportes sólidos se pueden eliminar de la solución y, por lo tanto, lo deja una biblioteca normalizada.

#### Amplificación

[0077] Una o más reacciones de amplificación de ácido nucleico se pueden realizar para crear múltiples copias de las moléculas de ácidos nucleicos diana normalizadas. La amplificación se puede realizar de manera multiplexada, en la que se amplifican simultáneamente múltiples secuencias de ácido nucleico diana. La reacción de amplificación se puede usar para agregar adaptadores de secuenciación a las moléculas de ácido nucleico. Las reacciones de amplificación pueden comprender amplificar al menos una parte de una etiqueta de muestra, si está presente. Las reacciones de amplificación pueden comprender amplificar al menos una parte del marcador celular y/o molecular. Las reacciones de amplificación pueden comprender amplificar al menos una parte de una etiqueta de muestra, una etiqueta celular, una etiqueta espacial, una etiqueta molecular, un ácido nucleico diana o una combinación de los mismos. Las reacciones de amplificación pueden comprender amplificar al menos 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. 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 una etiqueta de muestra, una etiqueta celular, una etiqueta espacial y/o una etiqueta molecular.



**[0078]** En algunas realizaciones, la amplificación se puede realizar usando una reacción en cadena de la polimerasa (PCR). Como se usa en este documento, la PCR puede referirse a una reacción para la amplificación *in vitro* de secuencias específicas de ADN por la extensión de cebador simultánea de cadenas complementarias de ADN. Como se usa en este documento, la PCR puede abarcar formas derivadas de la reacción, que incluyen pero no se limitan a, RT-PCR, PCR en tiempo real, PCR anidada, PCR cuantitativa, PCR multiplexada, PCR digital y PCR de ensamblaje.

**[0079]** La amplificación de los ácidos nucleicos marcados también puede comprender métodos no basados en PCR. Los ejemplos de métodos no basados en PCR incluyen, pero no se limitan a, amplificación de desplazamiento múltiple (MDA), amplificación mediada por transcripción (TMA), amplificación de transcriptoma completo (WTA), amplificación del genoma completo (WGA), amplificación basada en secuencias de ácido nucleico (NASBA), amplificación de 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 PCR incluyen múltiples ciclos de amplificación de la transcripción de ARN impulsada por la ARN polimerasa dependiente de ADN o síntesis y transcripción de ADN dirigida por ARN para amplificar dianas de ADN o ARN, una reacción en cadena de la ligasa (LCR) y un método de replicasa Q $\beta$  (Q $\beta$ ), uso de sondas palindrómicas, amplificación por desplazamiento de cadena, amplificación dirigida por oligonucleótidos usando una endonucleasa de restricción, un método de amplificación en el que un cebador se hibrida a una secuencia de ácido nucleico y el dúplex resultante se escinde antes de la reacción de extensión y la amplificación, amplifique el desplazamiento de la cadena utilizando una polimerasa de ácido nucleico que carece de actividad de exonucleasa 5', amplificación de círculo rodante y amplificación de extensión de ramificación (RAM). En algunos casos, la amplificación puede no producir transcripciones circularizadas.

**[0080]** La PCR de supresión se puede usar para los métodos de amplificación de la divulgación. La PCR de supresión puede referirse a la exclusión selectiva de moléculas de menos de un cierto tamaño flanqueadas por repeticiones terminales invertidas, debido a su ineficiente amplificación cuando el/los cebador(es) utilizado(s) para la(s) amplificación(es) corresponde(n) a la repetición completa o una fracción de la repetición. La razón de esto puede estar en el equilibrio entre el recocido productivo de cebadores de PCR y el auto-recocido no productivo de los extremos complementarios del fragmento. A un tamaño fijo de una repetición invertida del terminal de flanqueo, cuanto más corta sea la inserción, más fuerte será el efecto de supresión y viceversa. Del mismo modo, a un tamaño de inserción fijo, cuanto más larga sea la repetición invertida del terminal, más fuerte será el efecto de supresión.

**[0081]** La PCR de supresión puede usar adaptadores que se ligan al final de un fragmento de ADN antes de la amplificación de la PCR. Tras la fusión y el recocido, los fragmentos de ADN de una sola hebra que tienen adaptadores autocomplementarios en los extremos 5' y 3' de la hebra pueden formar estructuras supresivas en forma de "raqueta de tenis" que suprimen la amplificación de los fragmentos durante la PCR.

**[0082]** En algunos casos, los métodos descritos aquí comprenden además la realización de una reacción en cadena de la polimerasa en el ácido nucleico marcado (por ejemplo, marcado con ARN, marcado con ADN, marcado con ADNc) para producir un amplicon etiquetado estocásticamente. El amplicón etiquetado estocásticamente puede ser una molécula de doble cadena. La molécula de doble cadena puede comprender o ser una molécula de ARN de doble cadena, una molécula de ADN de doble cadena o una molécula de ARN hibridada a una molécula de ADN. Una o ambas cadenas de la molécula de doble cadena pueden comprender una etiqueta de muestra, una etiqueta espacial, una etiqueta celular y/o una etiqueta molecular. El amplicón etiquetado estocásticamente puede ser una molécula monocatenaria. La molécula monocatenaria puede comprender ADN, ARN o una combinación de los mismos. Los ácidos nucleicos de la divulgación pueden comprender ácidos nucleicos sintéticos o alterados.

**[0083]** La amplificación puede comprender el uso de uno o más nucleótidos no naturales. Los nucleótidos no naturales pueden comprender nucleótidos fotolabiles o disparables. Los ejemplos de nucleótidos no naturales pueden incluir, pero no se limitan a, ácido nucleico peptídico (PNA), morfolino y ácido nucleico encerrado (LNA), así como ácido nucleico de glicol (GNA) y ácido nucleico de treosa (TNA). Se pueden agregar nucleótidos no naturales a uno o más ciclos de una reacción de amplificación. La adición de los nucleótidos no naturales se puede usar para identificar productos como ciclos específicos o puntos de tiempo en la reacción de amplificación.

**[0084]** La realización de una o más reacciones de amplificación puede comprender el uso de uno o más cebadores. El uno o más cebadores pueden comprender al menos 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 o 15 o más nucleótidos. El uno o más cebadores pueden comprender al menos 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 o 15 anormonucleótidos. El uno o más cebadores pueden comprender menos de 12-15 nucleótidos. El uno o más cebadores pueden recocer al menos a una porción de la pluralidad de dianas etiquetadas estocásticamente. El uno o más cebadores pueden unirse al extremo 3' o al extremo 5' de la pluralidad de dianas etiquetadas estocásticamente. El uno o más cebadores pueden asociarse a una región interna de la pluralidad de dianas etiquetadas estocásticamente. La región interna puede ser al menos aproximadamente 50, 100, 150, 200, 220, 230, 240, 250, 260, 270, 280, 290, 300, 310, 320, 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 estocásticamente. El uno o más cebadores pueden comprender un panel fijo de cebadores. El uno o más cebadores pueden comprender al menos uno o más cebadores personalizados. El uno o más cebadores pueden comprender al menos uno o más cebadores de control. El uno o más cebadores pueden comprender al menos uno o más cebadores específicos de gen.

5 **[0085]** El uno o más cebadores pueden comprender cualquier cebador universal de la descripción. El cebador universal puede unirse a un sitio de unión del cebador universal. El uno o más cebadores personalizados pueden asociarse a una primera etiqueta de muestra, una segunda etiqueta de muestra, una etiqueta espacial, una etiqueta celular, una etiqueta molecular, una diana o cualquier combinación de los mismos. El uno o más cebadores pueden comprender un cebador universal y un cebador personalizado. El cebador personalizado se puede diseñar para amplificar una o más dianas. Las dianas pueden comprender un subconjunto del total de dianas etiquetadas estocásticamente en una o más muestras. El uno o más cebadores pueden comprender al menos 96 o más cebadores personalizados. El uno o más cebadores pueden comprender al menos 960 o más cebadores personalizados. El uno o más cebadores pueden comprender al menos 9600 o más cebadores personalizados. El uno o más cebadores personalizados pueden reconocerse con dos o más ácidos nucleicos marcados diferentes. Los dos o más ácidos nucleicos marcados diferentes pueden corresponder a uno o más genes.

15 **[0086]** Cualquier esquema de amplificación se puede utilizar en los métodos de la presente descripción. Por ejemplo, en un esquema, la primera ronda de PCR puede amplificar moléculas (por ejemplo, unidas a la perla) utilizando un cebador específico del gen y un cebador contra la secuencia del cebador 1 de secuenciación de Illumina universal. La segunda ronda de PCR puede amplificar los primeros productos de PCR utilizando un cebador específico del gen anidado flanqueado por la secuencia del cebador 2 de secuenciación de Illumina, y un cebador contra la secuencia del cebador 1 de la secuenciación de Illumina universal. La tercera ronda de PCR agrega P5 y P7 y el índice de muestra para convertir los productos de PCR en una biblioteca de secuenciación de Illumina. La secuenciación con 150bp x 2 puede revelar el marcador celular y el índice molecular en la lectura 1, el gen en la lectura 2 y el índice de muestra en la lectura de índice 1.

25 **[0087]** La amplificación se puede realizar en una o más rondas. En algunos casos hay múltiples rondas de amplificación. La amplificación puede comprender dos o más rondas de amplificación. La primera amplificación puede ser una extensión de X' para generar la región específica del gen. La segunda amplificación puede ocurrir cuando una muestra nucleica hibrida a la hebra recién generada.

30 **[0088]** En algunas realizaciones, no tiene que producirse la hibridación al final de una molécula de ácido nucleico. En algunas realizaciones, un ácido nucleico diana dentro de una cadena intacta de un ácido nucleico más largo se hibrida y amplifica. Por ejemplo, una diana dentro de una sección más larga de ADN genómico o ARNm. Una diana puede ser más de 50 nt, más de 100 nt, o más de 1000 nt desde un extremo de un polinucleótido.

#### 35 Preparación de la biblioteca que incluye la ligadura del adaptador

**[0089]** Las moléculas de cadena sencilla de la biblioteca no normalizada (o biblioteca normalizada) se pueden preparar para la secuenciación, que puede, por ejemplo, incluir la generación de una molécula de doble cadena y la incorporación de adaptadores de secuenciación de células de flujo (por ejemplo, por ligadura y/o hibridación y PCR).

40 **[0090]** En algunas realizaciones, los adaptadores se pueden ligar al ácido nucleico de doble cadena. Los adaptadores pueden comprender una primera secuencia de cebador universal de la divulgación, una segunda secuencia de cebador universal de la divulgación y un sitio de unión de endonucleasas de restricción, o cualquier combinación de los mismos. En algunos casos, el adaptador comprende una segunda secuencia de cebador universal de la divulgación y un sitio de unión a la endonucleasa de restricción.

50 **[0091]** El término "adaptador" que se utiliza aquí se refiere a un oligonucleótido de hebra única o de doble hebra de al menos 10, 15, 20 o 25 bases que pueden ser fijadas al extremo de un ácido nucleico. Las secuencias adaptadoras pueden sintetizarse utilizando, por ejemplo, sitios de cebado, el complemento de un sitio de cebado y sitios de reconocimiento para endonucleasas, secuencias comunes y promotores. El adaptador puede ser de cadena total o sustancialmente doble. Un adaptador de doble cadena puede comprender dos oligonucleótidos que son al menos parcialmente complementarios. El adaptador puede estar fosforilado o no fosforilado en una o ambas cadenas. El adaptador puede tener una sección bicatenaria y una sección saliente monocatenaria que es completa o parcialmente complementaria a una suspensión (por ejemplo, generada por una enzima de restricción o una enzima polimerasa).

55 El saliente en el adaptador puede ser, por ejemplo, de 4 a 8 bases. Por ejemplo, cuando el ADN se digiere con la enzima de restricción EcoRI, los fragmentos de doble hebra resultantes están flanqueados en cada extremo por el saliente monocatenario 5'-AATT-3', un adaptador que lleva un saliente monocatenario 5'-AATT-3' puede hibridar con el fragmento a través de la complementariedad entre las regiones sobresalientes. Esta hibridación de "extremo pegajoso" del adaptador al fragmento facilita la ligadura del adaptador al fragmento, sin embargo, también es posible una ligadura de extremos romos. Los extremos romos se pueden convertir en extremos pegajosos utilizando, por ejemplo, la actividad de exonucleasa del fragmento Klenow. Por ejemplo, cuando el ADN se divide con PvuII, los extremos romos se pueden convertir en un saliente de dos pares de bases incubando los fragmentos con Klenow en presencia de dTTP y dCTP. Los salientes también se pueden convertir en extremos romos rellenando un saliente o eliminando un saliente.

65

**[0092]** Los adaptadores pueden estar ligados a ADNc bicatenarias de la divulgación. Los métodos de ligación pueden incluir el uso de la ADN ligasa T4 que cataliza la formación de un enlace fosfodiéster entre los fosfatos 5' yuxtapuestos y los extremos hidroxilo 3' en ADN o ARN dúplex con extremos romos y pegajosos; Taq ADN Ligasa, que cataliza la formación de un enlace fosfodiéster entre el extremo 5' fosfato yuxtapuesto y el extremo 3' hidroxilo de dos oligonucleótidos adyacentes que se hibridan a un ADN diana complementario; La ADN ligasa de E. coli que cataliza la formación de un enlace fosfodiéster entre los extremos 5'-fosfato yuxtapuestos y los extremos 3'-hidroxilo en el ADN dúplex que contiene extremos cohesivos; y T4 ARN ligasa que cataliza la ligadura de un donante de ácido nucleico 5' fosforil terminado en un aceptor de ácido nucleico 3' terminado en hidroxilo 3' a través de la formación de un enlace fosfodiéster 3' a 5', los sustratos incluyen ARN de cadena simple y ADN así como los pirofosfatos de dinucleósidos; o cualquier otro método descrito en la técnica. Diferentes enzimas generan diferentes salientes y el saliente del adaptador puede dirigirse a ligarse a fragmentos generados por enzimas de restricción seleccionadas.

**[0093]** En algunas realizaciones, un adaptador de doble cadena se utiliza y sólo una hebra del adaptador se liga con el ADNc de doble cadena. La ligadura de una hebra de un adaptador puede bloquearse selectivamente. Para bloquear la ligadura, por ejemplo, una cadena del adaptador puede diseñarse para introducir un espacio de uno o más nucleótidos entre el extremo 5' de esa cadena del adaptador y el extremo 3' del ácido nucleico diana. La ausencia de un fosfato en el extremo 5' de un adaptador puede bloquear la ligadura de ese extremo 5' a un 3'OH disponible.

### Secuenciación

**[0094]** La determinación del número de diferentes ácidos nucleicos marcados estocásticamente puede comprender determinar la secuencia de la diana etiquetada, el marcador espacial, el marcador molecular, el marcador de la muestra y el marcador celular o cualquier producto del mismo (por ejemplo, amplicones etiquetados, moléculas de ADNc marcadas). Una diana amplificada puede ser sometida a secuenciación. La determinación de la secuencia del ácido nucleico estocásticamente marcado o cualquier producto del mismo puede comprender realizar una reacción de secuenciación para determinar la secuencia de al menos una parte de una etiqueta de muestra, una etiqueta espacial, una etiqueta celular, una etiqueta molecular y/o al menos una parte de la diana etiquetada estocásticamente, un complemento del mismo, un complemento inverso del mismo, o cualquier combinación de los mismos.

**[0095]** La determinación de la secuencia de un ácido nucleico (por ejemplo, ácido nucleico amplificado, ácido nucleico marcado, copia de ADNc de un ácido nucleico marcado, etc.) puede realizarse utilizando una variedad de métodos de secuenciación que incluyen, entre otros, la secuenciación por síntesis (SBS), secuenciación por hibridación (SBH), secuenciación por ligadura (SBL), secuenciación cuantitativa de adición de nucleótidos fluorescentes incrementales (QIFNAS), ligadura por etapas y escisión, transferencia de energía de resonancia de fluorescencia (FRET), balizas moleculares, digestión de la sonda reportera TaqMan, pirosecuenciación, secuenciación fluorescente in situ (FISSEQ), esferas FISSEQ, secuenciación de bamboleo, secuenciación múltiplex, secuenciación de colonias polimerizadas (POLONY); secuenciación de círculo rodante (ROLONY), ensayos de ligamiento de oligo específicos de alelo (por ejemplo, ensayo de ligamiento de oligo (OLA), OLA de molécula de plantilla única que utiliza una sonda lineal ligada y una lectura de amplificación de círculo rotatorio (RCA), sondas de candado ligadas, o OLA de molécula de plantilla única que utiliza una sonda de candado circular ligada y una lectura de amplificación de círculo rodante (RCA), y similares.

**[0096]** En algunos casos, la determinación de la secuencia del ácido nucleico marcada o cualquier producto de los mismos comprende secuenciación de extremo emparejado, secuenciación de nanoporos, secuenciación de alto rendimiento, secuenciación de escopeta, secuenciación de terminador tinte, secuenciación de ADN múltiple cebador, primer recorrido, secuenciación dideoxi de Sanger, secuenciación de Maxim-Gilbert, pirosecuenciación, secuenciación de molécula única verdadera, o cualquier combinación de las mismas. Alternativamente, la secuencia del ácido nucleico marcado o cualquier producto del mismo puede determinarse por microscopía electrónica o una matriz de transistores de efecto de campo (chem-FET) sensibles a los productos químicos.

**[0097]** Métodos de secuenciación de alto rendimiento, tales como la secuenciación de disposición cíclica utilizando plataformas como Roche 454, Illumina Solexa, ABI-SOLiD, ION Torrent, Complete Genomics, Pacific Bioscience, Helicos, o la plataforma Polonator, también se pueden utilizar. La secuenciación puede comprender la secuenciación de MiSeq. La secuenciación puede comprender la secuenciación de HiSeq.

**[0098]** Las dianas estocásticamente etiquetadas pueden comprender ácidos nucleicos que representan desde aproximadamente 0,01% de los genes del genoma de un organismo a aproximadamente 100% de los genes del genoma de un organismo. Por ejemplo, alrededor del 0,01% de los genes del genoma de un organismo a aproximadamente el 100% de los genes del genoma de un organismo se puede secuenciar utilizando una región complementaria de destino que comprende una pluralidad de multímeros al capturar los genes que contienen una secuencia complementaria de la muestra. En algunas realizaciones, los ácidos nucleicos marcados comprenden ácidos nucleicos que representan desde aproximadamente el 0,01% de las transcripciones del transcriptoma de un organismo hasta aproximadamente el 100% de las transcripciones del transcriptoma de un organismo. Por ejemplo, alrededor del 0,501% de las transcripciones de un transcriptoma de un organismo a aproximadamente el 100% de las transcripciones de un transcriptoma de un organismo puede secuenciarse utilizando una región complementaria diana que comprende una cola poli-T capturando los ARNm de los muestra.

**[0099]** La secuenciación puede comprender la secuenciación de al menos aproximadamente 10, 20, 30, 40, 50, 60, 70, 80, 90, 100 o más nucleótidos o pares de bases del ácido nucleico marcado y/o el código de barras estocástico. La secuenciación puede comprender la secuenciación a lo sumo de aproximadamente 10, 20, 30, 40, 50, 60, 70, 80, 90, 100 o más nucleótidos o pares de bases del ácido nucleico marcado y/o el código de barras estocástico. La secuenciación puede comprender la secuenciación de al menos aproximadamente 200, 300, 400, 500, 600, 700, 800, 900, 1.000 o más nucleótidos o pares de base del ácido nucleico marcado y/o del código de barras estocástico. La secuenciación puede comprender la secuenciación a lo sumo aproximadamente 200, 300, 400, 500, 600, 700, 800, 900, 1.000 o más nucleótidos o pares de bases del ácido nucleico marcado y/o el código de barras estocástico. La secuenciación puede comprender la secuenciación de al menos aproximadamente 1.500; 2.000; 3.000; 4.000; 5.000; 6.000; 7.000; 8.000; 9.000; o 10.000 o más nucleótidos o pares de bases del ácido nucleico marcado y/o el código de barras estocástico. La secuenciación puede comprender la secuenciación a lo sumo alrededor de 1.500; 2.000; 3.000; 4.000; 5.000; 6.000; 7.000; 8.000; 9.000; o 10.000 o más nucleótidos o pares de bases del ácido nucleico marcado y/o el código de barras estocástico.

**[0100]** La secuenciación puede comprender al menos aproximadamente 200, 300, 400, 500, 600, 700, 800, 900, 1.000 o más lecturas de secuencia por ejecución. La secuenciación puede comprender a lo sumo aproximadamente 200, 300, 400, 500, 600, 700, 800, 900, 1.000 o más lecturas de secuencia por ejecución. En algunos casos, la secuenciación comprende la secuenciación de al menos aproximadamente 1.500; 2.000; 3.000; 4.000; 5.000; 6.000; 7.000; 8.000; 9.000; o 10.000 o más lecturas de secuencia por ejecución. En algunos casos, la secuenciación comprende la secuencia a lo sumo alrededor de 1.500; 2.000; 3.000; 4.000; 5.000; 6.000; 7.000; 8.000; 9.000; o 10.000 o más lecturas de secuencia por ejecución. La secuenciación puede comprender la secuenciación de al menos 10, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 900, 950 o 1000 o más millones de secuencias de lecturas por ejecución. La secuenciación puede comprender la secuenciación a lo sumo 10, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950 o 1000 o más millones de secuencias de lecturas por ejecución. La secuenciación puede comprender la secuenciación de al menos 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 2000, 3000, 4000 o 5000 o más millones de la secuenciación lee en total. La secuenciación puede comprender la secuenciación como máximo 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 2000, 3000, 4000 o 5000 o más millones de las lecturas de secuenciación en total. La secuenciación puede comprender menos o igual a aproximadamente 1.600.000.000 lecturas de secuencia por ejecución. La secuenciación puede comprender menos o igual a aproximadamente 200.000.000 lecturas por ejecución.

**[0101]** En la biblioteca normalizada generada por los métodos de la descripción, transcripciones menos abundantes (por ejemplo, más raras) pueden ser identificadas más fácilmente que en una biblioteca no normalizada. Las lecturas de secuencia de los transcritos menos abundantes en una biblioteca normalizada pueden comprender una porción más grande de lecturas totales que en una biblioteca no normalizada. Las lecturas de secuencia de una transcripción menos abundante en una biblioteca normalizada pueden comprender al menos 50, 100, 150, 200, 250, 300, 350, 400, 450 o 500% o más en comparación con las lecturas de la misma transcripción en una biblioteca normalizada. Las lecturas de secuencia de una transcripción menos abundante en una biblioteca normalizada pueden ser al menos 1, 2, 3, 4, 5 o 6 veces o más que las lecturas de secuencia para la misma transcripción en una biblioteca no normalizada.

#### Códigos de barras estocásticos

**[0102]** Un código de barras estocástico utilizado aquí se refiere a una secuencia de polinucleótidos que se puede usar para marcar estocásticamente (por ejemplo, código de barras, etiqueta) una diana. Un código de barras estocástico puede comprender una o más etiquetas. Las etiquetas ejemplares incluyen, pero no se limitan a, una etiqueta universal, una etiqueta celular, una etiqueta molecular, una etiqueta de muestra, una etiqueta de placa, una etiqueta espacial y/o una etiqueta preespacial. Un código de barras estocástico puede comprender una amina 5' que puede vincular el código de barras estocástico a un soporte sólido. El código de barras estocástico puede comprender una o más etiquetas universales, una o más etiquetas de dimensión, una o más etiquetas espaciales, una o más etiquetas celulares y/o una o más etiquetas moleculares. La ubicación de cada una de las diferentes etiquetas en el código de barras estocástico puede variar. Por ejemplo, la etiqueta universal puede ser la etiqueta más 5'. La etiqueta molecular puede ser la etiqueta 3' más. La etiqueta espacial, la etiqueta de dimensión y la etiqueta celular pueden estar en cualquier orden. En algunas realizaciones, la etiqueta universal, la etiqueta espacial, la etiqueta de dimensión, la etiqueta celular y la etiqueta molecular están en cualquier orden. El código de barras estocástico puede comprender 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) en una muestra. Por ejemplo, una región de unión a la diana puede comprender una secuencia oligo dT que puede interactuar con colas de poli-A de los ARNm. En algunos casos, las etiquetas del código de barras estocástico (por ejemplo, etiqueta universal, etiqueta de dimensión, etiqueta espacial, etiqueta celular y etiqueta molecular) pueden estar separadas por 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 o 20 o más nucleótidos.

**[0103]** Un código de barras estocástico puede comprender una o más etiquetas universales. La una o más etiquetas universales pueden ser iguales para todos los códigos de barras estocásticos en el conjunto de códigos de barras estocásticos (por ejemplo, unidos a un soporte sólido dado). En algunas realizaciones, una o más etiquetas universales

pueden ser las mismas para todos los códigos de barras estocásticos unidos a una pluralidad de perlas. En algunas realizaciones, una etiqueta universal puede comprender una secuencia de ácido nucleico que es capaz de hibridar con un cebador de secuenciación. Los cebadores de secuenciación pueden usarse para secuenciar códigos de barras estocásticos que comprenden una etiqueta universal. Los cebadores de secuenciación (por ejemplo, los 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 que es capaz de hibridar con un cebador de PCR. En algunas realizaciones, el marcador universal puede comprender una secuencia de ácido nucleico que es capaz de hibridar con un cebador de secuenciación y un cebador de PCR. La secuencia de ácido nucleico de la etiqueta universal que es capaz de hibridar con una secuenciación o cebador de PCR puede denominarse un sitio de unión de cebador. Una etiqueta universal puede comprender una secuencia que puede usarse para iniciar la transcripción del código de barras estocástico. Una etiqueta universal puede comprender una secuencia que se puede usar para la extensión del código de barras estocástico o una región dentro del código de barras estocástico. Una etiqueta universal puede ser al menos aproximadamente 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50 o más nucleótidos de longitud. Una etiqueta universal puede comprender al menos aproximadamente 10 nucleótidos. Una etiqueta universal puede tener como máximo aproximadamente 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50 o más nucleótidos de longitud. En algunas realizaciones, un enlazador escindible o un nucleótido modificado puede ser parte de la secuencia del marcador universal para permitir que el código de barras estocástico se escinda del soporte. Como se usa en este documento, una etiqueta universal se puede usar indistintamente con "cebador de PCR universal".

**[0104]** Un código de barras estocástico puede comprender una etiqueta de dimensión. Una etiqueta de dimensión puede comprender una secuencia de ácido nucleico que proporciona información sobre una dimensión en la que se produjo el etiquetado estocástico. Por ejemplo, una etiqueta de dimensión puede proporcionar información sobre el momento en que una diana estaba estocásticamente codificada con barras. Una etiqueta de dimensión se puede asociar con un tiempo de código de barras estocástico en una muestra. Una etiqueta de dimensión puede activarse en el momento del etiquetado estocástico. Las diferentes etiquetas pueden ser activadas en diferentes momentos. La etiqueta de dimensión proporciona información sobre el orden en que las dianas, los grupos de dianas y/o las muestras se codificaron mediante códigos de barras estocásticamente. Por ejemplo, una población de células puede tener un código de barras estocástico en la fase G0 del ciclo celular. Las células pueden pulsarse nuevamente con códigos de barras estocásticos en la fase G1 del ciclo celular. Las células pueden pulsarse nuevamente con códigos de barras estocásticos en la fase S del ciclo celular, y así sucesivamente. Los códigos de barras estocásticos en cada pulso (por ejemplo, cada fase del ciclo celular), pueden comprender diferentes etiquetas de dimensión. De esta manera, la etiqueta de dimensión proporciona información sobre qué dianas se etiquetaron en qué fase del ciclo celular. Las etiquetas de dimensión pueden interrogar muchos tiempos biológicos diferentes. Los tiempos 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 (p. ej., una célula, una población de células) se puede marcar de forma estocástica antes y/o después del tratamiento con un medicamento y/o terapia. Los cambios en el número de copias de distintas dianas pueden ser indicativas de la respuesta de la muestra al fármaco y/o terapia.

**[0105]** Una etiqueta de dimensión puede ser activable. Una etiqueta de dimensión activable se puede activar en un punto de tiempo específico. La etiqueta de dimensión activable se puede activar de forma constitutiva (por ejemplo, no apagada). La etiqueta de dimensión activable puede activarse de manera reversible (por ejemplo, la etiqueta de dimensión activa puede activarse y desactivarse). La etiqueta de la dimensión se puede activar de forma reversible al menos 1, 2, 3, 4, 5, 6, 7, 8, 9, o 10 veces más. La etiqueta de la dimensión se puede activar de forma reversible al menos 1, 2, 3, 4, 5, 6, 7, 8, 9, o 10 o más veces. La etiqueta de la dimensión puede activarse con fluorescencia, luz, un evento químico (p. ej., escisión, ligadura de otra molécula, adición de modificaciones (p. ej., pegilado, sumoilado, acetilado, metilado, desacetilado, desmetilado), un evento fotoquímico (p. ej., fotoenvasado), e introducción de un nucleótido no natural.

**[0106]** La etiqueta de dimensión puede ser idéntica para todos los códigos de barras estocásticos unidos a un soporte sólido dado (por ejemplo, perla), pero diferentes para diferentes soportes sólidos (por ejemplo, perlas). En algunas realizaciones, al menos 60%, 70%, 80%, 85%, 90%, 95%, 97%, 99% o 100% de códigos de barras estocásticos en el mismo soporte sólido pueden comprender la misma etiqueta de dimensión. En algunas realizaciones, al menos el 60% de los códigos de barras estocásticos en el mismo soporte sólido pueden comprender la misma etiqueta de dimensión. En algunas realizaciones, al menos el 95% de los códigos de barras estocásticos en el mismo soporte sólido pueden comprender la misma etiqueta de dimensión.

**[0107]** Puede haber hasta  $10^6$  o más secuencias de etiquetas de dimensión únicas representadas en una pluralidad de soportes sólidos (por ejemplo, perlas). Una etiqueta de dimensión puede tener al menos aproximadamente 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50 o más nucleótidos de longitud. Una etiqueta de dimensión puede tener, a lo sumo, aproximadamente 300, 200, 100, 90, 80, 70, 60, 50, 40, 30, 20, 15, 12, 10, 9, 8, 7, 6, 5, 4, o menos o más nucleótidos en longitud. Una etiqueta de dimensión puede comprender de aproximadamente 5 a aproximadamente 200 nucleótidos. Una etiqueta de dimensión puede comprender de aproximadamente 10 a aproximadamente 150 nucleótidos. Una etiqueta de dimensión puede comprender de aproximadamente 20 a aproximadamente 125 nucleótidos de longitud.

**[0108]** Un código de barras estocástico puede comprender una etiqueta espacial. Una etiqueta 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 estocástico. Una etiqueta espacial puede asociarse con una coordenada en una muestra. La coordenada puede ser una coordenada fija. Por ejemplo, una coordenada se puede fijar en referencia a un sustrato. Una etiqueta espacial puede estar en referencia a una cuadrícula bidimensional o tridimensional. Una coordenada se puede fijar en referencia a un hito. El hito puede ser identificable en el espacio. Un hito puede ser una estructura que puede ser fotografiada. Un hito puede ser una estructura biológica, por ejemplo, un hito anatómico. Un hito puede ser un hito 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, código de barras, propiedad magnética, fluorescentes, radioactividad o un tamaño o forma únicos. Una etiqueta espacial puede asociarse con una partición física (por ejemplo, un pozo, un contenedor o una gota). En algunos casos, varias etiquetas espaciales se utilizan juntas para codificar una o más posiciones en el espacio.

**[0109]** La etiqueta espacial puede ser idéntica para todos los códigos de barras estocásticos unidos a un soporte sólido dado (por ejemplo, perla), pero diferente para diferentes soportes sólidos (por ejemplo, granos). En algunas realizaciones, al menos 60%, 70%, 80%, 85%, 90%, 95%, 97%, 99% o 100% de códigos de barras estocásticos en el mismo soporte sólido pueden comprender la misma etiqueta espacial. En algunas realizaciones, al menos el 60% de los códigos de barras estocásticos en el mismo soporte sólido pueden comprender la misma etiqueta espacial. En algunas realizaciones, al menos el 95% de los códigos de barras estocásticos en el mismo soporte sólido pueden comprender la misma etiqueta espacial.

**[0110]** Puede haber tantos como  $10^6$  o más únicas secuencias de etiqueta espaciales representadas en una pluralidad de soportes sólidos (por ejemplo, perlas). Una etiqueta espacial puede ser al menos aproximadamente 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50 o más nucleótidos en longitud. Una etiqueta espacial puede ser a lo sumo aproximadamente 300, 200, 100, 90, 80, 70, 60, 50, 40, 20, 15, 12, 10, 9, 8, 7, 6, 5, 4 o menos o más nucleótidos en longitud. Una etiqueta espacial puede comprender de aproximadamente 5 a aproximadamente 200 nucleótidos. Una etiqueta espacial puede comprender de aproximadamente 10 a aproximadamente 150 nucleótidos. Una etiqueta espacial puede comprender de aproximadamente 20 a aproximadamente 125 nucleótidos de longitud.

**[0111]** Códigos de barras estocásticos pueden comprender una etiqueta celular (es decir, etiqueta de la muestra). Como se usa en este documento, los términos "etiqueta de muestra" y "etiqueta celular" se pueden usar de manera intercambiable. Una etiqueta 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, la etiqueta celular es idéntica para todos los códigos de barras estocásticos unidos a un soporte sólido dado (por ejemplo, perla), pero diferente para diferentes soportes sólidos (por ejemplo, perlas). En algunas realizaciones, al menos el 60%, 70%, 80%, 85%, 90%, 95%, 97%, 99% o 100% de los códigos de barras estocásticos en el mismo soporte sólido pueden comprender la misma etiqueta celular. En algunas realizaciones, al menos el 60% de los códigos de barras estocásticos en el mismo soporte sólido pueden comprender la misma etiqueta celular. En alguna realización, al menos el 95% de los códigos de barras estocásticos en el mismo soporte sólido pueden comprender la misma etiqueta celular.

**[0112]** Puede haber tantas como  $10^6$  o más únicas secuencias de etiqueta celular representadas en una pluralidad de soportes sólidos (por ejemplo, perlas). Una etiqueta celular puede tener al menos aproximadamente 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50 o más nucleótidos de longitud. Una etiqueta celular puede ser a lo sumo aproximadamente 300, 200, 100, 90, 80, 70, 60, 50, 40, 30, 20, 12, 10, 9, 8, 7, 6, 5, 4 o menos o más nucleótidos en longitud. Una etiqueta celular puede comprender de aproximadamente 5 a aproximadamente 200 nucleótidos. Una etiqueta celular puede comprender de aproximadamente 10 a aproximadamente 150 nucleótidos. Una etiqueta celular puede comprender de aproximadamente 20 a aproximadamente 125 nucleótidos de longitud.

**[0113]** Códigos de barras estocásticos pueden comprender una etiqueta molecular. Un marcador molecular puede comprender una secuencia de ácido nucleico que proporciona información de identificación para el tipo específico de especies de ácido nucleico diana hibridadas con el código de barras estocástico. 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 al código de barras estocástico (por ejemplo, la región de unión a la diana). En algunas realizaciones, un conjunto diverso de etiquetas moleculares se adjunta a un soporte sólido dado (por ejemplo, una perla). En algunas realizaciones, puede haber hasta  $10^6$  o más secuencias de etiquetas moleculares únicas unidas a un soporte sólido dado (por ejemplo, una perla). En algunas realizaciones, puede haber hasta  $10^5$  o más secuencias de marcadores moleculares únicas unidas a un soporte sólido dado (por ejemplo, una perla). En algunas realizaciones, puede haber hasta  $10^4$  o más secuencias de etiquetas moleculares únicas unidas a un soporte sólido dado (por ejemplo, una perla). En algunas realizaciones, puede haber hasta  $10^3$  o más secuencias de marcaje molecular únicas unidas a un soporte sólido dado (por ejemplo, una perla). En algunas realizaciones, puede haber hasta  $10^2$  o más secuenciación única de marcadores moleculares está unido a un soporte sólido dado (por ejemplo, una perla). Una etiqueta molecular puede tener al menos aproximadamente 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50 o más nucleótidos de longitud. Un marcador molecular puede tener a lo sumo unos 300, 200, 100, 90, 80, 70, 60, 50, 40, 30, 20, 15, 12, 10, 9, 8, 7, 6, 5, 4 o menos nucleótidos en longitud.

**[0114]** Códigos de barras estocásticos pueden comprender una región de unión a la diana. En algunas realizaciones, las regiones de unión a diana pueden comprender una secuencia de ácido nucleico que se hibrida específicamente a una diana (por ejemplo, ácido nucleico diana, molécula diana, por ejemplo, un ácido nucleico celular para ser analizado), por ejemplo a una secuencia de genes. En algunas realizaciones, una región de unión a diana puede comprender una secuencia de ácido nucleico que puede unirse (por ejemplo, hibridar) a una ubicación específica de un ácido nucleico diana específico. En algunas realizaciones, la región de unión a diana puede comprender una secuencia de ácido nucleico que es capaz de hibridación específica a un saliente del sitio de restricción (por ejemplo, un saliente de extremo pegajoso EcoRI). El código de barras estocástico puede ligarse a cualquier molécula de ácido nucleico que comprenda una secuencia complementaria al saliente del sitio de restricción.

**[0115]** Un código de barras estocástico puede comprender una región de unión a diana. Una región de unión a diana puede hibridar con una diana de interés. Por ejemplo, una región de unión a diana puede comprender un oligo dT que puede hibridar con ARNm que comprende extremos poli-adenilados. Una región de unión a diana puede ser específica de gen. Por ejemplo, una región de enlace de destino puede configurarse para hibridar con una región específica de un destino. Una región de unión a diana puede ser al menos 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 o más nucleótidos de longitud. Una región de unión a diana puede ser 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 o más nucleótidos de longitud. Una región de unión a diana puede tener una longitud de 5 a 30 nucleótidos. Cuando un código de barras estocástico comprende una región de unión a diana específica del gen, el código de barras estocástico puede denominarse un código de barras estocástico específico del gen.

**[0116]** Una región diana de unión puede comprender una secuencia de ácido nucleico no diana específico. 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 diana puede comprender una secuencia de multímero aleatoria, o una secuencia de oligo-dT que se hibrida a la cola de poli-A en moléculas de ARNm. Una secuencia de multímeros aleatorios puede ser, por ejemplo, un dímero, trímero, cuatrímero, pentámero, hexámero, septámero, octámero, nonámero, decámero o secuencia de multímeros superior de cualquier longitud. En algunas realizaciones, la región de unión a diana es la misma para todos los códigos de barras estocásticos unidos a un cordón dado. En algunas realizaciones, las regiones de unión a diana para la pluralidad de códigos de barras estocásticos unidos a una perla dada pueden comprender dos o más secuencias de unión a diana diferentes. Una región de unión a diana puede tener al menos aproximadamente 5, 10, 15, 20, 25, 30, 35, 40, 45, 50 o más nucleótidos de longitud. Una región de enlace de destino puede estar en la mayoría de aproximadamente 5, 10, 15, 20, 25, 30, 35, 40, 45, 50 o más nucleótidos de longitud.

**[0117]** Un código de barras estocástico puede comprender una propiedad de orientación que puede ser utilizada para orientar (por ejemplo, alinear) los códigos de barras estocásticos. Un código de barras estocástico puede comprender un resto para el enfoque isoelectrónico. Diferentes códigos de barras estocásticos pueden comprender diferentes puntos de enfoque isoelectrónico. Cuando estos códigos de barras estocásticos se introducen en una muestra, la muestra puede experimentar un enfoque isoelectrónico para orientar los códigos de barras estocásticos en una forma conocida. De esta manera, la propiedad de orientación se puede utilizar para desarrollar un mapa conocido de códigos de barras estocásticos en una muestra. Las propiedades de orientación ejemplares pueden incluir movilidad electroforética (por ejemplo, según el tamaño del código de barras estocástico), punto isoelectrónico, espín, conductividad y/o autoensamblaje. Por ejemplo, los códigos de barras estocásticos pueden comprender una propiedad de orientación del autoensamblaje, o pueden autoensamblarse en una orientación específica (p. ej., nanoestructura de ácido nucleico) tras la activación.

**[0118]** Un código de barras estocástico puede comprender una propiedad de afinidad. Una etiqueta espacial puede comprender una propiedad de afinidad. Una propiedad de afinidad puede incluir un resto químico y/o biológico que puede facilitar la unión del código de barras estocástico a otra entidad (p. ej., receptor celular). Por ejemplo, una propiedad de afinidad puede comprender un anticuerpo. Un anticuerpo puede ser específico para un grupo específico (p. ej., un receptor) en una muestra. Un anticuerpo puede guiar el código de barras estocástico a un tipo de célula o molécula específica. Las dianas en y/o cerca del tipo de célula o molécula específicas pueden ser etiquetadas de manera estocástica. Una propiedad de afinidad también puede proporcionar información espacial además de la secuencia de nucleótidos de la marca espacial porque el anticuerpo puede guiar el código de barras estocástico a una ubicación específica. Un anticuerpo puede ser un anticuerpo terapéutico. Un anticuerpo puede ser un anticuerpo monoclonal. Un anticuerpo puede ser un anticuerpo policlonal. Un anticuerpo puede ser humanizado. Un anticuerpo puede ser quimérico. Un anticuerpo puede ser un anticuerpo desnudo. Un anticuerpo puede ser un anticuerpo de fusión.

**[0119]** Un anticuerpo, puede referirse a una molécula de inmunoglobulina (por ejemplo, un anticuerpo IgG) o una inmunoglobulina activa (es decir, específicamente unida a la porción de una molécula de inmunoglobulina, como un fragmento de anticuerpo).

**[0120]** Un anticuerpo puede ser un fragmento de anticuerpo. Un fragmento de anticuerpo puede ser una porción de un anticuerpo tal como F(ab')<sub>2</sub>, Fab', Fab, Fv, sFv y similares. Un fragmento de anticuerpo puede unirse con el mismo antígeno que es reconocido por el anticuerpo de longitud completa. Un fragmento de anticuerpo puede incluir

fragmentos aislados que consisten en las regiones variables de los anticuerpos, tales como los fragmentos "Fv" que consisten en las regiones variables de las cadenas pesada y ligera y las moléculas de polipéptidos de cadena simple recombinante en las cuales las regiones variables ligeras y pesadas están conectadas por un conector peptídico ("proteínas scFv"). Los anticuerpos ejemplares pueden incluir, pero no se limitan a, anticuerpos para anticuerpos para células cancerosas, anticuerpos para virus, anticuerpos que se unen a los receptores de la superficie celular (CDS, CD34, CD45), y anticuerpos terapéuticos.

**[0121]** La etiqueta celular y/o cualquier etiqueta de la descripción pueden comprender además un conjunto único de sub-secuencias de ácido nucleico de longitud definida, por ejemplo, 7 nucleótidos cada uno (equivalente al número de bits utilizados en algunos códigos de corrección de error de Hamming), que están diseñados para proporcionar la capacidad de corrección de errores. El conjunto de subsecuencias de corrección de errores comprende 7 secuencias de nucleótidos que pueden diseñarse de manera tal que cualquier combinación de secuencias por pares en el conjunto muestre una "distancia genética" definida (o un número de bases no coincidentes), por ejemplo, un conjunto de subsecuencias de corrección de errores puede diseñarse para exhibir una distancia genética de 3 nucleótidos. En algunas realizaciones, la longitud de las subsecuencias de ácido nucleico utilizadas para crear códigos de corrección de errores puede variar, por ejemplo, pueden ser al menos 3 nucleótidos, al menos 7 nucleótidos, al menos 15 nucleótidos o al menos 31 nucleótidos en longitud. En algunas realizaciones, se pueden usar sub-secuencias de ácidos nucleicos de otras longitudes para crear códigos de corrección de errores.

**[0122]** Códigos de barras estocásticos de la divulgación pueden comprender secuencias de corrección de errores (por ejemplo, códigos de Hamming) en ellos para corrección de errores. Un código de Hamming puede referir un proceso aritmético que identifica códigos binarios únicos basados en redundancia inherente que son capaces de corregir errores de un solo bit. Por ejemplo, un código de Hamming puede combinarse con un código de barras de ácido nucleico para detectar errores de un solo nucleótido que ocurren durante la amplificación del ácido nucleico. La identificación de un solo error de nucleótidos mediante el uso de un código de Hamming permite, por lo tanto, la corrección del código de barras de ácido nucleico.

**[0123]** Cuando un código de barras estocástico comprende más de uno de un tipo de etiqueta (p. ej., más de una etiqueta celular o más de una etiqueta molecular), las etiquetas pueden estar intercaladas con una secuencia de etiqueta enlazadora. Una secuencia de marcador de enlace puede tener al menos aproximadamente 5, 10, 15, 20, 25, 30, 35, 40, 45, 50 o más nucleótidos de longitud. Una secuencia de etiqueta enlazadora puede tener a lo sumo aproximadamente 5, 10, 15, 20, 25, 30, 35, 40, 45, 50 o más nucleótidos de longitud. En algunos casos, una secuencia de etiqueta de enlace tiene una longitud de 12 nucleótidos. Se puede usar una secuencia de etiqueta de enlace para facilitar la síntesis del código de barras estocástico. La etiqueta del enlazador puede comprender un código de corrección de errores (por ejemplo, Hamming).

#### Soportes sólidos

**[0124]** Los códigos de barras estocásticos descritos en este documento pueden estar unidos a un soporte sólido (por ejemplo, perla, sustrato). Como se usa en este documento, los términos "atado", "unido" e "inmovilizado" se usan de manera intercambiable, y pueden referirse a medios covalentes o no covalentes para unir códigos de barras estocásticos a un soporte sólido. Se puede usar cualquiera de una variedad de soportes sólidos diferentes como soportes sólidos para unir códigos de barras estocásticos pre-sintetizados o para la síntesis *in situ* en fase sólida de códigos de barras estocásticos.

**[0125]** En algunos casos, un soporte sólido es una perla. Una perla puede abarcar cualquier tipo de esfera, bola, cojinete, cilindro u otra configuración similar sólida, porosa o hueca compuesta de plástico, cerámica, metal o material polimérico sobre el que se puede inmovilizar un ácido nucleico (por ejemplo, covalentemente o no covalentemente). Una perla puede comprender una partícula discreta que puede ser esférica (p. ej., microesferas) o tener una forma no esférica o irregular, tal como cúbica, cuboide, piramidal, cilíndrica, cónica, oblongada o en forma de disco, y similares. Una perla puede ser de forma no esférica.

**[0126]** Las perlas pueden comprender una variedad de materiales que incluyen, entre otros, materiales paramagnéticos (por ejemplo, magnesio, molibdeno, litio y tantalio), materiales superparamagnéticos (por ejemplo, ferrita (Fe<sub>3</sub>O<sub>4</sub>; magnetita) nanopartículas), materiales ferromagnéticos (por ejemplo, hierro, níquel, cobalto, algunas aleaciones de los mismos y algunos compuestos de metales de tierras raras), cerámica, plástico, vidrio, poliestireno, sílice, metilmetacrilato, polímeros acrílicos, titanio, látex, sefrosa, agarosa, hidrogel, polímero, celulosa, nailon y cualquier combinación de los mismos.

**[0127]** El diámetro de las perlas puede variar, por ejemplo, puede ser al menos aproximadamente 5 μM, 10 μM, 20 μM, 25 μM, 30 μM, 35 μM, 40 μM, 45 μM o 50 μM. El diámetro de las perlas puede ser como máximo de aproximadamente 5 μM, 10 μM, 20 μM, 25 μM, 30 μM, 35 μM, 40 μM, 45 μM o 50 μM. El diámetro de la perla puede estar relacionado con el diámetro de los pocillos del sustrato. Por ejemplo, el diámetro de la perla puede ser al menos 10, 20, 30, 40, 50, 60, 70, 80, 90 o 100% más largo o más corto que el diámetro del pozo. El diámetro de la perla puede ser a lo sumo 10, 20, 30, 40, 50, 60, 70, 80, 90 o 100% más largo o más corto que el diámetro del pozo. El diámetro de la perla puede estar relacionado con el diámetro de una célula (por ejemplo, una sola célula atrapada por



un pocillo del sustrato). El diámetro de la perla puede ser al menos 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250 o 300% o más más o más corto que el diámetro de la célula. El diámetro de la perla puede ser a lo sumo 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250 o 300% o más, más largo o más corto que el diámetro de la célula.

5 **[0128]** Se puede unir y/o incrustar una perla en un sustrato de la divulgación. Una perla se puede unir y/o incrustar en un gel, hidrogel, polímero y/o matriz. La posición espacial de una cuenta dentro de un sustrato (p. ej., gel, matriz, andamio o polímero) puede identificarse utilizando la etiqueta espacial presente en el código de barras estocástico en la perla que puede servir como una dirección de ubicación.

10 **[0129]** Los ejemplos de perlas pueden incluir, pero no se limitan a perlas de estreptavidina, perlas de agarosa, perlas magnéticas, Dynabeads®, microperlas MACS®, perlas conjugadas con anticuerpos (por ejemplo, micro-bolas anti-inmunoglobulina), perlas conjugadas a proteína A, perlas conjuntadas de proteína G, perlas conjugadas de proteína A/G, perlas conjugadas de proteína L, perlas conjugadas de oligodT, perlas de sílice, perlas similares a la sílice, microperlas anti-biotina, microperlas anti-fluorocromos y perlas magnéticas terminadas por carboxi BcMag™.

15 **[0130]** Una perla puede estar asociada con (por ejemplo, impregnada con) puntos cuánticos o tintes fluorescentes para hacerla fluorescente en un canal óptico de fluorescencia o canales ópticos múltiples. Una perla puede asociarse con óxido de hierro u óxido de cromo para que sea paramagnética o ferromagnética. Las perlas pueden ser identificables. Una perla puede ser fotografiada usando una cámara. Una perla puede tener un código detectable asociado con la perla. Por ejemplo, una perla puede comprender una etiqueta RFID. Una perla puede comprender cualquier etiqueta detectable (por ejemplo, código UPC, código de barras electrónico, identificador grabado). Una perla puede cambiar de tamaño, por ejemplo debido a la hinchazón en una solución orgánica o inorgánica. Una perla puede ser hidrofóbica. Una perla puede ser hidrófila. Una perla puede ser biocompatible.

20 **[0131]** Un soporte sólido (por ejemplo, perla) puede ser visualizado. El soporte sólido puede comprender una etiqueta de visualización (por ejemplo, tinte fluorescente). Un soporte sólido (por ejemplo, una perla) puede grabarse con un identificador (por ejemplo, un número). El identificador se puede visualizar a través de imágenes de los soportes sólidos (por ejemplo, perlas).

25 **[0132]** Un soporte sólido puede referirse a un material insoluble, semisoluble o insoluble. Un soporte sólido puede denominarse "funcionalizado" cuando incluye un enlazador, un andamio, un bloque de construcción u otro resto reactivo unido a él, mientras que un soporte sólido puede estar "no funcionalizado" cuando carece de dicho reactivo. resto adjunto al mismo. El soporte sólido puede emplearse libre en solución, tal como en un formato de pocillo de microtitulación; en un formato de flujo continuo, como en una columna; o en una varilla.

30 **[0133]** El soporte sólido puede comprender una membrana, papel, plástico, superficie recubierta, superficie plana, vidrio, portaobjetos, chips, o cualquier combinación de los mismos. Un soporte sólido puede tomar la forma de resinas, geles, microesferas u otras configuraciones geométricas. Un soporte sólido puede incluir chips de silicón, micropartículas, nanopartículas, placas, matrices, capilares, soportes planos como filtros de fibra de vidrio, superficies de vidrio, superficies metálicas (acero, plata, aluminio, silicio y cobre), soportes de vidrio, plástico soportes, siluetas, chips, filtros, membranas, placas de micropocillos, portaobjetos, materiales plásticos que incluyen placas o membranas de múltiples pocillos (por ejemplo, formadas de polietileno, polipropileno, poliamida, polivinilidendifluoruro) y/o obleas, peines, alfileres o agujas (por ejemplo, conjuntos de pines adecuados para la síntesis o el análisis combinatorios) o perlas en un conjunto de pozos o pocillos de nanolitros de superficies planas tales como obleas (por ejemplo, obleas de silicio), obleas con hoyos con o sin fondo de filtro.

35 **[0134]** El soporte sólido puede comprender una matriz de polímero (por ejemplo, gel, hidrogel). La matriz polimérica puede ser capaz de penetrar en el espacio intracelular (por ejemplo, alrededor de orgánulos). La matriz de polímero puede ser bombeada a través del sistema circulatorio.

40 **[0135]** Un soporte sólido puede ser, en algunas realizaciones, una molécula biológica. Por ejemplo, un soporte sólido puede ser un ácido nucleico, una proteína, un anticuerpo, una histona, un compartimento celular, un lípido, un carbohidrato y similares. Los soportes sólidos que son moléculas biológicas pueden ser amplificados, traducidos, transcritos, degradados y/o modificados (por ejemplo, pegilados, sumoyilados, acetilados, metilados). Un soporte sólido que es una molécula biológica puede proporcionar información espacial y de tiempo además de la etiqueta espacial que está unida a la molécula biológica. Por ejemplo, una molécula biológica puede comprender una primera confirmación cuando no está modificada, pero puede cambiar a una segunda confirmación cuando se modifica. Las diferentes conformaciones pueden exponer a las dianas los códigos de barras estocásticos de la divulgación. Por ejemplo, una molécula biológica puede comprender códigos de barras estocásticos que son inaccesibles debido al plegamiento de la molécula biológica. Tras la modificación de la molécula biológica (p. ej., acetilación), la molécula biológica puede cambiar la conformación para exponer las etiquetas estocásticas. El momento de la modificación puede proporcionar otra dimensión temporal al método de codificación de barras estocástica de la divulgación.

45 **[0136]** En otro ejemplo, la molécula biológica que comprende códigos de barras estocásticos de la divulgación puede ser localizada en el citoplasma de una célula. Tras la activación, la molécula biológica puede moverse al núcleo, con lo cual puede tener lugar el código de barras estocástico. De esta manera, la modificación de la molécula biológica

puede codificar información adicional de espacio-tiempo para las dianas identificadas por los códigos de barras estocásticos.

**[0137]** Una etiqueta de dimensión puede proporcionar información sobre el espacio-tiempo de un evento biológico (por ejemplo, división celular). Por ejemplo, se puede agregar una etiqueta de dimensión a una primera célula, la primera célula puede dividirse generando una segunda célula hija, la segunda célula hija puede comprender todas, algunas o ninguna de las etiquetas de dimensión. Las etiquetas de dimensión se pueden activar en la célula original y en la célula hija. De esta manera, la etiqueta de dimensión puede proporcionar información sobre el tiempo del código de barras estocástico en espacios distintos.

#### Muestras

**[0138]** Como se describe en el presente documento, la pluralidad de moléculas de ácido nucleico pueden obtenerse o derivarse de una muestra, por ejemplo, una muestra de células. Una muestra para uso en el método de la divulgación puede comprender una o más células. Una muestra puede referirse a una o más células. En algunas realizaciones, las células son células cancerosas extirpadas de un tejido canceroso, por ejemplo, cáncer de mama, cáncer de pulmón, cáncer de colon, cáncer de próstata, cáncer de ovario, cáncer de páncreas, cáncer de cerebro, melanoma y cánceres de la piel no melanoma, y similares. En algunos casos, las células se derivan de un cáncer pero se recolectan de un fluido corporal (por ejemplo, las células tumorales circulantes). Los ejemplos no limitantes de cánceres pueden incluir adenoma, adenocarcinoma, carcinoma de células escamosas, carcinoma de células basales, carcinoma de células pequeñas, carcinoma de células grandes indiferenciadas, condrosarcoma y fibrosarcoma.

**[0139]** En algunas realizaciones, las células son células que han sido infectadas con el virus y contener oligonucleótidos virales. En algunas realizaciones, la infección viral puede ser causada por un virus seleccionado del grupo que consiste en virus de ADN de doble cadena (por ejemplo, adenovirus, virus de herpes, virus de la viruela), virus de ADN de cadena sencilla (+ cadena o "sentido") (por ejemplo, parvovirus), virus de ARN de doble cadena (por ejemplo, reovirus), virus de ARN de cadena sencilla (+ cadena o sentido) (por ejemplo, picornavirus, togavirus), virus de ARN de cadena simple (- cadena o antisentido) (por ejemplo, ortomixovirus, rabdovirus), virus de ARN de cadena simple ((+ cadena o sentido) con un ADN intermedio en su ciclo de vida) virus de ARN-RT (por ejemplo, virus de retrovirus) y virus de doble cadena de ADN-RT (por ejemplo, hepadnavirus). Los virus ejemplares pueden incluir, entre otros, SARS, VIH, coronavirus, Ébola, Malaria, Dengue, Hepatitis C, Hepatitis B e Influenza.

**[0140]** En algunas realizaciones, las células son bacterias. Estas pueden incluir bacterias grampositivas o gramnegativas. Los ejemplos de bacterias que pueden analizarse utilizando los métodos, dispositivos y sistemas descritos incluyen, entre otros, *Actinomeduræ*, *Actinomyces israelii*, *Bacillus anthracis*, *Bacillus cereus*, *Clostridium botulinum*, *Clostridium difficile*, *Clostridium perfringens*, *Clostridium tetani*, *Corynebacterium*, *Enterococcus faecalis*, *Listeria monocytogenes*, *Nocardia*, *Propionibacterium acnes*, *Staphylococcus aureus*, *Staphylococcus epiderm*, *Streptococcus mutans*, *Streptococcus pneumoniae* y similares. Las bacterias gram negativas incluyen, pero no se limitan a, *Afipia felis*, *Bacteriodes*, *Bartonella bacilliformis*, *Bordetella pertussis*, *Borrelia burgdorferi*, *Borrelia recurrentis*, *Bruceella*, *Calymmatobacterium granulomatis*, *Campylobacter*, *Escherichia coli*, *Francisella tularensis*, *vaginalis*, *Haemophilus aegyptius*, *Haemophilus ducreyi*, *Haemophilus influenzae*, *Heliobacter pylori*, *Legionella pneumophila*, *Leptospira interrogans*, *Neisseria meningitidis*, *Porphyromonas gingivalis*, *Providencia stuarti*, *Pseudomonas aeruginosa*, *Salmonella enteridis*, *Salmonella typhi*, *Serratia marcescens*, *Shigella boydii*, *Streptobacillus moniliformis*, *Streptococcus pyogenes*, *Treponema pallidum*, *Vibrio cholerae*, *Yersinia enterocolitica*, *Yersinia pestis* y similares. Otras bacterias pueden incluir *Myobacterium avium*, *Myobacterium leprae*, *Myobacterium tuberculosis*, *Bartonella henselae*, *Chlamydia psittaci*, *Chlamydia trachomatis*, *Coxiella burnetii*, *Mycoplasma pneumoniae*, *Rickettsia akari*, *Rickettsia prowazekii*, *Rickettsia rickettsii*, *Rickettsia tsutsugamushi*, *Rickettsia typhi*, *Ureaplasma urealyticum*, *Diplococcus pneumoniae*, *Ehrlichia chaffeensis*, *Enterococcus faecium*, *Meningococci* y similares.

**[0141]** En algunas realizaciones, las células son hongos. Los ejemplos no limitativos de hongos que pueden analizarse usando los métodos, dispositivos y sistemas divulgados incluyen, entre otros, *Aspergilli*, *Candidae*, *Candida albicans*, *Coccidioides immitis*, *Cryptococci* y combinaciones de los mismos.

**[0142]** En algunas realizaciones, las células son protozoos u otros parásitos. Los ejemplos de parásitos que se analizarán utilizando los métodos, dispositivos y sistemas de la presente divulgación incluyen, pero no se limitan a, *Balantidium coli*, *Cryptosporidium parvum*, *Cyclospora cayatanensis*, *Encefalitozoa*, *Entamoeba histolytica*, *Enterocytozoon bieneusi*, *Giardia lamblia*, *Leishmaniae*, *Plasmodii*, *Toxo plasma gondii*, *Tripanosoma*, *ameba trapezoidal*, *gusanos* (por ejemplo, helmintos), en particular gusanos parásitos que incluyen, entre otros, *Nematoda* (gusanos redondos, por ejemplo, tricocéfalos, anquilostomas, oxiruos, ascáridos, filáridos y similares), *Cestoda* (por ejemplo, tenias).

**[0143]** Tal como se utiliza aquí, el término "célula" puede referirse a una o más células. En algunas realizaciones, las células son células normales, por ejemplo, células humanas en diferentes etapas de desarrollo, o células humanas de diferentes órganos o tipos de tejidos (por ejemplo, glóbulos blancos, glóbulos rojos, plaquetas, células epiteliales, células endoteliales, neuronas, células gliales, fibroblastos, células del músculo esquelético, células del músculo liso, gametos o células del corazón, pulmones, cerebro, hígado, riñón, bazo, páncreas, timo, vejiga, estómago, colon,

intestino delgado). En algunas realizaciones, las células pueden ser células madre humanas indiferenciadas, o células madre humanas que se han inducido a diferenciarse. En algunas realizaciones, las células pueden ser células fetales humanas. Las células humanas del feto pueden obtenerse de una madre embarazada del feto. En algunas realizaciones, las células son células raras. Una célula rara puede ser, por ejemplo, una célula tumoral circulante (CTC), una célula epitelial circulante, una célula endotelial circulante, una célula endometrial circulante, una célula madre circulante, una célula madre, una célula madre indiferenciada, una célula madre cancerosa, una célula médula ósea, una célula progenitora, célula de espuma, célula mesenquimática, trofoblasto, célula del sistema inmunitario (huésped o injerto), fragmento celular, orgánulo celular (por ejemplo, mitocondrias o núcleos), célula infectada con patógenos y similares.

**[0144]** En algunas realizaciones, las células son células no humanas, por ejemplo, otros tipos de células de mamíferos (por ejemplo, ratón, rata, cerdo, perro, vaca, o caballo). En algunas realizaciones, las células son otros tipos de células animales o vegetales. En otras realizaciones, las células pueden ser cualquier célula procariota o eucariota.

**[0145]** En algunas realizaciones, una primera muestra de células se obtiene de una persona que no tenga una enfermedad o condición, y una segunda muestra de células se obtiene de una persona que tiene la enfermedad o condición. En algunas realizaciones, las personas son diferentes. En algunas realizaciones, las personas son iguales pero las muestras de células se toman en diferentes puntos de tiempo. En algunas realizaciones, las personas son pacientes, y las muestras de células son muestras de pacientes. La enfermedad o afección puede ser un cáncer, una infección bacteriana, una infección viral, una enfermedad inflamatoria, una enfermedad neurodegenerativa, una enfermedad fúngica, una enfermedad parasitaria, un trastorno genético o cualquier combinación de estos.

**[0146]** En algunas realizaciones, las células adecuadas para uso en los métodos actualmente descritos pueden variar en tamaño desde aproximadamente 2 micrómetros hasta aproximadamente 100 micrómetros de diámetro. En algunas realizaciones, las células pueden tener diámetros de al menos 2 micrómetros, al menos 5 micrómetros, al menos 10 micrómetros, al menos 15 micrómetros, al menos 20 micrómetros, al menos 30 micrómetros, al menos 40 micrómetros, al menos 50 micrómetros, al menos 60 micrómetros, al menos 70 micrómetros, al menos 80 micrómetros, al menos 90 micrómetros o al menos 100 micrómetros. En algunas realizaciones, las células pueden tener diámetros de como máximo 100 micrómetros, a lo sumo 90 micrómetros, a lo sumo 80 micrómetros, a lo sumo 70 micrómetros, a lo sumo 60 micrómetros, a lo sumo 50 micrómetros, a lo sumo 40 micrómetros, a lo sumo 30 micrómetros, a lo sumo 20 micrómetros, a lo sumo 15 micrómetros, a lo sumo 10 micrómetros, a lo sumo 5 micrómetros o a lo sumo 2 micrómetros. Las células pueden tener un diámetro de cualquier valor dentro de un rango, por ejemplo de aproximadamente 5 micrómetros a aproximadamente 85 micrómetros. En algunas realizaciones, las células tienen diámetros de aproximadamente 10 micrómetros.

**[0147]** En algunas realizaciones, las células son ordenadas antes de la asociación de una célula con una perla y/o en un micropocillo. Por ejemplo, las células se pueden clasificar por clasificación de células activadas por fluorescencia o clasificación de células activadas por magnetismo, o por ejemplo, por citometría de flujo. Las células pueden ser filtradas por tamaño. En algunos casos, un retenido contiene las células que se asociarán con la perla. En algunos casos, el flujo continuo contiene las células que se asociarán con la perla.

**[0148]** En algunas realizaciones, la muestra comprende una célula inmune. Una célula inmunitaria puede incluir, por ejemplo, células T, células B, células madre linfoides, células progenitoras mieloides, linfocitos, granulocitos, células progenitoras de células B, células progenitoras de células T, células asesinas naturales, células Tc, células Thc, células Th, células plasmáticas, células de memoria, neutrófilos, eosinófilos, basófilos, mastocitos, monocitos, células dendríticas y/o macrófagos, o cualquier combinación de los mismos.

**[0149]** La célula AT puede ser un clon de células T, que puede referirse a células T derivadas de una sola célula T o aquellas que tienen TCR idénticas. Las células AT pueden formar parte de una línea de células T que puede incluir clones de células T y poblaciones mixtas de células T con diferentes TCR, todas las cuales pueden reconocer la misma diana (por ejemplo, antígeno, tumor, virus). Las células T pueden obtenerse de varias fuentes, incluidas las células mononucleares de sangre periférica, la médula ósea, el tejido de los ganglios linfáticos, el tejido del bazo y los tumores. Las células T se pueden obtener de una unidad de sangre extraída de un sujeto, como el uso de la separación Ficoll. Las células de la sangre circulante de un individuo pueden obtenerse mediante aféresis o leucaféresis. El producto de la aféresis puede comprender linfocitos, incluidos linfocitos T, monocitos, granulocitos, linfocitos B, otros glóbulos blancos nucleados, glóbulos rojos y plaquetas. Las células pueden lavarse y resuspenderse en medios para aislar la célula de interés.

**[0150]** Las células T se pueden aislar a partir de linfocitos de sangre periférica mediante la lisis de las células rojas de la sangre y agotando los monocitos, por ejemplo, por centrifugación a través de un gradiente PERCOLL™. Una subpoblación específica de células T, como las células T CD28+, CD4+, CDC, CD45RA + y CD45RO+, se puede aislar aún más mediante técnicas de selección positiva o negativa. Por ejemplo, las células T pueden aislarse mediante incubación con perlas conjugadas con anti-CD3/anti-CD28 (es decir, 3 X 28), como DYNABEADS® M-450 CD3/CD28 T, o XCYTE DYNABEADS™ durante un período de tiempo suficiente para la selección positiva de las células T deseadas. Las células inmunitarias (p. ej., células T y células B) pueden ser específicas de antígeno (p. ej., específicas para un tumor).

**[0151]** En algunas realizaciones, la célula puede ser una célula presentadora de antígeno (APC), como una célula B, una célula B activada de un ganglio linfático, una célula linfoblastoide, una célula B en reposo o una célula B neoplásica, por ejemplo, de un linfoma. Una APC puede referirse a una célula B o una célula dendrítica folicular que expresan al menos una de las proteínas BCRC en su superficie.

#### Métodos de codificación de barras estocástica y normalización de bibliotecas

**[0152]** La descripción proporciona métodos para la normalización de la biblioteca de una muestra. Los métodos de normalización de bibliotecas se pueden combinar con cualquier método de preparación de muestras de bibliotecas. Los métodos de normalización de bibliotecas se pueden combinar con los métodos de codificación de barras estocásticas de la divulgación. Los códigos de barras estocásticos se pueden usar para indexar moléculas de ácido nucleico individuales (p. ej., moléculas de ARNm) con códigos de barras únicos, de modo que las moléculas de dianas específicas puedan rastrearse y/o contarse.

**[0153]** Los métodos de códigos de barras estocásticos pueden comprender colocación de los códigos de barras estocásticos en estrecha proximidad con la muestra, la lisis de la muestra, asociando dianas distintas con los códigos de barras estocásticos, amplificando las dianas y/o digitalmente contando las dianas. **La Figura 2** ilustra una realización ejemplar del método de codificación de barras estocástica de la divulgación. Una muestra (p. ej., sección de una muestra, corte fino y/o célula) puede ponerse en contacto con un soporte sólido que comprende un código de barras estocástico. Las dianas en la muestra pueden asociarse con los códigos de barras estocásticos. Los soportes sólidos se pueden recoger. La síntesis de ADNc se puede realizar en el soporte sólido. La síntesis de ADNc se puede realizar fuera del soporte sólido. La síntesis de ADNc puede incorporar la información de la etiqueta de las etiquetas en el código de barras estocástico en la nueva molécula diana de ADNc que se está sintetizando, generando así una molécula de código de barras diana. Las moléculas de código de barras diana pueden amplificarse utilizando PCR. La secuencia de las dianas y las etiquetas del código de barras estocástico en la molécula de código de barras diana puede determinarse mediante métodos de secuenciación.

#### Contacto con una muestra y un código de barras estocástico

**[0154]** Una muestra que comprende, por ejemplo, una célula, órgano o sección delgada de tejido, puede contactarse a códigos de barra estocásticos. Los soportes sólidos pueden ser flotantes. Los soportes sólidos se pueden incrustar en una matriz semisólida o sólida. Los códigos de barras estocásticos pueden no estar asociados con soportes sólidos. Los códigos de barras estocásticos pueden ser nucleótidos individuales. Los códigos de barras estocásticos pueden estar asociados con un sustrato. Cuando los códigos de barras estocásticos están cerca de las dianas, las dianas pueden hibridar con el código de barras estocástico. Los códigos de barras estocásticos pueden ser contactados en una proporción no agotable de tal manera que cada diana distinta pueda asociarse con un código de barras estocástico distinto de la divulgación. Para asegurar una asociación eficiente entre la diana y el código de barras estocástico, las dianas pueden estar entrecruzadas con el código de barras estocástico.

**[0155]** La probabilidad de que dos dianas distintos de una muestra puedan entrar en contacto con el mismo código de barras estocástico único puede ser al menos  $10^{-6}$ ,  $10^{-5}$ ,  $10^{-4}$ ,  $10^{-3}$ ,  $10^{-2}$  o  $10^{-1}$  o más. La probabilidad de que dos dianas distintas de una muestra puedan entrar en contacto con el mismo código de barras estocástico único puede ser a lo sumo  $10^{-6}$ ,  $10^{-5}$ ,  $10^{-4}$ ,  $10^{-3}$ ,  $10^{-2}$  o  $10^{-1}$  o más. La probabilidad de que dos dianas del mismo gen de la misma célula puedan entrar en contacto con el mismo código de barras estocástico puede tener al menos  $10^{-6}$ ,  $10^{-5}$ ,  $10^{-4}$ ,  $10^{-3}$ ,  $10^{-2}$  o  $10^{-1}$  o más. La probabilidad de que dos dianas del mismo gen de la misma célula puedan entrar en contacto con el mismo código de barras estocástico puede ser a lo sumo  $10^{-6}$ ,  $10^{-5}$ ,  $10^{-4}$ ,  $10^{-3}$ ,  $10^{-2}$  o  $10^{-1}$  o más.

**[0156]** En algunos casos, las células de una población de células se pueden separar (por ejemplo, aislar) en pocillos de un sustrato de la descripción. La población de células se puede diluir antes de separar. La población de células se puede diluir de tal manera que al menos 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95 o el 100% de los pocillos del sustrato reciben una sola célula. La población de células se puede diluir de modo que a lo sumo 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95 o el 100% de los pocillos del sustrato reciben una sola célula. La población de células se puede diluir de tal manera que el número de células en la población diluida sea al menos 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95 o 100% del número de pocillos en el sustrato. La población de células se puede diluir de tal manera que el número de células en la población diluida sea al menos 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95 o 100% del número de pocillos en el sustrato. En algunos casos, la población de células se diluye de tal manera que el número de células es aproximadamente el 10% del número de pocillos en el sustrato.

**[0157]** La distribución de células individuales en los pocillos del sustrato puede seguir una distribución de Poisson. Por ejemplo, hay una combinación de 10% de probabilidad de que un pocillo del sustrato tenga más de una célula. Puede haber al menos un 0,1, 0,5, 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10% o más de probabilidad de que un pocillo del sustrato tenga más de una célula. La distribución de células individuales en los pocillos del sustrato puede ser aleatoria. La distribución de células individuales en los pocillos del sustrato puede ser no aleatoria. Las células se pueden separar de manera que un pocillo del sustrato reciba solo una célula.

Lisis celular

5 **[0158]** Siguiendo la distribución de las células y los códigos de barras estocásticos, las células se pueden lisar para liberar las moléculas diana. La lisis celular puede realizarse por cualquiera de una variedad de medios, por ejemplo, por medios químicos o bioquímicos, por choque osmótico o por lisis térmica, lisis mecánica o lisis óptica. Las células se pueden lisar mediante la adición de un tampón de lisis celular que comprende un detergente (por ejemplo, SDS, dodecil sulfato de Li, Triton X-100, Tween-20 o NP-40), un disolvente orgánico (por ejemplo, metanol o acetona) o enzimas digestivas (p. ej., proteinasa K, pepsina o tripsina), o cualquier combinación de las mismas. Para aumentar la asociación de una diana y un código de barras estocástico, la velocidad de difusión de las moléculas diana se puede alterar, por ejemplo, reduciendo la temperatura y/o aumentando la viscosidad del lisado.

Fijación de códigos de barras estocásticos a moléculas de ácido nucleico diana

15 **[0159]** Después de la lisis de las células y la liberación de moléculas de ácido nucleico de la misma, las moléculas de ácido nucleico pueden asociarse al azar con los códigos de barras estocásticos del soporte sólido co-localizado. La asociación puede comprender la hibridación de la región de reconocimiento de la diana de un código de barras estocástico con una porción complementaria de la molécula de ácido nucleico diana (por ejemplo, el oligo dT del código de barras estocástico puede interactuar con una cola poli-A de una diana). Las condiciones de ensayo utilizadas para la hibridación (p. ej., pH del tampón, fuerza iónica, temperatura, etc.) se pueden elegir para promover la formación de híbridos específicos y estables.

20 **[0160]** La unión puede comprender además la ligadura de una región de reconocimiento de diana del código de barras estocástico 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 una hibridación específica con un saliente del 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 del sitio de restricción. El código de barras estocástico se puede ligar a cualquier molécula de ácido nucleico que comprenda una secuencia complementaria al saliente del sitio de restricción. Se puede usar una ligasa (por ejemplo, T4 ADN ligasa) para unir los dos fragmentos.

25 **[0161]** Las dianas marcadas de una pluralidad de células (o una pluralidad de muestras) (por ejemplo, las moléculas diana de código de barras) pueden agruparse posteriormente, por ejemplo mediante la recuperación de los códigos de barras estocásticos y/o las perlas a las que están adjuntas las moléculas diana de código de barras. La recuperación de colecciones sólidas basadas en soporte de moléculas de códigos de barras diana adjuntas puede implementarse mediante el uso de perlas magnéticas y un campo magnético aplicado externamente. Una vez que las moléculas de código de barras diana se han agrupado, todo el procesamiento adicional puede proceder en un solo recipiente de reacción. El procesamiento adicional puede incluir, por ejemplo, reacciones de transcripción inversa, reacciones de amplificación, reacciones de escisión, reacciones de disociación y/o reacciones de extensión de ácido nucleico. Se pueden realizar reacciones de procesamiento adicionales dentro de los micropocillos, es decir, sin primero agrupar las moléculas de ácido nucleico diana marcadas de una pluralidad de células.

Transcripción inversa

35 **[0162]** La descripción proporciona un método para crear un conjugado diana-código de barras estocástico mediante transcripción inversa. El conjugado de código de barras diana-estocástico puede comprender el código de barras estocástico y una secuencia complementaria de todo o una parte del ácido nucleico diana (es decir, una molécula de ADNc con código de barras estocástico). La transcripción inversa de la molécula de ARN asociada puede ocurrir mediante la adición de un cebador de transcripción inversa junto con la transcriptasa inversa. El cebador de transcripción inversa puede ser un cebador de oligo-dT, un cebador de hexanucleótido aleatorio o un cebador de oligonucleótido específico de la diana. Los cebadores oligo-dT pueden tener, por ejemplo, 12-18 nucleótidos de longitud y se unen a la cola 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 oligonucleotídicos específicos de la diana típicamente ceban selectivamente el ARNm de interés.

40 **[0163]** La molécula de ADNc con código de barras estocástico puede ser sometida a métodos de aguas abajo tales como amplificación (por ejemplo, por cebadores universales y/o específicos de genes) y los métodos de normalización de la biblioteca de la descripción.

Kits

45 **[0164]** En este documento se describen kits para realizar los métodos de normalización de bibliotecas de la divulgación. Un kit puede comprender un cebador de síntesis de segunda hebra que comprende un resto de unión. Un kit puede comprender un soporte sólido que comprende restos de captura que pueden unirse al fragmento de unión en el cebador de síntesis de la segunda cadena. Un kit puede componer un imán para capturar el soporte sólido. Un kit puede comprender reactivos para limpiar una reacción de amplificación (p. ej., perlas AmpureXP y/o una columna

de purificación). Un kit puede comprender adaptadores y/o cebadores que comprenden secuencias de células de flujo de secuenciación. El kit puede comprender además reactivos (por ejemplo, enzimas, cebadores, dNTP, NTP, inhibidores de la ARNasa o tampones) para realizar reacciones de extensión de ácido nucleico, por ejemplo, reacciones de transcripción inversa y reacciones de extensión de cebador. El kit puede comprender además reactivos (por ejemplo, enzimas, cebadores universales, cebadores de secuenciación, cebadores específicos de diana o tampones) para realizar reacciones de amplificación para preparar bibliotecas de secuenciación.

**[0165]** En este documento se describen kits para realizar ensayos estocásticos de códigos de barras. Los kits pueden comprender una o más suspensiones de soporte sólido, en donde los soportes sólidos individuales dentro de una suspensión comprenden una pluralidad de códigos de barras estocásticos adjuntos de la divulgación. Los kits pueden comprender códigos de barras estocásticos que pueden no estar unidos a un soporte sólido. El kit puede comprender además reactivos, por ejemplo, tampones de lisis, tampones de enjuague o tampones de hibridación, para realizar el ensayo de código de barras estocástico. El kit puede comprender además reactivos (por ejemplo, enzimas, cebadores, dNTP, NTP, inhibidores de la ARNasa o tampones) para realizar reacciones de extensión de ácido nucleico, por ejemplo, reacciones de transcripción inversa y reacciones de extensión del cebador. El kit puede comprender además reactivos (por ejemplo, enzimas, cebadores universales, cebadores de secuenciación, cebadores específicos de diana o tampones) para realizar reacciones de amplificación para preparar bibliotecas de secuenciación.

**[0166]** Los kits de la divulgación pueden incluir generalmente instrucciones para llevar a cabo uno o más de los métodos descritos en el presente documento. Las instrucciones incluidas en los kits se pueden adjuntar al material de embalaje o se pueden incluir como un inserto para el embalaje. Si bien las instrucciones suelen ser materiales escritos o impresos, no se limitan a los mismos. Cualquier medio capaz de almacenar dichas instrucciones y comunicárselas a un usuario final está contemplado por el cierre. Dichos medios pueden incluir, entre otros, medios de almacenamiento electrónico (por ejemplo, discos magnéticos, cintas, cartuchos, chips), medios ópticos (por ejemplo, CD ROM), etiquetas de RF y similares. Como se usa en el presente documento, el término "instrucciones" puede incluir la dirección de un sitio de Internet que proporciona las instrucciones.

**[0167]** Aunque las realizaciones preferidas de la presente invención se han mostrado y descrito en el presente documento, será obvio para los expertos en la técnica que tales realizaciones se proporcionan solo a modo de ejemplo. A los expertos en la técnica se les ocurrirán numerosas variaciones, cambios y sustituciones sin apartarse de la invención. Debe entenderse que diversas prácticas a las realizaciones de la invención descritas en el presente documento pueden emplearse en la práctica de la invención. Se pretende que las siguientes reivindicaciones definan el alcance de la invención y que los métodos y estructuras dentro del alcance de estas reivindicaciones y sus equivalentes se cubran de ese modo.

## EJEMPLOS

### Ejemplo 1: Normalización de bibliotecas con PCR anidada

**[0168]** Este ejemplo proporciona métodos para la normalización de bibliotecas con PCR anidada. Una pluralidad de ARNm se transcribe a la inversa en una serie de ADNc utilizando un cebador que comprende una secuencia oligo dT, una etiqueta molecular, una etiqueta de muestra y una etiqueta universal. Los ADNc se amplifican en una primera reacción de amplificación utilizando un cebador inverso específico del gen y un cebador universal (por ejemplo, que se une al marcador universal), generando así un primer conjunto de amplicones. El cebador universal puede comprender un resto de biotina. El primer conjunto de amplicones se puede amplificar en una segunda reacción de amplificación usando una segunda PCR anidada específica del gen. cebador y el cebador universal que comprende el resto de biotina. Esta reacción genera un amplicón etiquetado asimétricamente que comprende un óxido de biotina en un extremo. La biblioteca está desnaturalizada por calor. La biblioteca se enfría para inducir un recocido parcial. Durante el recocido parcial, los amplicones altamente abundantes se recocerán más rápido que los amplicones más abundantes.

**[0169]** La biblioteca parcialmente recocida puede ponerse en contacto con un soporte sólido que comprende restos de estreptavidina. La estreptavidina puede unirse a los restos de biotina en las cadenas de la biblioteca. Los amplicones recocidos compondrán la biotina y serán eliminados por la estreptavidina. Los filamentos que no se han recocido que comprenden la biotina también serán eliminados por la estreptavidina. Las cadenas restantes representan secuencias que son más bajas en abundancia y son el complemento de las cadenas con la biotina. Estas hebras representan una biblioteca normalizada.

**[0170]** La biblioteca se regenera con cebadores de PCR. Los cebadores de PCR pueden comprender secuencias de cebadores de células de flujo de secuenciación. La biblioteca normalizada está secuenciada.

### Ejemplo 2: Normalización de bibliotecas con amplificación de transcripción completa

**[0171]** La normalización de la biblioteca se puede realizar en una biblioteca generada a partir de la amplificación del transcriptoma completo. Se puede realizar una amplificación completa del transcriptoma utilizando un método de ligadura del adaptador. Una diana comprende una cola poli-A. La diana es un ARNm. La diana está hibridada a un

código de barras estocástico. El código de barras estocástico comprende una serie de etiquetas. Por ejemplo, el código de barras estocástico comprende una región específica de la diana (por ejemplo, oligo dT para unirse a las colas poli-A de los ARNm), un marcador molecular, un marcador celular y un primer marcador universal. El código de barras estocástico se transcribe a la inversa utilizando una transcriptasa inversa, generando así una molécula de ADNc marcada. El exceso de códigos de barras estocásticos se tratan con una enzima de degradación. La enzima de degradación es una exonucleasa.

**[0172]** La molécula de ADNc marcado se somete a síntesis de la segunda cadena generando de ese modo una molécula de ADNc marcado de doble cadena. La síntesis de la segunda hebra se realiza al poner en contacto el híbrido de molécula de ADNc marcado-ARNm con una enzima de mella (por ejemplo, RNasaH) que puede hibridar el ARNm con la molécula de ADNc marcado, generando así ARNm recortado. El ARNm mellado se utiliza como cebador y se extiende utilizando una polimerasa (p. ej., ADN Pol I), incorporando así la secuencia de la primera cadena. La polimerasa comprende la actividad exonucleasa 5'-3'. La polimerasa degrada las mellas de ARNm de flujo descendente que sirven como cebadores para la síntesis de la segunda cadena. Se utiliza una ligasa para unir las secuencias extendidas, generando así una segunda hebra (p. ej., una molécula de ADNc marcado de doble cadena).

**[0173]** La molécula de ADNc de doble cadena marcado comprende una secuencia que es complementaria a la primera etiqueta universal. La molécula de ADNc marcado de doble cadena se pone en contacto con un adaptador. El adaptador es de doble cadena. El adaptador comprende un sitio de escisión de endonucleasas de restricción. El adaptador comprende una segunda secuencia de cebador universal (que es la misma que la primera). El adaptador comprende un saliente 3'. El adaptador comprende un fosfato 5' libre (P) que puede unirse al hidroxilo 3' de la molécula de ADNc marcado de doble cadena. El adaptador se une a ambas cadenas de la molécula de ADNc marcado de doble cadena 1558.

**[0174]** El producto se amplifica usando uno o más cebadores de amplificación de la WTA. Uno de los cebadores de amplificación WTA comprende un resto de biotina. Uno de los cebadores de amplificación WTA no comprende un resto de biotina. El producto se amplifica de modo que una hebra se amplifica linealmente y una hebra se amplifica exponencialmente. La cadena linealmente amplificada comprende la secuencia universal amplificable en un extremo. La cadena exponencialmente amplificable comprende secuencias universales en ambos extremos, con uno de los extremos que comprende un resto de biotina. El producto amplificado WTA se somete al protocolo de normalización de la biblioteca como se describe en el Ejemplo 1.

#### Ejemplo 3: Uso de bloqueadores en la normalización de bibliotecas

**[0175]** Este ejemplo proporciona métodos para la normalización de bibliotecas con bloqueadores. Una pluralidad de ARNm se transcribe a la inversa en una serie de ADNc utilizando un cebador que comprende una secuencia oligo dT, una etiqueta molecular, una etiqueta de muestra y una etiqueta universal. Los ADNc se amplifican en una primera reacción de amplificación utilizando un cebador inverso específico del gen y un cebador universal (por ejemplo, que se une al marcador universal), generando así un primer conjunto de amplicones. El cebador universal puede comprender un resto de biotina. El primer conjunto de amplicones se amplifica en una segunda reacción de amplificación utilizando un segundo cebador de PCR anidado específico del gen y el cebador universal que comprende el resto de biotina. Esta reacción genera un amplicón etiquetado asimétricamente que comprende un óxido de biotina en un extremo. Los bloqueadores se agregan a la biblioteca. La biblioteca está desnaturalizada por calor. Los bloqueadores se unen a las etiquetas universales de los amplicones. La biblioteca se enfría para inducir un recocido parcial. Durante el recocido parcial, los amplicones muy abundantes se recocerán más rápido que los amplicones más abundantes. El recocido parcial será impulsado más por las secuencias diana que por cualquiera de las secuencias en el cebador (p. ej., molecular, muestra, etiqueta universal).

**[0176]** La biblioteca parcialmente re-recocida se pone en contacto con un soporte sólido que comprende restos de estreptavidina. La estreptavidina puede unirse a los restos de biotina en las cadenas de la biblioteca. Los amplicones recocidos comprenderán la biotina y serán eliminados por la estreptavidina. Los filamentos que no se han recocido que comprenden la biotina también serán eliminados por la estreptavidina. Las cadenas restantes representan secuencias que son más bajas en abundancia y son el complemento de las cadenas con la biotina. Estas hebras representan una biblioteca normalizada.

**[0177]** La biblioteca se regenera con cebadores de PCR. Los cebadores de PCR pueden comprender secuencias de cebadores de células de flujo de secuenciación. La biblioteca normalizada está secuenciada.

#### Ejemplo 4: Normalización de bibliotecas en un soporte sólido

**[0178]** Este ejemplo proporciona métodos para la normalización de bibliotecas con un soporte sólido. Una pluralidad de ARNm se transcribe de forma inversa en una pluralidad de ADNc utilizando un cebador que comprende una secuencia oligo dT, una etiqueta molecular, una etiqueta de muestra y una etiqueta universal. Los ADNc se amplifican en una primera reacción de amplificación utilizando un cebador inverso específico del gen y un cebador universal (por ejemplo, que se une al marcador universal), generando así un primer conjunto de amplicones. El cebador universal puede comprender un resto azida o alquino. El primer conjunto de amplicones se puede amplificar en una segunda

reacción de amplificación utilizando un segundo cebador de PCR anidado específico del gen y el cebador universal que comprende el resto azida o alquino. Esta reacción genera un amplicón etiquetado asimétricamente que comprende un resto azida o alquino en un extremo.

5 **[0179]** La biblioteca está unida a un soporte sólido usando la química de clic. El soporte sólido comprende una molécula de complemento implicada en la química del clic. Por ejemplo, si el amplicón comprende un alquino, entonces el soporte sólido comprende una azida. La biblioteca está decorada con calor. En algunos casos, se introducen bloqueadores. La biblioteca se enfría para inducir un recocido parcial. Durante recocidos parciales, los amplicones muy abundantes se recocerán más rápido que los amplicones menos abundantes.

10 **[0180]** Los amplicones re-recocidos se unirán al soporte sólido a través de la química de clic. El soporte sólido se elimina (por ejemplo, mediante centrifugación o magnetismo). Los filamentos restantes (p. ej., que no se han recocido) no comprenderán un resto azida o alquino. Las cadenas restantes representan secuencias que son más bajas en abundancia y son el complemento de las cadenas con la azida o el alquino. Estas hebras representan una biblioteca normalizada.

15 **[0181]** La biblioteca se regenera con cebadores de PCR. Los cebadores de PCR pueden comprender secuencias de cebadores de células de flujo de secuenciación. La biblioteca normalizada está secuenciada.

20 **[0182]** En al menos algunas de las realizaciones descritas anteriormente, uno o más elementos utilizados en una forma de realización de manera intercambiable se puede utilizar en otra forma de realización a menos que dicha sustitución no es técnicamente factible. Se apreciará por los expertos en la técnica que varias otras omisiones, adiciones y modificaciones pueden realizarse con los métodos y las estructuras descritas anteriormente sin apartarse del alcance de la materia reivindicada. Todas estas modificaciones y cambios están destinados a caer dentro del alcance del tema, tal como se define en las reivindicaciones adjuntas.

25 **[0183]** Con respecto al uso de prácticamente cualquier término plural y/o singular en el presente documento, los expertos en la técnica pueden traducir del plural al singular y/o del singular al plural según sea apropiado para el contexto y/o aplicación. Las diversas permutaciones singulares/plurales se pueden establecer expresamente en este documento por el bien de la claridad.

30 **[0184]** Se entenderá por aquellos dentro de la técnica que, en general, los términos utilizados en el presente documento, y especialmente en las reivindicaciones adjuntas (p. ej., cuerpos de las reivindicaciones adjuntas) se referirán generalmente como términos "abiertos" (p. ej., el término "incluido" debe interpretarse como "incluido pero no limitado a", el término "que tiene" debe interpretarse como "tener al menos", el término "incluye" debe interpretarse como "incluye pero no se limita a," etc.). Se entenderá por aquellos dentro de la técnica que si un número específico está destinado de una reivindicación introducida, tal pretensión se indicará explícitamente en la reivindicación, y en ausencia de tal recitación esta intención no está presente. Por ejemplo, como una ayuda para la comprensión, las siguientes reivindicaciones pendientes pueden contener el uso de las frases introductorias "al menos una" y "una o más" para introducir recitaciones de reivindicaciones. Sin embargo, el uso de tales frases no debe interpretarse para dar a entender que la introducción de una recitación de reivindicación por los artículos indefinidos "un" o "una" limita cualquiera de las reivindicaciones particulares que contiene tal recitación de reivindicación introducida a realizaciones que contienen sólo una de tales recitaciones incluso cuando la misma reivindicación incluye las frases introductorias "uno o más" o "al menos uno" y los artículos indefinidos como "un" o "una" (p. ej., "un" y/o "una" deben interpretarse como "al menos uno" o "uno o más"); lo mismo es válido para el uso de artículos definidos utilizados para introducir recitaciones de reivindicaciones. Además, incluso si se recita explícitamente un número específico de una recitación de una reivindicación introducida, los expertos en la técnica reconocerán que dicha recitación debe interpretarse como que significa al menos el número recitado (p. ej., la simple recitación de "dos recitaciones," sin otros modificadores, significa al menos dos recitaciones, o dos o más recitaciones). Además, en aquellas instancias donde se utiliza una convención análoga a "al menos uno de A, B y C, etc", en general, una construcción de este tipo se entiende en el sentido de que un experto en la técnica entendería la convención (p. ej., "un sistema que tenga al menos uno de A, B y C" incluiría, entre otros, sistemas que tenga A solo, B solo, C solo, A y B juntos, A y C juntos, B y C juntos, y/o A, B y C juntos, etc.). En aquellos casos donde se utiliza una convención análoga a "al menos uno de A, B o C, etc", en general, una construcción de este tipo se entiende en el sentido de que alguien con experiencia en la técnica entendería la convención (p. ej., "un sistema que tenga al menos uno de A, B o C" incluiría, entre otros, sistemas que tenga A solo, B solo, C solo, A y B juntos, A y C juntos, B y C juntos, y/o A, B y C juntos, etc.). Los expertos en la técnica entenderán que prácticamente cualquier palabra disyuntiva y/o frase que presente dos o más términos alternativos, ya sea en la descripción, las reivindicaciones o los dibujos, debe entenderse para contemplar las posibilidades de incluir uno de los términos, cualquiera de los términos, o ambos términos. Por ejemplo, se entenderá que la frase "A o B" incluye las posibilidades de "A" o "B" o "A y B".

55 **[0185]** Además, cuando se describen características o aspectos de la divulgación en términos de grupos de Markush, los expertos en la técnica reconocerán que la divulgación también se describe de este modo en términos de cualquier miembro individual o subgrupo de miembros del grupo de Markush.

65



**[0186]** Como entenderá un experto en la técnica, para cualquiera y para todos los propósitos, tal como en términos de proporcionar una descripción escrita, todos los rangos descritos en este documento también abarcan todos los sub-rangos y combinaciones posibles de sus sub-rangos. Cualquier rango listado puede reconocerse fácilmente por describir y permitir que el mismo rango se divida en al menos mitades iguales, tercios, cuartos, quintos, décimos, etc. Como un ejemplo no limitativo, cada rango discutido aquí se puede dividir fácilmente en un tercio inferior, un tercio medio y un tercio superior, etc. Como también entenderá un experto en la técnica todo lenguaje como "hasta", "al menos", "mayor que", "menos que," y similares incluyen el número recitado y se refieren a rangos que se pueden dividir posteriormente en sub-rangos como se discutió anteriormente. Finalmente, como entenderá un experto en la técnica, un intervalo incluye cada miembro individual. Así, por ejemplo, un grupo que tiene 1-3 artículos se refiere a grupos que tienen 1, 2 o 3 artículos. Del mismo modo, un grupo que tiene 1-5 artículos se refiere a grupos que tienen 1, 2, 3, 4 o 5 artículos, y así sucesivamente.

5

10

15

20

25

30

35

40

45

50

55

60

65

## REIVINDICACIONES

1. Un método para remover especies de alta abundancia de una pluralidad de moléculas de ácido nucleico, que comprende:

5                    hibridar una pluralidad de primeros oligonucleótidos que comprenden un resto de unión con una primera pluralidad de moléculas de ácido nucleico, en donde la primera pluralidad de moléculas de ácido nucleico comprende al menos una especie de alta abundancia;  
 10                   extender la pluralidad de primeros oligonucleótidos para generar una pluralidad de cadenas complementarias de la primera pluralidad de moléculas de ácido nucleico que comprenden el resto de unión;  
 15                   desnaturalizar una pluralidad de moléculas de ácido nucleico de doble cadena que comprenden la pluralidad de cadenas complementarias de la primera pluralidad de moléculas de ácido nucleico;  
 20                   volver a recocer parcialmente la pluralidad de cadenas complementarias de la primera pluralidad de moléculas de ácido nucleico; y  
 25                   eliminar las cadenas complementarias recocidas de la primera pluralidad de moléculas de ácido nucleico mediante una molécula de captura inmovilizada en uno o más soportes sólidos para generar una segunda pluralidad de moléculas de ácido nucleico, en donde las moléculas de captura se unen específicamente a la porción de unión,  
 30                   por lo que el contenido de al menos una especie de alta abundancia en la segunda pluralidad de moléculas de ácido nucleico se reduce en comparación con el contenido de la al menos una especie de alta abundancia en la primera pluralidad de moléculas de ácido nucleico.

2. El método de la reivindicación 1, en el que:

25                   (a) el resto de unión es un grupo funcional seleccionado del grupo que consiste en biotina, estreptavidina, heparina, un aptámero, un resto de química de clic, digoxigenina, amina(s) primaria(s), carboxilina(s), hidroxilo(s), aldehído(s), cetona(s) y cualquier combinación de los mismos, opcionalmente en donde el resto de unión es biotina, opcionalmente en donde la molécula de captura es estreptavidina; y/o  
 30                   (b) que comprende además sintetizar una segunda cadena para al menos una de la pluralidad de cadenas complementarias de la primera pluralidad de moléculas de ácido nucleico para generar una o más de la pluralidad de moléculas de ácido nucleico bicatenario que comprenden la pluralidad de cadenas complementarias de la primera pluralidad de moléculas de ácido nucleico,  
 35                   opcionalmente, en el que la síntesis comprende hibridar una pluralidad de segundos oligonucleótidos a la pluralidad de cadenas complementarias de la primera pluralidad de moléculas de ácido nucleico y extiende la pluralidad de segundo oligonucleótido.  
 40                   opcionalmente, en donde la pluralidad de primeros oligonucleótidos o la pluralidad de segundos oligonucleótidos comprende un sitio de unión del cebador universal; y/o  
 45                   (c) el método comprende además amplificar la pluralidad de moléculas de ácido nucleico de doble cadena.

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

40                   (a) la primera pluralidad de moléculas de ácido nucleico comprende una pluralidad de especies de alta abundancia; y/o  
 45                   (b) la al menos una especie de alta abundancia representa al menos el 50%, al menos el 60% o al menos el 70% de la primera pluralidad de moléculas de ácido nucleico y/o  
 50                   (c) la reducción del contenido de al menos una especie de alta abundancia es de al menos el 80%, al menos el 90%, al menos el 95% o al menos el 99%.

4. El método de la reivindicación 3, en el que la segunda pluralidad de moléculas de ácido nucleico comprende la pluralidad de especies de alta abundancia, opcionalmente en donde la pluralidad de especies de alta abundancia en la segunda pluralidad de moléculas de ácido nucleico representa menos del 50%, menos que 40% o menos del 30% de la segunda pluralidad de moléculas de ácido nucleico.

5. El método de las reivindicaciones 1-4, en el que la primera pluralidad de moléculas de ácido nucleico comprende una pluralidad de especies de baja abundancia, opcionalmente en donde la pluralidad de especies de baja abundancia representa menos del 10%, menos del 5% o menos del 1% de la primera pluralidad de moléculas de ácido nucleico, opcionalmente en donde la segunda pluralidad de moléculas de ácido nucleico comprende la pluralidad de especies de baja abundancia, opcionalmente en donde la pluralidad de especies de baja abundancia en la segunda pluralidad de moléculas de ácido nucleico representa al menos el 5%, al menos el 10% o al menos el 20% de la segunda pluralidad de moléculas de ácido nucleico.

6. El método de una cualquiera de las reivindicaciones 1-5, en el que:

65                   (a) cada una de la primera pluralidad de moléculas de ácido nucleico o cada una de la segunda pluralidad de moléculas de ácido nucleico comprende un código de barras estocástico; y/o

(b) el método comprende además la secuenciación de la segunda pluralidad de moléculas de ácido nucleico para generar una pluralidad de lecturas de secuenciación, opcionalmente en las que la secuenciación se lee para la pluralidad de especies de alta abundancia es inferior al 50%, inferior al 40% o inferior al 30% de las lecturas de secuenciación totales, en donde opcionalmente la lectura de secuenciación para la pluralidad de especies de baja abundancia es al menos del 5%, al menos el 10% o al menos el 20% del total de lecturas de secuenciación; y/o (c) el método comprende además la adición de una pluralidad de bloqueadores durante la etapa de resurtido parcial, opcionalmente en donde la pluralidad de bloqueadores híbridos al sitio de unión del cebador universal del primer oligonucleótido o el sitio de unión del cebador universal del segundo oligonucleótido, opcionalmente en donde la pluralidad de bloqueadores previene la hibridación entre el sitio de unión del cebador universal del primer oligonucleótido o el sitio de unión del cebador universal del segundo oligonucleótido y su secuencia complementaria.

**7. Un método para generar una biblioteca de ácidos nucleicos normalizada, que comprende:**

hibridar una pluralidad de primeros oligonucleótidos que comprenden un resto de unión con una pluralidad de dianas de ácido nucleico;  
extender la pluralidad de primeros oligonucleótidos para generar una pluralidad de cadenas complementarias de la pluralidad de dianas de ácido nucleico que comprenden el resto de unión;  
desnaturalizar una pluralidad de moléculas de ácido nucleico de doble cadena que comprenden la pluralidad de cadenas complementarias de la pluralidad de dianas de ácido nucleico;  
reanudación parcial de la pluralidad de cadenas complementarias de la pluralidad de dianas del ácido nucleico; y  
eliminar las cadenas complementarias recocidas de la pluralidad de dianas de ácido nucleico mediante una molécula de captura inmovilizada en uno o más soportes sólidos, en donde las moléculas de captura se unen específicamente al resto de unión, por lo que se genera una biblioteca de ácidos nucleicos normalizada de la pluralidad de dianas de ácido nucleico.

**8. El método de la reivindicación 7, en el que:**

(a) el resto de unión es un grupo funcional seleccionado del grupo que consiste en biotina, estreptavidina, heparina, un aptámero, un resto de química de clic, digoxigenina, amina(s) primaria(s), carboxilina(s), hidroxilo(s), aldehído(s), cetona(s) y cualquier combinación de los mismos, opcionalmente en donde el resto de unión es biotina, opcionalmente en donde la molécula de captura es estreptavidina; y/o  
(b) el método comprende además sintetizar una segunda cadena para una o más de la pluralidad de cadenas complementarias de la pluralidad de dianas de ácido nucleico para generar una o más de la pluralidad de moléculas de ácido nucleico bicatenario que comprenden la pluralidad de cadenas complementarias de la pluralidad de dianas de ácidos nucleicos, opcionalmente en donde la síntesis comprende hibridar una pluralidad de segundos oligonucleótidos con la pluralidad de cadenas complementarias de la pluralidad de dianas de ácidos nucleicos y extender la pluralidad de segundos oligonucleótidos, opcionalmente, en donde la pluralidad de primeros oligonucleótidos o la pluralidad de segundos oligonucleótidos comprende un sitio de unión del cebador universal; y/o  
(c) el método comprende además amplificar la pluralidad de moléculas de ácido nucleico de doble cadena.

**9. El método de la reivindicación 7 o la reivindicación 8, en el que:**

(a) la pluralidad de dianas de ácido nucleico comprende una pluralidad de dianas de ácido nucleico de baja abundancia, opcionalmente en donde la pluralidad de dianas de ácido nucleico de baja abundancia representa menos del 10%, menos del 5% o menos del 1% de la pluralidad de dianas de ácido nucleico; y/o  
(b) la biblioteca de ácidos nucleicos normalizada de la pluralidad de dianas de ácidos nucleicos comprende la pluralidad de dianas de ácidos nucleicos de baja abundancia, en donde opcionalmente la pluralidad de dianas de ácidos nucleicos de baja abundancia en la biblioteca de ácidos nucleicos normalizada representa al menos 5%, al menos el 10% o al menos el 20% de la pluralidad de dianas de ácido nucleico en la biblioteca de ácido nucleico normalizada.

**10. El método de una cualquiera de las reivindicaciones 7-9, en el que:**

(a) la pluralidad de dianas de ácido nucleico comprende una pluralidad de dianas de ácido nucleico de alta abundancia, opcionalmente en donde la pluralidad de dianas de ácido nucleico de alta abundancia representa al menos el 50%, al menos el 60% o al menos el 70% de la pluralidad de dianas de ácido nucleico; y/o  
(b) el contenido de la pluralidad de especies de alta abundancia en la biblioteca de ácidos nucleicos normalizada se reduce en al menos un 80%, al menos un 90%, al menos un 95% o al menos un 99%; y/o  
(c) la biblioteca de ácidos nucleicos normalizada de la pluralidad de dianas de ácidos nucleicos comprende la pluralidad de dianas de ácidos nucleicos de alta abundancia, donde opcionalmente la pluralidad de dianas de ácidos nucleicos de alta abundancia en la biblioteca de ácidos nucleicos normalizada representa menos del 50%, menos del 40% o menos del 30% de la pluralidad de dianas de ácido nucleico en la biblioteca de ácido nucleico normalizada; y/o

cada uno de la pluralidad de primeros oligonucleótidos o cada uno de la pluralidad de segundos oligonucleótidos comprende un código de barras estocástico.

5 11. El método de una cualquiera de las reivindicaciones 7-10, en el que:

- (a) el método que comprende además la secuenciación de la biblioteca de ácidos nucleicos normalizada para generar una pluralidad de lecturas de secuenciación, opcionalmente en la que las lecturas de secuenciación para la pluralidad de dianas de ácidos nucleicos de alta abundancia son menores del 50%, menores del 40% o menores del 30% de las lecturas de secuenciación total; y/o  
 10 (b) las lecturas de secuencia para la pluralidad de dianas de ácido nucleico de baja abundancia son al menos el 5%, al menos el 10% o al menos el 20% de las lecturas de secuenciación total.

12. El método de una cualquiera de las reivindicaciones 7-11, en el que:

- 15 (a) el método comprende además la adición de una pluralidad de bloqueadores durante la etapa de recocido parcial, opcionalmente en donde la pluralidad de bloqueadores híbridos al sitio de unión del cebador universal del primer oligonucleótido o el sitio de unión del cebador universal del segundo oligonucleótido, opcionalmente en el que la pluralidad de bloqueadores evita la hibridación entre el sitio de unión del cebador universal del primer oligonucleótido o el sitio de unión del cebador universal del segundo oligonucleótido y su secuencia complementaria; y/o  
 20 (b) la pluralidad de dianas de ácido nucleico comprende ARNm o ARNm mitocondrial o ARNm de proteína ribosomal, opcionalmente en donde las dianas de ácido nucleico de baja abundancia comprenden 7.000 genes, 4.000 genes o 2.000 genes con el menor número de transcripciones; y/o  
 25 (c) la pluralidad de primeros oligonucleótidos comprende cebadores específicos de diana o cebadores no específicos de diana; y/o  
 (d) la pluralidad de dianas de ácido nucleico es de una sola célula.

30 13. El método de la reivindicación 12 (a), en el que la pluralidad de dianas de ácido nucleico comprende ADNc o ADN genómico, opcionalmente en donde la pluralidad de dianas de ácido nucleico comprende ADN genómico, las dianas de ácido nucleico de alta abundancia comprenden secuencias de repetición de corto tándem, secuencias centroméricas; y/o en donde la pluralidad de dianas de ácido nucleico es de una sola célula.

14. Un método para generar una biblioteca de ácidos nucleicos normalizada, que comprende:

- 35 hibridar una pluralidad de primeros oligonucleótidos que comprenden un resto de unión con una pluralidad de dianas de ácido nucleico en una biblioteca de ácido nucleico no normalizada;  
 extender la pluralidad de primeros oligonucleótidos para generar una pluralidad de cadenas complementarias de la pluralidad de dianas de ácido nucleico que comprenden el resto de unión;  
 40 desnaturalizar una pluralidad de moléculas de ácido nucleico de doble cadena que comprenden la pluralidad de cadenas complementarias de la pluralidad de dianas de ácido nucleico;  
 reanudación parcial de la pluralidad de cadenas complementarias de la pluralidad de dianas del ácido nucleico;  
 y  
 45 eliminar las cadenas complementarias recocidas de la pluralidad de dianas de ácido nucleico, por lo que se genera una biblioteca de ácidos nucleicos normalizada de la pluralidad de dianas de ácido nucleico.

15. El método de la reivindicación 14, en el que:

- 50 (a) la biblioteca de ácido nucleico no normalizada comprende una o más dianas de ácido nucleico de alta abundancia y una o más dianas de ácido nucleico de baja abundancia, donde opcionalmente la una o más dianas de ácido nucleico de alta abundancia representa al menos el 50%, al menos el 60% o al menos el 70% de la biblioteca de ácido nucleico no normalizada; y/o  
 (b) el contenido de una o más dianas de ácido nucleico de alta abundancia en la biblioteca de ácido nucleico normalizada se reduce en al menos un 80%, al menos un 90%, al menos un 95% o al menos un 99%; y/o  
 55 (c) la una o más dianas de ácido nucleico de baja abundancia representan menos del 10%, menos del 5% o menos del 1% de la biblioteca de ácido nucleico no normalizada; y/o  
 (d) la una o más dianas de ácido nucleico de baja abundancia representan al menos el 5%, al menos el 10% o al menos el 20% de la biblioteca de ácido nucleico normalizada; y/o  
 (e) la una o más dianas de ácido nucleico de alta abundancia representan menos del 50%, menos del 40% o menos del 30% de la biblioteca de ácidos nucleicos normalizada.

60 16. El método de una cualquiera de las reivindicaciones 14 o 15, en el que:

- (a) la biblioteca de ácido nucleico no normalizada es una biblioteca de ADNc o una biblioteca genómica; y/o  
 65 (b) la biblioteca de ácido nucleico no normalizada es una biblioteca de ácido nucleico de células simples.

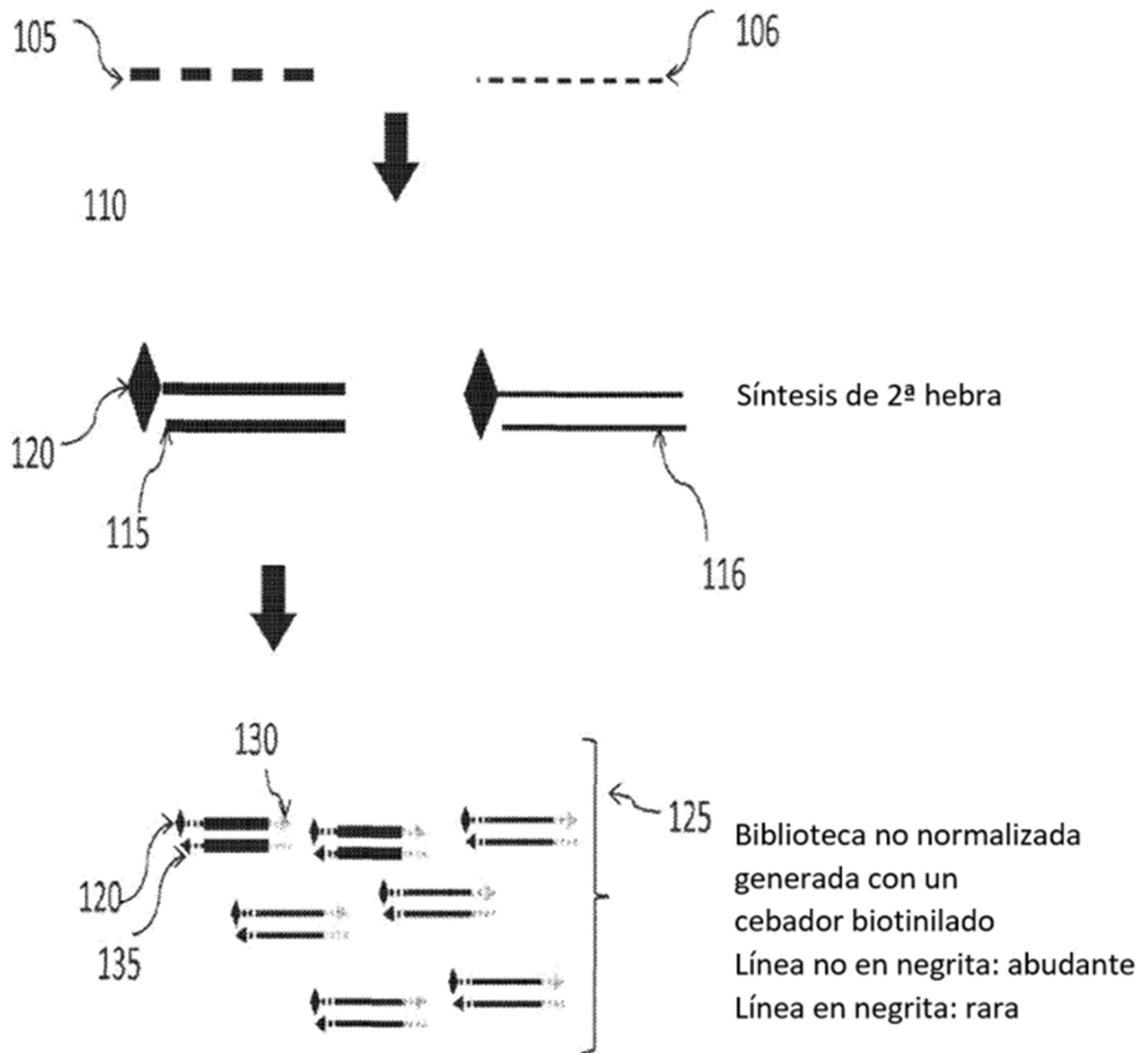


FIG. 1

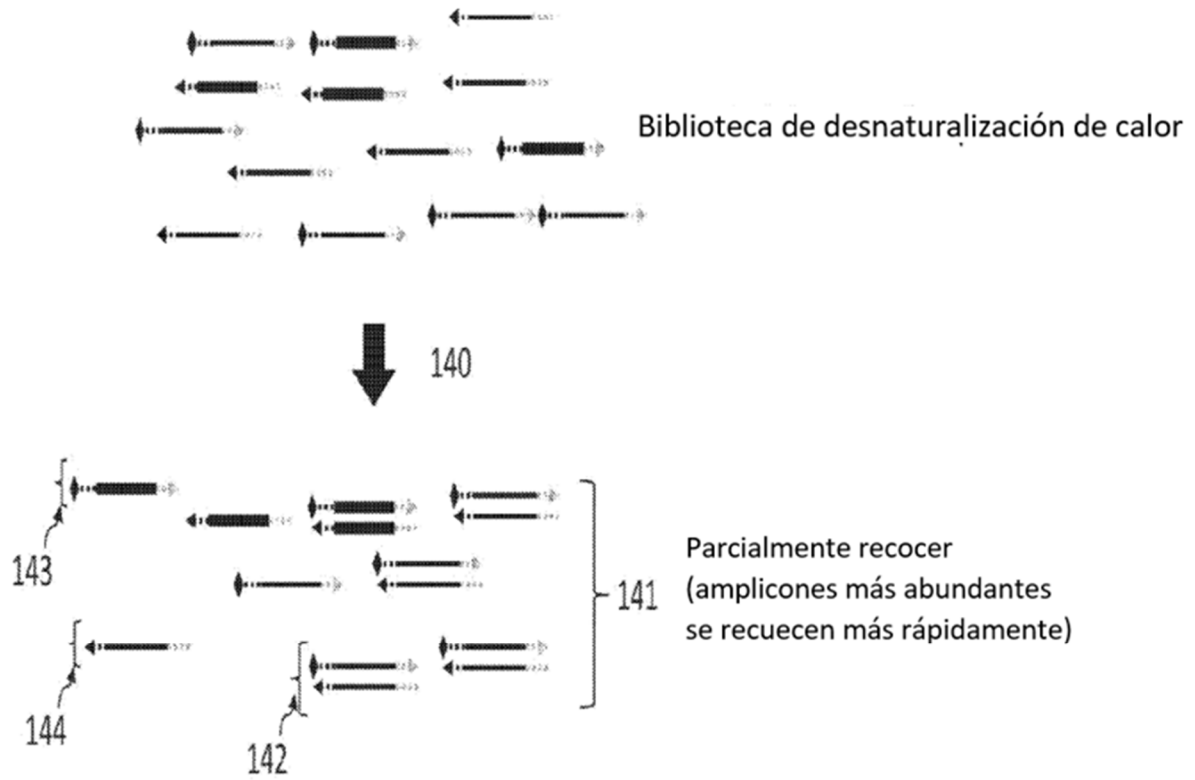


FIG. 1, continuado

FIG. 1, continuado

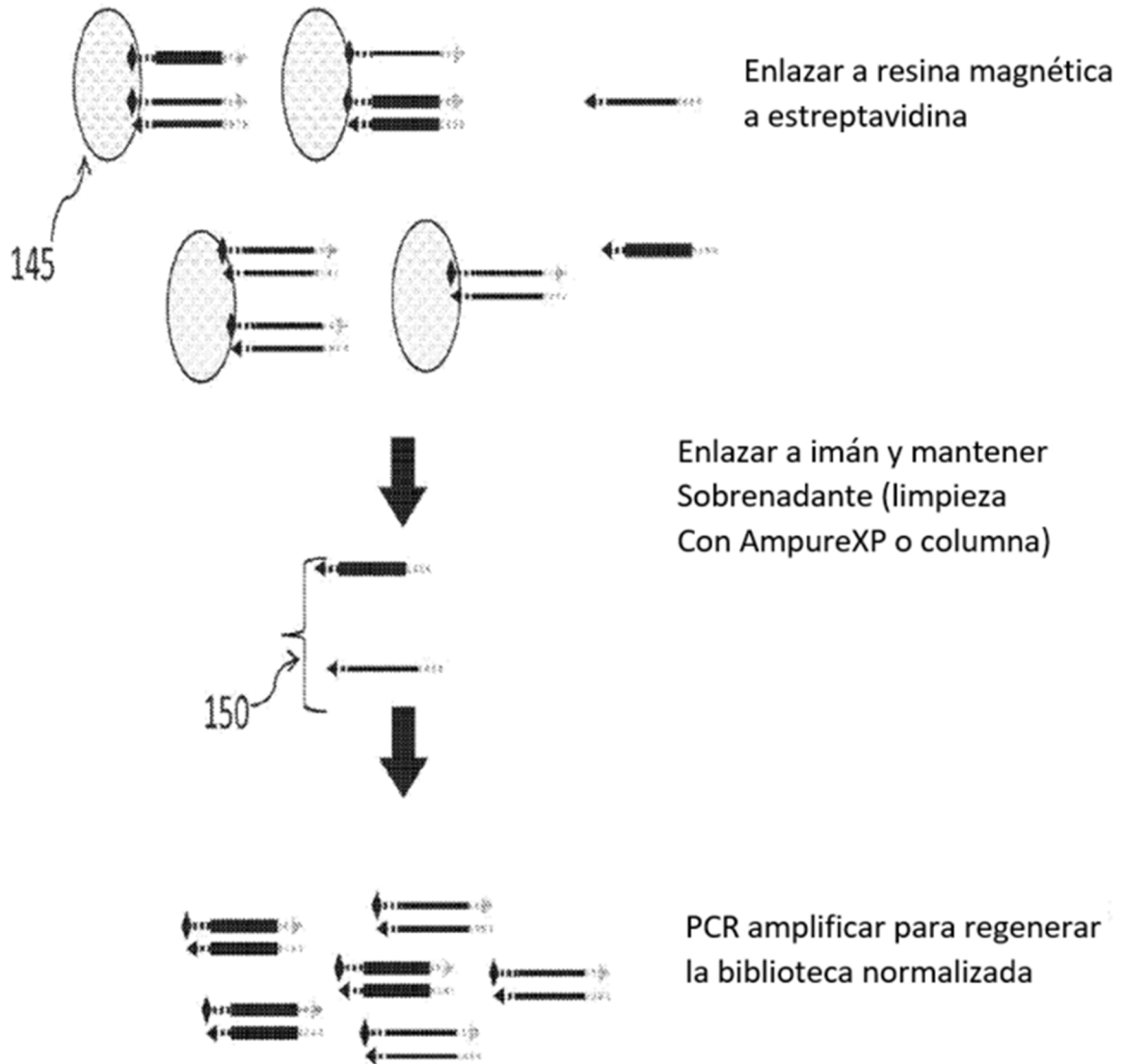
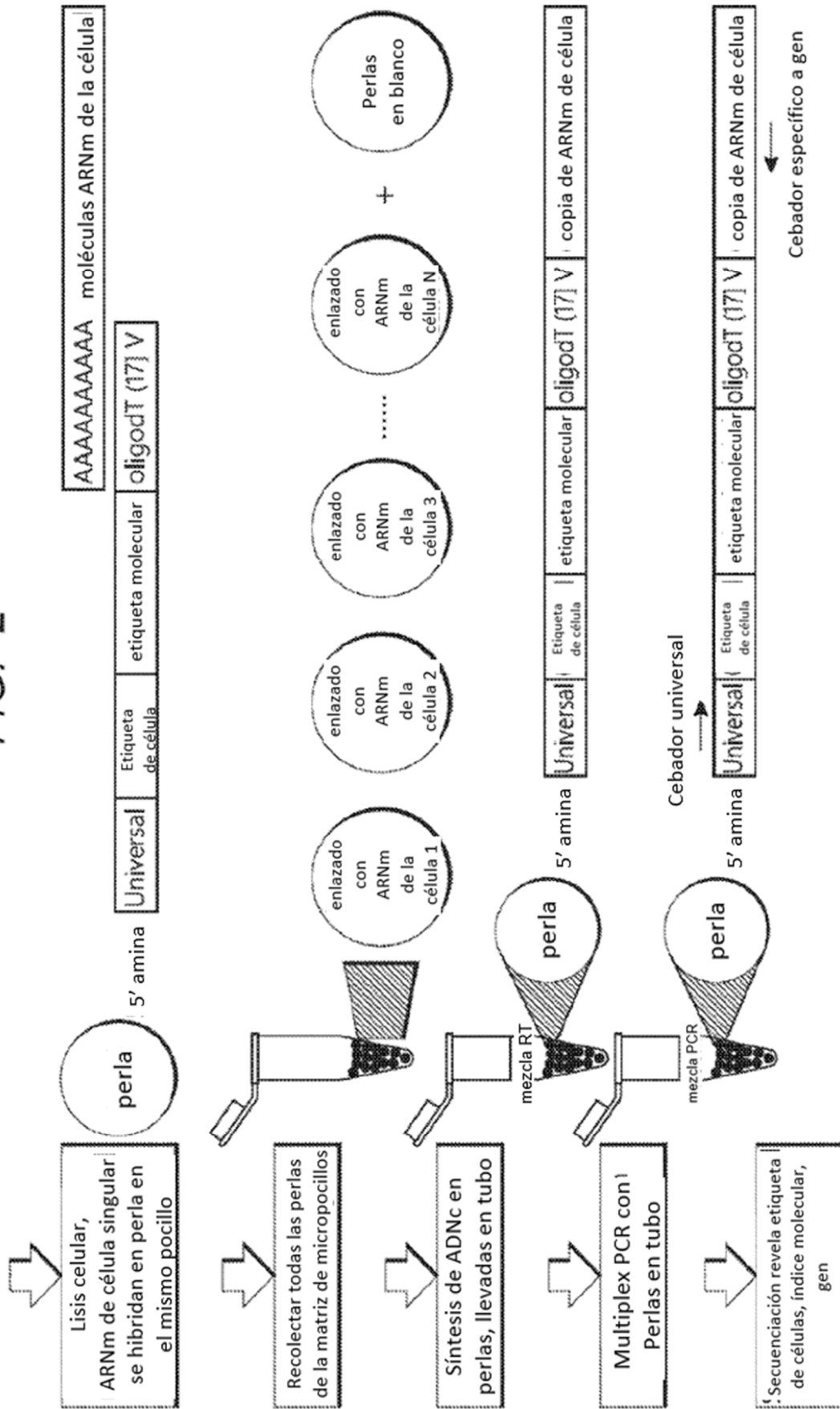


FIG. 2





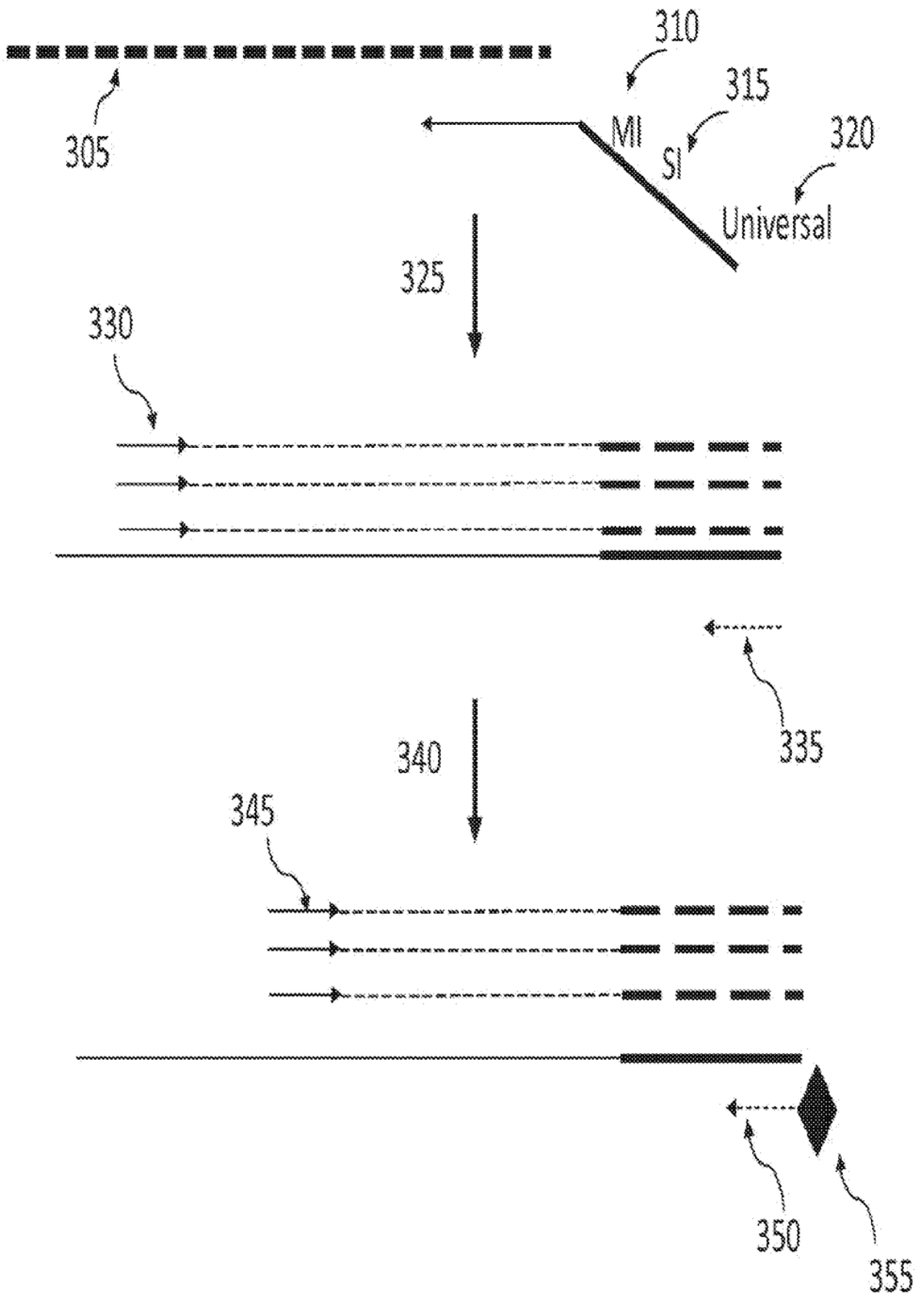


FIG. 3

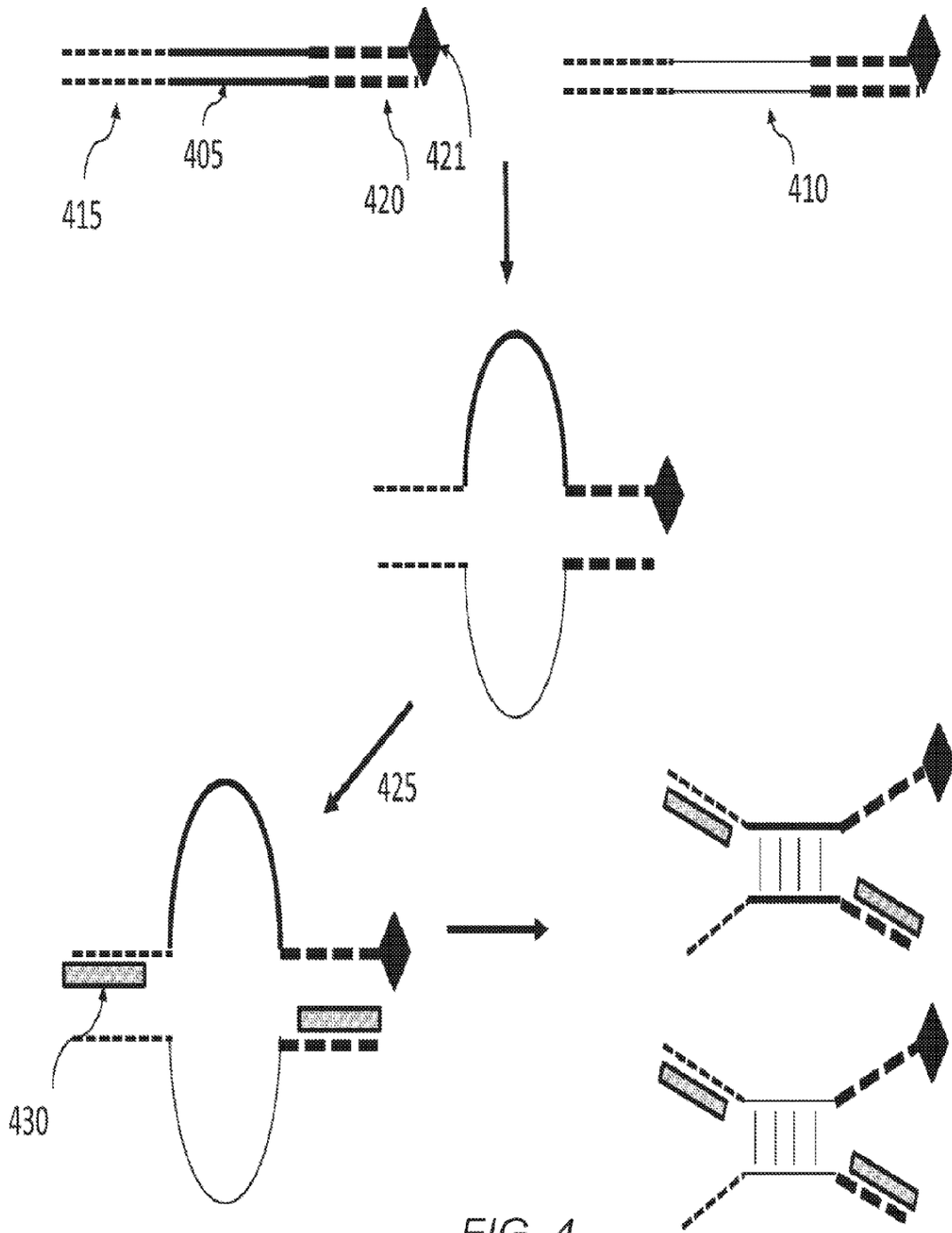


FIG. 4

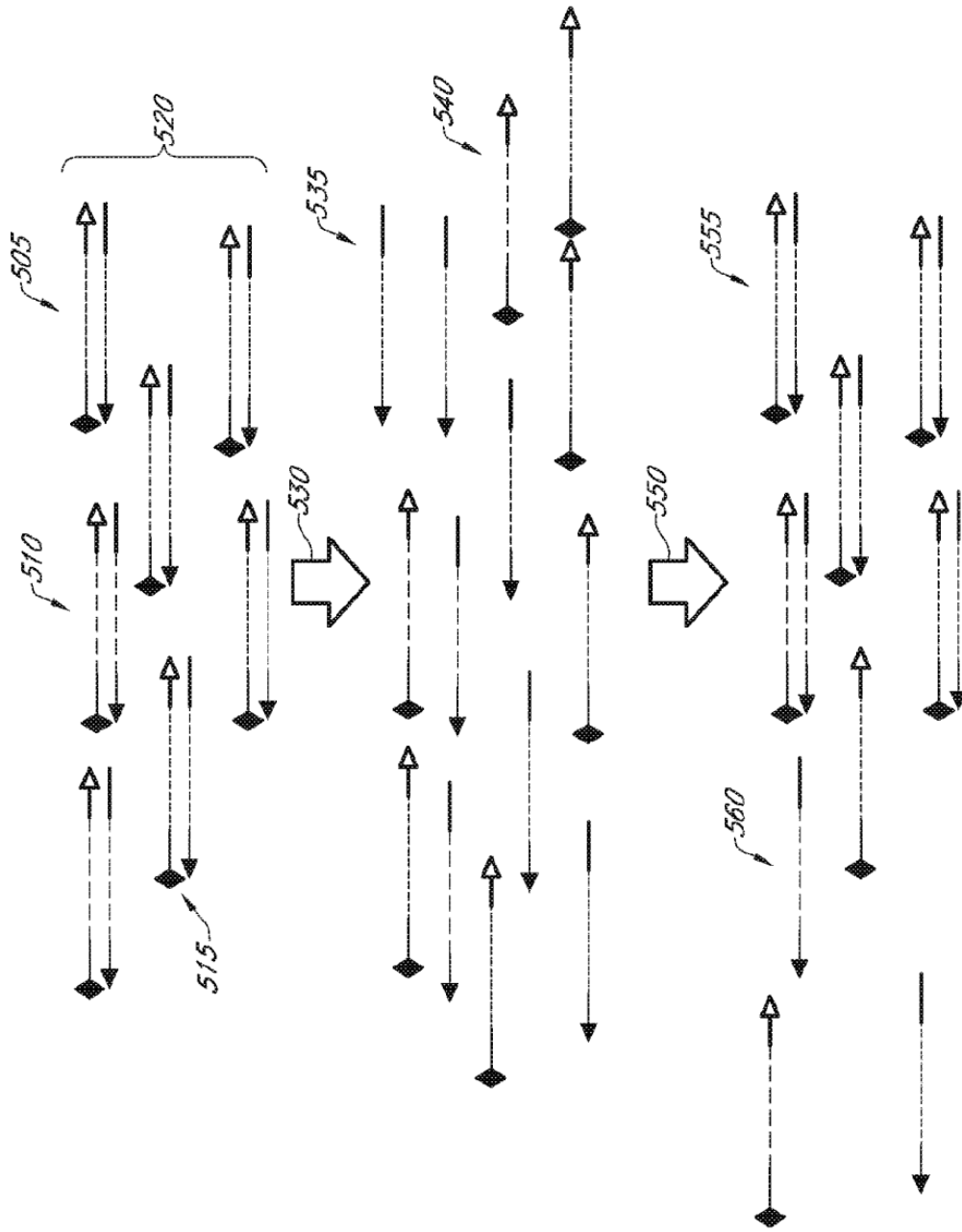


FIG. 5A

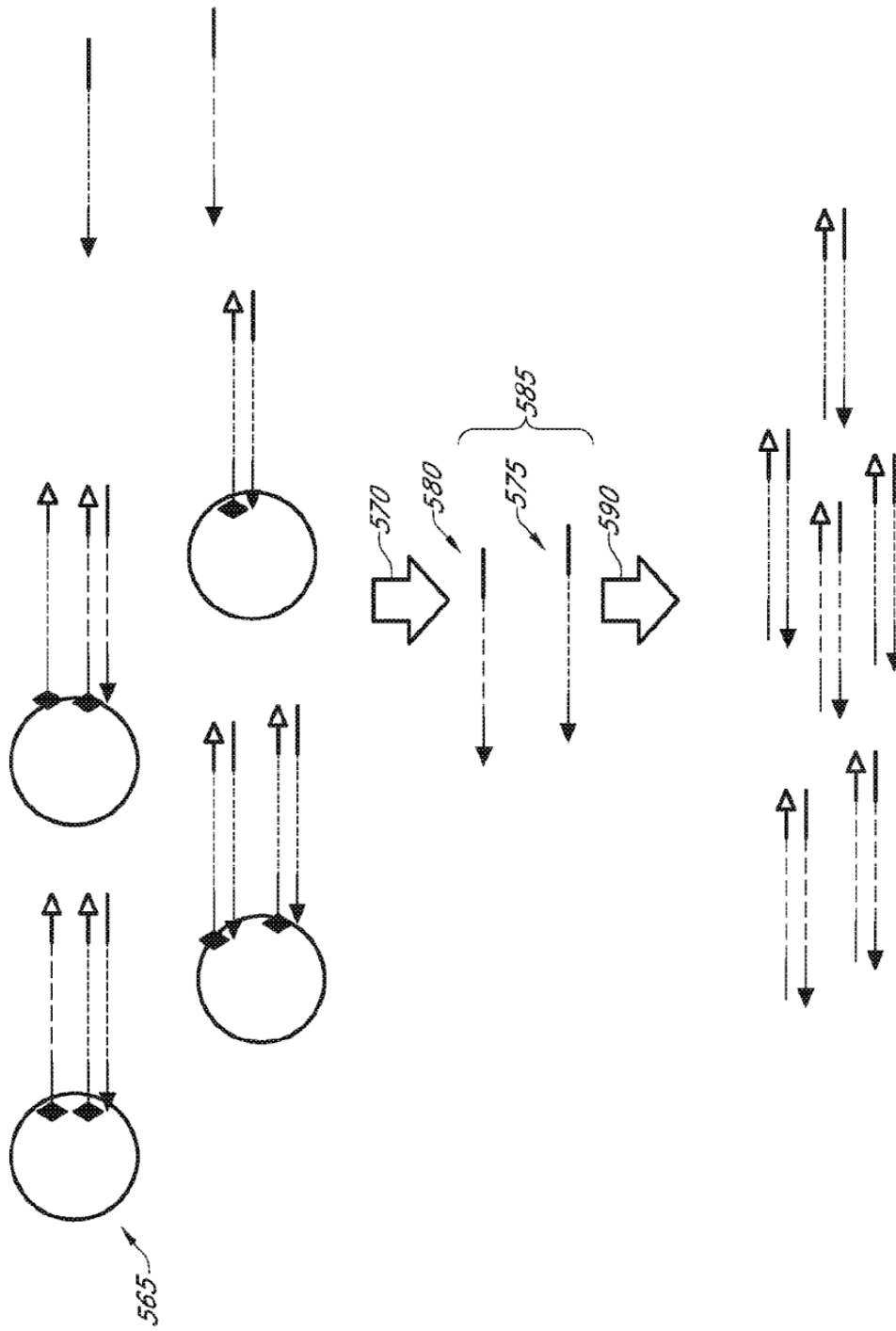


FIG. 5B