

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 955 997**

51 Int. Cl.:

C12Q 1/6804 (2008.01)

C12Q 1/6806 (2008.01)

C12N 15/10 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **19.12.2017** **E 21168174 (7)**

97 Fecha y número de publicación de la concesión europea: **28.06.2023** **EP 3957744**

54 Título: **Reactivos y métodos para código de barras moleculares de ácidos nucleicos de células individuales**

30 Prioridad:

23.12.2016 GB 201622222

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

11.12.2023

73 Titular/es:

**CS GENETICS LIMITED (100.0%)
Botanic House 100 Hills Road
Cambridge, Cambridgeshire CB2 1PH, GB**

72 Inventor/es:

EDELMAN, LUCAS BRANDON

74 Agente/Representante:

SÁEZ MAESO, Ana

Observaciones:

Véase nota informativa (Remarks, Remarques o Bemerkungen) en el folleto original publicado por la Oficina Europea de Patentes

ES 2 955 997 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

Reactivos y métodos para código de barras moleculares de ácidos nucleicos de células individuales

Campo técnico

5 La presente invención se refiere a códigos de barras moleculares. Se proporcionan métodos para codificar con barras ácidos nucleicos de células individuales utilizando bibliotecas de reactivos de codificación de barras multiméricos.

Antecedentes

El 'código de barras molecular' se desarrolló para abordar los problemas generados por las tasas de error sin procesar intrínsecas a las máquinas de secuencia de ADN (precisión sintética), y también los problemas relacionados con el recuento de moléculas de ácido nucleico individuales dentro de una muestra (recuento molecular).

10 El código de barras molecular generalmente implica unir (por ejemplo, por ligadura o por extensión de cebador) una etiqueta de ácido nucleico única (un 'código de barras') a varias moléculas diana únicas (ADN o ARN) en una solución que contiene una gran cantidad de tales moléculas. Estas moléculas etiquetadas luego se secuencian, lo que para cada uno revela tanto la secuencia del código de barras molecular como al menos parte de la secuencia de la propia molécula diana etiquetada.

15 Este código de barras se utiliza normalmente para dos fines diferentes. En primer lugar, se puede utilizar para habilitar la secuenciación redundante. Por ejemplo, imagina una muestra de ácido nucleico que contiene 1000 copias de un gen en particular en una muestra de ADN; 999 de las copias contienen secuencias idénticas entre sí, pero una sola copia tiene una mutación particular de un solo nucleótido. Sin código de barras, el secuenciador no podrá detectar esta copia mutada, dado que el secuenciador comete errores aleatorios a una velocidad superior a 1:1000, es decir, la mutación es tan rara en la población de moléculas secuenciadas que cae por debajo del umbral de ruido de fondo intrínseco del secuenciador.

25 Sin embargo, si las 1000 copias han sido etiquetadas cada una con un código de barras molecular único, y cada molécula etiquetada individual es secuenciada varias veces por la máquina de secuenciación (secuenciación redundante), observaría que cada vez (o, al menos el 99 % de las veces, equivalente a la precisión bruta del secuenciador) que la molécula mutada etiquetada fue secuenciada de manera redundante (es decir, cada vez que se observó que la secuencia del gen diana estaba etiquetada con ese código de barras único particular que se adjuntó a la molécula inicial mutada), que de hecho se observaría la misma mutación aparente. Por el contrario, esa mutación en particular solo se observaría aproximadamente el 1 % del tiempo (la tasa de error sin procesar del secuenciador) cuando las copias del gen etiquetado pero no mutado se secuenciaron de forma redundante, según sus respectivos

30 códigos de barras alternativos.

Por tanto, el código de barras sirve para identificar moléculas de entrada individuales en todas sus respectivas copias múltiples dentro de la reacción de secuenciación, permitiendo que un algoritmo de detección de secuencia se centre específicamente en sus lecturas respectivas dentro de un conjunto de datos de secuenciación, y así evitar la gran cantidad de ruido de secuencia estocástico (en forma de errores de secuencia) que está presente en el resto del

35 conjunto de datos. Esto permite, por tanto, la 'precisión sintética', mediante secuenciación redundante, que es potencialmente mucho más alta que la precisión bruta del secuenciador en sí.

El código de barras también se puede utilizar para permitir el "recuento molecular" digital de las moléculas de ADN o ARN de entrada. En este proceso, una gran cantidad de códigos de barras únicos se adjuntan a las moléculas de entrada, por ejemplo, copias de ADNc que se han realizado a partir de una especie de ARNm particular. Cada molécula de ADNc de entrada está etiquetada (por ejemplo, por extensión del cebador) con un solo código de barras único. Luego se secuencian las moléculas, que, como con la secuenciación redundante, revelan el código de barras único y al menos parte de cada molécula de entrada etiquetada asociada; estas moléculas también se secuencian cada una más de una vez.

40 En lugar de utilizar esta secuenciación redundante para reducir los errores de secuenciación, en el recuento molecular se utiliza para cuantificar digitalmente cuántas moléculas individuales de la molécula diana dada (ADNc en este caso) estaban presentes en la muestra original, simplemente contando el número total de códigos de barras únicos que se secuenciaron y se encontraron asociados con la diana en particular. De esta manera, la secuenciación redundante dirigida por código de barras reduce la posibilidad de que la reacción de secuenciación deje estocásticamente sin secuenciar cualquier molécula de entrada (ya que, en promedio, cada molécula etiquetada se secuencia varias veces),

50 conservando una medida precisa de la cantidad de entrada (dado que las moléculas de partida secuenciadas de forma redundante solo se cuentan una vez, discriminado por copias repetidas de su código de barras único).

Se proporcionan ejemplos del uso de códigos de barras moleculares en los documentos US 8728766, US 8685678, US 8722368, Kinde *et al.*, 2011 (PNAS, 108, 23, 9530-9535), US 20140227705 A1 y WO2016/100976.

55 Se genera una 'lectura larga sintética' cuando una secuencia larga contigua de ADN (más larga que la longitud de lectura alcanzable en un secuenciador de ADN) se convierte en dos o más 'subsecuencias' más cortas que son lo

suficientemente cortas para ser leídas por un secuenciador de ADN, y que de alguna manera están etiquetadas de manera que se pueda deducir (después de la secuenciación) que las subsecuencias se generaron a partir de la misma secuencia de ADN larga original. Por ejemplo, si desea secuenciar un gen humano en particular que tiene 1000 nucleótidos de longitud, pero hacerlo con un secuenciador de ADN de lectura corta con una longitud de lectura de 100 nucleótidos, podría separar la secuencia larga en 10 subsecuencias diferentes de 100 nucleótidos de longitud, luego etiquetar cada una de estas 10 subsecuencias con una secuencia de ADN sintética informativa que identifica cada una de las 10 subsecuencias como provenientes de la misma molécula de ADN original de 1000 nucleótidos, luego realizar una secuenciación de ADN de alto rendimiento con estas 10 moléculas de ADN resultantes, y así (para cada una de las 10 moléculas de ADN resultantes) obtener tanto la subsecuencia de 100 nucleótidos como la etiqueta de identificación de ADN asociada. Con estos datos de ADN de alto rendimiento, se puede usar un algoritmo que detecta estas etiquetas de identificación y las usa para asociar las 10 subsecuencias diferentes de 100 nucleótidos entre sí como una "agrupación" de subsecuencia colectiva, y con ello estimar que las 10 subsecuencias vinieron de un gen de 1000 nucleótidos más largo, y con ello estimar la secuencia genética total de 1000 nucleótidos de longitud 'uniendo' las 10 subsecuencias juntas de manera computacional en un solo gen de 1000 nucleótidos de longitud.

En la bibliografía se han descrito al menos dos tecnologías generales de lectura larga sintética: un enfoque basado en particiones que se describe en los documentos US 20130079231 A1 y US 2014378345 A1; y un enfoque de copia de códigos de barras que se describe en Casbon *et al.*, 2013 (Nucleic Acids Research, 2013, 41, 10, e112), US 8679756 y US 8563274.

La 'secuenciación espacial' se considera la secuenciación de ácidos nucleicos con la inclusión de cierta información sobre dónde se encuentra cada ácido nucleico secuenciado dentro de un espacio particular (por ejemplo, dentro de una muestra en particular, o dentro de una célula en particular). Sin embargo, se conocen muy pocos métodos de secuenciación espacial. La principal tecnología conocida es la técnica de secuenciación de ARN fluorescente *in situ* (FISSEQ). En FISSEQ, una muestra de células se reticula y, mientras las células aún están intactas, el ARN se transcribe de forma inversa en ADNc y se amplifica mientras aún se encuentra en las células reticuladas. A continuación, cada molécula de ADNc amplificada se secuenciaría ópticamente mientras aún se encuentra en las células, con un sistema de detección óptica sensible y de alta potencia. Este método se describe en Lee *et al.*, 2014 (Science, 343, 6177, 1360-1363). Las técnicas actuales para realizar análisis de ácido nucleico de células individuales generalmente tienen un rendimiento limitado (es decir, el número de células que pueden analizarse simultáneamente dentro de un solo experimento, o analizarse por unidad de tiempo) y también requieren una instrumentación experimental relativamente compleja, tal como equipos de microfluidos, y además pueden implicar procedimientos experimentales relativamente complejos y/o largos para llevar a cabo.

La invención aborda dos tipos principales de problemas en el campo de la secuenciación: 1) limitaciones analíticas específicas de las máquinas de secuenciación de ADN; y 2) desafíos biofísicos asociados con tipos comunes de muestras de ADN experimentales.

Las máquinas de secuenciación de ADN de alto rendimiento actuales son plataformas poderosas que se utilizan para analizar grandes cantidades de material genético (de miles a miles de millones de moléculas de ADN) y funcionan como sistemas tanto para la investigación básica como para las aplicaciones médicas aplicadas. Sin embargo, todas las máquinas de secuenciación de ADN actuales están sujetas a ciertas limitaciones analíticas que limitan las aplicaciones científicas y médicas en las que se pueden utilizar de forma eficaz. Las principales limitaciones de este tipo incluyen longitudes de lectura brutas finitas y precisión bruta finita, ambas se describen a continuación.

Con respecto a las longitudes de lectura brutas finitas, cada plataforma de secuenciación de ADN se caracteriza por una 'longitud de lectura' típica que puede alcanzar, que es la 'longitud' en nucleótidos de ADN que puede 'leer' de cada molécula secuenciada. Para la mayoría de las máquinas de secuenciación, esto varía de 100 a ~ 500 nucleótidos.

Con respecto a la precisión bruta finita, cada plataforma de secuenciación también se caracteriza por una 'precisión bruta' alcanzable, típicamente definida como la probabilidad de que cada secuencia de nucleótidos dada se haya determinado correctamente. La precisión bruta típica para las plataformas de secuenciación más populares varía entre el 98 y el 99,5 %. La cantidad relacionada, la tasa de 'error bruto', es esencialmente lo opuesto a la precisión bruta, y es la probabilidad por nucleótido de que el secuenciador documente aleatoriamente un nucleótido incorrecto en una molécula de ADN secuenciada en particular.

Además, ciertas muestras de ADN experimentales comunes plantean desafíos biofísicos para la secuenciación. Estos desafíos surgen del estado molecular único (y problemático) del ADN en estas muestras, lo que dificulta secuenciarlos o extraer importantes piezas de información genética de ellos, independientemente de la máquina de secuenciación empleada. Por ejemplo, las muestras embebidas en parafina fijadas con formalina (EPFF) son la herramienta experimental estándar para realizar patología molecular a partir de muestras de biopsia humana. Sin embargo, el proceso de creación de una muestra EPFF, en el que la muestra de biopsia se fija (reticula y se mantiene físicamente unida y estable a nivel molecular) por una sustancia química fuerte, y luego se embebe en una cera, crea un daño significativo al ADN y al ARN contenidos en la misma. Por lo tanto, el ADN y el ARN de las muestras EPFF están muy fragmentados (generalmente en pequeños fragmentos de entre 50 y 200 nucleótidos) y también incluyen daños esporádicos en los nucleótidos individuales, lo que hace que sea esencialmente imposible de amplificar o aislar por mucho tiempo, secuencias contiguas.

Descripción

La divulgación proporciona reactivos de códigos de barras multiméricos y métodos para su uso en la preparación de muestras de ácido nucleico que contienen células para secuenciación. En los métodos, los reactivos de códigos de barras multiméricos se utilizan para generar códigos de barras de los ácidos nucleicos diana de las células en las muestras. Las secuencias de códigos de barras pueden añadirse a partir de un único reactivo de códigos de barras multimérico a subsecuencias de un ácido nucleico diana de una sola célula para producir un conjunto de moléculas de ácido nucleico diana con código de barras. Dichas moléculas pueden secuenciarse para producir conjuntos de lecturas de secuencia, cada conjunto de lecturas de secuencias se corresponde con las moléculas de ácido nucleico de una sola célula (es decir, secuenciación de una sola célula). Además, los métodos se pueden realizar en muchas células en paralelo, lo que permite una secuenciación unicelular de alto rendimiento.

La divulgación proporciona una biblioteca que comprende al menos dos reactivos de códigos de barras multiméricos, en donde cada reactivo de códigos de barras multimérico comprende: (a) el primer y segundo oligonucleótidos con código de barras unidos entre sí y un resto de unión a células, en donde los oligonucleótidos con código de barras comprenden, cada uno, una región de código de barras y en donde las regiones de código de barras del primer y segundo oligonucleótidos con código de barras de un primer reactivo de código de barras multimérico de la biblioteca son diferentes a las regiones de código de barras del primer y segundo oligonucleótidos con código de barras de un segundo reactivo de código de barras multimérico de la biblioteca.

La divulgación proporciona una biblioteca de reactivos de códigos de barras multiméricos que comprenden al menos 2 reactivos de códigos de barras multiméricos para marcar ácidos nucleicos diana para secuenciación, en donde cada reactivo de códigos de barras multimérico comprende: (a) primera y segunda moléculas de hibridación unidas entre sí, en donde cada una de las moléculas de hibridación comprende una secuencia de ácido nucleico que comprende una región de hibridación; (b) el primer y segundo oligonucleótidos con código de barras, en donde el primer oligonucleótido con código de barras hibrida con la región de hibridación de la primera molécula de hibridación y en donde el segundo oligonucleótido con código de barras hibrida con la región de hibridación de la segunda molécula de hibridación, en donde los oligonucleótidos con código de barras comprenden, cada uno, una región de código de barras; y (c) un resto de unión a células; en donde las regiones de código de barras del primer y segundo oligonucleótidos con código de barras de un primer reactivo de código de barras multimérico de la biblioteca son diferentes a las regiones de código de barras del primer y segundo oligonucleótidos con código de barras de un segundo reactivo de código de barras multimérico de la biblioteca.

Se puede unir un resto de unión a células a cada una de las moléculas de código de barras. Adicionalmente o como alternativa, se puede unir un resto de unión a células a cada uno de los oligonucleótidos con código de barras. Los reactivos de códigos de barras multiméricos pueden ser para etiquetar subsecuencias de un ácido nucleico diana en una célula.

Cada reactivo de código de barras multimérico en la biblioteca puede ser para marcar los ácidos nucleicos diana de una sola célula. Cada reactivo de código de barras multimérico en la biblioteca puede ser para marcar los ácidos nucleicos diana en una sola célula.

La primera y la segunda moléculas de hibridación pueden estar comprendidas dentro de una (única) molécula de ácido nucleico. Como alternativa, la primera y segunda moléculas de hibridación pueden estar unidas por un soporte, por ejemplo, una macromolécula, soporte sólido o soporte semisólido, como se describe en el presente documento.

El primer y segundo oligonucleótidos con código de barras pueden tomar cualquier forma descrita en el presente documento. Por ejemplo, cada oligonucleótido con código de barras puede comprender además una región diana.

La biblioteca puede comprender al menos 10 reactivos de códigos de barras multiméricos. Las regiones de código de barras de los oligonucleótidos con código de barras primero y segundo de cada reactivo de código de barras multimérico pueden ser diferentes a las regiones de código de barras de los oligonucleótidos con código de barras de al menos otros 9 reactivos de código de barras multiméricos en la biblioteca.

La divulgación proporciona una biblioteca de reactivos de códigos de barras multiméricos que comprenden al menos 10 reactivos de códigos de barras multiméricos para marcar ácidos nucleicos diana para secuenciación, en donde cada reactivo de códigos de barras multimérico comprende: (a) la primera y segunda moléculas de hibridación comprendidas dentro de una molécula de ácido nucleico, en donde cada una de las moléculas de hibridación comprende una secuencia de ácido nucleico que comprende una región de hibridación; (b) el primer y segundo oligonucleótidos con código de barras, en donde el primer oligonucleótido con código de barras hibrida con la región de hibridación de la primera molécula de hibridación y en donde el segundo oligonucleótido con código de barras hibrida con la región de hibridación de la segunda molécula de hibridación, en donde los oligonucleótidos con código de barras comprenden, cada uno, una región de código de barras; y (c) un resto de unión a células; en donde regiones de código de barras de los oligonucleótidos con código de barras primero y segundo de cada reactivo de código de barras multimérico de la biblioteca son diferentes a las regiones de código de barras de los oligonucleótidos con código de barras de al menos otros 9 reactivos de código de barras multiméricos de la biblioteca.

La biblioteca puede comprender al menos dos reactivos de códigos de barras multiméricos, cada uno de los cuales

comprende: (a) la primera y segunda moléculas de código de barras unidas entre sí, en donde cada una de las moléculas de código de barras comprende una secuencia de ácido nucleico que comprende una región de código de barras; (b) el primer y segundo oligonucleótidos con código de barras, en donde el primer oligonucleótido con código de barras comprende una región de código de barras hibridada con la región de código de barras de la primera molécula de código de barras, y en donde el segundo oligonucleótido con código de barras comprende una región de código de barras hibridada con la región de código de barras de la segunda molécula de código de barras; y (c) un resto de unión a células; en donde las regiones de código de barras del primer y segundo oligonucleótidos con código de barras de un primer reactivo de código de barras multimérico de la biblioteca son diferentes a las regiones de código de barras del primer y segundo oligonucleótidos con código de barras de un segundo reactivo de código de barras multimérico de la biblioteca.

Se puede unir un resto de unión a células a cada una de las moléculas de código de barras. Adicionalmente o como alternativa, se puede unir un resto de unión a células a cada uno de los oligonucleótidos con código de barras.

La biblioteca puede comprender al menos 10 reactivos de códigos de barras multiméricos, en donde cada reactivo de códigos de barras multimérico comprende: (a) la primera y segunda moléculas de código de barras comprendidas dentro de una molécula de ácido nucleico, en donde cada una de las moléculas de código de barras comprende una secuencia de ácido nucleico que comprende una región de código de barras; (b) el primer y segundo oligonucleótidos con código de barras, en donde el primer oligonucleótido con código de barras comprende una región de código de barras hibridada con la región de código de barras de la primera molécula de código de barras, y en donde el segundo oligonucleótido con código de barras comprende una región de código de barras hibridada con la región de código de barras de la segunda molécula de código de barras; y (c) un resto de unión a células; en donde regiones de código de barras de los oligonucleótidos con código de barras primero y segundo de cada reactivo de código de barras multimérico de la biblioteca son diferentes a las regiones de código de barras de los oligonucleótidos con código de barras de al menos otros 9 reactivos de código de barras multiméricos de la biblioteca.

En las bibliotecas, cada reactivo de código de barras multimérico puede estar comprendido dentro de un vehículo lipídico diferente (o separado). El vehículo lipídico puede ser una micela o un liposoma. Como alternativa, el vehículo lipídico puede adoptar cualquiera de las formas descritas en el presente documento.

La divulgación proporciona un kit para marcar ácidos nucleicos diana para secuenciar, en donde el kit comprende: a) una biblioteca de reactivos de códigos de barras multiméricos que comprenda al menos dos reactivos de códigos de barras multiméricos, en donde cada reactivo de código de barras multimérico comprende (i) las moléculas de código de barras primera y segunda unidas entre sí, en donde cada una de las moléculas de código de barras comprende una secuencia de ácido nucleico que comprende, opcionalmente en la dirección de 5' a 3', una región de adaptador y una región de código de barras, (ii) el primer y segundo oligonucleótidos con código de barras, en donde el primer oligonucleótido con código de barras comprende una región de código de barras hibridada con la región de código de barras de la primera molécula de código de barras, y en donde el segundo oligonucleótido con código de barras comprende una región de código de barras hibridada con la región de código de barras de la segunda molécula de código de barras; en donde las regiones de código de barras del primer y segundo oligonucleótidos con código de barras de un primer reactivo de código de barras multimérico de la biblioteca son diferentes a las regiones de código de barras del primer y segundo oligonucleótidos con código de barras de un segundo reactivo de código de barras multimérico de la biblioteca; y (b) los oligonucleótidos adaptadores primero y segundo para cada uno de los reactivos de códigos de barras multiméricos, en donde el primer oligonucleótido adaptador comprende una región adaptadora capaz de hibridar con la región adaptadora de la primera molécula de código de barras y en donde el segundo oligonucleótido adaptador comprende una región adaptadora capaz de hibridar con la región adaptadora de la segunda molécula de código de barras, y donde un resto está unido a cada uno de los oligonucleótidos adaptadores.

El kit puede ser para etiquetar ácidos nucleicos diana de (o en) al menos dos células para secuenciar.

Cada uno de los reactivos de códigos de barras multiméricos puede comprender un resto de unión a células. Se puede unir un resto de unión a células a cada una de las moléculas de código de barras. Se puede unir un resto de unión a células a cada uno de los oligonucleótidos con código de barras.

La divulgación proporciona un kit para marcar ácidos nucleicos diana para secuenciar, en donde el kit comprende: a) una biblioteca de reactivos de códigos de barras multiméricos que comprenda al menos dos reactivos de códigos de barras multiméricos, en donde cada reactivo de códigos de barras multimérico comprende un primer y un segundo oligonucleótidos con código de barras unidos por un soporte, en donde los oligonucleótidos con código de barras comprenden, cada uno, una región de código de barras y una región diana, y en donde las regiones de código de barras del primer y segundo oligonucleótidos con código de barras de un primer reactivo de código de barras multimérico de la biblioteca son diferentes a las regiones de código de barras del primer y segundo oligonucleótidos con código de barras de un segundo reactivo de código de barras multimérico de la biblioteca; y (b) un resto de unión a células para cada reactivo de código de barras multimérico en la biblioteca, en donde cada uno de dichos restos de unión a células es capaz de unirse a un reactivo de código de barras multimérico dentro de la biblioteca.

La divulgación proporciona un kit para marcar ácidos nucleicos diana para secuenciar, en donde el kit comprende: a) una biblioteca de reactivos de códigos de barras multiméricos que comprenda al menos dos reactivos de códigos de

barras multiméricos, en donde cada reactivo de códigos de barras multimérico comprende al menos un primer y un segundo oligonucleótidos con código de barras unidos por un soporte, en donde los oligonucleótidos con código de barras comprenden, cada uno, una región de código de barras y una región diana de poli(T), y en donde las regiones de código de barras del primer y segundo oligonucleótidos con código de barras de un primer reactivo de código de barras multimérico de la biblioteca son diferentes a las regiones de código de barras del primer y segundo oligonucleótidos con código de barras de un segundo reactivo de código de barras multimérico de la biblioteca; y (b) un resto de unión a células para cada reactivo de código de barras multimérico en la biblioteca, en donde cada uno de dichos restos de unión a células es capaz de unirse a un reactivo de código de barras multimérico dentro de la biblioteca.

- 10 La divulgación proporciona un kit para marcar ácidos nucleicos diana para secuenciar, en donde el kit comprende: a) una biblioteca de reactivos de códigos de barras multiméricos que comprenda al menos dos reactivos de códigos de barras multiméricos, en donde cada reactivo de códigos de barras multimérico comprende al menos un primer y un segundo oligonucleótidos con código de barras unidos por un soporte, en donde los oligonucleótidos con código de barras comprenden, cada uno, una región de código de barras y una región diana, y en donde las regiones de código de barras del primer y segundo oligonucleótidos con código de barras de un primer reactivo de código de barras multimérico de la biblioteca son diferentes a las regiones de código de barras del primer y segundo oligonucleótidos con código de barras de un segundo reactivo de código de barras multimérico de la biblioteca; (b) un resto de unión a células para cada reactivo de código de barras multimérico en la biblioteca, en donde cada uno de dichos restos de unión a células es capaz de unirse a un reactivo de código de barras multimérico dentro de la biblioteca; y (c) oligonucleótidos de bloqueo (por ejemplo, una solución de oligonucleótidos de bloqueo), en donde cada oligonucleótido de bloqueo comprende una secuencia complementaria a todo o parte de un oligonucleótido con código de barras, y/o comprende una secuencia complementaria a todo o parte de un ácido nucleico diana.

- 25 La divulgación proporciona un kit para marcar ácidos nucleicos diana para secuenciar, en donde el kit comprende: a) una biblioteca de reactivos de códigos de barras multiméricos que comprenda al menos dos reactivos de códigos de barras multiméricos, en donde cada reactivo de códigos de barras multimérico comprende al menos un primer y un segundo oligonucleótidos con código de barras unidos por un soporte, en donde los oligonucleótidos con código de barras comprenden, cada uno, una región de código de barras y una región diana de poli(T), y en donde las regiones de código de barras del primer y segundo oligonucleótidos con código de barras de un primer reactivo de código de barras multimérico de la biblioteca son diferentes a las regiones de código de barras del primer y segundo oligonucleótidos con código de barras de un segundo reactivo de código de barras multimérico de la biblioteca; (b) un resto de unión a células para cada reactivo de código de barras multimérico en la biblioteca, en donde cada uno de dichos restos de unión a células es capaz de unirse a un reactivo de código de barras multimérico dentro de la biblioteca; y (c) oligonucleótidos de bloqueo (por ejemplo, una solución de oligonucleótidos de bloqueo), en donde cada oligonucleótido de bloqueo comprende una secuencia complementaria a todo o parte de un oligonucleótido con código de barras, y/o comprende una secuencia complementaria a todo o parte de un ácido nucleico diana.

En cualquier kit que comprenda una biblioteca de reactivos de códigos de barras multiméricos y restos de unión a células, se pueden proporcionar dos o más restos de unión a células (por ejemplo, en una solución de restos de unión a células) por separado a una biblioteca de reactivos de códigos de barras multiméricos (por ejemplo, una solución de una biblioteca de reactivos de códigos de barras multiméricos).

- 40 En cualquier kit que comprenda una biblioteca de reactivos de códigos de barras multiméricos y restos de unión a células, la biblioteca de reactivos de códigos de barras multiméricos y los restos de unión a células se pueden proporcionar juntos en una única solución.

- 45 En cualquier kit que comprenda una biblioteca de reactivos de códigos de barras multiméricos, restos de unión celular y oligonucleótidos bloqueantes, cada uno de los tres componentes del kit puede administrarse por separado (por ejemplo, en una solución separada) a los otros dos componentes del kit. Opcionalmente, se pueden proporcionar dos componentes del kit juntos (por ejemplo, en una sola solución). Opcionalmente, se pueden proporcionar los tres componentes del kit juntos (por ejemplo, en una sola solución).

- 50 La divulgación proporciona un kit para marcar ácidos nucleicos diana para secuenciar, en donde el kit comprende: a) una biblioteca de reactivos de códigos de barras multiméricos que comprenda al menos 10 reactivos de códigos de barras multiméricos, en donde cada reactivo de código de barras multimérico comprende (i) las moléculas de código de barras primera y segunda comprendidas dentro de una molécula de ácido nucleico, en donde cada una de las moléculas de código de barras comprende una secuencia de ácido nucleico que comprende, opcionalmente en la dirección de 5' a 3', una región de adaptador y una región de código de barras, (ii) el primer y segundo oligonucleótidos con código de barras, en donde el primer oligonucleótido con código de barras comprende una región de código de barras hibridada con la región de código de barras de la primera molécula de código de barras, y en donde el segundo oligonucleótido con código de barras comprende una región de código de barras hibridada con la región de código de barras de la segunda molécula de código de barras; en donde regiones de código de barras de los oligonucleótidos con código de barras primero y segundo de cada reactivo de código de barras multimérico de la biblioteca son diferentes a las regiones de código de barras de los oligonucleótidos con código de barras de al menos otros 9 reactivos de código de barras multiméricos de la biblioteca; y (b) los oligonucleótidos adaptadores primero y segundo para cada uno de los reactivos de códigos de barras multiméricos, en donde el primer oligonucleótido adaptador comprende una

región adaptadora capaz de hibridar con la región adaptadora de la primera molécula de código de barras y en donde el segundo oligonucleótido adaptador comprende una región adaptadora capaz de hibridar con la región adaptadora de la segunda molécula de código de barras, y donde un resto está unido a cada uno de los oligonucleótidos adaptadores.

- 5 En los kits, los oligonucleótidos adaptadores para cada reactivo de código de barras multimérico pueden estar comprendidos dentro de un vehículo lipídico diferente (o separado). El vehículo lipídico puede ser una micela o un liposoma. Como alternativa, el vehículo lipídico puede adoptar cualquiera de las formas descritas en el presente documento. Cada uno de los vehículos de lípidos puede comprender además un reactivo de código de barras multimérico, por ejemplo, el primer vehículo de lípidos comprende el primer reactivo de código de barras multimérico y los oligonucleótidos adaptadores para el primer reactivo de código de barras multimérico.

En las bibliotecas o kits, los reactivos de códigos de barras pueden comprender, cada uno, un soporte sólido o un soporte semisólido, y en donde un resto de unión a células está unido al soporte sólido o semisólido (por ejemplo, mediante un enlace covalente o no covalente).

- 15 Se puede unir un resto de unión a células a cada oligonucleótido con código de barras, molécula de hibridación, molécula de código de barras y/u oligonucleótido adaptador mediante una molécula enlazadora. Opcionalmente, dicho enlazador puede ser un enlazador flexible. Opcionalmente, dicho enlazador puede estar compuesto por una o más unidades de etilenglicol y/o polietilenglicol, tales como hexaetilenglicol o pentaetilenglicol. Opcionalmente, dicho enlazador puede estar compuesto por uno o más grupos etilo, tales como un espaciador C3 (tres carbonos), C6, C12 o C18. Opcionalmente, se puede utilizar cualquier otro espaciador.

- 20 El resto (o restos) de unión a células puede ser capaz de iniciar la endocitosis al unirse a una membrana celular.

El resto de unión a células puede comprender uno o más restos seleccionados de: un péptido, un péptido que penetra en las células, un aptámero, un aptámero de ADN, un aptámero de ARN, un anticuerpo, un fragmento de anticuerpo, un fragmento de anticuerpo de cadena ligera, un fragmento variable monocatenario (scFv), un lípido, un derivado lipídico, un fosfolípido, un ácido graso, un triglicérido, un glicerolípido, un glicerofosfolípido, un esfingolípido, un sacarolípido, un policétido, un lípido catiónico, un polímero catiónico, poli(etilen)glicol, espermina, un derivado o análogo de espermina, una polilisina, un derivado o análogo de polilisina, polietilenimina, dietilaminoetil (DEAE)-dextrano, colesterol, un resto de esteroles, una molécula catiónica, una molécula hidrófoba y una molécula anfifílica.

- 30 El resto de unión celular puede interactuar con una o más moléculas específicas en la superficie celular (como en el caso de, por ejemplo, un anticuerpo, un fragmento de anticuerpo y un aptámero). Como alternativa o adicionalmente, el resto de unión a células puede alterar la carga total y/o la distribución de carga de los reactivos de códigos de barras multiméricos (como en el caso de, por ejemplo, un polímero catiónico). Como alternativa o adicionalmente, el resto de unión a células puede alterar el carácter lipófilo/lipófilo y/o hidrófilo/hidrófilo y/o el equilibrio de los reactivos de códigos de barras multiméricos (como en el caso de, por ejemplo, un lípido o colesterol).

- 35 El resto de unión a células puede ser una molécula que tiene una carga neta positiva en una solución que comprende una célula y que permite la unión de un reactivo de código de barras multimérico a la célula.

Un reactivo de códigos de barras multimérico, oligonucleótido adaptador, oligonucleótido con código de barras, molécula de hibridación o molécula de código de barras puede comprender al menos 2, al menos 3, al menos 5, al menos 10, al menos 20, al menos 50, al menos 100, al menos 500, o al menos 1000 restos de unión a células.

- 40 Se puede unir un resto de unión a células a un reactivo de códigos de barras multimérico, oligonucleótido adaptador, oligonucleótido con código de barras, molécula de hibridación o molécula de código de barras por un enlace covalente o por un enlace no covalente.

La invención proporciona un método para preparar una muestra de ácido nucleico para secuenciar tal como se define en las reivindicaciones adjuntas.

- 45 La divulgación proporciona un método para preparar una muestra de ácido nucleico para secuenciar, en donde la muestra comprende una célula, y donde el método comprende las etapas de:

(a) poner en contacto la muestra con un reactivo de código de barras multimérico, en donde el reactivo de código de barras multimérico comprende una primera y segunda regiones de código de barras unidas entre sí y un resto de unión a células, en donde cada región de código de barras comprende

- 50 una secuencia de ácido nucleico, en donde el resto de unión a células del reactivo de código de barras multimérico se une a la membrana celular de la célula y la primera y segunda regiones de código de barras del reactivo de código de barras multimérico se internalizan en la célula; y

(b) añadir secuencias de códigos de barras a cada una de las subsecuencias primera y segunda de un ácido nucleico diana de la célula para producir moléculas de ácido nucleico diana con código de barras primera y segunda para la célula, en donde la primera molécula de ácido nucleico diana con código de barras comprende la secuencia de ácido

nucleico de la primera región de código de barras del reactivo de código de barras multimérico y la segunda molécula de ácido nucleico diana con código de barras comprende la secuencia de ácido nucleico de la segunda región de código de barras del reactivo de código de barras multimérico.

5 La divulgación proporciona un método para preparar una muestra de ácido nucleico para secuenciar, en donde la muestra comprende al menos dos células, y donde el método comprende las etapas de:

(a) poner en contacto la muestra con una biblioteca que comprende al menos dos reactivos de códigos de barras multiméricos, en donde cada reactivo de código de barras multimérico comprende una primera y segunda regiones de código de barras unidas entre sí y un resto de unión a células, en donde cada región de código de barras comprende una secuencia de ácido nucleico y en donde la primera y segunda regiones de código de barras de un primer reactivo de código de barras multimérico son diferentes a la primera y segunda regiones de código de barras de un segundo reactivo de código de barras multimérico de la biblioteca, en donde el resto de unión a células del primer reactivo de código de barras multimérico de la biblioteca se une a la membrana celular de una primera celda de la muestra y la primera y segunda regiones de código de barras del primer reactivo de código de barras multimérico se internalizan en la primera celda, y donde el resto de unión a células del segundo reactivo de código de barras multimérico de la biblioteca se une a la membrana celular de una segunda célula de la muestra y la primera y segunda regiones de código de barras del segundo reactivo de código de barras multimérico se internalizan en la segunda célula; y

(b) añadir secuencias de códigos de barras a cada una de las subsecuencias primera y segunda de un ácido nucleico diana de la primera célula para producir moléculas de ácido nucleico diana con código de barras primera y segunda para la primera célula, en donde la primera molécula de ácido nucleico diana con código de barras comprende la secuencia de ácido nucleico de la primera región de código de barras del primer reactivo de código de barras multimérico y la segunda molécula de ácido nucleico diana con código de barras comprende la secuencia de ácido nucleico de la segunda región de código de barras del primer reactivo de código de barras multimérico, y agregar secuencias de códigos de barras a cada una de las subsecuencias primera y segunda de un ácido nucleico diana de la segunda célula para producir moléculas de ácido nucleico diana con código de barras primera y segunda a partir de la segunda célula, en donde la primera molécula de ácido nucleico diana con código de barras comprende la secuencia de ácido nucleico de la primera región de código de barras del segundo reactivo de código de barras multimérico y la segunda molécula de ácido nucleico diana con código de barras comprende la secuencia de ácido nucleico de la segunda región de código de barras del segundo reactivo de código de barras multimérico.

El método puede comprender las etapas de: (a) poner en contacto la muestra con una biblioteca que comprende el primer y segundo reactivos de códigos de barras multiméricos, en donde cada reactivo de código de barras multimérico comprende moléculas de código de barras primera y segunda unidas entre sí y un resto de unión a células, en donde cada una de las moléculas de código de barras comprende una secuencia de ácido nucleico que comprende una región de código de barras y una región adaptadora y en donde la primera y segunda regiones de código de barras de un primer reactivo de código de barras multimérico son diferentes a la primera y segunda regiones de código de barras de un segundo reactivo de código de barras multimérico de la biblioteca, y en donde el resto de unión a células del primer reactivo de código de barras multimérico de la biblioteca se une a la membrana celular de una primera célula de la muestra y la primera y segunda moléculas de código de barras del primer reactivo de código de barras multimérico se internalizan en la primera célula, y en donde el resto de unión a células del segundo reactivo de código de barras multimérico de la biblioteca se une a la membrana celular de una segunda célula de la muestra y la primera y segunda moléculas de código de barras del segundo reactivo de código de barras multimérico se internalizan en la segunda célula; (b) adjuntar una secuencia de acoplamiento a cada una de las subsecuencias primera y segunda de un ácido nucleico diana de una primera célula, y adjuntar una secuencia de acoplamiento a cada una de las subsecuencias primera y segunda de un ácido nucleico diana de una segunda célula; (c) para cada uno de los reactivos de códigos de barras multiméricos, hibridar la secuencia de acoplamiento de la primera subsecuencia con la región adaptadora de la primera molécula de código de barras, e hibridar la secuencia de acoplamiento de la segunda subsecuencia con la región adaptadora de la segunda molécula de código de barras; y (d) añadir secuencias de códigos de barras a cada una de las subsecuencias primera y segunda del ácido nucleico diana de la primera célula para producir moléculas de ácido nucleico diana con código de barras primera y segunda para la primera célula, donde la primera molécula de ácido nucleico diana con código de barras comprende la secuencia de ácido nucleico de la primera región de código de barras de la primera molécula de código de barras del primer reactivo de código de barras multimérico y la segunda molécula de ácido nucleico diana con código de barras comprende la secuencia de ácido nucleico de la segunda región de código de barras de la primera molécula de código de barras del primer reactivo de código de barras multimérico, y adjuntar secuencias de códigos de barras a cada una de las subsecuencias primera y segunda de un ácido nucleico diana de la segunda célula para producir moléculas de ácido nucleico diana con código de barras primera y segunda a partir de la segunda célula, en donde la primera molécula de ácido nucleico diana con código de barras comprende la secuencia de ácido nucleico de la primera región de código de barras de la segunda molécula de código de barras del primer reactivo de código de barras multimérico y la segunda molécula de ácido nucleico diana con código de barras comprende la secuencia de ácido nucleico de la segunda región de código de barras de la segunda molécula de código de barras del primer reactivo de código de barras multimérico.

60 La divulgación proporciona un método para preparar una muestra de ácido nucleico para secuenciar, en donde la muestra comprende una célula, y donde el método comprende las etapas de: (a) poner en contacto la muestra con un reactivo de código de barras multimérico, en donde el reactivo de códigos de barras multimérico comprende un primer

y un segundo oligonucleótidos con código de barras unidos entre sí y un resto de unión a células, en donde los oligonucleótidos con código de barras comprenden, cada uno, una región de código de barras, y en donde el resto de unión a células del reactivo de código de barras multimérico se une a la membrana celular de la célula y el primer y segundo oligonucleótidos con código de barras del reactivo de código de barras multimérico se internalizan en la célula; y (b) hibridar o ligar el primer y segundo oligonucleótidos con código de barras del reactivo de código de barras multimérico a la primera y segunda subsecuencia de un ácido nucleico diana de la célula para producir una primera y segunda moléculas de ácido nucleico diana con código de barras.

La divulgación proporciona un método para preparar una muestra de ácido nucleico para secuenciar, en donde la muestra comprende al menos dos células, y donde el método comprende las etapas de: (a) poner en contacto la muestra con una biblioteca que comprende al menos dos reactivos de códigos de barras multiméricos, en donde cada reactivo de códigos de barras multimérico comprende un primer y un segundo oligonucleótidos con código de barras unidos entre sí y un resto de unión a células, en donde los oligonucleótidos con código de barras comprenden, cada uno, una región de código de barras y en donde las regiones de código de barras del primer y segundo oligonucleótidos con código de barras de un primer reactivo de código de barras multimérico de la biblioteca son diferentes a las regiones de código de barras del primer y segundo oligonucleótidos con código de barras de un segundo reactivo de código de barras multimérico de la biblioteca, en donde el resto de unión a células de un primer reactivo de código de barras multimérico de la biblioteca se une a la membrana celular de una primera célula de la muestra y el primer y segundo oligonucleótidos con código de barras del primer reactivo de código de barras multimérico se internalizan en la primera célula, y donde el resto de unión celular de un segundo reactivo de código de barras multimérico de la biblioteca se une a la membrana celular de una segunda célula de la muestra y el primer y segundo oligonucleótidos con código de barras del segundo reactivo de código de barras multimérico se internalizan en la segunda célula; y (b) hibridar o ligar el primer y segundo oligonucleótidos con código de barras del primer reactivo de código de barras multimérico a la primera y segunda subsecuencia de un ácido nucleico diana de la primera célula para producir una primera y segunda moléculas de ácido nucleico diana con código de barras, e hibridar o ligar el primer y segundo oligonucleótidos con código de barras del segundo reactivo de código de barras multimérico a la primera y segunda subsecuencias de un ácido nucleico diana de la segunda célula para producir una primera y segunda moléculas de ácido nucleico diana con código de barras.

En los métodos, la etapa de unión e internalización celular puede comprender un período de incubación, en donde dicha incubación tiene lugar durante al menos 5 segundos, al menos 10 segundos, al menos 30 segundos, al menos 60 segundos, al menos 2 minutos, al menos 5 minutos, al menos 10 minutos, al menos 15 minutos, al menos 30 minutos, al menos 60 minutos, al menos 2 horas o al menos 4 horas, opcionalmente durante 5 segundos a 4 horas, 10 segundos a 2 horas, 30 segundos a 60 minutos, 60 segundos a 30 minutos, 2 a 15 minutos o 5 a 10 minutos. Opcionalmente, dicha incubación tiene lugar a una temperatura de al menos 4 grados Celsius, al menos 12 grados Celsius, al menos 20 grados Celsius, al menos 30 grados Celsius, al menos 37 grados Celsius, al menos 40 grados Celsius, al menos 45 grados Celsius, o al menos 50 grados Celsius, opcionalmente de 4 a 50 grados Celsius, 12 a 45 grados Celsius, 20 a 40 grados Celsius o 30 a 37 grados Celsius.

La etapa de hibridación o ligado (etapa (b)) puede comprender: (i) hibridar el primer y segundo oligonucleótidos con código de barras del primer reactivo de código de barras multimérico con la primera y segunda subsecuencia de un ácido nucleico diana de la primera célula, e hibridar el primer y segundo oligonucleótidos con código de barras del segundo reactivo de código de barras multimérico con la primera y segunda subsecuencias de un ácido nucleico diana de la segunda célula; y (ii) extender el primer y segundo oligonucleótidos con código de barras del primer reactivo de código de barras multimérico para producir una primera y segunda moléculas de ácido nucleico diana con código de barras diferentes y extender el primer y segundo oligonucleótidos con código de barras del segundo reactivo de código de barras multimérico para producir una primera y segunda molécula de ácido nucleico diana con código de barras diferente, en donde cada una de las moléculas de ácido nucleico diana con código de barras comprende al menos un nucleótido sintetizado a partir del ácido nucleico diana como molde.

Se puede unir un resto de unión a células a cada uno de los oligonucleótidos con código de barras.

Cada uno de los reactivos de códigos de barras multiméricos puede comprender: (i) la primera y segunda moléculas de hibridación unidas entre sí, en donde cada una de las moléculas de hibridación comprende una secuencia de ácido nucleico que comprende una región de hibridación; y (ii) primer y segundo oligonucleótidos con código de barras, en donde el primer oligonucleótido con código de barras hibrida con la región de hibridación de la primera molécula de hibridación y en donde el segundo oligonucleótido con código de barras hibrida con la región de hibridación de la segunda molécula de hibridación; opcionalmente, en donde el primer reactivo de código de barras multimérico se internaliza en la primera célula y el segundo reactivo de código de barras multimérico se internaliza en la segunda célula.

Se puede unir un resto de unión a células a cada una de las moléculas de hibridación.

Cada uno de los reactivos de códigos de barras multiméricos puede comprender: (i) la primera y segunda moléculas de código de barras unidas entre sí, en donde cada una de las moléculas de código de barras comprende una secuencia de ácido nucleico que comprende una región de código de barras; y (ii) primer y segundo oligonucleótidos con código de barras, en donde el primer oligonucleótido con código de barras comprende una región de código de

barras hibridada con la región de código de barras de la primera molécula de código de barras, y en donde el segundo oligonucleótido con código de barras comprende una región de código de barras hibridada con la región de código de barras de la segunda molécula de código de barras; opcionalmente, en donde el primer reactivo de código de barras multimérico se internaliza en la primera célula y el segundo reactivo de código de barras multimérico se internaliza en la segunda célula.

Se puede unir un resto de unión a células a cada una de las moléculas de código de barras.

En los métodos, el primer reactivo de código de barras multimérico puede estar comprendido dentro de un primer vehículo de lípidos y el segundo reactivo de código de barras multimérico puede estar comprendido dentro de un segundo vehículo de lípidos, opcionalmente, en donde en la etapa (a) el primer vehículo de lípidos se fusiona con la membrana celular de la primera célula y el primer y segundo oligonucleótidos con código de barras del primer reactivo de código de barras multimérico se internalizan en la primera célula, y el segundo vehículo de lípidos se fusiona con la membrana celular de la segunda célula y los oligonucleótidos con código de barras primero y segundo del primer reactivo de código de barras multimérico se internalizan en la segunda célula. Opcionalmente, los oligonucleótidos con código de barras se liberan en la célula, por ejemplo, en el citoplasma. El vehículo lipídico puede ser un liposoma o una micela. Como alternativa, el vehículo lipídico puede adoptar cualquiera de las formas descritas en el presente documento.

La divulgación proporciona un método para preparar una muestra de ácido nucleico para secuenciar, en donde la muestra comprende una célula, y donde el método comprende las etapas de: (a) poner en contacto la muestra con un reactivo de código de barras multimérico, en donde el reactivo de códigos de barras multimérico comprende: (i) la primera y segunda moléculas de código de barras unidas entre sí, en donde cada una de las moléculas de código de barras comprende una secuencia de ácido nucleico que comprende, opcionalmente en la dirección de 5' a 3', una región adaptadora y una región de código de barras, y (ii) el primer y segundo oligonucleótidos con código de barras, en donde el primer oligonucleótido con código de barras comprende una región de código de barras hibridada con la región de código de barras de la primera molécula de código de barras y en donde el segundo oligonucleótido con código de barras comprende una región de código de barras hibridada con la región de código de barras de la segunda molécula de código de barras; en donde la muestra se pone en contacto además con el primer y segundo oligonucleótidos adaptadores para el reactivo de códigos de barras multimérico, en donde el primer y segundo oligonucleótidos adaptadores comprenden, cada uno, una región adaptadora, en donde un resto de unión celular se une a cada uno de los oligonucleótidos adaptadores, y en donde los restos de unión celular del primer y segundo oligonucleótidos adaptadores se unen a la membrana celular de la célula y al primer y segundo oligonucleótidos adaptadores para el primer reactivo de código de barras multimérico se internalizan en la célula; (b) hibridar o ligar el primer y segundo oligonucleótidos adaptadores para el reactivo de código de barras multimérico a subsecuencias de un ácido nucleico diana de la primera célula; (c) hibridar la región adaptadora del primer oligonucleótido adaptador con la región adaptadora de la primera molécula de código de barras, e hibridar la región adaptadora del segundo oligonucleótido adaptador con la región adaptadora de la segunda molécula de código de barras; y (d) ligar el extremo 3' del primer oligonucleótido con código de barras al extremo 5' del primer oligonucleótido adaptador para producir una primera molécula de ácido nucleico diana con código de barras y ligar el extremo 3' del segundo oligonucleótido con código de barras al extremo 5' del segundo oligonucleótido adaptador para producir una segunda molécula de ácido nucleico diana con código de barras. En los métodos, la etapa (b) puede comprender hibridar el primer y segundo oligonucleótidos adaptadores con subsecuencias de un ácido nucleico diana de la célula, y en donde: (i) la etapa (d) comprende ligar el extremo 3' del primer oligonucleótido con código de barras al extremo 5' del primer oligonucleótido adaptador para producir un primer oligonucleótido adaptador con código de barras y ligar el extremo 3' del segundo oligonucleótido con código de barras al extremo 5' del segundo oligonucleótido adaptador para producir un segundo oligonucleótido adaptador con código de barras, y extender el primer y segundo oligonucleótidos adaptadores con código de barras para producir una primera y segunda moléculas de ácido nucleico diana con códigos de barras diferentes, cada una de las cuales comprende al menos un nucleótido sintetizado a partir del ácido nucleico diana como molde, o (ii) antes de la etapa (d), el método comprende extender el primer y segundo oligonucleótidos adaptadores para producir una primera y una segunda moléculas de ácido nucleico diana diferentes, cada una de las cuales comprende al menos un nucleótido sintetizado a partir del ácido nucleico diana como molde.

La divulgación proporciona un método para preparar una muestra de ácido nucleico para secuenciar, en donde la muestra comprende al menos dos células, y donde el método comprende las etapas de: (a) poner en contacto la muestra con una biblioteca que comprende el primer y segundo reactivos de códigos de barras multiméricos, en donde cada reactivo de códigos de barras multimérico comprende: (i) la primera y segunda moléculas de código de barras unidas entre sí, en donde cada una de las moléculas de código de barras comprende una secuencia de ácido nucleico que comprende, opcionalmente en la dirección de 5' a 3', una región adaptadora y una región de código de barras, y (ii) el primer y segundo oligonucleótidos con código de barras, en donde el primer oligonucleótido con código de barras comprende una región de código de barras hibridada con la región de código de barras de la primera molécula de código de barras y en donde el segundo oligonucleótido con código de barras comprende una región de código de barras hibridada con la región de código de barras de la segunda molécula de código de barras, y en donde las regiones de código de barras del primer y segundo oligonucleótidos con código de barras del primer reactivo de código de barras multimérico de la biblioteca son diferentes de las regiones de código de barras del primer y segundo oligonucleótidos con código de barras del segundo reactivo de código de barras multimérico de la biblioteca; en donde la muestra se pone en contacto además con el primer y segundo oligonucleótidos adaptadores para cada uno de los

reactivos de códigos de barras multiméricos, en donde el primer y segundo oligonucleótidos adaptadores comprenden, cada uno, una región adaptadora, en donde un resto de unión celular está unido a cada uno de los oligonucleótidos adaptadores, y en donde los restos de unión celular del primer y segundo oligonucleótidos adaptadores para el primer reactivo de código de barras multimérico se unen a la membrana celular de una primera célula de la muestra y el primer y segundo oligonucleótidos adaptadores para el primer reactivo de código de barras multimérico se internalizan en la primera célula, y en donde los restos de unión a células del primer y segundo oligonucleótidos adaptadores para el segundo reactivo de código de barras multimérico se unen a la membrana celular de una segunda célula de la muestra y el primer y segundo oligonucleótidos adaptadores para el segundo reactivo de código de barras multimérico se internalizan en la segunda célula; (b) hibridar o ligar el primer y segundo oligonucleótidos adaptadores con el primer reactivo de código de barras multimérico a subsecuencias de un ácido nucleico diana de la primera célula, e hibridar o ligar el primer y segundo oligonucleótidos adaptadores con el segundo reactivo de código de barras multimérico a subsecuencias de un ácido nucleico diana de la segunda célula; (c) para cada uno de los reactivos de códigos de barras multiméricos, hibridar la región adaptadora del primer oligonucleótido adaptador con la región adaptadora de la primera molécula de código de barras, e hibridar la región adaptadora del segundo oligonucleótido adaptador con la región adaptadora de la segunda molécula de código de barras; y (d) para cada uno de los reactivos de códigos de barras multiméricos, ligar el extremo 3' del primer oligonucleótido con código de barras al extremo 5' del primer oligonucleótido adaptador para producir una primera molécula de ácido nucleico diana con código de barras y ligar el extremo 3' del segundo oligonucleótido con código de barras al extremo 5' del segundo oligonucleótido adaptador para producir una segunda molécula de ácido nucleico diana con código de barras.

En los métodos, la etapa (b) puede comprender hibridar el primer y segundo oligonucleótidos adaptadores con el primer reactivo de código de barras multimérico a subsecuencias de un ácido nucleico diana de la primera célula, e hibridar el primer y segundo oligonucleótidos adaptadores con el segundo reactivo de código de barras multimérico a subsecuencias de un ácido nucleico diana de la segunda célula, en donde: (i) para cada uno de los reactivos de códigos de barras multiméricos, la etapa (d) comprende ligar el extremo 3' del primer oligonucleótido con código de barras al extremo 5' del primer oligonucleótido adaptador para producir un primer oligonucleótido adaptador con código de barras y ligar el extremo 3' del segundo oligonucleótido con código de barras al extremo 5' del segundo oligonucleótido adaptador para producir un segundo oligonucleótido adaptador con código de barras, y extender el primer y segundo oligonucleótidos adaptadores con código de barras para producir una primera y segunda moléculas de ácido nucleico diana con códigos de barras diferentes, cada una de las cuales comprende al menos un nucleótido sintetizado a partir del ácido nucleico diana como molde, o (ii) para cada uno de los reactivos de códigos de barras multiméricos, antes de la etapa (d), el método comprende extender el primer y segundo oligonucleótidos adaptadores para producir una primera y una segunda moléculas de ácido nucleico diana diferentes, cada una de las cuales comprende al menos un nucleótido sintetizado a partir del ácido nucleico diana como molde.

Cada uno de los reactivos de códigos de barras multiméricos puede comprender un resto de unión a células, en donde opcionalmente: (i) el resto de unión a células del primer reactivo de código de barras multimérico se une a la membrana celular de la primera célula de la muestra y el reactivo de código de barras multimérico se internaliza en la primera célula y (ii) el resto de unión a células del segundo reactivo de código de barras multimérico se une a la membrana celular de la segunda célula de la muestra y el segundo reactivo de código de barras multimérico se internaliza en la segunda célula.

Se puede unir un resto de unión a células a cada una de las moléculas de código de barras. Adicionalmente o como alternativa, se puede unir un resto de unión a células a cada uno de los oligonucleótidos con código de barras.

En los métodos, el primer y segundo oligonucleótidos adaptadores para el primer reactivo de códigos de barras multimérico pueden estar comprendidos dentro de un primer vehículo de lípidos y los oligonucleótidos adaptadores primero y segundo para el segundo reactivo de códigos de barras multimérico pueden estar comprendidos dentro de un segundo vehículo de lípidos, opcionalmente en donde en la etapa (a) el primer vehículo lipídico se fusiona con la membrana celular de la primera célula y el primer y segundo oligonucleótidos adaptadores para el primer reactivo de código de barras multimérico se internalizan en la primera célula, y el segundo vehículo lipídico se fusiona con la membrana celular de la segunda célula y los oligonucleótidos adaptadores primero y segundo para el segundo reactivo de código de barras multimérico se internalizan en la segunda célula. Opcionalmente, los oligonucleótidos adaptadores se liberan en la célula, por ejemplo, en el citoplasma.

El primer vehículo de lípidos puede comprender además el primer reactivo de código de barras multimérico y el segundo vehículo de lípidos puede comprender además el segundo reactivo de código de barras multimérico.

El vehículo lipídico puede ser un liposoma o una micela. Como alternativa, el vehículo lipídico puede adoptar cualquiera de las formas descritas en el presente documento.

Se puede unir un resto de unión a células a un reactivo de códigos de barras multimérico, oligonucleótido adaptador, oligonucleótido con código de barras, molécula de hibridación o molécula de código de barras por un enlace covalente o por un enlace no covalente.

Se puede unir un resto de unión a células a cada oligonucleótido con código de barras, molécula de hibridación, molécula de código de barras y/u oligonucleótido adaptador mediante una molécula enlazadora. Opcionalmente, dicho

enlazador puede ser un enlazador flexible. Opcionalmente, dicho enlazador puede estar compuesto por una o más unidades de etilenglicol y/o polietilenglicol, tales como hexaetilenglicol o pentaetilenglicol. Opcionalmente, dicho enlazador puede estar compuesto por uno o más grupos etilo, tales como un espaciador C3 (tres carbonos), C6, C12 o C18. Opcionalmente, se puede utilizar cualquier otro espaciador.

- 5 El resto (o restos) de unión a células puede ser capaz de iniciar la endocitosis al unirse a una membrana celular.

El resto de unión a células puede comprender uno o más restos seleccionados de: un péptido, un péptido que penetra en las células, un aptámero, un aptámero de ADN, un aptámero de ARN, un anticuerpo, un fragmento de anticuerpo, un fragmento de anticuerpo de cadena ligera, un fragmento variable monocatenario (scFv), un lípido, un derivado lipídico, un fosfolípido, un ácido graso, un triglicérido, un glicerolípido, un glicerofosfolípido, un esfingolípido, un sacarolípido, un policétido, un lípido catiónico, un polímero catiónico, poli(etilen)glicol, espermina, un derivado o análogo de espermina, una polilisina, un derivado o análogo de polilisina, polietilenimina, dietilaminoetil (DEAE)-dextrano, colesterol, un resto de esteroles, una molécula catiónica, una molécula hidrófoba y una molécula anfifílica.

- 10 El resto de unión a células puede interactuar con una o más moléculas específicas en la superficie o membrana celular (como en el caso de, por ejemplo, un anticuerpo, un fragmento de anticuerpo y un aptámero). Como alternativa o adicionalmente, el resto de unión a células puede alterar la carga total y/o la distribución de carga de los reactivos de códigos de barras multiméricos (como en el caso de, por ejemplo, un polímero catiónico). Como alternativa o adicionalmente, el resto de unión a células puede alterar el carácter lipófilo/lipófilo y/o hidrófilo/hidrófilo y/o el equilibrio de los reactivos de códigos de barras multiméricos (como en el caso de, por ejemplo, un lípido o colesterol).

- 15 El resto de unión a células puede ser una molécula que tiene una carga neta positiva en una solución que comprende una célula y que permite la unión de un reactivo de código de barras multimérico a la célula.

Un reactivo de códigos de barras multimérico, oligonucleótido adaptador, oligonucleótido con código de barras, molécula de hibridación o molécula de código de barras puede comprender al menos 2, al menos 3, al menos 5, al menos 10, al menos 20, al menos 50, al menos 100, al menos 500, o al menos 1000 restos de unión a células.

- 25 La divulgación proporciona un método para preparar una muestra de ácido nucleico para secuenciar, en donde la muestra comprende al menos dos células, y donde el método comprende las etapas de: (a) poner en contacto la muestra con una biblioteca que comprende al menos dos reactivos de códigos de barras multiméricos, en donde cada reactivo de código de barras multimérico comprende una primera y segunda regiones de código de barras unidas entre sí, en donde cada región de código de barras comprende una secuencia de ácido nucleico y en donde la primera y segunda regiones de código de barras de un primer reactivo de código de barras multimérico son diferentes a la primera y segunda regiones de código de barras de un segundo reactivo de código de barras multimérico de la biblioteca; (b) transferir la primera y segunda regiones de código de barras del primer reactivo de código de barras multimérico de la biblioteca a una primera célula de la muestra y transferir la primera y segunda regiones de código de barras del segundo reactivo de código de barras multimérico de la biblioteca a una segunda célula de la muestra; y (c) añadir secuencias de códigos de barras a cada una de las subsecuencias primera y segunda de un ácido nucleico diana de la primera célula para producir moléculas de ácido nucleico diana con código de barras primera y segunda para la primera célula, en donde la primera molécula de ácido nucleico diana con código de barras comprende la secuencia de ácido nucleico de la primera región de código de barras del primer reactivo de código de barras multimérico y la segunda molécula de ácido nucleico diana con código de barras comprende la secuencia de ácido nucleico de la segunda región de código de barras del primer reactivo de código de barras multimérico, y agregar secuencias de códigos de barras a cada una de las subsecuencias primera y segunda de un ácido nucleico diana de la segunda célula para producir moléculas de ácido nucleico diana con código de barras primera y segunda a partir de la segunda célula, en donde la primera molécula de ácido nucleico diana con código de barras comprende la secuencia de ácido nucleico de la primera región de código de barras del segundo reactivo de código de barras multimérico y la segunda molécula de ácido nucleico diana con código de barras comprende la secuencia de ácido nucleico de la segunda región de código de barras del segundo reactivo de código de barras multimérico.

- El método para preparar una muestra de ácido nucleico para secuenciar, en donde la muestra comprende al menos dos células, puede comprender las etapas de: (a) poner en contacto la muestra con una biblioteca que comprende al menos dos reactivos de códigos de barras multiméricos, en donde cada reactivo de códigos de barras multimérico comprende un primer y un segundo oligonucleótidos con código de barras unidos entre sí, en donde los oligonucleótidos con código de barras comprenden, cada uno, una región de código de barras y en donde las regiones de código de barras del primer y segundo oligonucleótidos con código de barras de un primer reactivo de código de barras multimérico de la biblioteca son diferentes a las regiones de código de barras del primer y segundo oligonucleótidos con código de barras de un segundo reactivo de código de barras multimérico de la biblioteca; (b) transferir el primer y segundo oligonucleótidos con código de barras del primer reactivo de código de barras multimérico de la biblioteca a una primera célula de la muestra y transferir el primer y segundo oligonucleótidos con código de barras del segundo reactivo de código de barras multimérico de la biblioteca a una segunda célula de la muestra; y (c) hibridar o ligar el primer y segundo oligonucleótidos con código de barras del primer reactivo de código de barras multimérico a la primera y segunda subsecuencia de un ácido nucleico diana de la primera célula para producir una primera y segunda moléculas de ácido nucleico diana con código de barras, e hibridar o ligar el primer y segundo oligonucleótidos con código de barras del segundo reactivo de código de barras multimérico a la primera y segunda

subsecuencias de un ácido nucleico diana de la segunda célula para producir una primera y segunda moléculas de ácido nucleico diana con código de barras.

En los métodos, la etapa de hibridación o ligado (etapa (c)) puede comprender: (i) hibridar el primer y segundo oligonucleótidos con código de barras del primer reactivo de código de barras multimérico con la primera y segunda subsecuencia de un ácido nucleico diana de la primera célula, e hibridar el primer y segundo oligonucleótidos con código de barras del segundo reactivo de código de barras multimérico con la primera y segunda subsecuencias de un ácido nucleico diana de la segunda célula; y (ii) extender el primer y segundo oligonucleótidos con código de barras del primer reactivo de código de barras multimérico para producir una primera y segunda moléculas de ácido nucleico diana con código de barras diferentes y extender el primer y segundo oligonucleótidos con código de barras del segundo reactivo de código de barras multimérico para producir una primera y segunda molécula de ácido nucleico diana con código de barras diferente, en donde cada una de las moléculas de ácido nucleico diana con código de barras comprende al menos un nucleótido sintetizado a partir del ácido nucleico diana como molde.

En los métodos, cada uno de los reactivos de códigos de barras multiméricos puede comprender: (i) la primera y segunda moléculas de hibridación unidas entre sí, en donde cada una de las moléculas de hibridación comprende una secuencia de ácido nucleico que comprende una región de hibridación; y (ii) primer y segundo oligonucleótidos con código de barras, en donde el primer oligonucleótido con código de barras hibrida con la región de hibridación de la primera molécula de hibridación y en donde el segundo oligonucleótido con código de barras hibrida con la región de hibridación de la segunda molécula de hibridación; opcionalmente, en donde la etapa (b) comprende transferir el primer reactivo de código de barras multimérico a la primera célula y transferir el segundo reactivo de código de barras multimérico a la segunda célula.

En los métodos, cada uno de los reactivos de códigos de barras multiméricos puede comprender: (i) la primera y segunda moléculas de código de barras unidas entre sí, en donde cada una de las moléculas de código de barras comprende una secuencia de ácido nucleico que comprende una región de código de barras; y (ii) primer y segundo oligonucleótidos con código de barras, en donde el primer oligonucleótido con código de barras comprende una región de código de barras hibridada con la región de código de barras de la primera molécula de código de barras, y en donde el segundo oligonucleótido con código de barras comprende una región de código de barras hibridada con la región de código de barras de la segunda molécula de código de barras; opcionalmente, en donde la etapa (b) comprende transferir el primer reactivo de código de barras multimérico a la primera célula y transferir el segundo reactivo de código de barras multimérico a la segunda célula.

La divulgación proporciona un método para preparar una muestra de ácido nucleico para secuenciar, en donde la muestra comprende al menos dos células, y donde el método comprende las etapas de: (a) poner en contacto la muestra con una biblioteca que comprende el primer y segundo reactivos de códigos de barras multiméricos, en donde cada reactivo de códigos de barras multimérico comprende: (i) la primera y segunda moléculas de código de barras unidas entre sí, en donde cada una de las moléculas de código de barras comprende una secuencia de ácido nucleico que comprende, opcionalmente en la dirección de 5' a 3', una región adaptadora y una región de código de barras, y (ii) el primer y segundo oligonucleótidos con código de barras, en donde el primer oligonucleótido con código de barras comprende una región de código de barras hibridada con la región de código de barras de la primera molécula de código de barras, en donde el segundo oligonucleótido con código de barras comprende una región de código de barras hibridada con la región de código de barras de la segunda molécula de código de barras, y en donde las regiones de código de barras del primer y segundo oligonucleótidos con código de barras del primer reactivo de código de barras multimérico de la biblioteca son diferentes a las regiones de código de barras del primer y segundo oligonucleótidos con código de barras del segundo reactivo de código de barras multimérico de la biblioteca; en donde la muestra se pone en contacto además con el primer y segundo oligonucleótidos adaptadores para cada uno de los reactivos de códigos de barras multiméricos, en donde el primer y segundo oligonucleótidos adaptadores comprenden, cada uno, una región adaptadora; (b) transferir el primer y segundo oligonucleótidos adaptadores para el primer reactivo de código de barras multimérico a la primera célula y transferir el primer y segundo oligonucleótidos adaptadores para el segundo reactivo de código de barras multimérico a la segunda célula, opcionalmente, en donde la etapa comprende además transferir el primer reactivo de código de barras multimérico a la primera célula y transferir el segundo reactivo de código de barras multimérico a la segunda célula; (c) hibridar o ligar el primer y segundo oligonucleótidos adaptadores con el primer reactivo de código de barras multimérico a subsecuencias de un ácido nucleico diana de la primera célula, e hibridar o ligar el primer y segundo oligonucleótidos adaptadores con el segundo reactivo de código de barras multimérico a subsecuencias de un ácido nucleico diana de la segunda célula; (d) para cada uno de los reactivos de códigos de barras multiméricos, hibridar la región adaptadora del primer oligonucleótido adaptador con la región adaptadora de la primera molécula de código de barras, e hibridar la región adaptadora del segundo oligonucleótido adaptador con la región adaptadora de la segunda molécula de código de barras; y (e) para cada uno de los reactivos de códigos de barras multiméricos, ligar el extremo 3' del primer oligonucleótido con código de barras al extremo 5' del primer oligonucleótido adaptador para producir una primera molécula de ácido nucleico diana con código de barras y ligar el extremo 3' del segundo oligonucleótido con código de barras al extremo 5' del segundo oligonucleótido adaptador para producir una segunda molécula de ácido nucleico diana con código de barras.

En los métodos, la etapa de hibridación o ligadura (etapa (c)) puede comprender hibridar el primer y segundo oligonucleótidos adaptadores con el primer reactivo de código de barras multimérico a subsecuencias de un ácido nucleico diana de la primera célula, e hibridar el primer y segundo oligonucleótidos adaptadores con el segundo

reactivo de código de barras multimérico a subsecuencias de un ácido nucleico diana de la segunda célula, en donde:

- (i) para cada uno de los reactivos de códigos de barras multiméricos, la etapa (e) comprende ligar el extremo 3' del primer oligonucleótido con código de barras al extremo 5' del primer oligonucleótido adaptador para producir un primer oligonucleótido adaptador con código de barras y ligar el extremo 3' del segundo oligonucleótido con código de barras al extremo 5' del segundo oligonucleótido adaptador para producir un segundo oligonucleótido adaptador con código de barras, y extender el primer y segundo oligonucleótidos adaptadores con código de barras para producir una primera y segunda moléculas de ácido nucleico diana con códigos de barras diferentes, cada una de las cuales comprende al menos un nucleótido sintetizado a partir del ácido nucleico diana como molde, o (ii) para cada uno de los reactivos de códigos de barras multiméricos, antes de la etapa (e), el método comprende extender el primer y segundo oligonucleótidos adaptadores para producir una primera y una segunda moléculas de ácido nucleico diana diferentes, cada una de las cuales comprende al menos un nucleótido sintetizado a partir del ácido nucleico diana como molde.

- En los métodos, antes de la etapa de transferencia (etapa (b)), la membrana celular de las células puede permeabilizarse por contacto con un tensioactivo químico. Opcionalmente, los oligonucleótidos con código de barras, los oligonucleótidos adaptadores y/o los reactivos de códigos de barras multiméricos se transfieren a las células a través de la membrana permeabilizada.

El tensioactivo químico puede ser un tensioactivo no iónico. El tensioactivo químico puede ser uno o más de Tritón X-100 ($C_{14}H_{22}O(C_2H_4O)_n$ ($n=9-10$)), Brij 35, Brij 58, Digitonina, IGEPAL CA-630, Saponina, TWEEN 20, TWEEN 40 y/o TWEEN 80.

- El tensioactivo químico puede estar en solución a una concentración de menos de 1,0 micromolar, menos de 5 micromolar, menos de 10 micromolar, menos de 25 micromolar, menos de 50 micromolar, menos de 100 micromolar, menos de 200 micromolar o menos de 500 micromolar, menos de 1,0 milimolar o menos de 5,0 milimolar.

La(s) célula(s) se pueden permeabilizar mediante una mezcla de dos o más tensioactivos químicos diferentes.

- En los métodos, después de la etapa de permeabilizar las membranas celulares, la concentración del tensioactivo químico en la solución se puede reducir mediante la adición de una segunda solución a la muestra que comprende las células y el tensioactivo químico. Opcionalmente, esta segunda solución puede no contener un tensioactivo químico.

- En los métodos, después de la etapa de permeabilizar las membranas celulares, la muestra de células se puede sedimentar mediante una etapa de centrifugación, el sobrenadante (que contiene el tensioactivo químico pero no las células) se puede eliminar y las células sedimentadas se pueden resuspender en una segunda solución. Opcionalmente, esta segunda solución puede no contener un tensioactivo químico.

En los métodos, antes de la etapa de transferencia (etapa (b)), la membrana celular de las células se puede permeabilizar por contacto con un disolvente o disolvente molecular (capaz de alterar la bicapa lipídica de la membrana celular). Opcionalmente, los oligonucleótidos con código de barras, los oligonucleótidos adaptadores y/o los reactivos de códigos de barras multiméricos se transfieren a las células a través de la membrana permeabilizada.

- El disolvente puede ser uno o más de betaína, formamida y/o dimetilsulfóxido (DMSO)

El disolvente se puede utilizar en una concentración de al menos el 1 % en peso o en volumen, al menos el 5 % en peso o en volumen, al menos el 10 % en peso o en volumen, al menos el 20 % en peso o en volumen, al menos el 30 % en peso o en volumen, al menos el 40 % en peso o en volumen, o al menos el 50 % en peso o en volumen.

- En los métodos, antes de la etapa de transferencia (etapa (b)), la membrana celular de las células se puede permeabilizar mediante una etapa de incubación térmica a alta temperatura. Opcionalmente, los oligonucleótidos con código de barras, los oligonucleótidos adaptadores y/o los reactivos de códigos de barras multiméricos se transfieren a las células a través de la membrana permeabilizada.

- la etapa de incubación térmica se puede realizar a una temperatura de al menos 37 grados Celsius, al menos 40 grados Celsius, al menos 45 grados Celsius, al menos 50 grados Celsius, al menos 55 grados Celsius, al menos 60 grados Celsius, al menos 65 grados Celsius, al menos 70 grados Celsius, al menos 75 grados Celsius, al menos 80 grados Celsius, o al menos 85 grados Celsius.

La etapa de permeabilizar las membranas celulares se puede realizar durante menos de 5 segundos, menos de 10 segundos, menos de 30 segundos, menos de 60 segundos, menos de 2 minutos, menos de 5 minutos, menos de 10 minutos, menos de 15 minutos, menos de 30 minutos, menos de 60 minutos o menos de 2 horas.

- En los métodos, los oligonucleótidos con código de barras, los oligonucleótidos adaptadores y/o reactivos de códigos de barras multiméricos se pueden transferir a las células mediante complejación con un reactivo de transfección o un vehículo lipídico (seguido de transfección, transferencia o liberación en las células). Este proceso puede implicar transfección, transferencia o liberación de los reactivos a la célula.

El reactivo de transfección puede ser un reactivo de transfección de lípidos, por ejemplo, un reactivo de transfección

de lípidos catiónicos. Opcionalmente, dicho reactivo de transfección de lípidos catiónicos comprende al menos dos cadenas de alquilo. Opcionalmente, dicho reactivo de transfección de lípidos catiónicos puede ser un reactivo de transfección de lípidos catiónicos disponible comercialmente, tal como lipofectamina.

5 Los oligonucleótidos con código de barras, los oligonucleótidos adaptadores y/o reactivos de códigos de barras multiméricos se pueden transferir a las células mediante complejación con un reactivo de polímero catiónico (seguido de transfección, transferencia o liberación en las células). Opcionalmente, dicho reactivo de polímero catiónico puede comprender un polímero catiónico lineal, tal como espermina o polilisina. Opcionalmente, dicho reactivo de polímero catiónico puede comprender un polímero de polietilenimina. Opcionalmente, dicho reactivo de polímero catiónico puede comprender un polímero de dietilaminoetil (DEAE)-dextrano. Opcionalmente, dicho reactivo de polímero catiónico puede comprender un polímero catiónico ramificado.

10 Los oligonucleótidos con código de barras, los oligonucleótidos adaptadores y/o reactivos de códigos de barras multiméricos se pueden transferir a las células mediante complejación con un dendrímero y/o un dendrímero activado (seguido de transfección, transferencia o liberación en las células). Opcionalmente, dicho dendrímero activado se activa con uno o más grupos amino; opcionalmente, dichos grupos amino están cargados positivamente. 15 Opcionalmente, cualquier dendrímero y/o dendrímero activado comprende al menos 2 generaciones, al menos 3 generaciones, al menos 5 generaciones, al menos 10 generaciones, al menos 20 generaciones o al menos 30 generaciones.

20 Los oligonucleótidos con código de barras, los oligonucleótidos adaptadores y/o reactivos de códigos de barras multiméricos se pueden transferir a las células mediante complejación con un reactivo liposomal o micelar (seguido de transfección, transferencia o liberación en las células). Opcionalmente, los oligonucleótidos con código de barras, los oligonucleótidos adaptadores y/o los reactivos de códigos de barras multiméricos se pueden cargar en una preparación de reactivos liposomales o micelares con una etapa de carga de reactivo. Opcionalmente, dichos reactivos liposomales o micelares pueden comprender uno o más anfífilos. Opcionalmente, dichos reactivos liposomales o micelares pueden comprender uno o más fosfolípidos. Opcionalmente, dichos fosfolípidos pueden comprender una o más moléculas de 25 fosfatidilcolinas. Opcionalmente, dichos fosfolípidos pueden comprender una o más moléculas de fosfatidiletanolamina. Opcionalmente, dichos reactivos liposomales o micelares pueden comprender copolímeros. Opcionalmente, dichos reactivos liposomales o micelares pueden comprender copolímeros de bloqueo. Opcionalmente, cada reactivo liposomal o micelar puede en promedio formar complejos con 1, o menos de 1, o más de 1, o cualquier otro número de reactivo(s) de código de barras multimérico dentro de una preparación de dicho reactivo(s) de código de barras multimérico complejo. Opcionalmente, cada reactivo liposomal o micelar puede en promedio formar complejos con al menos 2 oligonucleótidos con código de barras (y/o 2 oligonucleótidos adaptadores).

Los oligonucleótidos con código de barras, los oligonucleótidos adaptadores y/o los reactivos de códigos de barras multiméricos se pueden transferir a las células mediante complejación dentro de una solución de cloruro de calcio y fosfato para formar un precipitado y luego transfectarse en las células.

35 Los oligonucleótidos con código de barras, los oligonucleótidos adaptadores y/o reactivos de códigos de barras multiméricos pueden formar complejos con reactivos de transfección con una etapa de incubación de formación de complejos. Opcionalmente, esta etapa de incubación de formación de complejos puede durar al menos 5 segundos, al menos 10 segundos, al menos 30 segundos, al menos 60 segundos, al menos 2 minutos, al menos 5 minutos, al menos 10 minutos, al menos 30 minutos, al menos 60 minutos, al menos 2 horas de longitud o al menos 4 horas de longitud. Opcionalmente, esta etapa de incubación de formación de complejos puede tener lugar a aproximadamente 40 4 grados Celsius, aproximadamente 12 grados Celsius, aproximadamente 20 grados Celsius, aproximadamente 25 grados Celsius, aproximadamente 30 grados Celsius o aproximadamente 37 grados Celsius. Opcionalmente, los reactivos de códigos de barras multiméricos complejados se pueden procesar y/o almacenar adicionalmente, antes de transferirlos a las células.

45 En los métodos, después de los oligonucleótidos con código de barras, los oligonucleótidos adaptadores y/o los reactivos de códigos de barras multiméricos forman complejos con los reactivos de transfección, se puede realizar una etapa de incubación de transferencia. Opcionalmente, esta etapa de incubación de transferencia puede durar al menos 5 segundos, al menos 10 segundos, al menos 30 segundos, al menos 60 segundos, al menos 2 minutos, al menos 5 minutos, al menos 10 minutos, al menos 30 minutos, al menos 60 minutos, al menos 2 horas de longitud o al menos 4 horas de longitud. Opcionalmente, esta etapa de incubación de transferencia puede tener lugar a aproximadamente 50 4 grados Celsius, aproximadamente 12 grados Celsius, aproximadamente 20 grados Celsius, aproximadamente 25 grados Celsius, aproximadamente 30 grados Celsius o aproximadamente 37 grados Celsius.

Los oligonucleótidos con código de barras del primer reactivo de código de barras multimérico pueden estar comprendidos dentro de un primer vehículo de lípidos, y los oligonucleótidos con código de barras del segundo reactivo de código de barras multimérico pueden estar comprendidos dentro de un segundo vehículo de lípidos. Opcionalmente, tales oligonucleótidos con código de barras se pueden transferir a las células mediante un proceso que implica la fusión del liposoma o la micela con la membrana celular. Opcionalmente, este proceso de fusión puede liberar los oligonucleótidos con código de barras en el citoplasma de la célula. Opcionalmente, los oligonucleótidos con código de barras se pueden cargar en una preparación de reactivos liposomales o micelares con una etapa de 60 carga de oligonucleótidos. Opcionalmente, dichos liposomas o micelas pueden comprender uno o más anfífilos.

Opcionalmente, dichos liposomas o micelas pueden comprender uno o más fosfolípidos. Opcionalmente, dichos fosfolípidos pueden comprender una o más fosfatidilcolinas. Opcionalmente, dichos fosfolípidos pueden comprender una o más moléculas de fosfatidiletanolamina. Opcionalmente, dichos liposomas o micelas pueden comprender copolímeros. Opcionalmente, dichos liposomas o micelas pueden comprender copolímeros de bloque. Opcionalmente, cada liposoma o micela puede, en promedio, estar formando complejo o cargado con, al menos 2, al menos 3, al menos 5, al menos 10, al menos 50, al menos 100, al menos 500, al menos 1000, al menos 10.000, o al menos 100.000 oligonucleótidos con código de barras, o cualquier número mayor de oligonucleótidos con código de barras.

En los métodos, los oligonucleótidos con código de barras, los oligonucleótidos adaptadores y/o los reactivos de códigos de barras multiméricos se pueden transferir a las células mediante un proceso que comprende la compresión de las células.

En los métodos, la etapa de transferencia puede comprender la deformación mecánica de las células de la muestra para producir alteraciones transitorias de la membrana que permitan la transferencia de los oligonucleótidos con códigos de barras, oligonucleótidos adaptadores y/o reactivos de códigos de barras multiméricos en las células. La muestra puede ponerse en contacto con una biblioteca de reactivos de códigos de barras multiméricos (y/u oligonucleótidos adaptadores para cada reactivo de códigos de barras multiméricos) antes, durante o después de la etapa de deformación mecánica de las células.

Los métodos para exprimir células (*cell squeezing*) se proporcionan en Sharei et al, Cell Squeezing as a Robust, Microfluidic Intracellular Delivery Platform. J. Vis. Exp. (81, e50980, doi:10.3791/50980 (2013) y Sharei et al, Proc Natl Acad Sci EE.UU. 5 de febrero de 2013;110(6):2082-7).

En los métodos de compresión de células, las células intactas pueden obtenerse a través de un conducto mecánico (por ejemplo, un canal de microfluidos dentro de un circuito de microfluidos) que es más pequeño (es decir, de diámetro más pequeño) que una célula, y en donde, cuando una célula transita a través de este conducto o canal, la célula se "exprime" (es decir, se encuentra con un esfuerzo mecánico y/o deformación o esfuerzo de cizalla) y se deforma al menos parcialmente. En función de este proceso, la membrana celular se altera parcialmente y esto puede permitir que las moléculas (incluidos los oligonucleótidos con códigos de barras, los oligonucleótidos adaptadores y/o los reactivos con códigos de barras multiméricos) transiten desde la solución que rodea la célula, a la propia célula. La compresión celular comprende, por lo tanto, un medio mecánico, no químico, no biológico para transferir reactivos a las células.

Los métodos pueden comprender mezclar una biblioteca de reactivos de códigos de barras multiméricos con una muestra de células y pasar la mezcla resultante a través de un aparato exprimidor de células. Este proceso permite que los reactivos de códigos de barras multiméricos de la biblioteca de los mismos entren en una o más células en la muestra de células. Las células resultantes pueden luego procesarse más; por ejemplo, se pueden incubar durante un período de tiempo, por ejemplo, para permitir que los oligonucleótidos de códigos de barras hibriden con ácidos nucleicos afines dentro de las células a las que se han transferido.

En los métodos, los oligonucleótidos con código de barras, Los oligonucleótidos adaptadores y/o reactivos de códigos de barras multiméricos pueden transferirse a las células mediante un proceso que comprende electroporación (o electroporación).

La muestra puede ponerse en contacto con una biblioteca de reactivos de códigos de barras multiméricos (y/u oligonucleótidos adaptadores para cada reactivo de códigos de barras multiméricos) antes, durante o después del proceso de electroporación.

La electroporación puede utilizar una forma de onda de electroporación cuadrada. La electroporación puede utilizar una forma de onda de electroporación exponencial.

Durante el proceso de electroporación, el gradiente de voltaje máximo puede ser de al menos 1,0 kilovoltios por centímetro, al menos 2,0 kilovoltios por centímetro, al menos 5,0 kilovoltios por centímetro, al menos 10,0 kilovoltios por centímetro, al menos 15,0 kilovoltios por centímetro o al menos 20,0 kilovoltios por centímetro.

Durante el proceso de electroporación, los pulsos de electroporación pueden tener una duración de al menos 100 microsegundos, al menos 500 microsegundos de duración, al menos 1,0 milisegundos de duración, al menos 2,0 milisegundos de duración, al menos 3,0 milisegundos de duración, al menos 5,0 milisegundos de duración o al menos 10,0 milisegundos de duración.

Estos métodos a continuación describen técnicas particulares para su uso con cualquiera de los métodos anteriores en los que los reactivos de códigos de barras multiméricos se transfieren (o internalizan) a las células mediante cualquier método. Estos métodos describen realizaciones alternativas, así como las etapas experimentales posteriores, que posiblemente podrían ser aplicables a cualquiera de los protocolos anteriores.

En los métodos, después de una etapa de transferencia de reactivos de códigos de barras multiméricos u oligonucleótidos adaptadores a las células, las células se pueden incubar durante un período de tiempo para permitir que las regiones diana del reactivo o reactivos de códigos de barras multiméricos aparezcan con subsecuencias de

un ácido nucleico diana dentro de la célula. El período de incubación puede ser de al menos 1 minuto, o al menos 5 minutos, o al menos 15 minutos, o al menos 30 minutos, o al menos 60 minutos. La incubación puede tener lugar dentro de una solución que contiene un desnaturalizante de ácido nucleico, tal como DMSO o betaína. La incubación puede tener lugar a una temperatura de al menos 37 grados Celsius, al menos 45 grados Celsius, al menos 50 grados Celsius, al menos 55 grados Celsius, al menos 60 grados Celsius, al menos 65 grados Celsius, o al menos 70 grados Celsius.

En los métodos, después de una etapa de introducción de oligonucleótidos con códigos de barras y/o reactivos de códigos de barras multiméricos en las células, se puede realizar una etapa de división de reactivos en la que los reactivos de códigos de barras multiméricos se dividen en dos o más componentes de los mismos que se difunden independientemente. Opcionalmente, en realizaciones en las que un reactivo de códigos de barras multimérico comprende oligonucleótidos con códigos de barras hibridados con moléculas de códigos de barras, esta etapa de división de reactivo puede comprender una etapa de desnaturalización de uno o más oligonucleótidos con código de barras de las moléculas de código de barras con las que hibridan, de modo que dichos oligonucleótidos con código de barras se puedan difundir independientemente dentro de la(s) célula(s) a la(s) que se han transferido. Opcionalmente, dicha etapa de desnaturalización se puede realizar con una incubación a alta temperatura, en donde los oligonucleótidos con código de barras están desnaturalizados a una temperatura de al menos 37 grados Celsius, al menos 45 grados Celsius, al menos 50 grados Celsius, al menos 55 grados Celsius, al menos 60 grados Celsius, al menos 65 grados Celsius, o al menos 70 grados Celsius. Opcionalmente, esta etapa de desnaturalización tiene lugar dentro de una solución que contiene un desnaturalizante de ácido nucleico, tal como DMSO o betaína. Opcionalmente, esta etapa de desnaturalización puede tener lugar antes de una etapa de incubación tal como se describe anteriormente; u opcionalmente, esta etapa de desnaturalización puede tener lugar dentro de la misma etapa que una etapa de incubación.

En los métodos, tras la transferencia de oligonucleótidos con códigos de barras, oligonucleótidos adaptadores y/o reactivos de códigos de barras multiméricos en las células y, opcionalmente, siguiendo una etapa de incubación, las células se pueden poner en contacto con una solución de oligonucleótidos complementaria a la totalidad o parte de una o más regiones diana de los oligonucleótidos con códigos de barras dentro de reactivos de códigos de barras multiméricos.

En los métodos, tras la introducción de los oligonucleótidos con código de barras, oligonucleótidos adaptadores y/o reactivos de códigos de barras multiméricos en la célula y, opcionalmente, siguiendo una etapa de incubación, la(s) célula(s) se puede(n) aislar de una mezcla de reacción por centrifugación.

En los métodos, tras la transferencia de los oligonucleótidos con código de barras, oligonucleótidos adaptadores y/o reactivos de códigos de barras multiméricos en la célula y, opcionalmente, siguiendo una etapa de incubación, los oligonucleótidos con código de barras y/o moléculas de ácido nucleico diana con código de barras y/o reactivo(s) de código de barras multimérico se pueden aislar de la célula.

Los reactivos de códigos de barras multiméricos y/o los oligonucleótidos con códigos de barras pueden comprender uno o más restos de biotina.

En los métodos, tras la transferencia de oligonucleótidos con códigos de barras, oligonucleótidos adaptadores y/o reactivos de códigos de barras multiméricos en la célula y, opcionalmente, siguiendo una etapa de incubación, los oligonucleótidos con código de barras y/o moléculas de ácido nucleico diana con código de barras y/o reactivo(s) de código de barras multimérico se pueden aislar por un proceso de: (a) disolver y/o permeabilizar las membranas celulares, opcionalmente usando un tensioactivo químico, utilizando un disolvente (molecular) o mediante incubación a alta temperatura; (b) poner en contacto la mezcla resultante con un soporte sólido, opcionalmente en donde el soporte sólido comprende restos de estreptavidina; y (c) capturar los oligonucleótidos con código de barras y/o las moléculas de ácido nucleico diana con código de barras y/o los reactivos de códigos de barras multiméricos en el soporte sólido, opcionalmente a través de la interacción estreptavidina-biotina.

El soporte sólido puede ser una o más perlas magnéticas, opcionalmente en donde una o más perlas magnéticas comprenden moléculas de estreptavidina en su superficie. Las perlas magnéticas se pueden aislar de una mezcla de reacción con un imán.

En los métodos, cualquier etapa de permeabilización de células y/o transferencia de reactivos de códigos de barras multiméricos a células y/o incubación de células puede tener lugar en una solución hipotónica. En los métodos, cualquier etapa de permeabilización de células y/o transferencia de reactivos de códigos de barras multiméricos a células y/o incubación de células puede tener lugar en una solución hipertónica.

En los métodos, se puede proporcionar una biblioteca de reactivos de códigos de barras multiméricos en la misma solución que un tensioactivo químico y/o en la misma solución que un disolvente molecular y/o en la misma solución que un desnaturalizante.

La divulgación proporciona un método para preparar una muestra de ácido nucleico para secuenciar, en donde la muestra comprende al menos dos células, y donde el método comprende las etapas de: (a) poner en contacto la muestra con una biblioteca que comprende al menos dos reactivos de códigos de barras multiméricos, en donde cada

reactivo de código de barras multimérico comprende una primera y segunda regiones de código de barras unidas entre sí, en donde cada región de código de barras comprende una secuencia de ácido nucleico y en donde la primera y segunda regiones de código de barras de un primer reactivo de código de barras multimérico son diferentes a la primera y segunda regiones de código de barras de un segundo reactivo de código de barras multimérico de la biblioteca; (b) lisar las células o permeabilizar las membranas celulares de las células; y (c) añadir secuencias de códigos de barras a cada una de las subsecuencias primera y segunda de un ácido nucleico diana de la primera célula para producir moléculas de ácido nucleico diana con código de barras primera y segunda para la primera célula, en donde la primera molécula de ácido nucleico diana con código de barras comprende la secuencia de ácido nucleico de la primera región de código de barras del primer reactivo de código de barras multimérico y la segunda molécula de ácido nucleico diana con código de barras comprende la secuencia de ácido nucleico de la segunda región de código de barras del primer reactivo de código de barras multimérico, y agregar secuencias de códigos de barras a cada una de las subsecuencias primera y segunda de un ácido nucleico diana de la segunda célula para producir moléculas de ácido nucleico diana con código de barras primera y segunda para la segunda célula, en donde la primera molécula de ácido nucleico diana con código de barras comprende la secuencia de ácido nucleico de la primera región de código de barras del segundo reactivo de código de barras multimérico y la segunda molécula de ácido nucleico diana con código de barras comprende la secuencia de ácido nucleico de la segunda región de código de barras del segundo reactivo de código de barras multimérico.

La divulgación proporciona un método para preparar una muestra de ácido nucleico para secuenciar, en donde la muestra comprende al menos dos células, y donde el método comprende (en orden) las etapas de:

(a) poner en contacto la muestra con una biblioteca que comprende al menos dos reactivos de códigos de barras multiméricos, en donde cada reactivo de códigos de barras multimérico comprende un primer y un segundo oligonucleótidos con código de barras unidos entre sí, en donde los oligonucleótidos con código de barras comprenden, cada uno, una región de código de barras y en donde las regiones de código de barras del primer y segundo oligonucleótidos con código de barras de un primer reactivo de código de barras multimérico de la biblioteca son diferentes a las regiones de código de barras del primer y segundo oligonucleótidos con código de barras de un segundo reactivo de código de barras multimérico de la biblioteca;

(b) lisar las células o permeabilizar las membranas celulares de las células; y

(c) hibridar o ligar el primer y segundo oligonucleótidos con código de barras del primer reactivo de código de barras multimérico a la primera y segunda subsecuencia de un ácido nucleico diana de la primera célula para producir una primera y segunda moléculas de ácido nucleico diana con código de barras, e hibridar o ligar el primer y segundo oligonucleótidos con código de barras del segundo reactivo de código de barras multimérico a la primera y segunda subsecuencias de un ácido nucleico diana de la segunda célula para producir una primera y segunda moléculas de ácido nucleico diana con código de barras.

El método puede comprender (en orden) las etapas de: (a) poner en contacto la muestra con una biblioteca que comprende al menos dos reactivos de códigos de barras multiméricos, en donde cada reactivo de códigos de barras multimérico comprende un primer y un segundo oligonucleótidos con código de barras unidos entre sí, en donde los oligonucleótidos con código de barras comprenden, cada uno, una región de código de barras y en donde las regiones de código de barras del primer y segundo oligonucleótidos con código de barras de un primer reactivo de código de barras multimérico de la biblioteca son diferentes a las regiones de código de barras del primer y segundo oligonucleótidos con código de barras de un segundo reactivo de código de barras multimérico de la biblioteca; (b) lisar las células o permeabilizar las membranas celulares de las células; y (c) hibridar o ligar el primer y segundo oligonucleótidos con código de barras del primer reactivo de código de barras multimérico a la primera y segunda subsecuencia de un ácido nucleico diana de la primera célula para producir una primera y segunda moléculas de ácido nucleico diana con código de barras, e hibridar o ligar el primer y segundo oligonucleótidos con código de barras del segundo reactivo de código de barras multimérico a la primera y segunda subsecuencias de un ácido nucleico diana de la segunda célula para producir una primera y segunda moléculas de ácido nucleico diana con código de barras.

La divulgación proporciona un método para preparar una muestra de ácido nucleico para secuenciar, en donde la muestra comprende al menos 10 células, y donde el método comprende, en orden, las etapas de: (a) poner en contacto la muestra con una biblioteca que comprende al menos dos reactivos de códigos de barras multiméricos, en donde cada reactivo de códigos de barras multimérico comprende un primer y un segundo oligonucleótidos con código de barras unidos entre sí y un resto de unión a células, en donde los oligonucleótidos con código de barras comprenden, cada uno, una región de código de barras y en donde las regiones de código de barras del primer y segundo oligonucleótidos con código de barras de un primer reactivo de código de barras multimérico de la biblioteca son diferentes a las regiones de código de barras del primer y segundo oligonucleótidos con código de barras de un segundo reactivo de código de barras multimérico de la biblioteca, en donde el resto de unión celular del primer reactivo de código de barras multimérico se une a la membrana celular de una primera célula antes de la etapa (b), y en donde el resto de unión celular del segundo reactivo de código de barras multimérico se une a la membrana celular de una segunda célula antes de la etapa (b); (b) lisar las células o permeabilizar las membranas celulares de las células; y (c) hibridar o ligar el primer y segundo oligonucleótidos con código de barras del primer reactivo de código de barras multimérico a la primera y segunda subsecuencia de un ácido nucleico diana de la primera célula para producir una primera y segunda moléculas de ácido nucleico diana con código de barras, e hibridar o ligar el primer y segundo

oligonucleótidos con código de barras del segundo reactivo de código de barras multimérico a la primera y segunda subsecuencias de un ácido nucleico diana de la segunda célula para producir una primera y segunda moléculas de ácido nucleico diana con código de barras. Preferentemente, las células están comprendidas dentro de un único volumen acuoso contiguo durante las etapas (a), (b) y/o (c).

- 5 En los métodos, la etapa de hibridación o ligado (etapa (c)) puede comprender: (i) hibridar el primer y segundo oligonucleótidos con código de barras del primer reactivo de código de barras multimérico con la primera y segunda subsecuencia de un ácido nucleico diana de la primera célula, e hibridar el primer y segundo oligonucleótidos con código de barras del segundo reactivo de código de barras multimérico con la primera y segunda subsecuencias de un ácido nucleico diana de la segunda célula; y (ii) extender el primer y segundo oligonucleótidos con código de barras del primer reactivo de código de barras multimérico para producir una primera y segunda moléculas de ácido nucleico diana con código de barras diferentes y extender el primer y segundo oligonucleótidos con código de barras del segundo reactivo de código de barras multimérico para producir una primera y segunda molécula de ácido nucleico diana con código de barras diferente, en donde cada una de las moléculas de ácido nucleico diana con código de barras comprende al menos un nucleótido sintetizado a partir del ácido nucleico diana como molde.
- 10
- 15 En los métodos, cada uno de los reactivos de códigos de barras multiméricos puede comprender: (i) la primera y segunda moléculas de hibridación unidas entre sí, en donde cada una de las moléculas de hibridación comprende una secuencia de ácido nucleico que comprende una región de hibridación; y (ii) primer y segundo oligonucleótidos con código de barras, en donde el primer oligonucleótido con código de barras hibrida con la región de hibridación de la primera molécula de hibridación y en donde el segundo oligonucleótido con código de barras hibrida con la región de hibridación de la segunda molécula de hibridación.
- 20

Cada uno de los reactivos de códigos de barras multiméricos puede comprender: (i) la primera y segunda moléculas de código de barras unidas entre sí, en donde cada una de las moléculas de código de barras comprende una secuencia de ácido nucleico que comprende una región de código de barras; y (ii) primer y segundo oligonucleótidos con código de barras, en donde el primer oligonucleótido con código de barras comprende una región de código de barras hibridada con la región de código de barras de la primera molécula de código de barras y en donde el segundo oligonucleótido con código de barras comprende una región de código de barras hibridada con la región de código de barras de la segunda molécula de código de barras.

25

La divulgación proporciona un método para preparar una muestra de ácido nucleico para secuenciar, en donde la muestra comprende al menos dos células, y donde el método comprende las etapas de: (a) poner en contacto la muestra con una biblioteca que comprende el primer y segundo reactivos de códigos de barras multiméricos, en donde cada reactivo de códigos de barras multimérico comprende: (i) la primera y segunda moléculas de código de barras unidas entre sí, en donde cada una de las moléculas de código de barras comprende una secuencia de ácido nucleico que comprende, opcionalmente en la dirección de 5' a 3', una región adaptadora y una región de código de barras, y (ii) el primer y segundo oligonucleótidos con código de barras, en donde el primer oligonucleótido con código de barras comprende una región de código de barras hibridada con la región de código de barras de la primera molécula de código de barras y en donde el segundo oligonucleótido con código de barras comprende una región de código de barras hibridada con la región de código de barras de la segunda molécula de código de barras, y en donde las regiones de código de barras del primer y segundo oligonucleótidos con código de barras del primer reactivo de código de barras multimérico son diferentes de las regiones de código de barras del primer y segundo oligonucleótidos con código de barras del segundo reactivo de código de barras multimérico; en donde la muestra se pone en contacto además con el primer y segundo oligonucleótidos adaptadores para cada uno de los reactivos de códigos de barras multiméricos, en donde el primer y segundo oligonucleótidos adaptadores comprenden, cada uno, una región adaptadora; (b) lisar las células o permeabilizar las membranas celulares de las células; (c) hibridar o ligar el primer y segundo oligonucleótidos adaptadores con el primer reactivo de código de barras multimérico a subsecuencias de un ácido nucleico diana de la primera célula, e hibridar o ligar el primer y segundo oligonucleótidos adaptadores con el segundo reactivo de código de barras multimérico a subsecuencias de un ácido nucleico diana de la segunda célula; (d) para cada uno de los reactivos de códigos de barras multiméricos, hibridar la región adaptadora del primer oligonucleótido adaptador con la región adaptadora de la primera molécula de código de barras, e hibridar la región adaptadora del segundo oligonucleótido adaptador con la región adaptadora de la segunda molécula de código de barras; y (e) para cada uno de los reactivos de códigos de barras multiméricos, ligar el extremo 3' del primer oligonucleótido con código de barras al extremo 5' del primer oligonucleótido adaptador para producir una primera molécula de ácido nucleico diana con código de barras y ligar el extremo 3' del segundo oligonucleótido con código de barras al extremo 5' del segundo oligonucleótido adaptador para producir una segunda molécula de ácido nucleico diana con código de barras.

30

35

40

45

50

- 55 En los métodos, la etapa de hibridación o ligadura (etapa (c)) puede comprender hibridar el primer y segundo oligonucleótidos adaptadores con el primer reactivo de código de barras multimérico a subsecuencias de un ácido nucleico diana de la primera célula, e hibridar el primer y segundo oligonucleótidos adaptadores con el segundo reactivo de código de barras multimérico a subsecuencias de un ácido nucleico diana de la segunda célula, en donde: (i) para cada uno de los reactivos de códigos de barras multiméricos, la etapa (e) comprende ligar el extremo 3' del primer oligonucleótido con código de barras al extremo 5' del primer oligonucleótido adaptador para producir un primer oligonucleótido adaptador con código de barras y ligar el extremo 3' del segundo oligonucleótido con código de barras al extremo 5' del segundo oligonucleótido adaptador para producir un segundo oligonucleótido adaptador con código
- 60

de barras, y extender el primer y segundo oligonucleótidos adaptadores con código de barras para producir una primera y segunda moléculas de ácido nucleico diana con códigos de barras diferentes, cada una de las cuales comprende al menos un nucleótido sintetizado a partir del ácido nucleico diana como molde, o (ii) para cada uno de los reactivos de códigos de barras multiméricos, antes de la etapa (e), el método comprende extender el primer y segundo oligonucleótidos adaptadores para producir una primera y una segunda moléculas de ácido nucleico diana diferentes, cada una de las cuales comprende al menos un nucleótido sintetizado a partir del ácido nucleico diana como molde.

En los métodos, tras la etapa de lisar o permeabilizar (etapa (b)), los ácidos nucleicos diana de cada célula dentro de la muestra se pueden difundir fuera de la célula (es decir, fuera del espacio citoplasmático o del volumen celular).

Opcionalmente, los reactivos de códigos de barras multiméricos no pueden entrar en la célula. Opcionalmente, tras la etapa (b), la membrana celular está sustancial o totalmente disuelta. Opcionalmente, tras la etapa (b), la membrana celular permanece parcialmente intacta pero en donde las moléculas de ARN mensajero y/u otras moléculas de ácido nucleico pueden difundirse fuera de la célula (es decir, fuera del espacio citoplasmático o volumen celular) a través de poros y/u otras discontinuidades estructurales dentro de la membrana celular.

En los métodos, la etapa (b) se puede realizar aumentando la temperatura de la muestra. Opcionalmente, se puede realizar una etapa de incubación a alta temperatura, por ejemplo, la etapa de incubación a alta temperatura se puede realizar a una temperatura de al menos 37 grados Celsius, al menos 40 grados Celsius, al menos 50 grados Celsius, al menos 60 grados Celsius, al menos 70 grados Celsius, al menos 80 grados Celsius, al menos 90 grados Celsius, o al menos 95 grados Celsius.

En los métodos, la etapa (b) se puede realizar en presencia de un tensioactivo químico. El tensioactivo químico puede ser uno o más de Tritón X-100 ($C_{14}H_{22}O(C_2H_4O)_n$ (n=9-10)), Brij 35, Brij 58, Digitonina, IGEPAL CA-630, Saponina, TWEEN 20, TWEEN 40 y/o TWEEN 80.

En los métodos, la etapa (b) se puede realizar en presencia de un disolvente o disolvente molecular (capaz de alterar la bicapa lipídica de la membrana celular). El disolvente puede ser uno o más de betaína, formamida y/o dimetilsulfóxido (DMSO).

En los métodos, cualquier etapa puede tener lugar en condiciones hipotónicas o hipertónicas. Opcionalmente, la etapa (b) se puede realizar en condiciones hipotónicas o hipertónicas.

En los métodos, los reactivos de códigos de barras multiméricos y/o los oligonucleótidos adaptadores pueden comprender, cada uno, un resto de unión a células, opcionalmente, en donde el resto de unión a células une cada reactivo de código de barras multimérico y/u oligonucleótido adaptador a la membrana celular de una célula antes de la etapa (b). Opcionalmente, cada uno de los oligonucleótidos con código de barras, las moléculas de hibridación multiméricas y/o las moléculas de código de barras multiméricas comprenden un resto de unión a las células. La fracción de unión a las células de cada oligonucleótido con código de barras, la molécula de hibridación multimérica y/o la molécula de código de barras multimérica se puede unir a la membrana celular de una célula antes de la etapa (b).

En los métodos, la etapa de hibridación de oligonucleótidos con código de barras a ácidos nucleicos diana puede comprender una etapa de incubación, en donde la muestra se incuba durante un período de tiempo para permitir que las regiones diana de los oligonucleótidos con código de barras aparezcan con los ácidos nucleicos diana. Opcionalmente, este período de incubación es de al menos 1 minuto, o al menos 5 minutos, o al menos 15 minutos, o al menos 30 minutos, o al menos 60 minutos. Opcionalmente, esta incubación tiene lugar dentro de una solución que contiene un desnaturizante de ácido nucleico, tal como DMSO o betaína. Opcionalmente, esta incubación tiene lugar a una temperatura de al menos 37 grados Celsius, al menos 45 grados Celsius, al menos 50 grados Celsius, al menos 55 grados Celsius, al menos 60 grados Celsius, al menos 65 grados Celsius, o al menos 70 grados Celsius.

En los métodos, durante o antes de la etapa (c), se puede realizar una etapa de división de reactivos en la que los reactivos de códigos de barras multiméricos se dividen en dos o más componentes de los mismos que se difunden independientemente. Opcionalmente, en donde un reactivo de códigos de barras multimérico comprende oligonucleótidos con códigos de barras hibridados con moléculas de códigos de barras, esta etapa de división de reactivo comprende una etapa de desnaturización de uno o más oligonucleótidos con código de barras de las moléculas de código de barras con las que hibridan, de manera que dichos oligonucleótidos con código de barras puedan difundirse independientemente dentro de la solución. Opcionalmente, dicha etapa de desnaturización se puede realizar con una incubación a alta temperatura, en donde los oligonucleótidos con código de barras están desnaturizados a una temperatura de al menos 37 grados Celsius, al menos 45 grados Celsius, al menos 50 grados Celsius, al menos 55 grados Celsius, al menos 60 grados Celsius, al menos 65 grados Celsius, o al menos 70 grados Celsius. Opcionalmente, esta etapa de desnaturización tiene lugar dentro de una solución que contiene un desnaturizante de ácido nucleico, tal como DMSO o betaína. Opcionalmente, esta etapa de división de reactivo y/o etapa de desnaturización puede tener lugar antes de una etapa de hibridación como se describe anteriormente; u opcionalmente, esta etapa de división de reactivo y/o etapa de desnaturización puede tener lugar durante la etapa de hibridación. Adicionalmente, esta etapa de división de reactivo y/o etapa de desnaturización puede tener lugar durante la etapa de lisis celular. Por ejemplo, una sola etapa de incubación térmica a alta temperatura puede tener el

efecto de lisar las células a través de un proceso de lisis térmica y desnaturalizar los oligonucleótidos con códigos de barras a partir de moléculas de códigos de barras dentro de reactivos de códigos de barras multiméricos. Adicionalmente, tal etapa combinada de lisis celular a alta temperatura y división de reactivo puede tener lugar a la misma temperatura y/o durante la etapa de hibridación de los oligonucleótidos con código de barras con los ácidos nucleicos diana.

5

La muestra de ácido nucleico puede tener una concentración de células para la etapa (a) de menos de 10 picomolar, menos de 1 picomolar, menos de 100 femtomolar, menos de 10 femtomolar, menos de 1 femtomolar, menos de 100 attomolar, menos de 10 attomolar o menos de 1 attomolar. También se pueden usar concentraciones alternativas más altas o más bajas. Preferentemente, las células estarán en una concentración de menos de 10 femtomolar.

10 La muestra de ácido nucleico puede tener una concentración de células para la etapa (b) de menos de 10 picomolar, menos de 1 picomolar, menos de 100 femtomolar, menos de 10 femtomolar, menos de 1 femtomolar, menos de 100 attomolar, menos de 10 attomolar o menos de 1 attomolar. También se pueden usar concentraciones alternativas más altas o más bajas. Preferentemente, las células estarán en una concentración de menos de 10 femtomolar.

15 La muestra de ácido nucleico puede tener una concentración de células para la etapa (c) de menos de 10 picomolar, menos de 1 picomolar, menos de 100 femtomolar, menos de 10 femtomolar, menos de 1 femtomolar, menos de 100 attomolar, menos de 10 attomolar o menos de 1 attomolar. También se pueden usar concentraciones alternativas más altas o más bajas. Preferentemente, las células estarán en una concentración de menos de 10 femtomolar.

20 En los métodos, antes de la etapa (b), el método puede comprender diluir la muestra de ácido nucleico. La etapa de diluir la muestra se puede realizar después de una etapa de unión de restos de unión a células (de oligonucleótidos adaptadores, oligonucleótidos con códigos de barras y/o reactivos de códigos de barras multiméricos) a las membranas celulares de las células de la muestra. La muestra de ácido nucleico puede tener una concentración de células para la etapa (a) y/o la etapa (b) y/o la etapa (c) de menos de 10 picomolar, menos de 1 picomolar, menos de 100 femtomolar, menos de 10 femtomolar, menos de 1 femtomolar, menos de 100 attomolar, menos de 10 attomolar o menos de 1 attomolar. También se pueden usar concentraciones alternativas más altas o más bajas.

25 Preferentemente, las células estarán en una concentración de menos de 10 femtomolar. Tener una baja concentración de células en la muestra de ácido nucleico durante las etapas (b) y (c) puede reducir el 'código de barras cruzado' de dos células físicamente cercanas por el mismo reactivo de código de barras multimérico.

30 En los métodos, cualquiera de las etapas (a), (b) y/o (c) se puede realizar en una solución de alta viscosidad. Opcionalmente, tal solución de alta viscosidad puede estar compuesta por una solución de poli(etilenglicol) (PEG), tal como PEG 20.000. Opcionalmente, tal solución puede comprender al menos el 5 % de poli(etilenglicol), al menos el 10 % de poli(etilenglicol), al menos el 20 % de poli(etilenglicol), al menos el 25 % de poli(etilenglicol), al menos el 30 % de poli(etilenglicol), al menos el 40 % de poli(etilenglicol), o al menos el 50 % de poli(etilenglicol) en peso o en volumen. Opcionalmente, tal solución de alta viscosidad puede comprender un gel o hidrogel solidificado o semisolidificado, tal como un gel de agarosa, un gel de poliacrilamida, un gel reticulado tal como un hidrogel de PEG-acrilato/PEG-tiol reticulado o un gel de copolímero de bloqueo. Opcionalmente, tal solución de alta viscosidad puede comprender la solución empleada durante cualquier etapa de lisis celular y/o permeabilización celular. Opcionalmente, tal solución de alta viscosidad puede comprender la solución empleada durante cualquier etapa de hibridación de oligonucleótidos con códigos de barras a ácidos nucleicos diana. Opcionalmente, tal solución de alta viscosidad puede tener una viscosidad dinámica de al menos 1,0 centipoise, al menos 1,1 centipoise, al menos 1,2 centipoise, al menos 1,5 centipoise, al menos 2,0 centipoise, al menos 5,0 centipoise, al menos 10,0 centipoise, al menos 20,0 centipoise, al menos 50,0 centipoise, al menos 100,0 centipoise, o al menos 200,0 centipoise (por ejemplo, a 25 grados Celsius a la presión estándar al nivel del mar). Preferentemente, tal solución de alta viscosidad tendrá una viscosidad dinámica de al menos 2,0 centipoise. El uso de una solución de alta viscosidad puede ralentizar la difusión de los oligonucleótidos con código de barras y sus ácidos nucleicos diana alejándose entre sí, es decir, cuando un reactivo de código de barras multimérico se ha unido a la membrana de una sola célula en particular, y luego la membrana se lisa o permeabiliza, una solución de alta viscosidad tendrá el efecto de mantener los oligonucleótidos con código de barras y los ácidos nucleicos diana de las células en las proximidades de la célula original durante un período de tiempo más largo, manteniendo así la 'concentración' eficaz de ambos más alta durante un período de tiempo más largo (ya que ocuparán un volumen total menor durante un período de tiempo más largo). Esta difusión retardada también puede tener el efecto adicional de ralentizar la difusión de ácidos nucleicos diana de una célula a un volumen ocupado por ácidos nucleicos diana de otra célula.

55 En los métodos, después de poner en contacto una muestra que comprende células con una biblioteca de al menos 2 reactivos de códigos de barras multiméricos, los oligonucleótidos con código de barras se pueden digerir o digerir parcialmente con una etapa de digestión con exonucleasa. Opcionalmente, esta etapa de digestión de exonucleasas se puede realizar antes o después, una etapa de transferencia de reactivos de códigos de barras multiméricos a las células. Opcionalmente, esta etapa de digestión de exonucleasas se puede realizar antes o después, una etapa de hibridación de oligonucleótidos con código de barras para dirigir los ácidos nucleicos de las células. Opcionalmente, esta etapa de digestión con exonucleasa se puede realizar mediante Exonucleasa I de E. coli, o Lambda exonucleasa de E. coli.

60 En los métodos, una muestra que comprende células y/o una biblioteca de dos o más reactivos de códigos de barras

multiméricos puede ponerse en contacto con una solución de uno o más oligonucleótidos bloqueantes, en donde dichos oligonucleótidos de bloqueo pueden ser complementarios a todo o parte de uno o más oligonucleótidos con código de barras. Opcionalmente, dichos oligonucleótidos de bloqueo pueden ser complementarios a toda o parte de la región diana de uno o más oligonucleótidos con código de barras.

- 5 En los métodos, una muestra que comprende células y/o una biblioteca de dos o más reactivos de códigos de barras multiméricos puede ponerse en contacto con una solución de uno o más oligonucleótidos bloqueantes, en donde los oligonucleótidos de bloqueo pueden ser complementarios a todo o parte de uno o más ácidos nucleicos diana. Opcionalmente, los oligonucleótidos de bloqueo pueden ser complementarios a una o más secuencias específicas de ADN o ARN. Opcionalmente, los oligonucleótidos de bloqueo pueden ser complementarios a una o más secuencias de ARN mensajero (ARNm). Opcionalmente, los oligonucleótidos de bloqueo pueden ser complementarios a la
- 10 secuencia de cola poli(A) de las secuencias de ARN mensajero (ARNm). Opcionalmente, los oligonucleótidos de bloqueo pueden comprender una secuencia de poli(T) de al menos 2, al menos 3, al menos 5, al menos 10, al menos 20, al menos 30, o al menos 50 nucleótidos que son complementarios a la secuencia de cola poli(A) de las secuencias de ARN mensajero (ARNm).
- 15 Opcionalmente, cualquiera de dichos oligonucleótidos de bloqueo puede hibridar con las secuencias respectivas a las que son complementarias o parcialmente complementarias. Opcionalmente, la temperatura de hibridación a la que tales oligonucleótidos de bloqueo hibridan con sus respectivas secuencias complementarias puede ser menor que la temperatura a la que la región diana de los oligonucleótidos con código de barras hibrida con la región diana de sus ácidos nucleicos celulares diana. Opcionalmente, esta etapa de bloqueo de oligonucleótidos se puede realizar antes,
- 20 o se puede realizar después, una etapa de poner en contacto una muestra de células con una biblioteca de dos o más reactivos de códigos de barras multiméricos. Opcionalmente, esta etapa de bloqueo de oligonucleótidos se puede realizar antes, o se puede realizar después, una etapa de transferencia de reactivos de códigos de barras multiméricos a las células. Opcionalmente, esta etapa de bloqueo de oligonucleótidos se puede realizar antes, o se puede realizar después, una etapa de unión de reactivos de códigos de barras multiméricos a la superficie de las células, en donde
- 25 dichos reactivos de códigos de barras multiméricos comprenden restos de unión a células. Opcionalmente, esta etapa de bloqueo de oligonucleótidos se puede realizar antes, o se puede realizar después, una etapa de lisar o permeabilizar las células. Opcionalmente, esta etapa de bloqueo de oligonucleótidos se puede realizar antes, o se puede realizar después, una etapa de hibridación de oligonucleótidos con código de barras a los ácidos nucleicos de las células. Opcionalmente, esta etapa de bloqueo de oligonucleótidos se puede realizar después de una etapa de hibridación de
- 30 oligonucleótidos con códigos de barras a ácidos nucleicos de las células, en donde la etapa de bloqueo de oligonucleótidos comprende un proceso de reducción de la temperatura de la solución de muestra a una temperatura igual o inferior a la temperatura a la que los oligonucleótidos de bloqueo hibridan con sus respectivas secuencias. Opcionalmente, esta etapa de bloqueo de oligonucleótidos se puede realizar en una biblioteca de reactivos de códigos de barras multiméricos, antes de poner en contacto una muestra de células con dicha biblioteca. Opcionalmente, los
- 35 oligonucleótidos de bloqueo pueden comprender un resto de bloqueo en su extremo 3' que evita la extensión de dicho extremo 3' por una polimerasa. Cualquiera de estos oligonucleótidos bloqueantes puede estar presente en una concentración de al menos 1 nanomolar, al menos 10 nanomolar, al menos 100 nanomolar, o al menos 1 micromolar.

- Se pueden incluir uno o más oligonucleótidos de bloqueo juntos en la misma solución como un tensioactivo químico y/o dentro de la misma solución como un disolvente molecular y/o dentro de la misma solución como un
- 40 desnaturalizante de ácido nucleico y/o dentro de la misma solución como una biblioteca de reactivos de códigos de barras multiméricos.

- En los métodos, después de una etapa de hibridación de oligonucleótidos con código de barras a ácidos nucleicos diana, se puede realizar una incubación de bloqueo para hibridar oligonucleótidos de bloqueo con secuencias complementarias dentro de oligonucleótidos con código de barras. Opcionalmente, esta incubación de bloqueo se
- 45 puede realizar a una temperatura por debajo de la temperatura a la que los oligonucleótidos con códigos de barras hibridan con los ácidos nucleicos de las células. Opcionalmente, esta incubación de bloqueo se puede realizar a una temperatura por debajo de la temperatura a la que los oligonucleótidos de bloqueo hibridan con secuencias complementarias dentro de los oligonucleótidos con código de barras.

- En los métodos, se puede realizar una etapa de selección del tamaño de ácido nucleico después de una etapa de
- 50 hibridación de oligonucleótidos con código de barras a ácidos nucleicos diana. Opcionalmente, esta etapa se puede realizar mediante una etapa de selección de tamaño basada en gel. Opcionalmente, esta etapa de selección de tamaño se puede realizar con un proceso de inmovilización reversible en fase sólida, tal como una etapa de selección de tamaño que implica perlas magnéticas o superparamagnéticas. Opcionalmente, esta etapa de selección de tamaño se puede realizar con una etapa de purificación de ácido nucleico o de selección de tamaño basada en columna.
- 55 Opcionalmente, esta etapa de selección de tamaño puede eliminar de forma selectiva o preferencial los oligonucleótidos con códigos de barras que no están hibridados o unidos a ácidos nucleicos de las células. Opcionalmente, esta etapa de selección de tamaño puede eliminar preferentemente moléculas de ácido nucleico de menos de 50 nucleótidos de longitud, menos de 100 nucleótidos de longitud, menos de 150 nucleótidos de longitud, menos de 200 nucleótidos de longitud, menos de 300 nucleótidos de longitud, menos de 400 nucleótidos de longitud,
- 60 menos de 500 nucleótidos de longitud o menos de 1000 nucleótidos de longitud.

En los métodos, los reactivos de códigos de barras multiméricos, los oligonucleótidos de código de barras y/o las

moléculas de código de barras multiméricas pueden comprender uno o más restos de biotina.

En los métodos, después de una etapa de hibridación de oligonucleótidos con código de barras para el ácido nucleico diana de una muestra celular, los oligonucleótidos con código de barras y/o moléculas de ácido nucleico diana con código de barras y/o reactivo(s) de código de barras multimérico se pueden aislar por un proceso de: (a) poner en contacto la mezcla resultante con un soporte sólido, opcionalmente en donde el soporte sólido comprende restos de estreptavidina; y (b) capturar los oligonucleótidos con código de barras y/o las moléculas de ácido nucleico diana con código de barras y/o los reactivos de códigos de barras multiméricos en el soporte sólido, opcionalmente a través de la interacción estreptavidina-biotina.

El soporte sólido puede ser una o más perlas magnéticas, opcionalmente en donde una o más perlas magnéticas comprenden moléculas de estreptavidina en su superficie.

Las perlas magnéticas pueden aislarse de una mezcla de reacción con un imán.

En los métodos, la muestra de ácido nucleico puede comprender al menos 2, al menos 5, al menos 10, al menos 100 o al menos 10^3 , al menos 10^4 , al menos 10^5 , al menos 10^6 , al menos 10^7 , al menos 10^8 o al menos 10^9 células, en donde estas células están comprendidas dentro de un único volumen acuoso contiguo durante cualquier etapa de poner en contacto la muestra con una biblioteca de reactivo de código de barras multimérico (etapa (a)), y/o cualquier etapa de lisar o permeabilizar células (etapa (b)), y/o cualquier etapa de agregar secuencias de códigos de barras a los ácidos nucleicos diana (etapas (c), (d) y/o (e)). Preferentemente, en los métodos, la muestra de ácido nucleico comprende al menos 10 células, en donde estas células están comprendidas dentro de un único volumen acuoso contiguo durante cualquier etapa de poner en contacto la muestra con una biblioteca de reactivo de código de barras multimérico (etapa (a)), y cualquier etapa de lisar o permeabilizar células (etapa (b)), y cualquier etapa de agregar secuencias de códigos de barras a los ácidos nucleicos diana (etapas (c), (d) y/o (e)).

Opcionalmente, la muestra de ácido nucleico puede comprender al menos 2, al menos 5, al menos 10, al menos 100 o al menos 10^3 , al menos 10^4 , al menos 10^5 , al menos 10^6 , al menos 10^7 , al menos 10^8 o al menos 10^9 células, en donde estas células se dividen en dos o más volúmenes acuoso contiguo durante cualquier etapa de poner en contacto la muestra con una biblioteca de reactivo de código de barras multimérico (etapa (a)), y/o cualquier etapa de lisar o permeabilizar células (etapa (b)), y/o cualquier etapa de agregar secuencias de códigos de barras a los ácidos nucleicos diana (etapas (c), (d) y/o (e)).

En los métodos, la muestra de ácido nucleico puede comprender al menos 2, al menos 5, al menos 10, al menos 100 o al menos 10^3 , al menos 10^4 , al menos 10^5 , al menos 10^6 , al menos 10^7 , al menos 10^8 o al menos 10^9 células, en donde estas células no se dividen en dos o más volúmenes acuoso contiguos durante cualquier etapa de poner en contacto la muestra con una biblioteca de reactivo de código de barras multimérico (etapa (a)), y/o cualquier etapa de lisar o permeabilizar células (etapa (b)), y/o cualquier etapa de agregar secuencias de códigos de barras a los ácidos nucleicos diana (etapas (c), (d) y/o (e)).

Opcionalmente, las moléculas de ácido nucleico diana con código de barras se producen a partir de ácidos nucleicos diana de al menos 2, al menos 5, al menos 10, al menos 100 o al menos 10^3 , al menos 10^4 , al menos 10^5 , al menos 10^6 , al menos 10^7 , al menos 10^8 o al menos 10^9 células.

Opcionalmente, se determinan las secuencias de las moléculas de ácido nucleico diana con código de barras producidas durante al menos 10, al menos 100 o al menos 10^3 , al menos 10^4 , al menos 10^5 , al menos 10^6 , al menos 10^7 , al menos 10^8 o al menos 10^9 células.

En los métodos, la biblioteca puede comprender al menos 100, o al menos 10^3 , al menos 10^4 , al menos 10^5 , al menos 10^6 , al menos 10^7 , al menos 10^8 o al menos 10^9 reactivos de códigos de barras multiméricos. En los métodos, para cada reactivo de códigos de barras multimérico, se pueden producir al menos 2, al menos 3, al menos 5, al menos 10, al menos 25, al menos 50, al menos 100, al menos 500, al menos 1000, al menos 5.000, al menos 10.000, o al menos 50.000 moléculas de ácido nucleico diana con código de barras a partir de los ácidos nucleicos diana de una sola célula. Preferentemente, se pueden producir al menos 2 moléculas de ácido nucleico diana con código de barras a partir de los ácidos nucleicos diana de una sola célula para cada reactivo de código de barras multimérico.

En los métodos, cada reactivo de código de barras multimérico puede comprender al menos 2, al menos 3, al menos 5, al menos 10, al menos 20, al menos 50, al menos 100, al menos 200, al menos 500, al menos 1000, al menos 5000, al menos 10.000, al menos 100.000, o al menos 1.000.000 de oligonucleótidos con código de barras. Opcionalmente, diferentes reactivos de códigos de barras multiméricos dentro de una biblioteca de reactivos de códigos de barras multiméricos pueden comprender diferentes números de oligonucleótidos con códigos de barras.

En los métodos, en promedio, los oligonucleótidos de código de barras de un solo reactivo de código de barras multimérico pueden hibridar, de forma acumulativa, con al menos 1, al menos 2, al menos 3, al menos 5, al menos 10, al menos 20, al menos 50, al menos 100, al menos 200, al menos 500, al menos 1000, al menos 5000, al menos 10.000 o al menos 100.000 ácidos nucleicos diana de las células.

En los métodos, el grupo de secuencias de ácido nucleico diana complementarias a las regiones diana de diferentes

oligonucleótidos con código de barras dentro de un reactivo de código de barras multimérico o una biblioteca de reactivos de código de barras multimérico puede comprender al menos 2 secuencias de ácido nucleico diferentes, al menos 3 secuencias de ácidos nucleicos diferentes, al menos 4 secuencias de ácidos nucleicos diferentes, al menos 5 secuencias de ácidos nucleicos diferentes, al menos 10 secuencias de ácidos nucleicos diferentes, al menos 20 secuencias de ácidos nucleicos diferentes, al menos 50 secuencias de ácidos nucleicos diferentes, al menos 100 secuencias de ácidos nucleicos diferentes o al menos 1000 secuencias de ácidos nucleicos diferentes.

En los métodos, durante cualquier etapa(s), dentro de una solución, volumen, o reacción, las células pueden estar presentes en concentraciones particulares dentro del volumen de la solución, por ejemplo, a concentraciones de menos de 10 picomolar, menos de 1 picomolar, menos de 100 femtomolar, menos de 10 femtomolar, menos de 1 femtomolar, menos de 100 attomolar, menos de 10 attomolar o menos de 1 attomolar.

En los métodos, durante cualquier etapa(s), dentro de una solución, volumen, o reacción, los reactivos de códigos de barras multiméricos pueden estar presentes en concentraciones particulares dentro del volumen de la solución, por ejemplo, a concentraciones de al menos 100 nanomolar, al menos 10 nanomolar, al menos 1 nanomolar, al menos 100 picomolar, al menos 10 picomolar, al menos 1 picomolar, al menos 100 femtomolar, al menos 10 femtomolar o al menos 1 femtomolar.

En los métodos, una muestra que comprende células permeabilizadas, lisadas o intactas y/o que comprenden reactivos de códigos de barras multiméricos y/o que comprenden oligonucleótidos con códigos de barras y/o que comprenden otras secuencias de oligonucleótidos, puede dividirse en dos o más volúmenes de partición. Opcionalmente, cada uno de dichos volúmenes de partición puede comprender un recipiente de reacción física diferente. Opcionalmente, dichos volúmenes de partición pueden comprender, cada uno, una gota diferente dentro de una emulsión, tal como diferentes gotitas acuosas dentro de una emulsión de agua en aceite. Tal evento de partición puede tener lugar antes y/o durante una o más etapas dentro de cualquier protocolo. Siguiendo esta etapa de partición, las reacciones de dos o más de tales particiones se pueden fusionar para formar un solo volumen de reacción.

En los métodos, la muestra de ácido nucleico puede comprender células intactas. La muestra de ácido nucleico puede comprender células que se hayan degradado parcialmente. La muestra de ácido nucleico puede comprender células que han sido parcialmente permeabilizadas y/o fragmentadas. La muestra de ácido nucleico puede comprender células que se han reticulado con formalina y se han embebido en parafina (es decir, una muestra de EPFF).

La muestra de ácido nucleico puede comprender células que están contenidas dentro de una muestra o sección de tejido intacta, o una muestra o sección de tejido parcialmente intacta. La muestra de ácido nucleico puede comprender células que se han procesado mediante un proceso de disociación de tejido y/o digestión de tejido. Opcionalmente, tal proceso de disociación o digestión puede comprender la digestión con una proteinasa como la proteinasa K.

La muestra de ácido nucleico puede comprender células que se han procesado mediante un proceso de clasificación de células, tal como un proceso de clasificación de células activadas por fluorescencia (FACS). La muestra de ácido nucleico puede comprender células que se encuentran dentro de una única suspensión celular.

La muestra de ácido nucleico puede comprender linfocitos, tal como linfocitos T y/o linfocitos B, y/o una mezcla de células inmunitarias tal como una muestra de células mononucleares de sangre periférica (PBMC). Por ejemplo, se puede utilizar un único reactivo de código de barras multimérico para añadir secuencias de códigos de barras a las secuencias de un ARNm de inmunoglobulina de cadena pesada y un ARNm de inmunoglobulina de cadena ligera del mismo único. Como alternativa, se puede usar un único reactivo de código de barras multimérico para añadir secuencias de códigos de barras a las secuencias de un ARNm de cadena alfa y un ARNm de cadena beta de un receptor de linfocitos T.

La muestra de ácido nucleico puede comprender células tumorales. Opcionalmente, la muestra puede comprender linfocitos infiltrantes de tumores (TIL). Opcionalmente, la muestra puede comprender muestras tumorales que comprenden tanto células tumorales como linfocitos infiltrantes de tumores. Opcionalmente, la muestra puede comprender células tumorales en circulación (CTC). La muestra de ácido nucleico puede ser una muestra humana.

El ácido nucleico diana puede ser una molécula de ácido nucleico (única) intacta de una célula, dos o más fragmentos de una molécula de ácido nucleico de una célula (tales fragmentos pueden estar colocalizados en la muestra) o dos o más moléculas de ácido nucleico de una célula. Por lo tanto, las subsecuencias de un ácido nucleico diana de una célula pueden ser subsecuencias de la misma molécula de ácido nucleico, subsecuencias de diferentes fragmentos de la misma molécula de ácido nucleico, o secuencias o subsecuencias de diferentes moléculas de ácido nucleico (por ejemplo, secuencias de diferentes moléculas de ARN mensajero (o porciones de las mismas) de una célula; por ejemplo, la primera y segunda subsecuencia de un ácido nucleico diana de una célula pueden ser una primera y una segunda moléculas de ARN mensajero diferentes (o partes de las mismas) de una célula).

Tal como se usa en el presente documento, el término ácido nucleico diana se refiere a los ácidos nucleicos presentes dentro de las células y a copias o amplicones de los mismos. Por ejemplo, cuando el ácido nucleico diana es el ADN genómico, el término ácido nucleico diana significa ADN genómico presente en una célula y copias o amplicones del mismo, por ejemplo, moléculas de ADN que se pueden preparar a partir del ADN genómico mediante una reacción de extensión del cebador. Como ejemplo adicional, cuando el ácido nucleico diana es ARNm, el término ácido nucleico

diana significa ARNm presente en la célula y copias o amplicones del mismo, por ejemplo, ADNc sintetizado a partir del ARNm por transcripción inversa.

En cualquiera de los métodos, los ácidos nucleicos diana pueden ser ADN (por ejemplo, ADN genómico) o ARN (por ejemplo, ARNm). Dichos ácidos nucleicos diana pueden comprender ADN o ARN de cualquier origen; por ejemplo, pueden comprender ADN genómico natural o no modificado o ARN mensajero de una muestra de células *in vivo* o *in vitro*. Además, pueden comprender ADN o ARN de cualquier tipo de origen sintético, tal como ADN (y/o transcritos de ARN expresado asociado) de cualquier tipo de transfección o método de transducción, tal como plásmidos lineales o circulares, construcciones de transfección vírica, ADN administrado exógenamente de cualquier tipo, ARN administrado exógenamente de cualquier tipo (como ARN mensajero administrado exógenamente o ARN de interferencia corto o ARN de horquilla corto), o construcciones CRISPR y/o construcciones de expresión CRISPR y/o derivados de las mismas (por ejemplo, una nucleasa Cas9 y/o versión expresada del mismo, y/o un ARN guía y/o una versión expresada del mismo). Además, los ácidos nucleicos diana pueden comprender secuencias de ADN y/o ARN que comprenden secuencias de códigos de barras o identificadores, en donde una muestra de células (por ejemplo, una muestra de células *in vitro* o una población de células *in vivo*) se ha puesto en contacto y/o se ha modificado genéticamente con una biblioteca combinada de dos o más secuencias sintéticas diferentes, en donde cada una de dichas dos o más secuencias sintéticas comprende una secuencia de identificación como una secuencia de código de barras (como secuencias de 'Código de barras guía' (GBC) dentro de los transcritos GBC expresados dentro del protocolo Perturb-Seq [Dixit *et al.*, 2016, Cell 167, 1853-1866 y Adamson *et al.*, 2016, Cell 167, 1867-1882], o códigos de barras de secuencia de identificación de bibliotecas de expresión lentivírica [por ejemplo, las bibliotecas de ARNhc lentivírica DECIPHER murino, CELLECTA, Inc]). En tales enfoques, dichas secuencias de identificación, al estar codificadas con barras y secuenciadas por cualquier método descrito en el presente documento, se puede usar para determinar con qué una o más (si las hay) secuencias sintéticas con las que una célula dada dentro de la muestra o población de células se puso en contacto y/o se modificó genéticamente.

En cualquiera de los métodos, los ácidos nucleicos diana pueden comprender secuencias de ácidos nucleicos administradas exógenamente que comprenden secuencias de códigos de barras dentro de una sonda de afinidad con códigos de barras, en donde una sonda de afinidad con código de barras comprende al menos un resto de afinidad unido al menos a una secuencia de código de barras.

Opcionalmente, cualquier resto de afinidad puede comprender uno o más de: un anticuerpo, un fragmento de anticuerpo, un fragmento de anticuerpo de cadena ligera, un fragmento variable monocatenario (scFv), un péptido, un péptido que penetra en las células, un aptámero, un aptámero de ADN y/o un aptámero de ARN. Opcionalmente, uno o más restos de afinidad pueden comprender un resto capaz de unirse y/o comprender una afinidad alta y/o específica por, una proteína específica, glicoproteína, proteína modificada postraduccionalmente y/u otras especies químicas o moleculares. Opcionalmente, uno o más de tales restos de afinidad pueden comprender un resto capaz de unirse y/o comprender una afinidad alta y/o específica por, una proteína específica, glicoproteína, proteína modificada postraduccionalmente y/u otra especie química o molecular comprendida en la superficie de una célula, y/o comprendida dentro de la membrana celular de una célula, y/o comprendida dentro del citoplasma de una célula, y/o comprendida dentro del núcleo de una célula y/o cualquier combinación de los mismos.

Cualquier sonda de afinidad con código de barras puede comprender un oligonucleótido con código de barras de sonda, en donde dicho oligonucleótido con código de barras de sonda comprende una secuencia de código de barras asociada con y/o identificando el resto de afinidad al que está unido. Opcionalmente, cualquier secuencia de código de barras puede comprender una secuencia de al menos 1, al menos 2, al menos 3, al menos 5, al menos 10, al menos 20, o al menos 30 nucleótidos de longitud. Opcionalmente, todos los oligonucleótidos de código de barras de sonda unidos con el mismo resto de afinidad particular (por ejemplo, la misma especie de anticuerpo particular específica para la misma proteína diana) pueden comprender la misma secuencia (por ejemplo, la misma secuencia de código de barras de identificación). Opcionalmente, los oligonucleótidos con código de barras de sonda enlazados con el mismo resto de afinidad particular (por ejemplo, la misma especie de anticuerpo particular específica para la misma proteína diana) pueden comprender dos o más secuencias diferentes (por ejemplo, dos o más secuencias de códigos de barras de identificación diferentes). Opcionalmente, cualquier oligonucleótido con código de barras de sonda puede comprender un adaptador y/o una secuencia de acoplamiento, en donde dicha secuencia es al menos 1, al menos 2, al menos 3, al menos 5, al menos 10, al menos 20, o al menos 30 nucleótidos de longitud. Opcionalmente, cualquier adaptador y/o secuencia de acoplamiento dentro de un oligonucleótido con código de barras de sonda puede comprender una secuencia complementaria a una región diana de un oligonucleótido con código de barras comprendido dentro de cualquier reactivo de código de barras multimérico y/o biblioteca del mismo. Opcionalmente, cualquier adaptador y/o secuencia de acoplamiento dentro de un oligonucleótido con código de barras de sonda puede comprender una secuencia de poli(A) de 2 o más nucleótidos de longitud. Opcionalmente, cualquier adaptador y/o secuencia de acoplamiento dentro de un oligonucleótido con código de barras de sonda puede estar comprendido dentro del extremo 3' y/o dentro del extremo 5', de dicho oligonucleótido con código de barras de sonda.

Cualquier oligonucleótido con código de barras de sonda y resto de afinidad comprendido dentro de una sonda de afinidad con código de barras se puede unir por cualquier medio. Opcionalmente, un oligonucleótido con código de barras de sonda y un resto de afinidad se pueden unir mediante un enlace covalente (por ejemplo, como los kits de marcado de anticuerpos LighteningLink de Innova Biosciences). Opcionalmente, un oligonucleótido con código de barras de sonda y un resto de afinidad pueden estar unidos por un enlace no covalente (usando, por ejemplo, en

donde un resto de afinidad comprende un dominio de estreptavidina, y en donde un oligonucleótido con código de barras de sonda comprende un resto de biotina para generar un enlace de biotina/estreptavidina no covalente).

Cualquiera o más sondas de afinidad con código de barras se pueden poner en contacto y/o incubar con una muestra de células en la que dichas sondas de afinidad de código de barras se encuentran en cualquier concentración, por ejemplo, a concentraciones de al menos 100 nanomolar, al menos 10 nanomolar, al menos 1 nanomolar, al menos 100 picomolar, al menos 10 picomolar, al menos 1 picomolar, al menos 100 femtomolar, al menos 10 femtomolar o al menos 1 femtomolar. Las concentraciones pueden ser de 1 picomolar a 100 nanomolar, 10 picomolar a 10 nanomolar, o 100 picomolar a 1 nanomolar.

Opcionalmente, en los métodos se puede utilizar un grupo de dos o más sondas de afinidad de códigos de barras diferentes. El grupo (o biblioteca) puede comprender: una primera sonda de afinidad de código de barras que comprende un primer resto de afinidad y un primer oligonucleótido con código de barras de sonda, en donde el primer resto de afinidad es capaz de unirse y/o comprender una afinidad alta y/o específica por, una primera diana (por ejemplo, una proteína específica, una glucoproteína, una proteína modificada postraduccionalmente y/u otra especie química o molecular); y una segunda sonda de afinidad de código de barras que comprende un segundo resto de afinidad y una segunda sonda de oligonucleótido con código de barras, en donde el segundo resto de afinidad es capaz de unirse y/o comprender una afinidad alta y/o específica por, una segunda diana (por ejemplo, una proteína específica, una glucoproteína, una proteína modificada postraduccionalmente y/u otra especie química o molecular). El grupo (o biblioteca) de sondas de afinidad con código de barras se puede proporcionar dentro de una única solución. El grupo (o biblioteca) de sondas de afinidad con código de barras puede ponerse en contacto y/o incubarse con células. Opcionalmente, el grupo (o biblioteca) puede comprender al menos 3, al menos 5, al menos 10, al menos 20, o al menos 30 sondas de afinidad con códigos de barras diferentes (por ejemplo, dirigidas a al menos 3, al menos 5, al menos 10, al menos 20, o al menos 30 dianas diferentes (por ejemplo, proteínas específicas, glucoproteínas, proteínas modificadas postraduccionalmente y/u otras especies químicas o moleculares)).

Opcionalmente, los ácidos nucleicos diana pueden comprender un oligonucleótido con código de barras de sonda dentro de sondas de afinidad con código de barras, en donde una muestra de células (por ejemplo, una muestra de células *in vitro* o una población de células *in vivo*) se ha puesto en contacto y/o incubado con una o más de tales sondas de afinidad con código de barras. Opcionalmente, una muestra de células puede reticularse químicamente (por ejemplo, con formaldehído) antes de cualquier etapa de poner en contacto y/o incubar células con una o más sondas de afinidad con código de barras. Opcionalmente, una muestra de células se puede permeabilizar (por ejemplo, con un tensioactivo químico) antes de cualquier etapa de poner en contacto y/o incubar las células con una o más sondas de afinidad con código de barras. Opcionalmente, una muestra de células se puede reticular químicamente (por ejemplo, con formaldehído) y después permeabilizar (por ejemplo, con un tensioactivo químico) antes de cualquier etapa de poner en contacto y/o incubar las células con una o más sondas de afinidad con código de barras.

Opcionalmente, los ácidos nucleicos diana pueden comprender tanto ácidos nucleicos comprendidos dentro de una muestra de células como también oligonucleótido(s) de código de barras de sonda dentro de sondas de afinidad con código de barras, en donde la muestra de células (por ejemplo, una muestra de células *in vitro* o una población de células *in vivo*) se ha puesto en contacto y/o incubado con una o más de tales sondas de afinidad con código de barras. Opcionalmente, los ácidos nucleicos diana pueden comprender moléculas de ARN mensajero comprendidas dentro de una muestra de células como también oligonucleótido(s) de código de barras de sonda dentro de sondas de afinidad con código de barras, en donde la muestra de células (por ejemplo, una muestra de células *in vitro* o una población de células *in vivo*) se ha puesto en contacto y/o incubado con una o más de tales sondas de afinidad con código de barras.

En los métodos, los ácidos nucleicos diana de las células con las que hibridan los oligonucleótidos con códigos de barras pueden comprender secuencias de acoplamiento (por ejemplo, secuencias sintéticas de ácidos nucleicos). Opcionalmente, la región diana de los oligonucleótidos con códigos de barras dentro de los reactivos de códigos de barras multiméricos puede comprender secuencias complementarias a dichas secuencias de acoplamiento con las que pueden hibridar. Opcionalmente, cualquiera de dichas secuencias de acoplamiento puede comprender todos o porciones de oligonucleótidos sintéticos que se han transferido a células dentro de la muestra de ácido nucleico. Opcionalmente, tales oligonucleótidos sintéticos pueden comprender una región de hibridación de reactivo y una región de direccionamiento, en donde la región de hibridación de reactivo es total o parcialmente complementaria a una región diana dentro de un oligonucleótido con código de barras, y en donde la región de direccionamiento es completa o parcialmente complementaria a una secuencia de ácido nucleico que se encuentra dentro de la muestra de ácido nucleico. Opcionalmente, una región de direccionamiento puede ser completa o parcialmente complementaria a una secuencia dentro del ADN genómico, o a una secuencia dentro de una o más moléculas de ARN mensajero (ARNm). Opcionalmente, tales oligonucleótidos sintéticos pueden comprender una región de enlazador de al menos 1 nucleótido entre una región de hibridación de reactivo y una región de direccionamiento. Opcionalmente, la región de hibridación de reactivo puede ubicarse dentro del extremo 5' de un oligonucleótido sintético y una región de direccionamiento puede ubicarse dentro del extremo 3' del oligonucleótido sintético. Opcionalmente, una solución de uno o más oligonucleótidos sintéticos puede hibridar con uno o más ácidos nucleicos diana dentro de las células en una etapa de hibridación de oligonucleótidos sintéticos. Opcionalmente, tal etapa de hibridación de oligonucleótidos sintéticos se puede realizar antes de poner en contacto la muestra de células con una biblioteca de dos o más reactivos de códigos de barras multiméricos.

En los métodos, los ácidos nucleicos diana de las células con las que hibridan los oligonucleótidos con códigos de barras pueden ser moléculas de ARNm (ARN mensajero). Opcionalmente, la región diana de oligonucleótidos con códigos de barras dentro de reactivos de códigos de barras multiméricos puede comprender secuencias complementarias a secuencias dentro de una o más moléculas de ARN mensajero con las que pueden hibridar.

- 5 Opcionalmente, las regiones diana de oligonucleótidos con código de barras pueden ser complementarias a secuencias específicas dentro de dianas específicas de ARN mensajero. Opcionalmente, las regiones diana de los oligonucleótidos con código de barras pueden ser complementarias a las regiones de la cola poli(A) de las moléculas de ARN mensajero; en este caso, las regiones diana de oligonucleótidos con código de barras pueden comprender una región poli(T) de dos o más nucleótidos contiguos.

- 10 En los métodos, cada molécula de ácido nucleico diana con código de barras puede producirse después del aislamiento del oligonucleótido con código de barras hibridado con una molécula de ARNm diana extendiendo el oligonucleótido con código de barras usando una transcriptasa inversa y en donde la molécula de ARNm diana se emplea como molde para un proceso de transcripción inversa por dicha transcriptasa.

- 15 En los métodos, las moléculas de ARNm pueden ser moléculas de ARNm correspondientes a cadenas alfa y/o beta de una secuencia de receptor de linfocitos T, opcionalmente, en donde se determinan las secuencias de cadenas alfa y beta emparejadas dentro de una célula individual.

En los métodos, las moléculas de ARNm pueden ser moléculas de ARNm correspondientes a cadenas ligeras y/o pesadas de una secuencia de inmunoglobulina, en donde opcionalmente se determinan las secuencias de cadenas ligeras y pesadas emparejadas dentro de una célula individual.

- 20 A continuación se proporcionan más detalles de las bibliotecas de reactivos de códigos de barras multiméricos y métodos de la divulgación.

1. PROPIEDADES GENERALES DE LOS REACTIVOS DE CÓDIGOS DE BARRAS MULTIMÉRICOS

- 25 La divulgación proporciona reactivos de códigos de barras multiméricos para etiquetar uno o más ácidos nucleicos diana. Un reactivo de código de barras multimérico comprende dos o más regiones de código de barras que están unidas (directa o indirectamente).

Cada región de código de barras comprende una secuencia de ácido nucleico. La secuencia de ácido nucleico puede ser ADN monocatenario, ADN bicatenario o ADN monocatenario con una o más regiones bicatenarias.

- 30 Cada región de código de barras puede comprender una secuencia que identifica el reactivo de código de barras multimérico. Por ejemplo, esta secuencia puede ser una región constante compartida por todas las regiones de códigos de barras de un único reactivo de códigos de barras multimérico. Cada región de código de barras puede contener una secuencia única que no está presente en otras regiones y, por tanto, puede servir para identificar de forma única cada región de código de barras. Cada región de código de barras puede comprender al menos 5, al menos 10, al menos 15, al menos 20, al menos 25, al menos 50 o al menos 100 nucleótidos. Preferentemente, cada región de código de barras comprende al menos 5 nucleótidos. Preferentemente, cada región de código de barras son desoxirribonucleótidos, opcionalmente, todos los nucleótidos en una región de código de barras son desoxirribonucleótidos. Uno o más de los desoxirribonucleótidos puede ser un desoxirribonucleótido modificado (por ejemplo, un desoxirribonucleótido modificado con un resto de biotina o un nucleótido de desoxiuracilo). Las regiones de código de barras pueden comprender uno o más nucleótidos o secuencias degenerados. Las regiones de código de barras pueden no comprender nucleótidos o secuencias degenerados. El reactivo de códigos de barras multimérico puede comprender al menos 5, al menos 10, al menos 20, al menos 25, al menos 50, al menos 75, al menos 100, al menos 200, al menos 500, al menos 1000, al menos 5000, o al menos 10.000 regiones de código de barras. Preferentemente, el reactivo de códigos de barras multimérico comprende al menos 5 regiones de códigos de barras.

- 45 El reactivo de códigos de barras multimérico puede comprender al menos 2, al menos 3, al menos 4, al menos 5, al menos 10, al menos 20, al menos 25, al menos 50, al menos 75, al menos 100, al menos 200, al menos 500, al menos 1000, al menos 5000, al menos 10^4 , al menos 10^5 , o al menos 10^6 regiones de códigos de barras únicas o diferentes. Preferentemente, el reactivo de código de barras multimérico comprende al menos 5 regiones de código de barras únicas o diferentes.

- 50 Un reactivo de códigos de barras multimérico puede comprender: la primera y segunda moléculas de código de barras unidas (es decir, una molécula de código de barras multimérica), en donde cada una de las moléculas de código de barras comprende una secuencia de ácido nucleico que comprende una región de código de barras.

- 55 Las moléculas de código de barras de una molécula de código de barras multimérica se pueden unir a una molécula de ácido nucleico. Las moléculas de código de barras de una molécula de código de barras multimérica pueden estar comprendidas dentro de una (única) molécula de ácido nucleico. Una molécula de código de barras multimérica puede comprender una sola secuencia de ácido nucleico contigua que comprende dos o más moléculas de código de barras. Una molécula de código de barras multimérica puede ser una molécula de ácido nucleico monocatenario (por ejemplo, ADN monocatenario), una molécula de ácido nucleico bicatenario o una molécula monocatenaria que comprende una o más regiones bicatenarias. Una molécula de código de barras multimérica puede comprender uno o más extremos

5' fosforilados capaces de ligarse a extremos 3' de otras moléculas de ácido nucleico. Opcionalmente, en una región bicatenaria o entre dos regiones bicatenarias diferentes, una molécula de código de barras multimérica puede comprender uno o más cortes, o uno o más espacios, donde la propia molécula de código de barras multimérica se ha dividido o separado. Cualquiera de dichos huecos puede ser al menos uno, al menos 2, al menos 5, al menos 10, al menos 20, al menos 50, o al menos 100 nucleótidos de longitud. Dichos cortes y/o espacios pueden servir para aumentar la flexibilidad molecular de la molécula de código de barras multimérica y/o reactivo de código de barras multimérico, por ejemplo, para aumentar la accesibilidad de la molécula o reactivo para interactuar con moléculas de ácido nucleico diana. Dichos cortes y/o espacios también pueden permitir una purificación o eliminación más eficaz de dichas moléculas o reactivos. Una molécula y/o reactivo que comprende dicho(s) corte(s) y/o hueco(s) puede conservar enlaces entre diferentes moléculas de código de barras al tener una hebra de ADN complementaria que hibrida conjuntamente con regiones de dos o más partes divididas de una molécula de código de barras multimérica.

Las moléculas de código de barras pueden estar unidas por un soporte, por ejemplo, una macromolécula, soporte sólido o soporte semisólido. Se pueden conocer las secuencias de las moléculas del código de barras unidas a cada soporte. Las moléculas de código de barras se pueden unir al soporte directa o indirectamente (por ejemplo, a través de una molécula enlazadora). Las moléculas de código de barras se pueden unir uniéndose al soporte y/o uniéndose o hibridando a moléculas enlazadoras que están unidas al soporte. Las moléculas de código de barras pueden estar unidas al soporte (o a las moléculas enlazadoras) por enlace covalente, enlace no covalente (por ejemplo, una interacción proteína-proteína o un enlace estreptavidina-biotina) o hibridación de ácido nucleico. La molécula enlazadora puede ser un biopolímero (por ejemplo, una molécula de ácido nucleico) o un polímero sintético. La molécula enlazadora puede comprender una o más unidades de etilenglicol y/o polietilenglicol (por ejemplo, hexaetilenglicol o pentaetilenglicol). La molécula enlazadora puede comprender uno o más grupos etilo, tales como un espaciador C3 (tres carbonos), espaciador C6, espaciador C12 o espaciador C18.

Las moléculas de código de barras pueden estar unidas por una macromolécula uniéndose a la macromolécula y/o hibridando con la macromolécula.

Las moléculas de código de barras se pueden unir a la macromolécula directa o indirectamente (por ejemplo, a través de una molécula enlazadora). Las moléculas de código de barras se pueden unir uniéndose a la macromolécula y/o uniéndose o hibridando a moléculas enlazadoras que están unidas a la macromolécula. Las moléculas de código de barras pueden estar unidas a la macromolécula (o a las moléculas enlazadoras) por enlace covalente, enlace no covalente (por ejemplo, una interacción proteína-proteína o un enlace estreptavidina-biotina) o hibridación de ácido nucleico. La molécula enlazadora puede ser un biopolímero (por ejemplo, una molécula de ácido nucleico) o un polímero sintético. La molécula enlazadora puede comprender una o más unidades de etilenglicol y/o polietilenglicol (por ejemplo, hexaetilenglicol o pentaetilenglicol). La molécula enlazadora puede comprender uno o más grupos etilo, tales como un espaciador C3 (tres carbonos), espaciador C6, espaciador C12, o espaciador C18.

La macromolécula puede ser un polímero sintético (por ejemplo, un dendrímero) o un biopolímero como un ácido nucleico (por ejemplo, un ácido nucleico monocatenario como ADN monocatenario), un péptido, un polipéptido o una proteína (por ejemplo, una proteína multimérica).

El dendrímero puede comprender al menos 2, al menos 3, al menos 5, o al menos 10 generaciones.

La macromolécula puede ser un ácido nucleico que comprende dos o más nucleótidos, cada uno de los cuales puede unirse a una molécula de código de barras. Adicionalmente o como alternativa, el ácido nucleico puede comprender dos o más regiones, cada una de las cuales puede hibridar con una molécula de código de barras.

El ácido nucleico puede comprender un primer nucleótido modificado y un segundo nucleótido modificado, en donde cada nucleótido modificado comprende un resto de unión (por ejemplo, un resto de biotina o un resto de alquino que se puede usar para una reacción química de clic) capaz de unirse a una molécula de código de barras. Opcionalmente, el primer y segundo nucleótidos modificados pueden estar separados por una secuencia de ácidos nucleicos intermedia de al menos uno, al menos dos, al menos 5 o al menos 10 nucleótidos.

El ácido nucleico puede comprender una primera región de hibridación y una segunda región de hibridación, en donde cada región de hibridación comprende una secuencia complementaria y capaz de hibridar con una secuencia de al menos un nucleótido dentro de una molécula de código de barras. La secuencia complementaria puede ser de al menos 5, al menos 10, al menos 15, al menos 20, al menos 25 o al menos 50 nucleótidos contiguos. Preferentemente, la secuencia complementaria es de al menos 10 nucleótidos contiguos. Opcionalmente, la primera y segunda regiones de hibridación pueden estar separadas por una secuencia de ácido nucleico intermedia de al menos una, al menos dos, al menos 5 o al menos 10 nucleótidos.

La macromolécula puede ser una proteína tal como una proteína multimérica, por ejemplo, una proteína homomérica o una proteína heteromérica. Por ejemplo, la proteína puede comprender estreptavidina, por ejemplo, estreptavidina tetramérica.

El soporte puede ser un soporte sólido o un soporte semisólido. El soporte puede comprender una superficie plana. El soporte puede ser un portaobjetos, por ejemplo, un portaobjetos de vidrio. El portaobjetos puede ser una celda de flujo para secuenciar. Si el soporte es un portaobjetos, la primera y la segunda moléculas de código de barras pueden

inmovilizarse en una región discreta del portaobjetos. Opcionalmente, las moléculas de código de barras de cada reactivo de código de barras multimérico en una biblioteca se inmovilizan en una región discreta diferente en el portaobjetos a las moléculas de código de barras de los otros reactivos de código de barras multimérico en la biblioteca. El soporte puede ser una placa que comprende pocillos, en donde opcionalmente la primera y la segunda moléculas de código de barras están inmovilizadas en el mismo pocillo. Opcionalmente, las moléculas de código de barras de cada reactivo de código de barras multimérico en la biblioteca se inmovilizan en un pocillo diferente de la placa a las moléculas de código de barras de los otros reactivos de código de barras multimérico en la biblioteca.

Preferentemente, el soporte es una perla (por ejemplo, una perla de gel). La perla puede ser una perla de agarosa, una perla de sílice, una perla de espuma de poliestireno, una perla de gel (como las disponibles en 10x Genomics®), una perla conjugada de anticuerpo, una perla conjugada oligo-dT, una perla de estreptavidina o una perla magnética (por ejemplo, una perla superparamagnética). La perla puede ser de cualquier tamaño y/o estructura molecular. Por ejemplo, la perla puede tener un diámetro de 10 nanómetros a 100 micrómetros, de 100 nanómetros a 10 micrómetros de diámetro, o de 1 micrómetro a 5 micrómetros de diámetro. Opcionalmente, la perla tiene aproximadamente 10 nanómetros de diámetro, aproximadamente 100 nanómetros de diámetro, aproximadamente 1 micrómetro de diámetro, aproximadamente 10 micrómetros de diámetro o aproximadamente 100 micrómetros de diámetro. La perla puede ser sólida o, como alternativa, la perla puede ser hueca o parcialmente hueca o porosa. Las perlas de ciertos tamaños pueden ser más preferibles para ciertos métodos de códigos de barras. Por ejemplo, perlas de menos de 5,0 micrómetros, o menos de 1,0 micrómetro, pueden ser más útiles para poner códigos de barras a dianas de ácidos nucleicos dentro de células individuales. Preferentemente, las moléculas de código de barras de cada reactivo de códigos de barras multimérico en una biblioteca están unidas en una perla diferente a las moléculas de código de barras de los otros reactivos de códigos de barras multiméricos en la biblioteca.

El soporte se puede funcionalizar para permitir la unión de dos o más moléculas de código de barras. Esta funcionalización se puede habilitar mediante la adición de restos químicos (por ejemplo, grupos carboxilados, alquinos, azidas, grupos de acrilato, grupos amino, grupos sulfato o grupos succinimida) y/o restos basados en proteínas (por ejemplo, estreptavidina, avidina o proteína G) al soporte. Las moléculas de código de barras se pueden unir a los restos directa o indirectamente (por ejemplo, mediante una molécula de enlazador).

Los soportes funcionalizados (por ejemplo, perlas) se pueden poner en contacto con una solución de moléculas de código de barras en condiciones que promuevan la unión de dos o más moléculas de código de barras a cada perla en la solución (generando reactivos de códigos de barras multiméricos).

En una biblioteca de reactivos de códigos de barras multiméricos, las moléculas de código de barras de cada reactivo de código de barras multimérico en una biblioteca se pueden unir entre sí en un soporte diferente a las moléculas de código de barras de los otros reactivos de código de barras multimérico en la biblioteca.

El reactivo de códigos de barras multimérico puede comprender: al menos 2, al menos 3, al menos 4, al menos 5, al menos 10, al menos 20, al menos 25, al menos 50, al menos 75, al menos 100, al menos 200, al menos 500, al menos 1000, al menos 5000, al menos 10^4 , al menos 10^5 , o al menos 10^6 moléculas de código de barras unidas entre sí, en donde cada molécula de código de barras es tal como se define en el presente documento; y un oligonucleótido con código de barras hibridado con cada molécula de código de barras, en donde cada oligonucleótido con código de barras es tal como se define en el presente documento. Preferentemente, el reactivo de código de barras multimérico comprende al menos 5 moléculas de código de barras unidas entre sí, en donde cada molécula de código de barras es tal como se define en el presente documento; y un oligonucleótido con código de barras hibridado con cada molécula de código de barras, en donde cada oligonucleótido con código de barras es tal como se define en el presente documento.

El reactivo de códigos de barras multimérico puede comprender: al menos 2, al menos 3, al menos 4, al menos 5, al menos 10, al menos 20, al menos 25, al menos 50, al menos 75, al menos 100, al menos 200, al menos 500, al menos 1000, al menos 5000, al menos 10^4 , al menos 10^5 , o al menos 10^6 moléculas de códigos de barras únicas o diferentes unidas entre sí, en donde cada molécula de código de barras es tal como se define en el presente documento; y un oligonucleótido con código de barras hibridado con cada molécula de código de barras, en donde cada oligonucleótido con código de barras es tal como se define en el presente documento. Preferentemente, el reactivo de código de barras multimérico comprende al menos 5 moléculas de código de barras únicas o diferentes unidas entre sí, en donde cada molécula de código de barras es tal como se define en el presente documento; y un oligonucleótido con código de barras hibridado con cada molécula de código de barras, en donde cada oligonucleótido con código de barras es tal como se define en el presente documento.

Un reactivo de código de barras multimérico puede comprender dos o más oligonucleótidos con código de barras como se define en el presente documento, en donde los oligonucleótidos con código de barras comprenden, cada uno, una región de código de barras. Un reactivo de códigos de barras multimérico puede comprender: al menos 2, al menos 3, al menos 4, al menos 5, al menos 10, al menos 20, al menos 25, al menos 50, al menos 75, al menos 100, al menos 200, al menos 500, al menos 1000, al menos 5000, al menos 10.000, al menos 100.000, o al menos 1.000.000 de oligonucleótidos con códigos de barras únicos o diferentes. Preferentemente, el reactivo de código de barras multimérico comprende al menos 5 oligonucleótidos con código de barras únicos o diferentes.

Los oligonucleótidos con código de barras de un reactivo de código de barras multimérico están unidos (directa o indirectamente). Los oligonucleótidos con código de barras de un reactivo de código de barras multimérico están unidos por un soporte, por ejemplo, una macromolécula, soporte sólido o soporte semisólido, como se describe en el presente documento. El reactivo de códigos de barras multimérico puede comprender uno o más polímeros con los que hibridan o se unen los oligonucleótidos con códigos de barras. Por ejemplo, los oligonucleótidos con código de barras de un reactivo de código de barras multimérico pueden hibridar con una molécula de hibridación multimérica, por ejemplo, una molécula de código de barras multimérica. Como alternativa, los oligonucleótidos con código de barras de un reactivo de código de barras multimérico pueden estar unidos por una macromolécula (como un polímero sintético, por ejemplo, un dendrímero, o un biopolímero, por ejemplo, una proteína) o un soporte (como un soporte sólido o un soporte semisólido, por ejemplo, una perla de gel). Adicionalmente o como alternativa, los oligonucleótidos con código de barras de un reactivo de código de barras (único) multimérico pueden unirse entre sí estando comprendidos dentro de un vehículo lipídico (único) (por ejemplo, un liposoma o una micela).

Un reactivo de códigos de barras multimérico puede comprender: primera y segunda moléculas de hibridación unidas entre sí (es decir, una molécula de hibridación multimérica), en donde cada una de las moléculas de hibridación comprende una secuencia de ácido nucleico que comprende una región de hibridación; y primer y segundo oligonucleótidos con código de barras, en donde el primer oligonucleótido con código de barras hibrida con la región de hibridación de la primera molécula de hibridación y en donde el segundo oligonucleótido con código de barras hibrida con la región de hibridación de la segunda molécula de hibridación.

Las moléculas de hibridación comprenden o consisten en desoxirribonucleótidos. Uno o más de los desoxirribonucleótidos puede ser un desoxirribonucleótido modificado (por ejemplo, un desoxirribonucleótido modificado con un resto de biotina o un nucleótido de desoxiuracilo). Las moléculas de hibridación pueden comprender uno o más nucleótidos o secuencias degenerados. Las moléculas de hibridación pueden no comprender nucleótidos o secuencias degenerados.

Las moléculas de hibridación de una molécula de hibridación multimérica se pueden unir a una molécula de ácido nucleico. Una molécula de ácido nucleico de este tipo puede proporcionar la estructura principal con la que pueden hibridar oligonucleótidos con código de barras monocatenario. Las moléculas de hibridación de una molécula de hibridación multimérica pueden estar comprendidas dentro de una (única) molécula de ácido nucleico. Una molécula de hibridación multimérica puede comprender una sola secuencia de ácido nucleico contigua que comprende dos o más moléculas de hibridación. Una molécula de hibridación multimérica puede ser una molécula de ácido nucleico monocatenario (por ejemplo, ADN monocatenario) que comprende dos o más moléculas de hibridación. Una molécula de hibridación multimérica puede comprender una o más regiones bicatenarias. Opcionalmente, en una región bicatenaria o entre dos regiones bicatenarias diferentes, una molécula de hibridación multimérica puede comprender uno o más cortes, o uno o más huecos, en donde la propia molécula de hibridación multimérica se ha dividido o separado. Cualquiera de dichos huecos puede ser al menos uno, al menos 2, al menos 5, al menos 10, al menos 20, al menos 50, o al menos 100 nucleótidos de longitud. Dichos cortes y/o espacios pueden servir para aumentar la flexibilidad molecular de la molécula de hibridación multimérica y/o el reactivo de código de barras multimérico, por ejemplo, para aumentar la accesibilidad de la molécula o reactivo para interactuar con moléculas de ácido nucleico diana. Dichos cortes y/o espacios también pueden permitir una purificación o eliminación más eficaz de dichas moléculas o reactivos. Una molécula y/o reactivo que comprende dicho(s) corte(s) y/o hueco(s) puede conservar enlaces entre diferentes moléculas de hibridación al tener una hebra de ADN complementaria que hibrida conjuntamente con regiones de dos o más partes divididas de una molécula de hibridación multimérica.

Las moléculas de hibridación pueden estar unidas por una macromolécula uniéndose a la macromolécula y/o hibridando con la macromolécula.

Las moléculas de hibridación se pueden unir a la macromolécula directa o indirectamente (por ejemplo, a través de una molécula enlazadora). Las moléculas de hibridación se pueden unir uniéndose a la macromolécula y/o uniéndose o hibridando a moléculas enlazadoras que están unidas a la macromolécula. Las moléculas de hibridación pueden estar unidas a la macromolécula (o a las moléculas enlazadoras) por enlace covalente, enlace no covalente (por ejemplo, una interacción proteína-proteína o un enlace estreptavidina-biotina) o hibridación de ácido nucleico. La molécula enlazadora puede ser un biopolímero (por ejemplo, una molécula de ácido nucleico) o un polímero sintético. La molécula enlazadora puede comprender una o más unidades de etilenglicol y/o polietilenglicol (por ejemplo, hexaetilenglicol o pentaetilenglicol). La molécula enlazadora puede comprender uno o más grupos etilo, tales como un espaciador C3 (tres carbonos), espaciador C6, espaciador C12 o espaciador C18.

La macromolécula puede ser un polímero sintético (por ejemplo, un dendrímero) o un biopolímero como un ácido nucleico (por ejemplo, un ácido nucleico monocatenario como ADN monocatenario), un péptido, un polipéptido o una proteína (por ejemplo, una proteína multimérica).

El dendrímero puede comprender al menos 2, al menos 3, al menos 5, o al menos 10 generaciones.

La macromolécula puede ser un ácido nucleico que comprende dos o más nucleótidos, cada uno de los cuales puede unirse a una molécula de hibridación. Adicionalmente o como alternativa, el ácido nucleico puede comprender dos o más regiones, cada una de las cuales puede hibridar con una molécula de hibridación.

El ácido nucleico puede comprender un primer nucleótido modificado y un segundo nucleótido modificado, en donde cada nucleótido modificado comprende un resto de unión (por ejemplo, un resto de biotina o un resto de alquino que se puede usar para una reacción química de clic) capaz de unirse a una molécula de hibridación. Opcionalmente, el primer y segundo nucleótidos modificados pueden estar separados por una secuencia de ácidos nucleicos intermedia de al menos uno, al menos dos, al menos 5 o al menos 10 nucleótidos.

El ácido nucleico puede comprender una primera región de hibridación y una segunda región de hibridación, en donde cada región de hibridación comprende una secuencia complementaria y capaz de hibridar con una secuencia de al menos un nucleótido dentro de una molécula de hibridación. La secuencia complementaria puede ser de al menos 5, al menos 10, al menos 15, al menos 20, al menos 25 o al menos 50 nucleótidos contiguos. Opcionalmente, la primera y segunda regiones de hibridación pueden estar separadas por una secuencia de ácido nucleico intermedia de al menos una, al menos dos, al menos 5 o al menos 10 nucleótidos.

La macromolécula puede ser una proteína tal como una proteína multimérica, por ejemplo, una proteína homomérica o una proteína heteromérica. Por ejemplo, la proteína puede comprender estreptavidina, por ejemplo, estreptavidina tetramérica.

Las moléculas de hibridación se pueden unir mediante un soporte. Las moléculas de hibridación se pueden unir al soporte directa o indirectamente (por ejemplo, a través de una molécula enlazadora). Las moléculas de hibridación se pueden unir uniéndose al soporte y/o uniéndose o hibridando a moléculas enlazadoras que están unidas al soporte. Las moléculas de hibridación pueden unirse al soporte (o a las moléculas enlazadoras) mediante enlace covalente, enlace no covalente (por ejemplo, una interacción proteína-proteína o un enlace estreptavidina-biotina) o hibridación de ácido nucleico. La molécula enlazadora puede ser un biopolímero (por ejemplo, una molécula de ácido nucleico) o un polímero sintético. La molécula enlazadora puede comprender una o más unidades de etilenglicol y/o polietilenglicol (por ejemplo, hexaetilenglicol o pentaetilenglicol). La molécula enlazadora puede comprender uno o más grupos etilo, tales como un espaciador C3 (tres carbonos), espaciador C6, espaciador C12 o espaciador C18.

El soporte puede ser un soporte sólido o un soporte semisólido. El soporte puede comprender una superficie plana. El soporte puede ser un portaobjetos, por ejemplo, un portaobjetos de vidrio. El portaobjetos puede ser una celda de flujo para secuenciar. Si el soporte es un portaobjetos, la primera y la segunda moléculas de hibridación pueden inmovilizarse en una región discreta del portaobjetos. Opcionalmente, las moléculas de hibridación de cada reactivo de código de barras multimérico en una biblioteca se inmovilizan en una región discreta diferente en el portaobjetos a las moléculas de hibridación de los otros reactivos de código de barras multimérico en la biblioteca. El soporte puede ser una placa que comprende pocillos, en donde opcionalmente la primera y la segunda moléculas de hibridación están inmovilizadas en el mismo pocillo. Opcionalmente, las moléculas de hibridación de cada reactivo de código de barras multimérico en la biblioteca se inmovilizan en un pocillo diferente de la placa a las moléculas de hibridación de los otros reactivos de código de barras multimérico en la biblioteca.

Preferentemente, el soporte es una perla (por ejemplo, una perla de gel). La perla puede ser una perla de agarosa, una perla de sílice, una perla de espuma de poliestireno, una perla de gel (como las disponibles en 10x Genomics®), una perla conjugada de anticuerpo, una perla conjugada oligo-dT, una perla de estreptavidina o una perla magnética (por ejemplo, una perla superparamagnética). La perla puede ser de cualquier tamaño y/o estructura molecular. Por ejemplo, la perla puede tener un diámetro de 10 nanómetros a 100 micrómetros, de 100 nanómetros a 10 micrómetros de diámetro, o de 1 micrómetro a 5 micrómetros de diámetro. Opcionalmente, la perla tiene aproximadamente 10 nanómetros de diámetro, aproximadamente 100 nanómetros de diámetro, aproximadamente 1 micrómetro de diámetro, aproximadamente 10 micrómetros de diámetro o aproximadamente 100 micrómetros de diámetro. La perla puede ser sólida o, como alternativa, la perla puede ser hueca o parcialmente hueca o porosa. Las perlas de ciertos tamaños pueden ser más preferibles para ciertos métodos de códigos de barras. Por ejemplo, perlas de menos de 5,0 micrómetros, o menos de 1,0 micrómetro, pueden ser más útiles para poner códigos de barras a dianas de ácidos nucleicos dentro de células individuales. Preferentemente, las moléculas de hibridación de cada reactivo de códigos de barras multimérico en una biblioteca están unidas en una perla diferente a las moléculas de hibridación de los otros reactivos de códigos de barras multiméricos en la biblioteca.

El soporte puede funcionalizarse para permitir la unión de dos o más moléculas de hibridación. Esta funcionalización se puede habilitar mediante la adición de restos químicos (por ejemplo, grupos carboxilados, alquinos, azidas, grupos de acrilato, grupos amino, grupos sulfato o grupos succinimida) y/o restos basados en proteínas (por ejemplo, estreptavidina, avidina o proteína G) al soporte. Las moléculas de hibridación se pueden unir a los restos directa o indirectamente (por ejemplo, mediante una molécula de enlazador).

Los soportes funcionalizados (por ejemplo, perlas) se pueden poner en contacto con una solución de moléculas de hibridación en condiciones que promuevan la unión de dos o más moléculas de hibridación a cada perla en la solución (generando reactivos de códigos de barras multiméricos).

En una biblioteca de reactivos de códigos de barras multiméricos, las moléculas de hibridación de cada reactivo de código de barras multimérico en una biblioteca se pueden unir entre sí en un soporte diferente a las moléculas de hibridación de los otros reactivos de código de barras multimérico en la biblioteca.

Opcionalmente, las moléculas de hibridación se unen a las perlas mediante enlace covalente, enlace no covalente (por ejemplo, un enlace estreptavidina-biotina) o hibridación de ácido nucleico.

El reactivo de códigos de barras multimérico puede comprender: al menos 2, al menos 3, al menos 4, al menos 5, al menos 10, al menos 20, al menos 25, al menos 50, al menos 75, al menos 100, al menos 200, al menos 500, al menos 1000, al menos 5000 o al menos 10000 moléculas de hibridación unidas entre sí, en donde cada molécula de hibridación es tal como se define en el presente documento; y un oligonucleótido con código de barras hibridado con cada molécula de hibridación, en donde cada oligonucleótido con código de barras es tal como se define en el presente documento. Preferentemente, el reactivo de código de barras multimérico comprende al menos 5 moléculas de hibridación unidas entre sí, en donde cada molécula de hibridación es tal como se define en el presente documento; y un oligonucleótido con código de barras hibridado con cada molécula de hibridación, en donde cada oligonucleótido con código de barras es tal como se define en el presente documento.

El reactivo de códigos de barras multimérico puede comprender: al menos 2, al menos 3, al menos 4, al menos 5, al menos 10, al menos 20, al menos 25, al menos 50, al menos 75, al menos 100, al menos 200, al menos 500, al menos 1000, al menos 5000, o al menos 10000 moléculas de hibridación únicas o diferentes unidas entre sí, en donde cada molécula de hibridación es tal como se define en el presente documento; y un oligonucleótido con código de barras hibridado con cada molécula de hibridación, en donde cada oligonucleótido con código de barras es tal como se define en el presente documento. Preferentemente, el reactivo de código de barras multimérico comprende al menos 5 moléculas de hibridación únicas o diferentes unidas entre sí, en donde cada molécula de hibridación es tal como se define en el presente documento; y un oligonucleótido con código de barras hibridado con cada molécula de hibridación, en donde cada oligonucleótido con código de barras es tal como se define en el presente documento.

La molécula de hibridación multimérica puede ser una molécula de código de barras multimérica, en donde la primera molécula de hibridación es una primera molécula de código de barras y la segunda molécula de hibridación es una segunda molécula de código de barras. Un reactivo de códigos de barras multimérico puede comprender: la primera y segunda moléculas de código de barras unidas (es decir, una molécula de código de barras multimérica), en donde cada una de las moléculas de código de barras comprende una secuencia de ácido nucleico que comprende una región de código de barras; y primer y segundo oligonucleótidos con código de barras, en donde el primer oligonucleótido con código de barras hibrida con la región de código de barras de la primera molécula de código de barras y en donde el segundo oligonucleótido con código de barras hibrida con la región de código de barras de la segunda molécula de código de barras.

Los oligonucleótidos con código de barras de un reactivo de código de barras multimérico pueden comprender: un primer oligonucleótido con código de barras que comprende, opcionalmente en la dirección de 5' a 3', una región de código de barras y una región diana capaz de hibridar o ligar a una primera subsecuencia del ácido nucleico diana; y un segundo oligonucleótido con código de barras que comprende, opcionalmente en la dirección de 5' a 3', una región de código de barras y una región diana capaz de hibridar o ligar a una segunda subsecuencia del ácido nucleico diana.

Los oligonucleótidos con código de barras de un reactivo de código de barras multimérico pueden comprender: un primer oligonucleótido con código de barras que comprende una región de código de barras y una región diana capaz de ligarse a una primera subsecuencia del ácido nucleico diana; y un segundo oligonucleótido con código de barras que comprende una región de código de barras y una región diana capaz de ligarse a una segunda subsecuencia del ácido nucleico diana.

Los oligonucleótidos con código de barras de un reactivo de código de barras multimérico pueden comprender: un primer oligonucleótido con código de barras que comprende, en la dirección 5' a 3', una región de código de barras y una región diana capaz de hibridar con una primera subsecuencia del ácido nucleico diana; y un segundo oligonucleótido con código de barras que comprende, en la dirección 5' a 3', una región de código de barras y una región diana capaz de hibridar con una segunda subsecuencia del ácido nucleico diana.

2. PROPIEDADES GENERALES DE LOS OLIGONUCLEÓTIDOS CON CÓDIGO DE BARRAS

Un oligonucleótido con código de barras comprende una región de código de barras. Los oligonucleótidos con código de barras pueden comprender, opcionalmente en la dirección de 5' a 3', una región de código de barras y una región diana. La región diana es capaz de hibridar o ligar a una subsecuencia del ácido nucleico diana. Como alternativa, un oligonucleótido con código de barras puede consistir esencialmente en o consistir en una región de código de barras.

El extremo 5' de un oligonucleótido con código de barras se puede fosforilar. Esto puede permitir que el extremo 5' del oligonucleótido con código de barras se ligue al extremo 3' de un ácido nucleico diana. Como alternativa, el extremo 5' de un oligonucleótido con código de barras puede no estar fosforilado.

Un oligonucleótido con código de barras puede ser una molécula de ácido nucleico monocatenario (por ejemplo, ADN monocatenario). Un oligonucleótido con código de barras puede comprender una o más regiones bicatenarias. Un oligonucleótido con código de barras puede ser una molécula de ácido nucleico bicatenario (por ejemplo, ADN bicatenario).

Los oligonucleótidos con código de barras pueden comprender o consistir en desoxirribonucleótidos. Uno o más de

los desoxirribonucleótidos puede ser un desoxirribonucleótido modificado (por ejemplo, un desoxirribonucleótido modificado con un resto de biotina o un nucleótido de desoxiuracilo). Los oligonucleótidos con código de barras pueden comprender uno o más nucleótidos o secuencias degenerados. Los oligonucleótidos con código de barras pueden no comprender nucleótidos o secuencias degenerados.

- 5 Las regiones de código de barras de cada oligonucleótido con código de barras pueden comprender secuencias diferentes. Cada región de código de barras puede comprender una secuencia que identifica el reactivo de código de barras multimérico. Por ejemplo, esta secuencia puede ser una región constante compartida por todas las regiones de códigos de barras de un único reactivo de códigos de barras multimérico. La región de código de barras de cada oligonucleótido con código de barras puede contener una secuencia única que no está presente en otros oligonucleótidos con código de barras y, por tanto, puede servir para identificar de forma única cada oligonucleótido con código de barras. Cada región de código de barras puede comprender al menos 5, al menos 10, al menos 15, al menos 20, al menos 25, al menos 50 o al menos 100 nucleótidos. Preferentemente, cada región de código de barras comprende al menos 5 nucleótidos. Preferentemente, cada región de código de barras comprende desoxirribonucleótidos, opcionalmente, todos los nucleótidos en una región de código de barras son desoxirribonucleótidos. Uno o más de los desoxirribonucleótidos puede ser un desoxirribonucleótido modificado (por ejemplo, un desoxirribonucleótido modificado con un resto de biotina o un nucleótido de desoxiuracilo). Las regiones de código de barras pueden comprender uno o más nucleótidos o secuencias degenerados. Las regiones de código de barras pueden no comprender nucleótidos o secuencias degenerados.

- 20 Las regiones diana de cada oligonucleótido con código de barras pueden comprender secuencias diferentes. Cada región diana puede comprender una secuencia capaz de hibridar solo con una única subsecuencia de un ácido nucleico diana dentro de una muestra de ácidos nucleicos (es decir, una secuencia específica diana). Cada región diana puede comprender uno o más secuencias aleatorias o una o más degeneradas, para permitir que la región diana hibride con más de una subsecuencia de un ácido nucleico diana. Cada región diana puede comprender al menos 5, al menos 10, al menos 15, al menos 20, al menos 25, al menos 50 o al menos 100 nucleótidos. Preferentemente, cada región diana comprende al menos 5 nucleótidos. Cada región diana puede comprender de 5 a 100 nucleótidos, de 5 a 10 nucleótidos, de 10 a 20 nucleótidos, de 20 a 30 nucleótidos, de 30 a 50 nucleótidos, de 50 a 100 nucleótidos, de 10 a 90 nucleótidos, de 20 a 80 nucleótidos, de 30 a 70 nucleótidos o de 50 a 60 nucleótidos. Preferentemente, cada región diana comprende de 30 a 70 nucleótidos. Preferentemente, cada región diana comprende desoxirribonucleótidos, opcionalmente, todos los nucleótidos en una región diana son desoxirribonucleótidos. Uno o más de los desoxirribonucleótidos puede ser un desoxirribonucleótido modificado (por ejemplo, un desoxirribonucleótido modificado con un resto de biotina o un nucleótido de desoxiuracilo). Cada región diana puede comprender una o más bases universales (por ejemplo, inosina), uno o más nucleótidos modificados y/o uno o más análogos de nucleótidos.

- 35 Las regiones diana se pueden usar para hibridar los oligonucleótidos con código de barras a subsecuencias de ácidos nucleicos diana, y luego se pueden usar como cebadores para una reacción de extensión de cebador o una reacción de amplificación, por ejemplo, una reacción en cadena de la polimerasa. Como alternativa, las regiones diana se pueden usar para ligar los oligonucleótidos con código de barras a subsecuencias de ácidos nucleicos diana. La región diana puede estar en el extremo 5' de un oligonucleótido con código de barras. Tal región diana se puede fosforilar. Esto puede permitir que el extremo 5' de la región diana se ligue al extremo 3' de una subsecuencia de un ácido nucleico diana.

- Los oligonucleótidos con código de barras pueden comprender además una o más regiones adaptadoras. Una región adaptadora puede estar entre la región del código de barras y la región diana. Un oligonucleótido con código de barras puede, por ejemplo, comprender una región de adaptador 5' de una región de código de barras (una región de adaptador de 5') y/o una región de adaptador 3' de la región de código de barras (una región de adaptador de 3'). Opcionalmente, los oligonucleótidos con código de barras comprenden, en la dirección 5' a 3', una región de código de barras, una región adaptadora y una región diana.

- 50 La región o regiones adaptadoras de los oligonucleótidos con código de barras pueden comprender una secuencia complementaria a una región adaptadora de una molécula de código de barras multimérica o una secuencia complementaria a una región de hibridación de una molécula de hibridación multimérica. La región o regiones adaptadoras de los oligonucleótidos con código de barras pueden permitir que los oligonucleótidos con código de barras se unan a una macromolécula o soporte (por ejemplo, una perla). La(s) región(es) del adaptador se pueden utilizar para manipular, purificar, recuperar, amplificar o detectar oligonucleótidos con código de barras y/o ácidos nucleicos diana con los que se pueden hibridar o ligar.

- 55 La región adaptadora de cada oligonucleótido con código de barras puede comprender una región constante. Opcionalmente, todas las regiones adaptadoras de oligonucleótidos con código de barras de cada reactivo de código de barras multimérico son sustancialmente idénticas. La región adaptadora puede comprender al menos 1, al menos 2, al menos 3, al menos 4, al menos 5, al menos 6, al menos 8, al menos 10, al menos 15, al menos 20, al menos 25, al menos 50, al menos 100 o al menos 250 nucleótidos. Preferentemente, la región adaptadora comprende al menos 4 nucleótidos. Preferentemente, cada región adaptadora comprende desoxirribonucleótidos, opcionalmente, todos los nucleótidos en una región adaptadora son desoxirribonucleótidos. Uno o más de los desoxirribonucleótidos puede ser un desoxirribonucleótido modificado (por ejemplo, un desoxirribonucleótido modificado con un resto de biotina o un

nucleótido de desoxiuracilo). Cada región adaptadora puede comprender una o más bases universales (por ejemplo, inosina), uno o más nucleótidos modificados y/o uno o más análogos de nucleótidos.

Los oligonucleótidos con código de barras se pueden sintetizar mediante un proceso de síntesis química de oligonucleótidos. El proceso de síntesis de oligonucleótidos con código de barras puede incluir una o más etapas de un proceso de producción enzimática, un proceso de amplificación enzimática, o un procedimiento de modificación enzimática, tal como un proceso de transcripción *in vitro*, un proceso de transcripción inversa, un proceso de extensión del cebador o un proceso de reacción en cadena de la polimerasa.

Estas propiedades generales de los oligonucleótidos de códigos de barras son aplicables a cualquiera de los reactivos de códigos de barras multiméricos descritos en el presente documento.

3. PROPIEDADES GENERALES DE LAS BIBLIOTECAS DE REACTIVOS DE CÓDIGO DE BARRAS MULTIMÉRICOS

La divulgación proporciona una biblioteca de reactivos de códigos de barras multiméricos que comprenden reactivos de códigos de barras multiméricos primero y segundo tal como se define en el presente documento, en donde las regiones de código de barras del primer reactivo de código de barras multimérico son diferentes a las regiones de código de barras del segundo reactivo de código de barras multimérico.

La biblioteca de reactivos de códigos de barras multiméricos puede comprender al menos 5, al menos 10, al menos 20, al menos 25, al menos 50, al menos 75, al menos 100, al menos 250, al menos 500, al menos 10^3 , al menos 10^4 , al menos 10^5 , al menos 10^6 , al menos 10^7 , al menos 10^8 o al menos 10^9 reactivos de códigos de barras multiméricos tal como se define en el presente documento. Preferentemente, la biblioteca comprende al menos 10 reactivos de códigos de barras multiméricos tal como se define en el presente documento. Preferentemente, las regiones de código de barras primera y segunda de cada reactivo de código de barras multimérico son diferentes a las regiones de código de barras de al menos otros 9 reactivos de código de barras multiméricos en la biblioteca.

Las regiones de código de barras primera y segunda de cada reactivo de código de barras multimérico pueden ser diferentes a las regiones de código de barras de al menos 4, al menos 9, al menos 19, al menos 24, al menos 49, al menos 74, al menos 99, al menos 249, al menos 499, al menos 999 (es decir, 10^3-1), al menos 10^4-1 , al menos 10^5-1 , al menos 10^6-1 , al menos 10^7-1 , al menos 10^8-1 o al menos 10^9-1 otros reactivos de códigos de barras multiméricos en la biblioteca. Las regiones de código de barras primera y segunda de cada reactivo de código de barras multimérico pueden ser diferentes a las regiones de código de barras de todos los demás reactivos de código de barras multiméricos de la biblioteca. Preferentemente, las regiones de código de barras primera y segunda de cada reactivo de código de barras multimérico son diferentes a las regiones de código de barras de al menos otros 9 reactivos de código de barras multiméricos en la biblioteca.

Las regiones de código de barras de cada reactivo de código de barras multimérico pueden ser diferentes a las regiones de código de barras de al menos 4, al menos 9, al menos 19, al menos 24, al menos 49, al menos 74, al menos 99, al menos 249, al menos 499, al menos 999 (es decir, 10^3-1), al menos 10^4-1 , al menos 10^5-1 , al menos 10^6-1 , al menos 10^7-1 , al menos 10^8-1 o al menos 10^9-1 otros reactivos de códigos de barras multiméricos en la biblioteca. Las regiones de código de barras 11919231;OMB;OMB de cada reactivo de código de barras multimérico pueden ser diferentes a las regiones de código de barras de todos los demás reactivos de código de barras multiméricos de la biblioteca. Preferentemente, las regiones de código de barras de cada reactivo de código de barras multimérico son diferentes a las regiones de código de barras de al menos otros 9 reactivos de código de barras multiméricos en la biblioteca.

La divulgación proporciona una biblioteca de reactivos de códigos de barras multiméricos que comprenden reactivos de códigos de barras multiméricos primero y segundo tal como se define en el presente documento, en donde las regiones de código de barras de los oligonucleótidos con código de barras del primer reactivo de código de barras multimérico son diferentes a las regiones de código de barras de los oligonucleótidos con código de barras del segundo reactivo de código de barras multimérico.

Diferentes reactivos de códigos de barras multiméricos dentro de una biblioteca de reactivos de códigos de barras multiméricos pueden comprender diferentes números de oligonucleótidos con códigos de barras.

La biblioteca de reactivos de códigos de barras multiméricos puede comprender al menos 5, al menos 10, al menos 20, al menos 25, al menos 50, al menos 75, al menos 100, al menos 250, al menos 500, al menos 10^3 , al menos 10^4 , al menos 10^5 , al menos 10^6 , al menos 10^7 , al menos 10^8 o al menos 10^9 reactivos de códigos de barras multiméricos tal como se define en el presente documento. Preferentemente, la biblioteca comprende al menos 10 reactivos de códigos de barras multiméricos tal como se define en el presente documento. Preferentemente, las regiones de código de barras de los oligonucleótidos con código de barras primero y segundo de cada reactivo de código de barras multimérico son diferentes a las regiones de código de barras de los oligonucleótidos con código de barras de al menos otros 9 reactivos de código de barras multiméricos en la biblioteca.

Las regiones de código de barras de los oligonucleótidos con código de barras primero y segundo de cada reactivo de código de barras multimérico pueden ser diferentes a las regiones de código de barras de los oligonucleótidos con

- código de barras de al menos 4, al menos 9, al menos 19, al menos 24, al menos 49, al menos 74, al menos 99, al menos 249, al menos 499, al menos 999 (es decir, 10^3-1), al menos 10^4-1 , al menos 10^5-1 , al menos 10^6-1 , al menos 10^7-1 , al menos 10^8-1 o al menos 10^9-1 otros reactivos de códigos de barras multiméricos en la biblioteca. Las regiones de código de barras de los oligonucleótidos con código de barras primero y segundo de cada reactivo de código de barras multimérico pueden ser diferentes a las regiones de código de barras de los oligonucleótidos con código de barras de todos los otros reactivos de código de barras multiméricos en la biblioteca. Preferentemente, las regiones de código de barras de los oligonucleótidos con código de barras primero y segundo de cada reactivo de código de barras multimérico son diferentes a las regiones de código de barras de los oligonucleótidos con código de barras de al menos otros 9 reactivos de código de barras multiméricos en la biblioteca.
- 10 Las regiones de código de barras de los oligonucleótidos con código de barras de cada reactivo de código de barras multimérico pueden ser diferentes a las regiones de código de barras de los oligonucleótidos con código de barras de al menos 4, al menos 9, al menos 19, al menos 24, al menos 49, al menos 74, al menos 99, al menos 249, al menos 499, al menos 999 (es decir, 10^3-1), al menos 10^4-1 , al menos 10^5-1 , al menos 10^6-1 , al menos 10^7-1 , al menos 10^8-1 o al menos 10^9-1 otros reactivos de códigos de barras multiméricos en la biblioteca. Las regiones de código de barras de los oligonucleótidos con código de barras de cada reactivo de código de barras multimérico pueden ser diferentes a las regiones de código de barras de los oligonucleótidos con código de barras de todos los otros reactivos de código de barras multiméricos en la biblioteca. Preferentemente, las regiones de código de barras de los oligonucleótidos con código de barras de cada reactivo de código de barras multimérico son diferentes a las regiones de código de barras de los oligonucleótidos con código de barras de al menos otros 9 reactivos de código de barras multiméricos en la biblioteca.

Estas propiedades generales de las bibliotecas de reactivos de códigos de barras multiméricos son aplicables a cualquiera de los reactivos de códigos de barras multiméricos descritos en el presente documento.

4. REACTIVOS DE CÓDIGO DE BARRAS MULTIMÉRICOS QUE COMPRENDEN OLIGONUCLEÓTIDOS CON CÓDIGO DE BARRAS HIBRIDADOS CON UNA MOLÉCULA DE CÓDIGO DE BARRAS MULTIMÉRICA

- 25 La divulgación proporciona un reactivo de código de barras multimérico para etiquetar un ácido nucleico diana, en donde el reactivo comprende: la primera y segunda moléculas de código de barras unidas (es decir, una molécula de código de barras multimérica), en donde cada una de las moléculas de código de barras comprende una secuencia de ácido nucleico que comprende una región de código de barras; y primer y segundo oligonucleótidos con código de barras, en donde el primer oligonucleótido con código de barras comprende, opcionalmente en la dirección de 5' a 3', una región de código de barras hibridada con la región de código de barras de la primera molécula de código de barras y una región diana capaz de hibridar o ligar a una primera subsecuencia del ácido nucleico diana, y en donde el segundo oligonucleótido con código de barras comprende, opcionalmente en la dirección de 5' a 3', una región de código de barras hibridada con la región de código de barras de la segunda molécula de código de barras y una región diana capaz de hibridar o ligar a una segunda subsecuencia del ácido nucleico diana.
- 30
- 35 La divulgación proporciona un reactivo de código de barras multimérico para etiquetar un ácido nucleico diana, en donde el reactivo comprende: la primera y segunda moléculas de código de barras unidas (es decir, una molécula de código de barras multimérica), en donde cada una de las moléculas de código de barras comprende una secuencia de ácido nucleico que comprende una región de código de barras; y primer y segundo oligonucleótidos con código de barras, en donde el primer oligonucleótido con código de barras comprende una región de código de barras hibridada con la región de código de barras de la primera molécula de código de barras y una región diana capaz de ligarse a una primera subsecuencia del ácido nucleico diana, y en donde el segundo oligonucleótido con código de barras comprende una región de código de barras hibridada con la región de código de barras de la segunda molécula de código de barras y una región diana capaz de ligarse a una segunda subsecuencia del ácido nucleico diana.
- 40
- 45 La divulgación proporciona un reactivo de código de barras multimérico para etiquetar un ácido nucleico diana, en donde el reactivo comprende: la primera y segunda moléculas de código de barras unidas (es decir, una molécula de código de barras multimérica), en donde cada una de las moléculas de código de barras comprende una secuencia de ácido nucleico que comprende una región de código de barras; y primer y segundo oligonucleótidos con código de barras, en donde el primer oligonucleótido con código de barras comprende, en dirección 5' a 3', una región de código de barras hibridada con la región de código de barras de la primera molécula de código de barras y una región diana capaz de hibridar con una primera subsecuencia del ácido nucleico diana, y en donde el segundo oligonucleótido con código de barras comprende, en dirección 5' a 3', una región de código de barras hibridada con la región de código de barras de la segunda molécula de código de barras y una región diana capaz de hibridar con una segunda subsecuencia del ácido nucleico diana.
- 50
- 55 La divulgación proporciona un reactivo de código de barras multimérico para etiquetar un ácido nucleico diana, en donde el reactivo comprende: la primera y segunda moléculas de código de barras unidas (es decir, una molécula de código de barras multimérica), en donde cada una de las moléculas de código de barras comprende una secuencia de ácido nucleico que comprende una región de código de barras; y primer y segundo oligonucleótidos con código de barras, en donde el primer oligonucleótido con código de barras comprende una región de código de barras hibridada con la región de código de barras de la primera molécula de código de barras y capaz de ligarse a una primera subsecuencia del ácido nucleico diana, y en donde el segundo oligonucleótido con código de barras comprende una
- 60

región de código de barras hibridada con la región de código de barras de la segunda molécula de código de barras y capaz de ligarse a una segunda subsecuencia del ácido nucleico diana.

Cada oligonucleótido con código de barras puede consistir esencialmente en o consistir en una región de código de barras.

- 5 Preferentemente, las moléculas de código de barras comprenden o consisten en desoxirribonucleótidos. Uno o más de los desoxirribonucleótidos puede ser un desoxirribonucleótido modificado (por ejemplo, un desoxirribonucleótido modificado con un resto de biotina o un nucleótido de desoxiuracilo). Las moléculas de código de barras pueden comprender uno o más nucleótidos o secuencias degenerados. Las moléculas de código de barras pueden no comprender nucleótidos o secuencias degenerados.
- 10 Las regiones del código de barras pueden identificar de forma única cada una de las moléculas de código de barras. Cada región de código de barras puede comprender una secuencia que identifica el reactivo de código de barras multimérico. Por ejemplo, esta secuencia puede ser una región constante compartida por todas las regiones de códigos de barras de un único reactivo de códigos de barras multimérico. Cada región de código de barras puede comprender al menos 5, al menos 10, al menos 15, al menos 20, al menos 25, al menos 50 o al menos 100 nucleótidos.
- 15 Preferentemente, cada región de código de barras comprende al menos 5 nucleótidos. Preferentemente, cada región de código de barras comprende desoxirribonucleótidos, opcionalmente, todos los nucleótidos en una región de código de barras son desoxirribonucleótidos. Uno o más de los desoxirribonucleótidos puede ser un desoxirribonucleótido modificado (por ejemplo, un desoxirribonucleótido modificado con un resto de biotina o un nucleótido de desoxiuracilo). Las regiones de código de barras pueden comprender uno o más nucleótidos o secuencias degenerados. Las regiones de código de barras pueden no comprender nucleótidos o secuencias degenerados.
- 20

Preferentemente, la región de código de barras del primer oligonucleótido con código de barras comprende una secuencia que es complementaria e hibridada con la región de código de barras de la primera molécula de código de barras y la región de código de barras del segundo oligonucleótido con código de barras comprende una secuencia que es complementaria e hibridada con la región de código de barras de la segunda molécula de código de barras. La secuencia complementaria de cada oligonucleótido con código de barras puede ser al menos 5, al menos 10, al menos 15, al menos 20, al menos 25, al menos 50 o al menos 100 nucleótidos contiguos.

- 25 Las regiones diana de los oligonucleótidos con código de barras (que no hibridan con la(s) molécula(s) de código de barras multimérica(s)) pueden no ser complementarias a la(s) molécula(s) de código de barras multiméricas.

- 30 Los oligonucleótidos con código de barras pueden comprender una región enlazadora entre la región del código de barras y la región diana. La región enlazadora puede comprender uno o más nucleótidos contiguos que no están hibridados con la molécula de código de barras multimérica y no son complementarios a las subsecuencias del ácido nucleico diana. El enlazador puede comprender de 1 a 100, de 5 a 75, de 10 a 50, de 15 a 30 o de 20 a 25 nucleótidos no complementarios. Preferentemente, el enlazador comprende de 15 a 30 nucleótidos no complementarios. El uso de una región enlazadora de este tipo mejora la eficacia de las reacciones de códigos de barras realizadas utilizando los reactivos de códigos de barras multiméricos.
- 35

- 40 Las moléculas de código de barras pueden comprender además una o más secuencias de ácidos nucleicos que no son complementarias a las regiones de códigos de barras de los oligonucleótidos con códigos de barras. Por ejemplo, las moléculas de código de barras pueden comprender una o más regiones adaptadoras. Una molécula de código de barras, puede, por ejemplo, comprender una región de adaptador 5' de una región de código de barras (una región de adaptador de 5') y/o una región de adaptador 3' de la región de código de barras (una región de adaptador de 3'). La región o regiones adaptadoras (y/o una o más porciones de una región adaptadora) pueden ser complementarias e hibridar con oligonucleótidos, por ejemplo, las regiones adaptadoras de oligonucleótidos con código de barras. Como alternativa, la región o regiones adaptadoras (y/o una o más porciones de una región adaptadora) de la molécula de código de barras pueden no ser complementarias a las secuencias de oligonucleótidos con código de barras. La(s) región(es) del adaptador se pueden utilizar para manipular, purificar, recuperar, amplificar y/o detectar moléculas de códigos de barras.
- 45

- 50 El reactivo de códigos de barras multimérico se puede configurar de manera que: cada una de las moléculas de código de barras comprende una secuencia de ácido nucleico que comprende, en la dirección 5' a 3', una región adaptadora y una región de código de barras; el primer oligonucleótido con código de barras comprende, opcionalmente en la dirección de 5' a 3', una región de código de barras hibridada con la región de código de barras de la primera molécula de código de barras, una región adaptadora hibridada con la región adaptadora de la primera molécula de código de barras y una región diana capaz de hibridar con una primera subsecuencia del ácido nucleico diana; y el segundo oligonucleótido con código de barras comprende, opcionalmente en la dirección de 5' a 3', una región de código de barras hibridada con la región de código de barras de la segunda molécula de código de barras, una región adaptadora hibridada con la región adaptadora de la segunda molécula de código de barras y una región diana capaz de hibridar con una segunda subsecuencia del ácido nucleico diana.
- 55

La región adaptadora de cada molécula de código de barras puede comprender una región constante. Opcionalmente, todas las regiones adaptadoras de un reactivo de códigos de barras multimérico son sustancialmente idénticas. La

- región adaptadora puede comprender al menos 1, al menos 2, al menos 3, al menos 4, al menos 5, al menos 6, al menos 8, al menos 10, al menos 15, al menos 20, al menos 25, al menos 50, al menos 100 o al menos 250 nucleótidos. Preferentemente, la región adaptadora comprende al menos 4 nucleótidos. Preferentemente, cada región adaptadora comprende desoxirribonucleótidos, opcionalmente, todos los nucleótidos en una región adaptadora son desoxirribonucleótidos. Uno o más de los desoxirribonucleótidos puede ser un desoxirribonucleótido modificado (por ejemplo, un desoxirribonucleótido modificado con un resto de biotina o un nucleótido de desoxiuracilo). Cada región adaptadora puede comprender una o más bases universales (por ejemplo, inosina), uno o más nucleótidos modificados y/o uno o más análogos de nucleótidos.
- Los oligonucleótidos con código de barras pueden comprender una región enlazadora entre la región adaptadora y la región diana. La región enlazadora puede comprender uno o más nucleótidos contiguos que no están hibridados con la molécula de código de barras multimérica y no son complementarios a las subsecuencias del ácido nucleico diana. El enlazador puede comprender de 1 a 100, de 5 a 75, de 10 a 50, de 15 a 30 o de 20 a 25 nucleótidos no complementarios. Preferentemente, el enlazador comprende de 15 a 30 nucleótidos no complementarios. El uso de una región enlazadora de este tipo mejora la eficacia de las reacciones de códigos de barras realizadas utilizando los reactivos de códigos de barras multiméricos.
- Las moléculas de código de barras de una molécula de código de barras multimérica se pueden unir a una molécula de ácido nucleico. Una molécula de ácido nucleico de este tipo puede proporcionar la estructura principal con la que pueden hibridar oligonucleótidos con código de barras monocatenario. Como alternativa, las moléculas de código de barras de una molécula de código de barras multimérica pueden unirse mediante cualquiera de los otros medios descritos en el presente documento.
- El reactivo de códigos de barras multimérico puede comprender: al menos 2, al menos 3, al menos 4, al menos 5, al menos 10, al menos 20, al menos 25, al menos 50, al menos 75, al menos 100, al menos 200, al menos 500, al menos 1000, al menos 5000, o al menos 10000 moléculas de código de barras unidas entre sí, en donde cada molécula de código de barras es tal como se define en el presente documento; y un oligonucleótido con código de barras hibridado con cada molécula de código de barras, en donde cada oligonucleótido con código de barras es tal como se define en el presente documento. Preferentemente, el reactivo de código de barras multimérico comprende al menos 5 moléculas de código de barras unidas entre sí, en donde cada molécula de código de barras es tal como se define en el presente documento; y un oligonucleótido con código de barras hibridado con cada molécula de código de barras, en donde cada oligonucleótido con código de barras es tal como se define en el presente documento.
- El reactivo de códigos de barras multimérico puede comprender: al menos 2, al menos 3, al menos 4, al menos 5, al menos 10, al menos 20, al menos 25, al menos 50, al menos 75, al menos 100, al menos 200, al menos 500, al menos 1000, al menos 5000, al menos 10^4 , al menos 10^5 , o al menos 10^6 moléculas de códigos de barras únicas o diferentes unidas entre sí, en donde cada molécula de código de barras es tal como se define en el presente documento; y un oligonucleótido con código de barras hibridado con cada molécula de código de barras, en donde cada oligonucleótido con código de barras es tal como se define en el presente documento. Preferentemente, el reactivo de código de barras multimérico comprende al menos 5 moléculas de código de barras únicas o diferentes unidas entre sí, en donde cada molécula de código de barras es tal como se define en el presente documento; y un oligonucleótido con código de barras hibridado con cada molécula de código de barras, en donde cada oligonucleótido con código de barras es tal como se define en el presente documento.
- El reactivo de códigos de barras multimérico puede comprender: al menos 5, al menos 10, al menos 20, al menos 25, al menos 50, al menos 75, al menos 100, al menos 200, al menos 500, al menos 1000, al menos 5000, o al menos 10.000 regiones de código de barras, en donde cada región de código de barras es tal como se define en el presente documento; y un oligonucleótido con código de barras hibridado con cada región de código de barras, en donde cada oligonucleótido con código de barras es tal como se define en el presente documento. Preferentemente, el reactivo de código de barras multimérico comprende al menos 5 regiones de códigos de barras, en donde cada región de código de barras es tal como se define en el presente documento; y un oligonucleótido con código de barras hibridado con cada región de código de barras, en donde cada oligonucleótido con código de barras es tal como se define en el presente documento.
- El reactivo de códigos de barras multimérico puede comprender: al menos 2, al menos 3, al menos 4, al menos 5, al menos 10, al menos 20, al menos 25, al menos 50, al menos 75, al menos 100, al menos 200, al menos 500, al menos 1000, al menos 5000, al menos 10^4 , al menos 10^5 , o al menos 10^6 regiones de códigos de barras únicas o diferentes, en donde cada región de código de barras es tal como se define en el presente documento; y un oligonucleótido con código de barras hibridado con cada región de código de barras, en donde cada oligonucleótido con código de barras es tal como se define en el presente documento. Preferentemente, el reactivo de código de barras multimérico comprende al menos 5 regiones de código de barras únicas o diferentes, en donde cada región de código de barras es tal como se define en el presente documento; y un oligonucleótido con código de barras hibridado con cada región de código de barras, en donde cada oligonucleótido con código de barras es tal como se define en el presente documento.
- La Figura 1 muestra un reactivo de código de barras multimérico, incluyendo las moléculas de código de barras primera (D1, E1 y F1) y segunda (D2, E2 y F2), cada una de las cuales incluye una secuencia de ácido nucleico que comprende

una región de código de barras (E1 y E2). Estas moléculas de código de barras primera y segunda están unidas entre sí, por ejemplo, mediante una secuencia de ácido nucleico de conexión (S). El reactivo de código de barras multimérico también comprende los oligonucleótidos con código de barras primero (A1, B1, C1 y G1) y segundo (A2, B2, C2 y G2). Cada uno de estos oligonucleótidos con código de barras comprende una región de código de barras (B1 y B2) y una región diana (G1 y G2).

Las regiones de códigos de barras dentro de los oligonucleótidos con códigos de barras pueden contener, cada una, una secuencia única que no está presente en otros oligonucleótidos con códigos de barras y, por tanto, pueden servir para identificar, de forma única, cada una de tales moléculas de códigos de barras. Las regiones diana se pueden usar para hibridar los oligonucleótidos con código de barras a subsecuencias de ácidos nucleicos diana, y luego se pueden usar como cebadores para una reacción de extensión de cebador o una reacción de amplificación, por ejemplo, una reacción en cadena de la polimerasa.

Cada molécula de código de barras puede incluir opcionalmente también una región adaptadora 5' (F1 y F2). Los oligonucleótidos con código de barras también pueden incluir una región adaptadora 3' (C1 y C2) que es complementaria a la región adaptadora 5' de las moléculas de código de barras.

Cada molécula de código de barras puede incluir opcionalmente también una región 3' (D1 y D2), que puede estar compuesta por secuencias idénticas dentro de cada molécula de código de barras. Los oligonucleótidos con código de barras también pueden incluir una región 5' (A1 y A2) que es complementaria a la región 3' de las moléculas de código de barras. Estas regiones 3' pueden ser útiles para la manipulación o amplificación de secuencias de ácidos nucleicos, por ejemplo, secuencias que se generan etiquetando una diana de ácido nucleico con un oligonucleótido con código de barras. La región 3' puede comprender al menos 4, al menos 5, al menos 6, al menos 8, al menos 10, al menos 15, al menos 20, al menos 25, al menos 50, al menos 100 o al menos 250 nucleótidos. Preferentemente, la región 3' comprende al menos 4 nucleótidos. Preferentemente, cada región 3' comprende desoxirribonucleótidos, opcionalmente, todos los nucleótidos en una región 3' son desoxirribonucleótidos. Uno o más de los desoxirribonucleótidos puede ser un desoxirribonucleótido modificado (por ejemplo, un desoxirribonucleótido modificado con un resto de biotina o un nucleótido de desoxiuracilo). Cada región 3' puede comprender una o más bases universales (por ejemplo, inosina), uno o más nucleótidos modificados y/o uno o más análogos de nucleótidos.

La divulgación proporciona una biblioteca de reactivos de códigos de barras multiméricos que comprenden al menos 10 reactivos de códigos de barras multiméricos para marcar un ácido nucleico diana para secuenciación, en donde cada reactivo de códigos de barras multimérico comprende: la primera y segunda moléculas de código de barras comprendidas dentro de una (única) molécula de ácido nucleico, en donde cada una de las moléculas de código de barras comprende una secuencia de ácido nucleico que comprende una región de código de barras; y primer y segundo oligonucleótidos con código de barras, en donde el primer oligonucleótido con código de barras comprende, opcionalmente en la dirección de 5' a 3', una región de código de barras complementaria e hibridada con la región de código de barras de la primera molécula de código de barras y una región diana capaz de hibridar o ligar a una primera subsecuencia del ácido nucleico diana, y en donde el segundo oligonucleótido con código de barras comprende, opcionalmente en la dirección de 5' a 3', una región de código de barras complementaria e hibridada con la región de código de barras de la segunda molécula de código de barras y una región diana capaz de hibridar o ligar a una segunda subsecuencia del ácido nucleico diana. Preferentemente, las regiones de código de barras de los oligonucleótidos con código de barras primero y segundo de cada reactivo de código de barras multimérico son diferentes a las regiones de código de barras de los oligonucleótidos con código de barras de al menos otros 9 reactivos de código de barras multiméricos en la biblioteca.

5. REACTIVOS DE CÓDIGO DE BARRAS MULTIMÉRICOS QUE COMPRENDEN OLIGONUCLEÓTIDOS CON CÓDIGO DE BARRAS HIBRIDADOS CON UNA MOLÉCULA DE HIBRIDACIÓN MULTIMÉRICA

La divulgación proporciona un reactivo de código de barras multimérico para etiquetar un ácido nucleico diana, en donde el reactivo comprende: primera y segunda moléculas de hibridación unidas entre sí (es decir, una molécula de hibridación multimérica), en donde cada una de las moléculas de hibridación comprende una secuencia de ácido nucleico que comprende una región de hibridación; y primer y segundo oligonucleótidos con código de barras, en donde el primer oligonucleótido con código de barras comprende, opcionalmente en la dirección de 5' a 3', una región adaptadora hibridada con la región de hibridación de la primera molécula de hibridación, una región de código de barras, y una región diana capaz de hibridar o ligar a una primera subsecuencia del ácido nucleico diana, y en donde el segundo oligonucleótido con código de barras comprende, opcionalmente en la dirección de 5' a 3', una región adaptadora hibridada con la región de hibridación de la segunda molécula de hibridación, una región de código de barras y una región diana capaz de hibridar o ligar a una segunda subsecuencia del ácido nucleico diana.

Opcionalmente, el primer y el segundo oligonucleótidos con código de barras comprenden, cada uno, una región adaptadora y una región diana en una única secuencia contigua que es complementaria e hibridada con una región de hibridación de una molécula de hibridación, y también capaz de hibridar o ligar a una subsecuencia de un nucleico diana ácido.

La divulgación proporciona un reactivo de código de barras multimérico para etiquetar un ácido nucleico diana, en donde el reactivo comprende: primera y segunda moléculas de hibridación unidas entre sí (es decir, una molécula de

hibridación multimérica), en donde cada una de las moléculas de hibridación comprende una secuencia de ácido nucleico que comprende una región de hibridación; y primer y segundo oligonucleótidos con código de barras, en donde el primer oligonucleótido con código de barras comprende, opcionalmente en la dirección de 5' a 3', una región de código de barras, una región adaptadora hibridada con la región de hibridación de la primera molécula de hibridación y una región diana capaz de hibridar o ligar a una primera subsecuencia del ácido nucleico diana, y en la que el segundo oligonucleótido con código de barras comprende, opcionalmente en la dirección de 5' a 3', una región de código de barras, una región adaptadora hibridada con la región de hibridación de la segunda molécula de hibridación y una región diana capaz de hibridar o ligar a una segunda subsecuencia del ácido nucleico diana.

Opcionalmente, el primer y el segundo oligonucleótidos con código de barras comprenden, cada uno, una región adaptadora y una región diana en una única secuencia contigua que es complementaria e hibridada con una región de hibridación de una molécula de hibridación, y también capaz de hibridar o ligar a una subsecuencia de un nucleico diana ácido.

La divulgación proporciona un reactivo de código de barras multimérico para etiquetar un ácido nucleico diana, en donde el reactivo comprende: primera y segunda moléculas de hibridación unidas entre sí (es decir, una molécula de hibridación multimérica), en donde cada una de las moléculas de hibridación comprende una secuencia de ácido nucleico que comprende una región de hibridación; y primer y segundo oligonucleótidos con código de barras, en donde el primer oligonucleótido con código de barras comprende (en la dirección 5'-3' o 3'-5') una región adaptadora hibridada con la región de hibridación de la primera molécula de hibridación, una región de código de barras y una región diana capaz de ligarse a una primera subsecuencia del ácido nucleico diana, y donde el segundo oligonucleótido con código de barras comprende (en la dirección 5'-3' o 3'-5') una región adaptadora hibridada con la región de hibridación de la segunda molécula de hibridación, una región de código de barras y una región diana capaz de ligarse a una segunda subsecuencia del ácido nucleico diana.

Opcionalmente, el primer y el segundo oligonucleótidos con código de barras comprenden cada uno una región adaptadora y una región diana en una única secuencia contigua que es complementaria e hibridada a una región de hibridación de una molécula de hibridación, y también capaz de ligar a una subsecuencia de un ácido nucleico diana.

La divulgación proporciona un reactivo de código de barras multimérico para etiquetar un ácido nucleico diana, en donde el reactivo comprende: primera y segunda moléculas de hibridación unidas entre sí (es decir, una molécula de hibridación multimérica), en donde cada una de las moléculas de hibridación comprende una secuencia de ácido nucleico que comprende una región de hibridación; y primer y segundo oligonucleótidos con código de barras, en donde el primer oligonucleótido con código de barras comprende (en la dirección 5'-3' o 3'-5') una región de código de barras, una región adaptadora hibridada con la región de hibridación de la primera molécula de hibridación y una región diana capaz de ligarse a una primera subsecuencia del ácido nucleico diana, y en donde el segundo oligonucleótido con código de barras comprende (en la dirección 5'-3' o 3'-5') una región de código de barras, una región adaptadora hibridada con la región de hibridación de la segunda molécula de hibridación y una región diana capaz de ligarse a una segunda subsecuencia del ácido nucleico diana.

Opcionalmente, el primer y el segundo oligonucleótidos con código de barras comprenden cada uno una región adaptadora y una región diana en una única secuencia contigua que es complementaria e hibridada a una región de hibridación de una molécula de hibridación, y también capaz de ligar a una subsecuencia de un ácido nucleico diana.

La divulgación proporciona un reactivo de código de barras multimérico para etiquetar un ácido nucleico diana, en donde el reactivo comprende: primera y segunda moléculas de hibridación unidas entre sí (es decir, una molécula de hibridación multimérica), en donde cada una de las moléculas de hibridación comprende una secuencia de ácido nucleico que comprende una región de código de barras; y primer y segundo oligonucleótidos con código de barras, en donde el primer oligonucleótido con código de barras comprende, en la dirección 5' a 3', una región adaptadora hibridada con la región de hibridación de la primera molécula de hibridación, una región de código de barras y una región diana capaz de hibridar con una primera subsecuencia del ácido nucleico diana, y en donde el segundo oligonucleótido con código de barras comprende, en la dirección 5' a 3', una región adaptadora hibridada con la región de hibridación de la segunda molécula de hibridación, una región de código de barras y una región diana capaz de hibridar con una segunda subsecuencia del ácido nucleico diana.

La divulgación proporciona un reactivo de código de barras multimérico para etiquetar un ácido nucleico diana, en donde el reactivo comprende: primera y segunda moléculas de hibridación unidas entre sí (es decir, una molécula de hibridación multimérica), en donde cada una de las moléculas de hibridación comprende una secuencia de ácido nucleico que comprende una región de código de barras; y primer y segundo oligonucleótidos con código de barras, en donde el primer oligonucleótido con código de barras comprende, en la dirección 5' a 3', una región de código de barras, una región adaptadora hibridada con la región de hibridación de la primera molécula de hibridación y una región diana capaz de hibridar con una primera subsecuencia del ácido nucleico diana, y en donde el segundo oligonucleótido con código de barras comprende, en la dirección 5' a 3', una región de código de barras, una región adaptadora hibridada con la región de hibridación de la segunda molécula de hibridación y una región diana capaz de hibridar con una segunda subsecuencia del ácido nucleico diana.

Opcionalmente, el primer y el segundo oligonucleótidos con código de barras comprenden cada uno una región

adaptadora y una región diana en una única secuencia contigua que es complementaria e hibridada a una región de hibridación de una molécula de hibridación, y también capaz de hibridar con una subsecuencia de un ácido nucleico diana.

5 Preferentemente, la región adaptadora del primer oligonucleótido con código de barras comprende una secuencia que es complementaria e hibridada con la región de hibridación de la primera molécula de hibridación y la región adaptadora del segundo oligonucleótido con código de barras comprende una secuencia que es complementaria e hibridada con la región de hibridación de la segunda molécula de hibridación. La secuencia complementaria de cada oligonucleótido con código de barras puede ser al menos 5, al menos 10, al menos 15, al menos 20, al menos 25, al menos 50 o al menos 100 nucleótidos contiguos.

10 La región de hibridación de cada molécula de hibridación puede comprender una región constante. Preferentemente, todas las regiones de hibridación de un reactivo de código de barras multimérico son sustancialmente idénticas. Opcionalmente, todas las regiones de hibridación de una biblioteca de reactivos de códigos de barras multiméricos son sustancialmente idénticas. La región de hibridación puede comprender al menos 1, al menos 2, al menos 3, al menos 4, al menos 5, al menos 6, al menos 8, al menos 10, al menos 15, al menos 20, al menos 25, al menos 50, al menos 100 o al menos 250 nucleótidos. Preferentemente, la región de hibridación comprende al menos 4 nucleótidos. Preferentemente, cada región de hibridación comprende desoxirribonucleótidos, opcionalmente, todos los nucleótidos en una región de hibridación son desoxirribonucleótidos. Uno o más de los desoxirribonucleótidos puede ser un desoxirribonucleótido modificado (por ejemplo, un desoxirribonucleótido modificado con un resto de biotina o un nucleótido de desoxiuracilo). Cada región de hibridación puede comprender una o más bases universales (por ejemplo, inosina), uno o más nucleótidos modificados y/o uno o más análogos de nucleótidos.

Las regiones diana de los oligonucleótidos con código de barras pueden no hibridar con la(s) molécula(s) de hibridación multimérica(s). Las regiones diana de los oligonucleótidos con código de barras pueden no ser complementarias a la(s) molécula(s) de hibridación multimérica(s).

25 La divulgación proporciona un reactivo de código de barras multimérico para etiquetar un ácido nucleico diana, en donde el reactivo comprende: primera y segunda moléculas de hibridación unidas entre sí (i.e. comprende de 1 a 100, de 5 a 75, de 10 a 50, de 15 a 30 o de 20 a 25 nucleótidos no complementarios. Preferentemente, el enlazador comprende de 15 a 30 nucleótidos no complementarios. El uso de una región enlazadora de este tipo mejora la eficacia de las reacciones de códigos de barras realizadas utilizando los reactivos de códigos de barras multiméricos.

30 Las moléculas de hibridación pueden comprender además una o más secuencias de ácidos nucleicos que no son complementarias a los oligonucleótidos con códigos de barras. Por ejemplo, las moléculas de hibridación pueden comprender una o más regiones adaptadoras. Una molécula de hibridación, puede, por ejemplo, comprender una región de adaptador 5' de una región de hibridación (una región de adaptador de 5') y/o una región de adaptador 3' de la región de hibridación (una región de adaptador de 3'). La(s) región(es) del adaptador se pueden utilizar para manipular, purificar, recuperar, amplificar y/o detectar moléculas de hibridación.

35 La región adaptadora de cada molécula de hibridación puede comprender una región constante. Opcionalmente, todas las regiones adaptadoras de un reactivo de hibridación multimérico son sustancialmente idénticas. La región adaptadora puede comprender al menos 1, al menos 2, al menos 3, al menos 4, al menos 5, al menos 6, al menos 8, al menos 10, al menos 15, al menos 20, al menos 25, al menos 50, al menos 100 o al menos 250 nucleótidos. Preferentemente, la región adaptadora comprende al menos 4 nucleótidos. Preferentemente, cada región adaptadora comprende desoxirribonucleótidos, opcionalmente, todos los nucleótidos en una región adaptadora son desoxirribonucleótidos. Uno o más de los desoxirribonucleótidos puede ser un desoxirribonucleótido modificado (por ejemplo, un desoxirribonucleótido modificado con un resto de biotina o un nucleótido de desoxiuracilo). Cada región adaptadora puede comprender una o más bases universales (por ejemplo, inosina), uno o más nucleótidos modificados y/o uno o más análogos de nucleótidos.

45 Los oligonucleótidos con código de barras pueden comprender una región enlazadora entre la región adaptadora y la región diana. La región enlazadora puede comprender uno o más nucleótidos contiguos que no están hibridados con la molécula de código de hibridación y no son complementarios a las subsecuencias del ácido nucleico diana. El enlazador puede comprender de 1 a 100, de 5 a 75, de 10 a 50, de 15 a 30 o de 20 a 25 nucleótidos no complementarios. Preferentemente, el enlazador comprende de 15 a 30 nucleótidos no complementarios. El uso de una región enlazadora de este tipo mejora la eficacia de las reacciones de códigos de barras realizadas utilizando los reactivos de códigos de barras multiméricos.

55 La divulgación proporciona una biblioteca de reactivos de códigos de barras multiméricos que comprenden al menos 10 reactivos de códigos de barras multiméricos para marcar un ácido nucleico diana para secuenciación, en donde cada reactivo de códigos de barras multimérico comprende: primera y segunda moléculas de hibridación comprendidas dentro de una (única) molécula de ácido nucleico, en donde cada una de las moléculas de hibridación comprende una secuencia de ácido nucleico que comprende una región de hibridación; y primer y segundo oligonucleótidos con código de barras, en donde el primer oligonucleótido con código de barras comprende, opcionalmente en la dirección de 5' a 3', una región adaptadora complementaria e hibridada con la región de hibridación de la primera molécula de hibridación, una región de código de barras, y una región diana capaz de hibridar o ligar a una primera subsecuencia

del ácido nucleico diana, y en donde el segundo oligonucleótido con código de barras comprende, opcionalmente en la dirección de 5' a 3', una región adaptadora complementaria e hibridada con la región de hibridación de la segunda molécula de hibridación, una región de código de barras y una región diana capaz de hibridar o ligar a una segunda subsecuencia del ácido nucleico diana.

- 5 Preferentemente, las regiones de código de barras de los oligonucleótidos con código de barras primero y segundo de cada reactivo de código de barras multimérico son diferentes a las regiones de código de barras de los oligonucleótidos con código de barras de al menos otros 9 reactivos de código de barras multiméricos en la biblioteca.

La divulgación proporciona una biblioteca de reactivos de códigos de barras multiméricos que comprenden al menos 10 reactivos de códigos de barras multiméricos para marcar un ácido nucleico diana para secuenciación, en donde

- 10 cada reactivo de códigos de barras multimérico comprende: primera y segunda moléculas de hibridación comprendidas dentro de una (única) molécula de ácido nucleico, en donde cada una de las moléculas de hibridación comprende una secuencia de ácido nucleico que comprende una región de hibridación; y primer y segundo oligonucleótidos con código de barras, en donde el primer oligonucleótido con código de barras comprende, opcionalmente en la dirección de 5' a 3', una región de código de barras, una región adaptadora complementaria e hibridada con la región de hibridación de la primera molécula de hibridación y una región diana capaz de hibridar o ligar a una primera subsecuencia del ácido nucleico diana, y en donde el segundo oligonucleótido con código de barras comprende, opcionalmente en la dirección de 5' a 3', una región de código de barras, una región adaptadora complementaria e hibridada con la región de hibridación de la segunda molécula de hibridación y una región diana capaz de hibridar o ligar a una segunda subsecuencia del ácido nucleico diana. Preferentemente, las regiones de código de barras de los oligonucleótidos con
- 20 código de barras primero y segundo de cada reactivo de código de barras multimérico son diferentes a las regiones de código de barras de los oligonucleótidos con código de barras de al menos otros 9 reactivos de código de barras multiméricos en la biblioteca.

6. REACTIVOS DE CÓDIGO DE BARRAS MULTIMÉRICOS QUE COMPRENDEN OLIGONUCLEÓTIDOS CON CÓDIGO DE BARRAS UNIDOS POR UNA MACROMOLÉCULA

- 25 La divulgación proporciona un reactivo de código de barras multimérico para etiquetar un ácido nucleico diana, en donde el reactivo comprende un primer y segundo oligonucleótidos con código de barras unidos entre sí por una macromolécula, y en donde los oligonucleótidos con código de barras comprenden, cada uno, una región de código de barras.

- 30 El primer oligonucleótido con código de barras puede comprender, además, una región diana capaz de hibridar o ligar a una primera subsecuencia del ácido nucleico diana, y el segundo oligonucleótido con código de barras puede comprender, además, una región diana capaz de hibridar o ligar a una segunda subsecuencia del ácido nucleico diana.

- 35 El primer oligonucleótido con código de barras puede comprender, en la dirección 5'-3', una región de código de barras y una región diana capaz de hibridar con una primera subsecuencia del ácido nucleico diana, y el segundo oligonucleótido con código de barras puede comprender, en la dirección 5'-3', una región de código de barras y una región diana capaz de hibridar con una segunda subsecuencia del ácido nucleico diana.

Los oligonucleótidos con código de barras pueden comprender además cualquiera de las características descritas en el presente documento.

Los oligonucleótidos con código de barras pueden estar unidos por una macromolécula uniéndose a la macromolécula y/o hibridando con la macromolécula.

- 40 Los oligonucleótidos con código de barras pueden unirse a la macromolécula directa o indirectamente (por ejemplo, mediante una molécula enlazadora). Los oligonucleótidos con código de barras se pueden enlazar uniéndose a la macromolécula y/o uniéndose o hibridando con moléculas enlazadoras que están unidas a la macromolécula. Los oligonucleótidos con código de barras se pueden unir a la macromolécula (o a las moléculas enlazadoras) mediante enlace covalente, enlace no covalente (por ejemplo, una interacción proteína-proteína o un enlace estreptavidina-biotina) o hibridación de ácido nucleico. La molécula enlazadora puede ser un biopolímero (por ejemplo, una molécula de ácido nucleico) o un polímero sintético. La molécula enlazadora puede comprender una o más unidades de etilenglicol y/o polietilenglicol (por ejemplo, hexaetilenglicol o pentaetilenglicol). La molécula enlazadora puede comprender uno o más grupos etilo, tales como un espaciador C3 (tres carbonos), espaciador C6, espaciador C12 o espaciador C18.

- 50 La macromolécula puede ser un polímero sintético (por ejemplo, un dendrímero) o un biopolímero como un ácido nucleico (por ejemplo, un ácido nucleico monocatenario como ADN monocatenario), un péptido, un polipéptido o una proteína (por ejemplo, una proteína multimérica).

El dendrímero puede comprender al menos 2, al menos 3, al menos 5, o al menos 10 generaciones.

- 55 La macromolécula puede ser un ácido nucleico que comprende dos o más nucleótidos, cada uno de los cuales puede unirse a un oligonucleótido con código de barras. Adicionalmente o como alternativa, el ácido nucleico puede comprender dos o más regiones, cada una de las cuales puede hibridar con un oligonucleótido con código de barras.

El ácido nucleico puede comprender un primer nucleótido modificado y un segundo nucleótido modificado, en donde cada nucleótido modificado comprende un resto de unión (por ejemplo, un resto de biotina o un resto de alquino que se puede usar para una reacción química de clic) capaz de unirse a un oligonucleótido con código de barras. Opcionalmente, el primer y segundo nucleótidos modificados pueden estar separados por una secuencia de ácidos nucleicos intermedia de al menos uno, al menos dos, al menos 5 o al menos 10 nucleótidos.

El ácido nucleico puede comprender una primera región de hibridación y una segunda región de hibridación, en donde cada región de hibridación comprende una secuencia complementaria y capaz de hibridar con una secuencia de al menos un nucleótido dentro de un oligonucleótido con código de barras. La secuencia complementaria puede ser de al menos 5, al menos 10, al menos 15, al menos 20, al menos 25 o al menos 50 nucleótidos contiguos. Opcionalmente, la primera y segunda regiones de hibridación pueden estar separadas por una secuencia de ácido nucleico intermedia de al menos una, al menos dos, al menos 5 o al menos 10 nucleótidos.

La macromolécula puede ser una proteína tal como una proteína multimérica, por ejemplo, una proteína homomérica o una proteína heteromérica. Por ejemplo, la proteína puede comprender estreptavidina, por ejemplo, estreptavidina tetramérica.

También se proporcionan bibliotecas de reactivos de códigos de barras multiméricos que comprenden oligonucleótidos con códigos de barras unidos por una macromolécula. Tales bibliotecas pueden basarse en las propiedades generales de las bibliotecas de reactivos de códigos de barras multiméricos descritos en el presente documento. En las bibliotecas, cada reactivo de códigos de barras multimérico puede comprender una macromolécula diferente.

7. REACTIVOS DE CÓDIGO DE BARRAS MULTIMÉRICOS QUE COMPRENDEN OLIGONUCLEÓTIDOS CON CÓDIGO DE BARRAS UNIDOS POR UN SOPORTE SÓLIDO O SEMISÓLIDO

La divulgación proporciona un reactivo de código de barras multimérico para etiquetar un ácido nucleico diana, en donde el reactivo comprende un primer y segundo oligonucleótidos con código de barras unidos entre sí por un soporte sólido o un soporte semisólido, y en donde los oligonucleótidos con código de barras comprenden, cada uno, una región de código de barras.

El primer oligonucleótido con código de barras puede comprender, además, una región diana capaz de hibridar o ligar a una primera subsecuencia del ácido nucleico diana, y el segundo oligonucleótido con código de barras puede comprender, además, una región diana capaz de hibridar o ligar a una segunda subsecuencia del ácido nucleico diana.

El primer oligonucleótido con código de barras puede comprender, en la dirección 5'-3', una región de código de barras y una región diana capaz de hibridar con una primera subsecuencia del ácido nucleico diana, y el segundo oligonucleótido con código de barras puede comprender, en la dirección 5'-3', una región de código de barras y una región diana capaz de hibridar con una segunda subsecuencia del ácido nucleico diana.

Los oligonucleótidos con código de barras pueden comprender además cualquiera de las características descritas en el presente documento.

Los oligonucleótidos con código de barras se pueden unir mediante un soporte sólido o un soporte semisólido. Los oligonucleótidos con código de barras se pueden unir al soporte directa o indirectamente (por ejemplo, mediante una molécula enlazadora). Los oligonucleótidos con código de barras se pueden unir uniéndose al soporte y/o uniéndose o hibridando a moléculas enlazadoras que están unidas al soporte. Los oligonucleótidos con código de barras se pueden unir al soporte (o a las moléculas enlazadoras) mediante enlace covalente, enlace no covalente (por ejemplo, una interacción proteína-proteína o un enlace estreptavidina-biotina) o hibridación de ácido nucleico. La molécula enlazadora puede ser un biopolímero (por ejemplo, una molécula de ácido nucleico) o un polímero sintético. La molécula enlazadora puede comprender una o más unidades de etilenglicol y/o polietilenglicol (por ejemplo, hexaetilenglicol o pentaetilenglicol). La molécula enlazadora puede comprender uno o más grupos etilo, tales como un espaciador C3 (tres carbonos), espaciador C6, espaciador C12 o espaciador C18.

El soporte puede comprender una superficie plana. El soporte puede ser un portaobjetos, por ejemplo, un portaobjetos de vidrio. El portaobjetos puede ser una celda de flujo para secuenciar. Si el soporte es un portaobjetos, el primer y segundo oligonucleótidos con código de barras se pueden inmovilizar en una región discreta del portaobjetos. Opcionalmente, los oligonucleótidos con código de barras de cada reactivo de código de barras multimérico en una biblioteca se inmovilizan en una región discreta diferente en el portaobjetos a los oligonucleótidos con código de barras de los otros reactivos de código de barras multimérico en la biblioteca. El soporte puede ser una placa que comprende pocillos, opcionalmente en donde el primer y segundo oligonucleótidos con código de barras están inmovilizados en el mismo pocillo. Opcionalmente, los oligonucleótidos con código de barras de cada reactivo de código de barras multimérico en la biblioteca se inmovilizan en un pocillo diferente de la placa a los oligonucleótidos con código de barras de los otros reactivos de código de barras multimérico en la biblioteca.

Preferentemente, el soporte es una perla (por ejemplo, una perla de gel). La perla puede ser una perla de agarosa, una perla de sílice, una perla de espuma de poliestireno, una perla de gel (como las disponibles en 10x Genomics®), una perla conjugada de anticuerpo, una perla conjugada oligo-dT, una perla de estreptavidina o una perla magnética (por ejemplo, una perla superparamagnética). La perla puede ser de cualquier tamaño y/o estructura molecular. Por

ejemplo, la perla puede tener un diámetro de 10 nanómetros a 100 micrómetros, de 100 nanómetros a 10 micrómetros de diámetro, o de 1 micrómetro a 5 micrómetros de diámetro. Opcionalmente, la perla tiene aproximadamente 10 nanómetros de diámetro, aproximadamente 100 nanómetros de diámetro, aproximadamente 1 micrómetro de diámetro, aproximadamente 10 micrómetros de diámetro o aproximadamente 100 micrómetros de diámetro. La perla puede ser sólida o, como alternativa, la perla puede ser hueca o parcialmente hueca o porosa. Las perlas de ciertos tamaños pueden ser más preferibles para ciertos métodos de códigos de barras. Por ejemplo, perlas de menos de 5,0 micrómetros, o menos de 1,0 micrómetro, pueden ser más útiles para poner códigos de barras a dianas de ácidos nucleicos dentro de células individuales. Preferentemente, los oligonucleótidos con código de barras de cada reactivo de códigos de barras multimérico en una biblioteca están unidas en una perla diferente a los oligonucleótidos con código de barras de los otros reactivos de códigos de barras multiméricos en la biblioteca.

El soporte se puede funcionalizar para permitir la unión de dos o más oligonucleótidos con código de barras. Esta funcionalización se puede habilitar mediante la adición de restos químicos (por ejemplo, grupos carboxilados, alquinos, azidas, grupos de acrilato, grupos amino, grupos sulfato o grupos succinimida) y/o restos basados en proteínas (por ejemplo, estreptavidina, avidina o proteína G) al soporte. Los oligonucleótidos con código de barras se pueden unir a los restos directa o indirectamente (por ejemplo, mediante una molécula de enlazador).

Los soportes funcionalizados (por ejemplo, perlas) se pueden poner en contacto con una solución de oligonucleótidos con código de barras en condiciones que promuevan la unión de dos o más oligonucleótidos con código de barras a cada perla en la solución (generando reactivos de códigos de barras multiméricos).

También se proporcionan bibliotecas de reactivos de códigos de barras multiméricos que comprenden oligonucleótidos con códigos de barras unidos por soporte. Tales bibliotecas pueden basarse en las propiedades generales de las bibliotecas de reactivos de códigos de barras multiméricos descritos en el presente documento. En las bibliotecas, cada reactivo de código de barras multimérico puede comprender un soporte diferente (por ejemplo, una perla etiquetada de manera diferente). En una biblioteca de reactivos de códigos de barras multiméricos, los oligonucleótidos con código de barras de cada reactivo de código de barras multimérico en una biblioteca se pueden unir entre sí en un soporte diferente a los oligonucleótidos con código de barras de los otros reactivos de código de barras multiméricos en la biblioteca.

8. REACTIVOS DE CÓDIGOS DE BARRAS MULTIMÉRICOS QUE COMPRENDEN OLIGONUCLEÓTIDOS CON CÓDIGO DE BARRAS UNIDOS ENTRE SÍ POR ESTAR COMPRENDIDOS DENTRO DE UN VEHÍCULO DE LÍPIDOS

La divulgación proporciona un reactivo de código de barras multimérico para etiquetar un ácido nucleico diana, en donde el reactivo comprende un primer y un segundo oligonucleótidos con código de barras y un vehículo de lípidos, en donde el primer y segundo oligonucleótidos con código de barras están unidos entre sí al estar comprendidos dentro del vehículo de lípidos, y en donde los oligonucleótidos con código de barras comprenden, cada uno, una región de código de barras.

El primer oligonucleótido con código de barras puede comprender, además, una región diana capaz de hibridar o ligar a una primera subsecuencia del ácido nucleico diana, y el segundo oligonucleótido con código de barras puede comprender, además, una región diana capaz de hibridar o ligar a una segunda subsecuencia del ácido nucleico diana.

El primer oligonucleótido con código de barras puede comprender, en la dirección 5'-3', una región de código de barras y una región diana capaz de hibridar con una primera subsecuencia del ácido nucleico diana, y el segundo oligonucleótido con código de barras puede comprender, en la dirección 5'-3', una región de código de barras y una región diana capaz de hibridar con una segunda subsecuencia del ácido nucleico diana.

Los oligonucleótidos con código de barras pueden comprender además cualquiera de las características descritas en el presente documento.

La divulgación proporciona una biblioteca de reactivos de códigos de barras multiméricos que comprenden reactivos de códigos de barras multiméricos primero y segundo tal como se define en el presente documento, en donde los oligonucleótidos con código de barras del primer reactivo de código de barras multimérico están comprendidos dentro de un primer vehículo de lípidos, y en donde los oligonucleótidos con código de barras del segundo reactivo de código de barras multimérico están compuestos por un segundo vehículo de lípidos, y en donde las regiones de código de barras de los oligonucleótidos con código de barras del primer reactivo de código de barras multimérico son diferentes a las regiones de códigos de barras de los oligonucleótidos con códigos de barras del segundo reactivo de códigos de barras multimérico.

La biblioteca de reactivos de códigos de barras multiméricos puede comprender al menos 5, al menos 10, al menos 20, al menos 25, al menos 50, al menos 75, al menos 100, al menos 250, al menos 500, al menos 10^3 , al menos 10^4 , al menos 10^5 , al menos 10^6 , al menos 10^7 , al menos 10^8 o al menos 10^9 reactivos de códigos de barras multiméricos tal como se define en el presente documento. Preferentemente, la biblioteca comprende al menos 10 reactivos de códigos de barras multiméricos tal como se define en el presente documento. Preferentemente, las regiones de código de barras de los oligonucleótidos con código de barras primero y segundo de cada reactivo de código de barras multimérico son diferentes a las regiones de código de barras de los oligonucleótidos con código de barras de al menos otros 9 reactivos de código de barras multiméricos en la biblioteca.

Los oligonucleótidos con código de barras de cada reactivo de código de barras multimérico están comprendidos dentro de un vehículo lipídico diferente.

El vehículo lipídico puede ser un liposoma o una micela. El vehículo de lípidos puede ser un vehículo de fosfolípidos. El vehículo lipídico puede comprender una o más moléculas anfifílicas. El vehículo lipídico puede comprender uno o más fosfolípidos. El fosfolípido puede ser fosfatidilcolina. El vehículo lipídico puede comprender uno o más de los siguientes constituyentes: fosfatidiletanolamina, fosfatidilserina, colesterol, cardiolipina, dicetilfosfato, estearilamina, fosfatidilglicerol, dipalmitoilfosfatidilcolina, diestearilfosfatidilcolina y/o cualquier molécula relacionada y/o derivada de los mismos. Opcionalmente, el vehículo lipídico puede comprender cualquier combinación de dos o más constituyentes descritos anteriormente, con o sin otros componentes.

- 10 El vehículo lipídico (por ejemplo, un liposoma o una micela) puede ser unilaminar o multilaminar. Una biblioteca de reactivos de códigos de barras multiméricos puede comprender vehículos lipídicos tanto unilaminares como multilaminares. El vehículo lipídico puede comprender un copolímero, por ejemplo, un copolímero de bloques.

El vehículo lipídico puede comprender al menos 2, al menos 3, al menos 5, al menos 10, al menos 50, al menos 100, al menos 500, al menos 1000, al menos 10.000, o al menos 100.000 oligonucleótidos con código de barras, o cualquier número mayor de oligonucleótidos con código de barras.

- 15 Cualquier vehículo lipídico (por ejemplo, liposoma o micela y/o reactivo liposómico o micelar) puede en promedio formar complejo con 1, o menos de 1, o más de 1 reactivo(s) de código de barras multimérico(s) para formar una biblioteca de dicho(s) reactivo(s) de código de barras multimérico(s).

- 20 La divulgación proporciona una biblioteca de reactivos de códigos de barras multiméricos que comprenden al menos 10 reactivos de códigos de barras multiméricos tal como se define en el presente documento, en donde cada reactivo de código de barras multimérico comprende oligonucleótidos con código de barras primero y segundo comprendidos dentro de un vehículo de lípidos diferente, y en donde las regiones de código de barras del primer y segundo oligonucleótidos con código de barras de cada reactivo de código de barras multimérico son diferentes a las regiones de código de barras de los oligonucleótidos con código de barras de al menos otros 9 reactivos de códigos de barras multiméricos en la biblioteca.

- 25 Un método para preparar reactivos de códigos de barras multiméricos comprende cargar oligonucleótidos con códigos de barras y/o reactivo(s) de códigos de barras multiméricos en vehículos lipídicos (por ejemplo, liposomas o micelas). El método puede comprender una etapa de carga pasiva, activa y/o remota. Los vehículos de lípidos preformados (por ejemplo, liposomas y/o micelas) se pueden cargar poniéndolos en contacto con una solución de oligonucleótidos con códigos de barras y/o reactivo(s) de códigos de barras multiméricos. Los vehículos de lípidos (por ejemplo, liposomas y/o micelas) se pueden cargar poniéndolos en contacto con una solución de oligonucleótidos con códigos de barras y/o reactivo(s) de códigos de barras multiméricos antes y/o durante la formación o síntesis de los vehículos de lípidos. El método puede comprender la encapsulación pasiva y/o la captura de oligonucleótidos con códigos de barras y/o reactivo(s) de códigos de barras multiméricos en vehículos lipídicos.

- 35 Los vehículos de lípidos (por ejemplo, liposomas y/o micelas) se pueden preparar mediante un método basado en ultrasonidos, un método basado en la prensa francesa, un método de fase inversa, un método de evaporación de disolventes, un método basado en extrusión, un método basado en la mezcla mecánica, un método basado en congelación/descongelación, un método basado en deshidratar/rehidratar y/o cualquier combinación de los mismos.

- 40 Los vehículos de lípidos (por ejemplo, liposomas y/o micelas) se pueden estabilizar y/o almacenar antes de su uso usando métodos conocidos.

Cualquiera de los reactivos o kits de códigos de barras multiméricos descritos en el presente documento puede estar compuesto por un vehículo lipídico.

9. KITS QUE COMPRENDEN REACTIVOS DE CÓDIGO DE BARRAS MULTIMÉRICOS Y OLIGONUCLEÓTIDOS ADAPTADORES

- 45 La divulgación proporciona además kits que comprenden uno o más de los componentes definidos en el presente documento. La divulgación también proporciona kits específicamente adaptados para realizar cualquiera de los métodos definidos en el presente documento.

La divulgación proporciona además un kit para etiquetar un ácido nucleico diana, en donde el kit comprende:

- 50 (a) un reactivo de código de barras multimérico que comprende (i) moléculas de código de barras primera y segunda unidas entre sí (es decir, una molécula de código de barras multimérica), en donde cada una de las moléculas de código de barras comprende una secuencia de ácido nucleico que comprende, opcionalmente en la dirección de 5' a 3', una región adaptadora y una región de código de barras, y (ii) el primer y segundo oligonucleótidos con código de barras, en donde el primer oligonucleótido con código de barras comprende una región de código de barras hibridada con la región de código de barras de la primera molécula de código de barras, y en donde el segundo oligonucleótido con código de barras comprende una región de código de barras hibridada con la región de código de barras de la

segunda molécula de código de barras; y (b) oligonucleótidos adaptadores primero y segundo, en donde el primer oligonucleótido adaptador comprende, opcionalmente en la dirección de 5' a 3', una región adaptadora capaz de hibridar con la región adaptadora de la primera molécula de código de barras y una región diana capaz de hibridar o ligar a una primera subsecuencia del ácido nucleico diana, y en donde el segundo oligonucleótido adaptador

comprende, opcionalmente en la dirección de 5' a 3', una región adaptadora capaz de hibridar con la región adaptadora de la segunda molécula de código de barras y una región diana capaz de hibridar o ligar con una segunda subsecuencia del ácido nucleico diana.

La divulgación proporciona además un kit para etiquetar un ácido nucleico diana, en donde el kit comprende:

(a) un reactivo de código de barras multimérico que comprende (i) moléculas de código de barras primera y segunda unidas entre sí (es decir, una molécula de código de barras multimérica), en donde cada una de las moléculas de código de barras comprende una secuencia de ácido nucleico que comprende una región adaptadora y una región de código de barras, y (ii) un primer y un segundo oligonucleótidos con código de barras, en donde el primer oligonucleótido con código de barras comprende una región de código de barras hibridada con la región de código de barras de la primera molécula de código de barras, y en donde el segundo oligonucleótido con código de barras comprende una región de código de barras hibridada con la región de código de barras de la segunda molécula de código de barras; y (b) oligonucleótidos adaptadores primero y segundo, en donde el primer oligonucleótido adaptador comprende una región adaptadora capaz de hibridar con la región adaptadora de la primera molécula de código de barras y una región diana capaz de ligarse a una primera subsecuencia del ácido nucleico diana, y en donde el segundo oligonucleótido adaptador comprende una región adaptadora capaz de hibridar con la región adaptadora de la segunda molécula de código de barras y una región diana capaz de ligarse a una segunda subsecuencia del ácido nucleico diana.

La divulgación proporciona además un kit para etiquetar un ácido nucleico diana, en donde el kit comprende:

(a) un reactivo de código de barras multimérico que comprende (i) moléculas de código de barras primera y segunda unidas entre sí (es decir, una molécula de código de barras multimérica), en donde cada una de las moléculas de código de barras comprende una secuencia de ácido nucleico que comprende, en la dirección 5' a 3', una región adaptadora y una región de código de barras, y (ii) un primer y un segundo oligonucleótidos con código de barras, en donde el primer oligonucleótido con código de barras comprende una región de código de barras hibridada con la región de código de barras de la primera molécula de código de barras, y en donde el segundo oligonucleótido con código de barras comprende una región de código de barras hibridada con la región de código de barras de la segunda molécula de código de barras; y (b) oligonucleótidos adaptadores primero y segundo, en donde el primer oligonucleótido adaptador comprende, en la dirección 5' a 3', una región adaptadora capaz de hibridar con la región adaptadora de la primera molécula de código de barras y una región diana capaz de hibridar con una primera subsecuencia del ácido nucleico diana, y en donde el segundo oligonucleótido adaptador comprende, en la dirección 5' a 3', una región adaptadora capaz de hibridar con la región adaptadora de la segunda molécula de código de barras y una región diana capaz de hibridar con una segunda subsecuencia del ácido nucleico diana.

La divulgación proporciona además un kit para etiquetar un ácido nucleico diana, en donde el kit comprende:

(a) un reactivo de código de barras multimérico que comprende (i) moléculas de código de barras primera y segunda unidas entre sí (es decir, una molécula de código de barras multimérica), en donde cada una de las moléculas de código de barras comprende una secuencia de ácido nucleico que comprende, opcionalmente en la dirección de 5' a 3', una región adaptadora y una región de código de barras, y (ii) el primer y segundo oligonucleótidos con código de barras, en donde el primer oligonucleótido con código de barras comprende una región de código de barras hibridada con la región de código de barras de la primera molécula de código de barras, y en donde el segundo oligonucleótido con código de barras comprende una región de código de barras hibridada con la región de código de barras de la segunda molécula de código de barras; y (b) oligonucleótidos adaptadores primero y segundo, en donde el primer oligonucleótido adaptador comprende una región adaptadora capaz de hibridar con la región adaptadora de la primera molécula de código de barras y capaz de ligarse a una primera subsecuencia del ácido nucleico diana, y en donde el segundo oligonucleótido adaptador comprende una región adaptadora capaz de hibridar a la región adaptadora de la segunda molécula de código de barras y capaz de ligarse a una segunda subsecuencia del ácido nucleico diana.

Cada oligonucleótido adaptador puede consistir esencialmente en o consistir en una región adaptadora. Cada oligonucleótido adaptador puede no comprender una región diana.

Preferentemente, la región adaptadora del primer oligonucleótido adaptador comprende una secuencia que es complementaria y capaz de hibridar con la región adaptadora de la primera molécula de código de barras y la región adaptadora del segundo oligonucleótido adaptador comprende una secuencia que es complementaria y capaz de hibridar con la región adaptadora de la segunda molécula de código de barras. La secuencia complementaria de cada oligonucleótido adaptador puede ser al menos 5, al menos 10, al menos 15, al menos 20, al menos 25, al menos 50 o al menos 100 nucleótidos contiguos.

Las regiones diana de los oligonucleótidos adaptadores pueden no ser capaces de hibridar con la(s) molécula(s) de código de barras multimérica(s). Las regiones diana de los oligonucleótidos adaptadores pueden no ser

complementarias a la(s) molécula(s) de código de barras multimérica(s).

Las regiones diana de cada oligonucleótido adaptador pueden comprender secuencias diferentes. Cada región diana puede comprender una secuencia capaz de hibridar solo con una única subsecuencia de un ácido nucleico diana dentro de una muestra de ácidos nucleicos. Cada región diana puede comprender uno o más secuencias aleatorias o una o más degeneradas, para permitir que la región diana hibride con más de una subsecuencia de un ácido nucleico diana. Cada región diana puede comprender al menos 5, al menos 10, al menos 15, al menos 20, al menos 25, al menos 50 o al menos 100 nucleótidos. Preferentemente, cada región diana comprende al menos 5 nucleótidos. Cada región diana puede comprender de 5 a 100 nucleótidos, de 5 a 10 nucleótidos, de 10 a 20 nucleótidos, de 20 a 30 nucleótidos, de 30 a 50 nucleótidos, de 50 a 100 nucleótidos, de 10 a 90 nucleótidos, de 20 a 80 nucleótidos, de 30 a 70 nucleótidos o de 50 a 60 nucleótidos. Preferentemente, cada región diana comprende de 30 a 70 nucleótidos. Preferentemente, cada región diana comprende desoxirribonucleótidos, opcionalmente, todos los nucleótidos en una región diana son desoxirribonucleótidos. Uno o más de los desoxirribonucleótidos puede ser un desoxirribonucleótido modificado (por ejemplo, un desoxirribonucleótido modificado con un resto de biotina o un nucleótido de desoxiuracilo). Cada región diana puede comprender una o más bases universales (por ejemplo, inosina), uno o más nucleótidos modificados y/o uno o más análogos de nucleótidos.

Las regiones diana se pueden usar para hibridar los oligonucleótidos adaptadores a subsecuencias de ácidos nucleicos diana, y luego se pueden usar como cebadores para una reacción de extensión de cebador o una reacción de amplificación, por ejemplo, una reacción en cadena de la polimerasa. Como alternativa, las regiones diana se pueden usar para ligar los oligonucleótidos adaptadores a subsecuencias de ácidos nucleicos diana. La región diana puede estar en el extremo 5' de un oligonucleótido adaptador. Tal región diana se puede fosforilar. Esto puede permitir que el extremo 5' de la región diana se ligue al extremo 3' de una subsecuencia de un ácido nucleico diana.

Los oligonucleótidos adaptadores pueden comprender una región enlazadora entre la región adaptadora y la región diana. La región enlazadora puede comprender uno o más nucleótidos contiguos que no están hibridados con la primera y la segunda moléculas de código de barras (es decir, la molécula de código de barras multimérica) y no son complementarios a las subsecuencias del ácido nucleico diana. El enlazador puede comprender de 1 a 100, de 5 a 75, de 10 a 50, de 15 a 30 o de 20 a 25 nucleótidos no complementarios. Preferentemente, el enlazador comprende de 15 a 30 nucleótidos no complementarios. El uso de una región enlazadora de este tipo mejora la eficacia de las reacciones de códigos de barras realizadas utilizando los kits descritos en el presente documento.

Cada uno de los componentes del kit puede adoptar cualquiera de las formas definidas en el presente documento.

El o los reactivos de códigos de barras multiméricos y los oligonucleótidos adaptadores se pueden proporcionar en el kit como componentes separados físicamente.

El kit puede comprender: (a) un reactivo de códigos de barras multimérico que comprende al menos 5, al menos 10, al menos 20, al menos 25, al menos 50, al menos 75 o al menos 100 moléculas de código de barras unidas entre sí, en donde cada molécula de código de barras es tal como se define en el presente documento; y (b) un oligonucleótido adaptador capaz de hibridar con cada molécula de código de barras, en donde cada oligonucleótido adaptador es tal como se define en el presente documento.

La Figura 2 muestra un kit que comprende un reactivo de código de barras multimérico y oligonucleótidos adaptadores para etiquetar un ácido nucleico diana. Con más detalle, el kit comprende las moléculas de código de barras primera (D1, E1 y F1) y segunda (D2, E2 y F2), cada una incorporando una región de código de barras (E1 y E2) y también una región adaptadora 5' (F1 y F2). Estas moléculas de código de barras primera y segunda están unidas entre sí, en esta realización mediante una secuencia de ácido nucleico de conexión (S).

El kit comprende además oligonucleótidos con códigos de barras primero (A1 y B1) y segundo (A2 y B2), cada uno de los cuales comprende una región de código de barras (B1 y B2), así como regiones 5' (A1 y A2). La región 5' de cada oligonucleótido con código de barras es complementaria y, por lo tanto, puede hibridar con, las regiones 3' de las moléculas de código de barras (D1 y D2). Las regiones de código de barras (B1 y B2) son complementarias y, por lo tanto, pueden hibridar con, las regiones de código de barras (E1 y E2) de las moléculas de código de barras.

El kit comprende además oligonucleótidos adaptadores primero (C1 y G1) y segundo (C2 y G2), en donde cada oligonucleótido adaptador comprende una región adaptadora (C1 y C2) que es complementaria y, por tanto, capaz de hibridar con, la región adaptadora de 5' de una molécula de código de barras (F1 y F2). Estos oligonucleótidos adaptadores se pueden sintetizar para incluir un grupo fosfato 5' terminal. Cada oligonucleótido adaptador también comprende una región diana (G1 y G2), que se puede usar para hibridar los oligonucleótidos adaptadores con código de barras (A1, B1, C1 y G1 y A2, B2, C2 y G2) para el direccionamiento a ácidos nucleicos, y luego se pueden usar como cebadores para una reacción de extensión de cebador o una reacción en cadena de la polimerasa.

El kit puede comprender una biblioteca de dos o más reactivos de código de barras multiméricos, en donde cada reactivo de código de barras multimérico es tal como se define en el presente documento, y los oligonucleótidos adaptadores para cada uno de los reactivos de códigos de barras multiméricos, en donde cada oligonucleótido adaptador es tal como se define en el presente documento. Las regiones de código de barras del primer y segundo oligonucleótidos con código de barras del primer reactivo de código de barras multimérico son diferentes a las regiones

de código de barras del primer y segundo oligonucleótidos con código de barras del segundo reactivo de código de barras multimérico.

El kit puede comprender una biblioteca que comprende al menos 5, al menos 10, al menos 20, al menos 25, al menos 50, al menos 75, al menos 100, al menos 250, al menos 500, al menos 10^3 , al menos 10^4 , al menos 10^5 , al menos 10^6 , al menos 10^7 , al menos 10^8 o al menos 10^9 reactivos de códigos de barras multiméricos tal como se define en el presente documento. Preferentemente, el kit comprende una biblioteca que comprende al menos 10 reactivos de códigos de barras multiméricos tal como se define en el presente documento. El kit puede comprender además oligonucleótidos adaptadores para cada uno de los reactivos de códigos de barras multiméricos, en donde cada oligonucleótido adaptador puede tomar la forma de cualquiera de los oligonucleótidos adaptadores definidos en el presente documento. Preferentemente, las regiones de código de barras de los oligonucleótidos con código de barras primero y segundo de cada reactivo de código de barras multimérico son diferentes a las regiones de código de barras de los oligonucleótidos con código de barras de al menos otros 9 reactivos de código de barras multiméricos en la biblioteca.

Las regiones de código de barras de los oligonucleótidos con código de barras primero y segundo de cada reactivo de código de barras multimérico pueden ser diferentes a las regiones de código de barras de los oligonucleótidos con código de barras de al menos 4, al menos 9, al menos 19, al menos 24, al menos 49, al menos 74, al menos 99, al menos 249, al menos 499, al menos 999 (es decir, 10^3-1), al menos 10^4-1 , al menos 10^5-1 , al menos 10^6-1 , al menos 10^7-1 , al menos 10^8-1 o al menos 10^9-1 otros reactivos de códigos de barras multiméricos en la biblioteca. Las regiones de código de barras de los oligonucleótidos con código de barras primero y segundo de cada reactivo de código de barras multimérico pueden ser diferentes a las regiones de código de barras de los oligonucleótidos con código de barras de todos los otros reactivos de código de barras multiméricos en la biblioteca. Preferentemente, las regiones de código de barras de los oligonucleótidos con código de barras primero y segundo de cada reactivo de código de barras multimérico son diferentes a las regiones de código de barras de los oligonucleótidos con código de barras de al menos otros 9 reactivos de código de barras multiméricos en la biblioteca.

Las regiones de código de barras de los oligonucleótidos con código de barras de cada reactivo de código de barras multimérico pueden ser diferentes a las regiones de código de barras de los oligonucleótidos con código de barras de al menos 4, al menos 9, al menos 19, al menos 24, al menos 49, al menos 74, al menos 99, al menos 249, al menos 499, al menos 999 (es decir, 10^3-1), al menos 10^4-1 , al menos 10^5-1 , al menos 10^6-1 , al menos 10^7-1 , al menos 10^8-1 o al menos 10^9-1 otros reactivos de códigos de barras multiméricos en la biblioteca. Las regiones de código de barras de los oligonucleótidos con código de barras de cada reactivo de código de barras multimérico pueden ser diferentes a las regiones de código de barras de los oligonucleótidos con código de barras de todos los otros reactivos de código de barras multiméricos en la biblioteca. Preferentemente, las regiones de código de barras de los oligonucleótidos con código de barras de cada reactivo de código de barras multimérico son diferentes a las regiones de código de barras de los oligonucleótidos con código de barras de al menos otros 9 reactivos de código de barras multiméricos en la biblioteca.

La divulgación proporciona un kit para etiquetar un ácido nucleico diana para secuenciar, en donde el kit comprende: a) una biblioteca de reactivos de códigos de barras multiméricos que comprenda al menos 10 reactivos de códigos de barras multiméricos, en donde cada reactivo de códigos de barras multimérico comprende: (i) la primera y segunda moléculas de código de barras comprendidas dentro de una (única) molécula de ácido nucleico, en donde cada una de las moléculas de código de barras comprende una secuencia de ácido nucleico que comprende, opcionalmente en la dirección de 5' a 3', una región adaptadora y una región de código de barras, y (ii) el primer y segundo oligonucleótidos con código de barras, en donde el primer oligonucleótido con código de barras comprende una región de código de barras complementaria e hibridada con la región de código de barras de la primera molécula de código de barras y en donde el segundo oligonucleótido con código de barras comprende una región de código de barras complementaria e hibridada con la región de código de barras de la segunda molécula de código de barras; y (b) los oligonucleótidos adaptadores primero y segundo para cada uno de los reactivos de códigos de barras multiméricos, en donde el primer oligonucleótido adaptador comprende, opcionalmente en la dirección de 5' a 3', una región adaptadora capaz de hibridar con la región adaptadora de la primera molécula de código de barras y una región diana capaz de hibridar o ligar a una primera subsecuencia del ácido nucleico diana, y en donde el segundo oligonucleótido adaptador comprende, opcionalmente en la dirección de 5' a 3', una región adaptadora capaz de hibridar con la región adaptadora de la segunda molécula de código de barras y una región diana capaz de hibridar o ligar con una segunda subsecuencia del ácido nucleico diana.

10. KITS QUE COMPRENDEN REACTIVOS DE CÓDIGO DE BARRAS MULTIMÉRICOS, OLIGONUCLEÓTIDOS ADAPTADORES Y CEBADORES DE EXTENSIÓN

La divulgación proporciona además un kit para etiquetar un ácido nucleico diana para secuenciar, en donde el kit comprende: (a) una molécula de código de barras multimérica que comprende una primera y una segunda moléculas de código de barras unidas entre sí, en donde cada una de las moléculas de código de barras comprende una secuencia de ácido nucleico que comprende, opcionalmente en la dirección de 5' a 3', una región adaptadora, una región de código de barras y una región de cebador; (b) el primer y segundo cebadores de extensión para la molécula de código de barras multimérica, en donde el primer cebador de extensión comprende una secuencia capaz de hibridar con la región de cebador de la primera molécula de código de barras, y en donde el segundo cebador de extensión

comprende una secuencia capaz de hibridar con la región de cebador de la segunda molécula de código de barras; y (c) primer y segundo oligonucleótidos adaptadores para la molécula de código de barras multimérica, en donde el primer oligonucleótido adaptador comprende, opcionalmente en la dirección de 5' a 3', una región adaptadora capaz de hibridar con la región adaptadora de la primera molécula de código de barras y una región diana capaz de hibridar o ligar a una primera subsecuencia del ácido nucleico diana, y en donde el segundo oligonucleótido adaptador comprende, opcionalmente en la dirección de 5' a 3', una región adaptadora capaz de hibridar con la región adaptadora de la segunda molécula de código de barras y una región diana capaz de hibridar o ligar con una segunda subsecuencia del ácido nucleico diana.

La divulgación proporciona además un kit para etiquetar un ácido nucleico diana para secuenciar, en donde el kit comprende: (a) una molécula de código de barras multimérica que comprende una primera y una segunda moléculas de código de barras unidas entre sí, en donde cada una de las moléculas de código de barras comprende una secuencia de ácido nucleico que comprende, opcionalmente en la dirección de 5' a 3', una región adaptadora, una región de código de barras y una región de cebador; (b) el primer y segundo cebadores de extensión para la molécula de código de barras multimérica, en donde el primer cebador de extensión comprende una secuencia capaz de hibridar con la región de cebador de la primera molécula de código de barras, y en donde el segundo cebador de extensión comprende una secuencia capaz de hibridar con la región de cebador de la segunda molécula de código de barras; y (c) primer y segundo oligonucleótidos adaptadores para la molécula de código de barras multimérica, en donde el primer oligonucleótido adaptador comprende una región adaptadora capaz de hibridar con la región adaptadora de la primera molécula de código de barras y capaz de ligarse a una primera subsecuencia del ácido nucleico diana, y en donde el segundo oligonucleótido adaptador comprende una región adaptadora capaz de hibridar a la región adaptadora de la segunda molécula de código de barras y capaz de ligarse a una segunda subsecuencia del ácido nucleico diana.

Cada oligonucleótido adaptador puede consistir esencialmente en o consistir en una región adaptadora.

Los componentes del kit pueden adoptar cualquiera de las formas descritas en el presente documento.

Preferentemente, el primer cebador de extensión comprende una secuencia que es complementaria y capaz de hibridar con la región de cebador de la primera molécula de código de barras y el segundo cebador de extensión comprende una secuencia que es complementaria y capaz de hibridar con la región de cebador de la segunda molécula de código de barras. La secuencia complementaria de cada cebador de extensión puede ser al menos 5, al menos 10, al menos 15, al menos 20, al menos 25, al menos 50 o al menos 100 nucleótidos contiguos.

Los cebadores de extensión primero y segundo pueden ser capaces de ser extendidos usando las regiones de código de barras de la primera y segunda moléculas de código de barras como moldes para producir oligonucleótidos con código de barras primero y segundo, en donde el primer oligonucleótido con código de barras comprende una secuencia complementaria con la región de código de barras de la primera molécula de código de barras y el segundo oligonucleótido con código de barras comprende una secuencia complementaria con la región de código de barras de la segunda molécula de código de barras.

Los cebadores de extensión primero y segundo pueden ser idénticos en secuencia. Como alternativa, los primeros y segundos cebadores de extensión pueden tener una secuencia diferente.

El primer y/o segundo cebadores de extensión pueden comprender, además, una o más regiones con secuencias de ácido nucleico que no son complementarias a la primera molécula de código de barras y la segunda molécula de código de barras, respectivamente. Opcionalmente, tal región no complementaria puede incluir un sitio de unión para uno o más cebadores de amplificación. Opcionalmente, tal región no complementaria puede colocarse dentro de la región 5' de la molécula. Opcionalmente, el primer y segundo cebadores de extensión pueden comprender un grupo fosfato 5' terminal capaz de ligarse a un extremo 3' de una molécula de ácido nucleico.

Los cebadores de extensión primero y/o segundo pueden comprender además una o más regiones de código de barras secundarias. Opcionalmente, una región de código de barras secundaria puede estar comprendida dentro de una región del cebador de extensión que no es complementaria a una molécula de código de barras. Opcionalmente, una región de código de barras secundaria puede estar comprendida dentro de una región del cebador de extensión que se encuentra entre una región 3' del cebador de extensión que es complementaria a una molécula de código de barras y una región 5' del cebador de extensión que comprende un sitio de unión para un cebador de amplificación.

Una región de código de barras secundaria puede comprender una secuencia de uno o más nucleótidos, en donde las secuencias de las regiones de código de barras secundarias del primer cebador de extensión y el segundo cebador de extensión son diferentes. Opcionalmente, dichos uno o más nucleótidos pueden comprender nucleótidos aleatorios o degenerados. Opcionalmente, dichos uno o más nucleótidos pueden comprender nucleótidos diferentes pero no aleatorios. Cualquier región secundaria de código de barras puede comprender al menos 2, al menos 3, al menos 5, al menos 10, al menos 15, al menos 20 o al menos 30 nucleótidos. Cualquier región de código de barras secundaria puede comprender una secuencia contigua de oligonucleótidos de código de barras, o puede comprender dos o más segmentos diferentes separados por al menos un nucleótido invariante o sin código de barras. Opcionalmente, cualquier región de código de barras secundaria puede comprender un identificador molecular único (IMU).

El kit puede comprender una biblioteca de dos o más moléculas de código de barras multiméricas, en donde cada molécula de código de barras multimérica es tal como se define en el presente documento, y un primer y segundo cebadores de extensión, y un primer y segundo oligonucleótidos adaptadores, para cada una de las moléculas de código de barras multiméricas. Los cebadores de extensión y los oligonucleótidos adaptadores pueden adoptar cualquiera de las formas descritas en el presente documento. Las regiones de código de barras de la primera y segunda moléculas de código de barras de la primera molécula de código de barras multimérica son diferentes a las regiones de código de barras de la primera y segunda moléculas de código de barras de la segunda molécula de código de barras multimérica.

El kit puede comprender una biblioteca que comprende al menos 5, al menos 10, al menos 20, al menos 25, al menos 50, al menos 75, al menos 100, al menos 250, al menos 500, al menos 10^3 , al menos 10^4 , al menos 10^5 , al menos 10^6 , al menos 10^7 , al menos 10^8 o al menos 10^9 moléculas de código de barras multiméricas tal como se define en el presente documento. Preferentemente, el kit comprende una biblioteca que comprende al menos 10 moléculas de código de barras multiméricas tal como se define en el presente documento. El kit puede comprender además cebadores de extensión y/u oligonucleótidos adaptadores para cada una de las moléculas de código de barras multiméricas. Los cebadores de extensión y los oligonucleótidos adaptadores pueden adoptar cualquiera de las formas descritas en el presente documento. Preferentemente, las regiones de código de barras de las moléculas de código de barras primera y segunda de cada molécula de código de barras son diferentes a las regiones de código de barras de las moléculas de código de barras de al menos otras 9 moléculas de código de barras multiméricas en la biblioteca.

Las regiones de código de barras de la primera y segunda moléculas de código de barras de cada molécula de código de barras multimérica pueden ser diferentes a las regiones de código de barras de las moléculas de código de barras de al menos 4, al menos 9, al menos 19, al menos 24, al menos 49, al menos 74, al menos 99, al menos 249, al menos 499, al menos 999 (es decir, 10^3-1), al menos 10^4-1 , al menos 10^5-1 , al menos 10^6-1 , al menos 10^7-1 , al menos 10^8-1 o al menos 10^9-1 otras moléculas de código de barras en la biblioteca. Las regiones de código de barras de las moléculas de código de barras primera y segunda de cada molécula de código de barras pueden ser diferentes a las regiones de código de barras de las moléculas de código de barras de todas las otras moléculas de código de barras multiméricas en la biblioteca. Preferentemente, las regiones de código de barras de las moléculas de código de barras primera y segunda de cada molécula de código de barras son diferentes a las regiones de código de barras de las moléculas de código de barras de al menos otras 9 moléculas de código de barras multiméricas en la biblioteca.

Las regiones de código de barras de las moléculas de código de barras de cada molécula de código de barras multimérica pueden ser diferentes a las regiones de código de barras de las moléculas de código de barras de al menos 4, al menos 9, al menos 19, al menos 24, al menos 49, al menos 74, al menos 99, al menos 249, al menos 499, al menos 999 (es decir, 10^3-1), al menos 10^4-1 , al menos 10^5-1 , al menos 10^6-1 , al menos 10^7-1 , al menos 10^8-1 o al menos 10^9-1 otras moléculas de código de barras en la biblioteca. Las regiones de códigos de barras de las moléculas de códigos de barras de cada molécula de códigos de barras multiméricas pueden ser diferentes a las regiones de códigos de barras de las moléculas de códigos de barras de todas las demás moléculas de códigos de barras multiméricas en la biblioteca. Preferentemente, las regiones de código de barras de las moléculas de código de barras de cada molécula de código de barras son diferentes a las regiones de código de barras de las moléculas de código de barras de al menos otras 9 moléculas de código de barras multiméricas en la biblioteca.

La divulgación proporciona además un kit para etiquetar un ácido nucleico diana para secuenciar, en donde el kit comprende: a) una biblioteca de moléculas de código de barras multiméricas que comprenda al menos 10 moléculas de código de barras multiméricas, cada molécula de código de barras multimérica que comprende una primera y segunda moléculas de código de barras comprendidas dentro de una (única) molécula de ácido nucleico, en donde cada una de las moléculas de código de barras comprende una secuencia de ácido nucleico que comprende, opcionalmente en la dirección de 5' a 3', una región adaptadora, una región de código de barras y una región de cebador, y en donde las regiones de código de barras de la primera y segunda moléculas de código de barras de cada molécula de código de barras multimérica son diferentes a las regiones de código de barras de al menos otras 9 moléculas de código de barras multiméricas en la biblioteca; (b) el primer y segundo cebadores de extensión para cada una de las moléculas de código de barras multiméricas, en donde el primer cebador de extensión comprende una secuencia capaz de hibridar con la región de cebador de la primera molécula de código de barras, y en donde el segundo cebador de extensión comprende una secuencia capaz de hibridar con la región de cebador de la segunda molécula de código de barras; y (c) primer y segundo oligonucleótidos adaptadores para cada una de las moléculas de código de barras multiméricas, en donde el primer oligonucleótido adaptador comprende, opcionalmente en la dirección de 5' a 3', una región adaptadora capaz de hibridar con la región adaptadora de la primera molécula de código de barras y una región diana capaz de hibridar o ligar a una primera subsecuencia del ácido nucleico diana, y en donde el segundo oligonucleótido adaptador comprende, opcionalmente en la dirección de 5' a 3', una región adaptadora capaz de hibridar con la región adaptadora de la segunda molécula de código de barras y una región diana capaz de hibridar o ligar con una segunda subsecuencia del ácido nucleico diana.

11. MÉTODOS DE PREPARACIÓN DE UNA MUESTRA DE ÁCIDO NUCLEICO PARA LA SECUENCIACIÓN

Los métodos de preparación de una muestra de ácido nucleico para secuenciar pueden comprender (i) poner en contacto la muestra de ácido nucleico con un reactivo de código de barras multimérico que comprende una primera y segunda regiones de código de barras unidas entre sí, en donde cada región de código de barras comprende una

secuencia de ácido nucleico, y (ii) adjuntar secuencias de código de barras a la primera y segunda subsecuencias de un ácido nucleico diana para producir una primera y segunda moléculas de ácido nucleico diana con códigos de barras diferentes, en donde la primera molécula de ácido nucleico diana con código de barras comprende la secuencia de ácido nucleico de la primera región de código de barras y la segunda molécula de ácido nucleico diana con código de barras comprende la secuencia de ácido nucleico de la segunda región de código de barras.

En los métodos en los que el reactivo de códigos de barras multimérico comprende un primer y un segundo oligonucleótidos con código de barras unidos entre sí, las secuencias de códigos de barras pueden adjuntarse a la primera y segunda subsecuencias del ácido nucleico diana mediante cualquiera de los métodos descritos en el presente documento.

El primer y segundo oligonucleótidos con código de barras se pueden ligar a la primera y segunda subsecuencia del ácido nucleico diana para producir la primera y segunda moléculas de ácido nucleico diana diferentes con código de barras. Opcionalmente, antes de la etapa de ligadura, el método comprende añadir secuencias de acoplamiento primera y segunda al ácido nucleico diana, en donde la primera y la segunda secuencias de acoplamiento son la primera y la segunda subsecuencias del ácido nucleico diana al que se ligan el primer y el segundo oligonucleótidos con código de barras.

El primer y segundo oligonucleótidos con código de barras pueden hibridar con la primera y segunda subsecuencia del ácido nucleico diana extendida para producir la primera y segunda moléculas de ácido nucleico diana diferentes con código de barras. Opcionalmente, antes de la etapa de hibridación, el método comprende añadir secuencias de acoplamiento primera y segunda al ácido nucleico diana, en donde la primera y la segunda secuencias de acoplamiento son la primera y la segunda subsecuencias del ácido nucleico diana con el que hibridan el primer y el segundo oligonucleótidos con código de barras.

El primer y segundo oligonucleótidos con código de barras se pueden hibridar en sus extremos 5' con la primera y segunda subsecuencias del ácido nucleico diana y los primeros y segundos cebadores diana pueden hibridar con la tercera y cuarta secuencias del ácido nucleico diana, respectivamente, en donde la tercera subsecuencia es 3' de la primera subsecuencia y en donde la cuarta subsecuencia es 3' de la segunda subsecuencia. El método comprende además extender el primer cebador diana usando el ácido nucleico diana como molde hasta que alcanza la primera subsecuencia para producir un primer cebador diana extendido, y extender el segundo cebador diana usando el ácido nucleico diana como molde hasta que alcanza la segunda subsecuencia para producir un segundo cebador diana extendido, y ligar el extremo 3' del primer cebador diana extendido al extremo 5' del primer oligonucleótido con código de barras para producir una primera molécula de ácido nucleico diana con código de barras, y ligar el extremo 3' del segundo cebador diana extendido al extremo 5' del segundo oligonucleótido con código de barras para producir una segunda molécula de ácido nucleico diana con código de barras, en donde la primera y la segunda moléculas de ácido nucleico diana con código de barras son diferentes y cada una comprende al menos un nucleótido sintetizado a partir del ácido nucleico diana como molde. Opcionalmente, antes de una o ambas etapas de hibridación, el método comprende adjuntar la primera y segunda, y/o la tercera y la cuarta, secuencias de acoplamiento al ácido nucleico diana, en donde la primera y segunda secuencias de acoplamiento son la primera y segunda subsecuencias del ácido nucleico diana con la que hibridan el primer y segundo oligonucleótidos con código de barras, y/o donde la tercera y cuarta secuencias de acoplamiento son la tercera y cuarta subsecuencias del ácido nucleico diana con el que hibridan el primer y segundo cebadores diana.

Como se describe en el presente documento, antes de hibridar o ligar una molécula de hibridación multimérica, la molécula de código de barras multimérica, oligonucleótido con código de barras, oligonucleótido adaptador o cebador diana a un ácido nucleico diana, se puede adjuntar una secuencia de acoplamiento al ácido nucleico diana. La molécula de hibridación multimérica, la molécula de código de barras multimérica, el oligonucleótido con código de barras, el oligonucleótido adaptador o el cebador diana pueden hibridar o ligarse entonces a la secuencia de acoplamiento.

Se puede añadir una secuencia de acoplamiento al extremo 5' o al extremo 3' de dos o más ácidos nucleicos diana de la muestra de ácido nucleico (por ejemplo, una muestra de ADN EPFF). En este método, las regiones diana (de los oligonucleótidos con código de barras) pueden comprender una secuencia que es complementaria a la secuencia de acoplamiento.

Una secuencia de acoplamiento puede estar comprendida dentro de un oligonucleótido de acoplamiento bicatenario o dentro de un oligonucleótido de acoplamiento monocatenario. Se puede añadir un oligonucleótido de acoplamiento al ácido nucleico diana mediante una reacción de ligadura bicatenaria o una reacción de ligadura monocatenaria. Un oligonucleótido de acoplamiento puede comprender una región 5' o 3' monocatenaria capaz de ligarse a un ácido nucleico diana y la secuencia de acoplamiento puede unirse al ácido nucleico diana mediante una reacción de ligadura monocatenaria.

Un oligonucleótido de acoplamiento puede comprender una región 5' o 3' roma, rebajada o sobresaliente capaz de ligarse a un ácido nucleico diana y la secuencia de acoplamiento puede adjuntarse al ácido nucleico diana en una reacción de ligadura bicatenaria.

Los extremos de un ácido nucleico diana se pueden convertir en extremos romos monocatenarios en una reacción de embotamiento, y el oligonucleótido de acoplamiento puede comprender un extremo bicatenario romo, y donde el oligonucleótido de acoplamiento se puede ligar al ácido nucleico diana en una reacción de ligadura de extremos romos.

- 5 Los extremos de un ácido nucleico diana se pueden convertir en extremos romos bicatenarios en una reacción de embotamiento, y luego convertirse en una forma con (a) salientes de adenosina 3' simples, y en donde el oligonucleótido de acoplamiento puede comprender un extremo bicatenario con un único saliente de timina en 3' capaz de hibridar con el único saliente de adenosina 3' del ácido nucleico diana, y en donde el oligonucleótido de acoplamiento se liga al ácido nucleico diana en una reacción de ligadura A/T bicatenaria.

- 10 El ácido nucleico diana puede ponerse en contacto con una enzima de restricción, en donde la enzima de restricción digiere el ácido nucleico diana en los sitios de restricción para crear (a) unión(es) de ligadura en el(los) sitio(s) de restricción, y en donde el oligonucleótido de acoplamiento comprende un extremo compatible con la unión de ligadura, y en donde el oligonucleótido de acoplamiento luego se liga al ácido nucleico diana en una reacción de ligadura bicatenaria.

- 15 Se puede añadir un oligonucleótido de acoplamiento mediante una etapa de reacción en cadena de la polimerasa o extensión del cebador.

Se puede añadir un oligonucleótido de acoplamiento mediante una etapa de reacción en cadena de la polimerasa o extensión del cebador, utilizando uno o más oligonucleótidos que comprenden un segmento de cebador que incluye una o más bases degeneradas.

- 20 Se puede añadir un oligonucleótido de acoplamiento mediante una etapa de reacción en cadena de la polimerasa o extensión del cebador, utilizando uno o más oligonucleótidos que comprenden además un segmento de cebador o hibridación específico para una secuencia de ácido nucleico diana particular.

- 25 Se puede añadir una secuencia de acoplamiento mediante una reacción de colocación de cola de polinucleótidos. Se puede añadir una secuencia de acoplamiento mediante una enzima transferasa terminal (por ejemplo, una enzima desoxinucleotidil transferasa terminal). Se puede adjuntar una secuencia de acoplamiento mediante una reacción de cola de polinucleótidos realizada con una enzima desoxinucleotidil transferasa terminal, y en donde la secuencia de acoplamiento comprende al menos dos nucleótidos contiguos de una secuencia homopolimérica.

Una secuencia de acoplamiento puede comprender una cola 3' homopolimérica (por ejemplo, una cola de poli(A)). Opcionalmente, en dichos métodos, las regiones diana (de los oligonucleótidos con código de barras) comprenden una cola 3' homopolimérica complementaria (por ejemplo, una cola de poli(T)).

- 30 Una secuencia de acoplamiento puede estar comprendida dentro de un transposoma sintético y puede adjuntarse mediante una reacción de transposición *in vitro*.

- 35 Se puede adjuntar una secuencia de acoplamiento a un ácido nucleico diana, y en donde un oligonucleótido con código de barras se añade al ácido nucleico diana mediante al menos una etapa de extensión del cebador o una etapa de reacción en cadena de la polimerasa, y en donde dicho oligonucleótido con código de barras comprende una región de al menos un nucleótido de longitud que es complementario a dicha secuencia de acoplamiento. Opcionalmente, esta región de complementariedad está en el extremo 3' del oligonucleótido del código de barras. Opcionalmente, esta región de complementariedad tiene al menos 2 nucleótidos de longitud, al menos 5 nucleótidos de longitud, al menos 10 nucleótidos de longitud, al menos 20 nucleótidos de longitud o al menos 50 nucleótidos de longitud.

- 40 En los métodos en los que se adjunta un oligonucleótido adaptador (por ejemplo, ligado o hibridado) a un ácido nucleico diana, la región adaptadora del oligonucleótido adaptador proporciona una secuencia de acoplamiento capaz de hibridar con la región adaptadora de una molécula de hibridación multimérica o una molécula de código de barras multimérica.

- 45 La divulgación proporciona un método para preparar una muestra de ácido nucleico para secuenciar que comprende las etapas de: (a) añadir una secuencia de acoplamiento a la primera y segunda subsecuencias de un ácido nucleico diana; (b) poner en contacto la muestra de ácido nucleico con un reactivo de código de barras multimérico que comprende la moléculas de código de barras primera y segunda unidas entre sí, en donde cada una de las moléculas de código de barras comprende una secuencia de ácido nucleico que comprende (en la dirección 5' a 3' o 3' a 5'), una región de código de barras y una región adaptadora; (c) hibridar la secuencia de acoplamiento de la primera subsecuencia con la región adaptadora de la primera molécula de código de barras, e hibridar la secuencia de acoplamiento de la segunda subsecuencia con la región adaptadora de la segunda molécula de código de barras; y
- 50 (d) adjuntar secuencias de códigos de barras a cada una de las al menos dos subsecuencias del ácido nucleico diana para producir una primera y una segunda moléculas de ácido nucleico diana con códigos de barras diferentes, en donde la primera molécula de ácido nucleico diana con código de barras comprende la secuencia de ácido nucleico de la región de código de barras de la primera molécula de código de barras y la segunda molécula de ácido nucleico diana con código de barras comprende la secuencia de ácido nucleico de la región de código de barras de la segunda molécula de código de barras.
- 55

En el método, cada una de las moléculas de código de barras puede comprender una secuencia de ácido nucleico que comprende, en la dirección 5' a 3', una región de código de barras y una región adaptadora, y la etapa (d) puede comprender extender la secuencia de acoplamiento de la primera subsecuencia del ácido nucleico diana usando la región de código de barras de la primera molécula de código de barras como molde para producir una primera molécula de ácido nucleico diana con código de barras, y extender la secuencia de acoplamiento de la segunda subsecuencia del ácido nucleico diana usando la región de código de barras de la segunda molécula de código de barras como molde para producir una segunda molécula de ácido nucleico diana con código de barras, en donde la primera molécula de ácido nucleico diana con código de barras comprende una secuencia complementaria a la región de código de barras de la primera molécula de código de barras y la segunda molécula de ácido nucleico diana con código de barras comprende una secuencia complementaria a la región de código de barras de la segunda molécula de código de barras.

En el método, cada una de las moléculas de código de barras puede comprender una secuencia de ácido nucleico que comprende, en la dirección 5' a 3', una región adaptadora y una región de código de barras, y la etapa (d) puede comprender (i) hibridar y extender un primer cebador de extensión usando la región de código de barras de la primera molécula de código de barras como molde para producir un primer oligonucleótido con código de barras, e hibridar y extender un segundo cebador de extensión utilizando la región del código de barras de la segunda molécula de código de barras como molde para producir un segundo oligonucleótido con código de barras, en donde el primer oligonucleótido con código de barras comprende una secuencia complementaria con la región de código de barras de la primera molécula de código de barras y el segundo oligonucleótido con código de barras comprende una secuencia complementaria con la región de código de barras de la segunda molécula de código de barras, (ii) ligar el extremo 3' del primer oligonucleótido con código de barras al extremo 5' de la secuencia de acoplamiento de la primera subsecuencia del ácido nucleico diana para producir una primera molécula de ácido nucleico diana con código de barras y ligar el extremo 3' del segundo oligonucleótido con código de barras al extremo 5' de la secuencia de acoplamiento de la segunda subsecuencia del ácido nucleico diana para producir una segunda molécula de ácido nucleico diana con código de barras.

En el método, cada una de las moléculas de código de barras puede comprender una secuencia de ácido nucleico que comprende, en la dirección 5' a 3', una región adaptadora, una región de código de barras y una región de cebador donde la etapa (d) comprende (i) hibridar un primer cebador de extensión con la región de cebador de la primera molécula de código de barras y extender el primer cebador de extensión usando la región de código de barras de la primera molécula de código de barras como molde para producir un primer oligonucleótido con código de barras, y unir un segundo cebador de extensión a la región de cebador de la segunda molécula de código de barras y extender el segundo cebador de extensión usando la región de código de barras de la segunda molécula de código de barras como molde para producir un segundo oligonucleótido con código de barras, en donde el primer oligonucleótido con código de barras comprende una secuencia complementaria con la región de código de barras de la primera molécula de código de barras y el segundo oligonucleótido con código de barras comprende una secuencia complementaria con la región de código de barras de la segunda molécula de código de barras, (ii) ligar el extremo 3' del primer oligonucleótido con código de barras al extremo 5' de la secuencia de acoplamiento de la primera subsecuencia del ácido nucleico diana para producir una primera molécula de ácido nucleico diana con código de barras y ligar el extremo 3' del segundo oligonucleótido con código de barras al extremo 5' de la secuencia de acoplamiento de la segunda subsecuencia del ácido nucleico diana para producir una segunda molécula de ácido nucleico diana con código de barras.

Los métodos para preparar una muestra de ácido nucleico para secuenciar se pueden usar para preparar una gama de diferentes muestras de ácido nucleico para secuenciar. Los ácidos nucleicos diana pueden ser moléculas de ADN (por ejemplo, moléculas de ADN genómico) o moléculas de ARN (por ejemplo, moléculas de ARNm). Los ácidos nucleicos diana pueden ser de cualquier muestra. Por ejemplo, una célula individual (o células), un tejido, un fluido corporal (por ejemplo, sangre, plasma y/o suero), una biopsia o una muestra embebida en parafina fijada con formalina (EPFF).

La muestra puede comprender al menos 10, al menos 100 o al menos 10^3 , al menos 10^4 , al menos 10^5 , al menos 10^6 , al menos 10^7 , al menos 10^8 o al menos 10^9 ácidos nucleicos diana.

El ácido nucleico diana puede ser una molécula de ácido nucleico (única) intacta de una célula o dos o más fragmentos colocalizados de una molécula de ácido nucleico de una célula. Tal como se usa en el presente documento, el término ácido nucleico diana se refiere a los ácidos nucleicos presentes dentro de las células y a copias o amplicones de los mismos. Por ejemplo, cuando el ácido nucleico diana es el ADN genómico, el término ácido nucleico diana significa ADN genómico presente en una célula y copias o amplicones del mismo, por ejemplo, moléculas de ADN que se pueden preparar a partir del ADN genómico mediante una reacción de extensión del cebador. Como ejemplo adicional, cuando el ácido nucleico diana es ARNm, el término ácido nucleico diana significa ARNm presente en la célula y copias o amplicones del mismo, por ejemplo, ADNc sintetizado a partir del ARNm por transcripción inversa.

El método puede comprender producir al menos 2, al menos 5, al menos 10, al menos 20, al menos 25, al menos 50, al menos 75, al menos 100, al menos 250, al menos 500, al menos 10^3 , al menos 10^4 , al menos 10^5 , al menos 10^6 , al menos 10^7 , al menos

10⁸ o al menos 10⁹ diferentes moléculas de ácido nucleico diana con códigos de barras. Preferentemente, el método comprende producir al menos 5 moléculas de ácido nucleico diana con códigos de barras diferentes.

Cada molécula de ácido nucleico diana con código de barras puede comprender al menos 1, al menos 5, al menos 10, al menos 25, al menos 50, al menos 100, al menos 250, al menos 500, al menos 1000, al menos 2000, al menos 5000, o al menos 10000 nucleótidos sintetizados a partir del ácido nucleico diana como molde. Preferentemente, cada molécula de ácido nucleico diana con código de barras comprende al menos 20 nucleótidos sintetizados a partir del ácido nucleico diana como molde.

Como alternativa, cada molécula de ácido nucleico diana con código de barras puede comprender al menos 5, al menos 10, al menos 25, al menos 50, al menos 100, al menos 250, al menos 500, al menos 1000, al menos 2000, al menos 5000, o al menos 10000 nucleótidos del ácido nucleico diana. Preferentemente, cada molécula de ácido nucleico diana con código de barras comprende al menos 5 nucleótidos del ácido nucleico diana.

Se puede añadir una secuencia de cebador universal a las moléculas de ácido nucleico diana con código de barras. Esta secuencia puede permitir la posterior amplificación de al menos 5, al menos 10, al menos 20, al menos 25, al menos 50, al menos 75, al menos 100, al menos 250, al menos 500, al menos 10³, al menos 10⁴, al menos 10⁵, al menos 10⁶, al menos 10⁷, al menos 10⁸, o al menos 10⁹ diferentes moléculas de ácido nucleico diana con código de barras utilizando un cebador directo y un cebador inverso.

El método puede comprender preparar dos o más muestras de ácido nucleico independientes para secuenciar, en donde cada muestra de ácido nucleico se prepara usando una biblioteca diferente de reactivos de códigos de barras multiméricos (o una biblioteca diferente de moléculas de códigos de barras multiméricos), y en donde las regiones de código de barras de cada biblioteca de reactivos de códigos de barras multiméricos (o moléculas de códigos de barras multiméricos) comprenden una secuencia que es diferente a las regiones de códigos de barras de las otras bibliotecas de reactivos de códigos de barras multiméricos (o moléculas de códigos de barras multiméricos). Tras la preparación por separado de cada una de las muestras para su secuenciación, las moléculas de ácido nucleico diana con código de barras preparadas a partir de las diferentes muestras pueden agruparse y secuenciarse juntas. La secuencia de lectura generada para cada molécula de ácido nucleico diana con código de barras se puede usar para identificar la biblioteca de reactivos de códigos de barras multiméricos (o moléculas de código de barras multiméricas) que se usó en su preparación y, por lo tanto, para identificar la muestra de ácido nucleico a partir de la cual se preparó.

En cualquier método para preparar una muestra de ácido nucleico para secuenciar, las moléculas de ácido nucleico diana pueden estar presentes en concentraciones particulares dentro de la muestra de ácido nucleico, por ejemplo, a concentraciones de al menos 100 nanomolar, al menos 10 nanomolar, al menos 1 nanomolar, al menos 100 picomolar, al menos 10 picomolar, al menos 1 picomolar, al menos 100 femtomolar, al menos 10 femtomolar o al menos 1 femtomolar. Las concentraciones pueden ser de 1 picomolar a 100 nanomolar, 10 picomolar a 10 nanomolar, o 100 picomolar a 1 nanomolar. Preferentemente, las concentraciones son de 10 picomolar a 1 nanomolar.

En cualquier método para preparar una muestra de ácido nucleico para secuenciar, los reactivos de códigos de barras multiméricos pueden estar presentes en concentraciones particulares dentro de la muestra de ácido nucleico, por ejemplo, a concentraciones de al menos 100 nanomolar, al menos 10 nanomolar, al menos 1 nanomolar, al menos 100 picomolar, al menos 10 picomolar, al menos 1 picomolar, al menos 100 femtomolar, al menos 10 femtomolar o al menos 1 femtomolar. Las concentraciones pueden ser de 1 picomolar a 100 nanomolar, 10 picomolar a 10 nanomolar, o 100 picomolar a 1 nanomolar. Preferentemente, las concentraciones son de 1 picomolar a 100 picomolar.

En cualquier método para preparar una muestra de ácido nucleico para secuenciar, las moléculas de código de barras multiméricas pueden estar presentes en concentraciones particulares dentro de la muestra de ácido nucleico, por ejemplo, a concentraciones de al menos 100 nanomolar, al menos 10 nanomolar, al menos 1 nanomolar, al menos 100 picomolar, al menos 10 picomolar, al menos 1 picomolar, al menos 100 femtomolar, al menos 10 femtomolar o al menos 1 femtomolar. Las concentraciones pueden ser de 1 picomolar a 100 nanomolar, 10 picomolar a 10 nanomolar, o 100 picomolar a 1 nanomolar. Preferentemente, las concentraciones son de 1 picomolar a 100 picomolar.

En cualquier método para preparar una muestra de ácido nucleico para secuenciar, los oligonucleótidos con código de barras pueden estar presentes en concentraciones particulares dentro de la muestra de ácido nucleico, por ejemplo, a concentraciones de al menos 100 nanomolar, al menos 10 nanomolar, al menos 1 nanomolar, al menos 100 picomolar, al menos 10 picomolar, al menos 1 picomolar, al menos 100 femtomolar, al menos 10 femtomolar o al menos 1 femtomolar. Las concentraciones pueden ser de 1 picomolar a 100 nanomolar, 10 picomolar a 10 nanomolar, o 100 picomolar a 1 nanomolar. Preferentemente, las concentraciones son de 100 picomolar a 100 nanomolar.

12. MÉTODOS DE PREPARACIÓN DE UNA MUESTRA DE ÁCIDO NUCLEICO PARA SECUENCIAR UTILIZANDO REACTIVOS DE CÓDIGO DE BARRAS MULTIMÉRICOS

La divulgación proporciona un método para preparar una muestra de ácido nucleico para secuenciar, en donde el método comprende las etapas de: poner en contacto la muestra de ácido nucleico con un reactivo de código de barras multimérico tal como se define en el presente documento; hibridar la región diana del primer oligonucleótido con código de barras con una primera subsecuencia de un ácido nucleico diana, e hibridar la región diana del segundo oligonucleótido con código de barras con una segunda subsecuencia del ácido nucleico diana; y extender el primer y

segundo oligonucleótidos con código de barras para producir una primera y segunda moléculas de ácido nucleico diana con códigos de barras diferentes, en donde cada una de las moléculas de ácido nucleico diana con código de barras comprende al menos un nucleótido sintetizado a partir del ácido nucleico diana como molde.

- 5 En cualquier método para preparar una muestra de ácido nucleico para secuenciar, ya sea las moléculas de ácido nucleico dentro de la muestra de ácido nucleico y/o los reactivos de códigos de barras multiméricos, pueden estar presentes en concentraciones particulares dentro del volumen de la solución, por ejemplo, a concentraciones de al menos 100 nanomolar, al menos 10 nanomolar, al menos 1 nanomolar, al menos 100 picomolar, al menos 10 picomolar, o al menos 1 picomolar. Las concentraciones pueden ser de 1 picomolar a 100 nanomolar, 10 picomolar a 10 nanomolar, o 100 picomolar a 1 nanomolar. También se pueden usar concentraciones alternativas más altas o más bajas.

- 10 El método de preparación de una muestra de ácido nucleico para secuenciar puede comprender poner en contacto la muestra de ácido nucleico con una biblioteca de reactivos de códigos de barras multiméricos tal como se define en el presente documento, y en donde: los oligonucleótidos con código de barras del primer reactivo de código de barras multimérico hibridan con subsecuencias de un primer ácido nucleico diana y se producen una primera y una segunda moléculas de ácido nucleico diana con códigos de barras diferentes, en donde cada molécula de ácido nucleico diana con código de barras comprende al menos un nucleótido sintetizado a partir del primer ácido nucleico diana como molde; y los oligonucleótidos con código de barras del segundo reactivo de código de barras multimérico hibridan con subsecuencias de un segundo ácido nucleico diana y se producen una primera y segunda moléculas de ácido nucleico diana con códigos de barras diferentes, en donde cada molécula de ácido nucleico diana con código de barras comprende al menos un nucleótido sintetizado a partir del segundo ácido nucleico diana como molde.

En el método, los oligonucleótidos con código de barras pueden aislarse de la muestra de ácido nucleico después de la hibridación con las subsecuencias del ácido nucleico diana y antes de que se produzcan las moléculas de ácido nucleico diana con código de barras. Opcionalmente, los oligonucleótidos con código de barras se aíslan por captura en un soporte sólido a través de una interacción de estreptavidina-biotina.

- 25 Adicionalmente o como alternativa, las moléculas de ácido nucleico diana con código de barras se pueden aislar de la muestra de ácido nucleico. Opcionalmente, las moléculas de ácido nucleico diana con código de barras se aíslan mediante captura en un soporte sólido a través de una interacción de estreptavidina-biotina.

La etapa de extender los oligonucleótidos con código de barras se puede realizar mientras los oligonucleótidos con código de barras hibridan con las moléculas del código de barras.

- 30 La Figura 3 muestra un método para preparar una muestra de ácido nucleico para secuenciar, en la que un reactivo de código de barras multimérico definido en el presente documento (por ejemplo, tal como se ilustra en la Figura 1) se usa para etiquetar y extender dos o más subsecuencias de ácido nucleico en una muestra de ácido nucleico. En este método, se sintetiza un reactivo de códigos de barras multimérico que incorpora al menos un primer (A1, B1, C1 y G1) y un segundo (A2, B2, C2 y G2) oligonucleótido con código de barras, cada uno de los cuales comprende una región de código de barras (B1 y B2) y una región diana (G1 y G2 respectivamente).

- 35 Una muestra de ácido nucleico que comprende un ácido nucleico diana se pone en contacto o se mezcla con el reactivo de código de barras multimérico, y se permite que las regiones diana (G1 y G2) de dos o más oligonucleótidos con código de barras hibriden con dos o más subsecuencias correspondientes dentro del ácido nucleico diana (H1 y H2). Tras la etapa de hibridación, el primer y segundo oligonucleótidos con código de barras se extienden (por ejemplo, con las regiones diana que sirven como cebadores para una polimerasa) en la secuencia del ácido nucleico diana, de manera que al menos un nucleótido de una subsecuencia se incorpore en el extremo 3' extendido de cada uno de los oligonucleótidos con código de barras. Este método crea moléculas de ácido nucleico diana con código de barras, en donde dos o más subsecuencias del ácido nucleico diana están marcadas por un oligonucleótido con código de barras.

- 40 Como alternativa, el método puede comprender además la etapa de disociar los oligonucleótidos con código de barras de las moléculas del código de barras antes de hibridar las regiones diana de los oligonucleótidos con código de barras con subsecuencias del ácido nucleico diana.

- 45 La Figura 4 muestra un método para preparar una muestra de ácido nucleico para secuenciar, en donde un reactivo de código de barras multimérico descrito en el presente documento (por ejemplo, tal como se ilustra en la Figura 1) se usa para etiquetar y extender dos o más subsecuencias de ácido nucleico en una muestra de ácido nucleico, pero en donde los oligonucleótidos con código de barras del reactivo de código de barras multimérico se disocian de las moléculas de código de barras antes de la hibridación (y la extensión) de las secuencias de ácido nucleico diana. En este método, se sintetiza un reactivo de códigos de barras multimérico que incorpora al menos un primer (A1, B1, C1 y G1) y un segundo (A2, B2, C2 y G2) oligonucleótido con código de barras, cada uno de los cuales comprende una región de código de barras (B1 y B2) y una región diana (G1 y G2).

- 50 Una muestra de ácido nucleico que comprende un ácido nucleico diana se pone en contacto con el reactivo de código de barras multimérico, y luego los oligonucleótidos con código de barras se disocian de las moléculas de código de barras. Esta etapa se puede lograr, por ejemplo, mediante la exposición del reactivo a una temperatura elevada (por ejemplo, una temperatura de al menos 35 °C, al menos 40 °C, al menos 45 °C, al menos 50 °C, al menos 55 °C, al

menos 60 °C, al menos 65 °C, al menos 70 °C, al menos 75 °C, al menos 80 °C, al menos 85 °C o al menos 90 °C) o mediante un desnaturalizante químico, o una combinación de los mismos. Esta etapa también puede desnaturalizar los ácidos nucleicos bicatenarios dentro de la propia muestra. A continuación, se puede permitir que los oligonucleótidos con código de barras se difundan durante un cierto período de tiempo (por ejemplo, al menos 5 segundos, al menos 15 segundos, al menos 30 segundos, al menos 60 segundos, al menos 2 minutos, al menos 5 minutos, al menos 15 minutos, al menos 30 minutos, o al menos 60 minutos) (y, en consecuencia, para difundir una cierta distancia física dentro de la muestra).

Las condiciones de la mezcla de reactivo-muestra pueden cambiarse para permitir que las regiones diana (G1 y G2) de dos o más oligonucleótidos con código de barras hibriden con dos o más subsecuencias correspondientes dentro del ácido nucleico diana (H1 y H2). Esto podría comprender, por ejemplo, bajar la temperatura de la solución para permitir la hibridación (por ejemplo, bajar la temperatura a menos de 90 °C, menos de 85 °C, menos de 70 °C, menos de 65 °C, menos de 60 °C, menos de 55 °C, menos de 50 °C, menos de 45 °C, menos de 40 °C, menos de 35 °C, menos de 30 °C, menos de 25 °C, o menos de 20 °C). Después de esta etapa de hibridación (o por ejemplo, después de una etapa de purificación/preparación), el primer y segundo oligonucleótidos con código de barras se extienden (por ejemplo, con las regiones diana que sirven como cebadores para una polimerasa) en la secuencia del ácido nucleico diana, de manera que al menos un nucleótido de una subsecuencia se incorpore en el extremo 3' extendido de cada uno de los oligonucleótidos con código de barras.

Este método crea moléculas de ácido nucleico diana con código de barras en donde dos o más subsecuencias de la muestra de ácido nucleico están etiquetadas por un oligonucleótido con código de barras. Además, la etapa de disociar los oligonucleótidos con código de barras y permitir que se difundan a través de la muestra tiene ventajas para tipos particulares de muestras. Por ejemplo, muestras de ácidos nucleicos reticulados (por ejemplo, embebidas en parafina, fijadas con formalina (EPFF)) pueden ser susceptibles a la difusión de oligonucleótidos con código de barras individual relativamente pequeños. Este método puede permitir el etiquetado de muestras de ácido nucleico con poca accesibilidad (por ejemplo, muestras EPFF) u otras propiedades biofísicas, por ejemplo, cuando las subsecuencias de ácido nucleico diana están físicamente lejos unas de otras.

Se puede añadir una secuencia de cebador universal a las moléculas de ácido nucleico diana con código de barras. Esta secuencia puede permitir la posterior amplificación de al menos 5, al menos 10, al menos 20, al menos 25, al menos 50, al menos 75, al menos 100, al menos 250, al menos 500, al menos 10³, al menos 10⁴, al menos 10⁵, al menos 10⁶, al menos 10⁷, al menos 10⁸, o al menos 10⁹ diferentes moléculas de ácido nucleico diana con código de barras utilizando un cebador directo y un cebador inverso.

Antes de poner en contacto la muestra de ácido nucleico con un reactivo de códigos de barras multimérico, o una biblioteca de reactivos de códigos de barras multiméricos, tal como se define en el presente documento, se puede añadir una secuencia de acoplamiento al extremo 5' o al extremo 3' de dos o más ácidos nucleicos diana de la muestra de ácido nucleico (por ejemplo, una muestra de ADN EPFF). En este método, las regiones diana pueden comprender una secuencia que es complementaria a la secuencia de acoplamiento. La secuencia de acoplamiento puede comprender una cola 3' homopolimérica (por ejemplo, una cola de poli(A)). La secuencia de acoplamiento puede añadirse mediante una enzima transferasa terminal. En el método en el que la secuencia de acoplamiento comprende una cola de poli(A), las regiones diana pueden comprender una secuencia de poli(T). Dichas secuencias de acoplamiento se pueden agregar después de una incubación a alta temperatura de la muestra de ácido nucleico, para desnaturalizar los ácidos nucleicos contenidos en el mismo antes de añadir una secuencia de acoplamiento.

Como alternativa, se podría añadir una secuencia de acoplamiento mediante la digestión de una muestra de ácido nucleico diana (por ejemplo, una muestra de ADN EPFF) con una enzima de restricción, en cuyo caso, una secuencia de acoplamiento puede estar compuesta por uno o más nucleótidos de una secuencia de reconocimiento de enzima de restricción. En este caso, una secuencia de acoplamiento puede ser al menos parcialmente bicatenaria y puede comprender una secuencia de ADN bicatenario de extremos romos, o una secuencia con una región saliente 5' de 1 o más nucleótidos, o una secuencia con una región saliente 3' de 1 o más nucleótidos. En estos casos, las regiones diana en reactivos de códigos de barras multiméricos pueden entonces comprender secuencias que son bicatenarias y de extremos romos (y por lo tanto capaces de ligarse a productos de digestión de restricción de extremos romos), o las regiones diana pueden contener secuencias salientes 5' o 3' de 1 o más nucleótidos, que los hace cohesivos (y por lo tanto capaces de hibridar y ligar) frente a dichos productos de digestión de restricción.

El método puede comprender preparar dos o más muestras de ácido nucleico independientes para secuenciar, en donde cada muestra de ácido nucleico se prepara usando una biblioteca diferente de reactivos de códigos de barras multiméricos (o una biblioteca diferente de moléculas de códigos de barras multiméricos), y en donde las regiones de código de barras de cada biblioteca de reactivos de códigos de barras multiméricos (o moléculas de códigos de barras multiméricos) comprenden una secuencia que es diferente a las regiones de códigos de barras de las otras bibliotecas de reactivos de códigos de barras multiméricos (o moléculas de códigos de barras multiméricos). Tras la preparación por separado de cada una de las muestras para su secuenciación, las moléculas de ácido nucleico diana con código de barras preparadas a partir de las diferentes muestras pueden agruparse y secuenciarse juntas. La secuencia de lectura generada para cada molécula de ácido nucleico diana con código de barras se puede usar para identificar la biblioteca de reactivos de códigos de barras multiméricos (o moléculas de código de barras multiméricas) que se usó en su preparación y, por lo tanto, para identificar la muestra de ácido nucleico a partir de la cual se preparó.

La divulgación proporciona un método para preparar una muestra de ácido nucleico para secuenciar, en donde el método comprende las etapas de: (a) poner en contacto la muestra de ácido nucleico con un reactivo de código de barras multimérico, en donde cada oligonucleótido con código de barras comprende en la dirección 5' a 3' una región diana y una región de código de barras, y un primer y segundo cebadores diana; (b) hibridar la región diana del primer oligonucleótido con código de barras con una primera subsecuencia de un ácido nucleico diana e hibridar la región diana del segundo oligonucleótido con código de barras con una segunda subsecuencia del ácido nucleico diana; (c) hibridar el primer cebador diana con una tercera subsecuencia del ácido nucleico diana, en donde la tercera subsecuencia es 3' de la primera subsecuencia, e hibridar el segundo cebador diana con una cuarta subsecuencia del ácido nucleico diana, en donde la cuarta subsecuencia es 3' de la segunda subsecuencia; (d) extender el primer cebador diana usando el ácido nucleico diana como molde hasta que alcance la primera subsecuencia para producir un primer cebador diana extendido, y extender el segundo cebador diana usando el ácido nucleico diana como molde hasta que alcance la segunda subsecuencia para producir un segundo cebador diana extendido; y (e) ligar el extremo 3' del primer cebador diana extendido al extremo 5' del primer oligonucleótido con código de barras para producir una primera molécula de ácido nucleico diana con código de barras, y ligar el extremo 3' del segundo cebador diana extendido al extremo 5' del segundo oligonucleótido con código de barras para producir una segunda molécula de ácido nucleico diana con código de barras, en donde la primera y la segunda moléculas de ácido nucleico diana con código de barras son diferentes, y en donde cada una de las moléculas de ácido nucleico diana con código de barras comprende al menos un nucleótido sintetizado a partir del ácido nucleico diana como molde.

En el método, las etapas (b) y (c) se pueden realizar al mismo tiempo.

13. MÉTODOS PARA PREPARAR UNA MUESTRA DE ÁCIDO NUCLEICO PARA SECUENCIAR UTILIZANDO REACTIVOS DE CÓDIGO DE BARRAS MULTIMÉRICOS Y OLIGONUCLEÓTIDOS ADAPTADORES

Los métodos proporcionados a continuación se pueden realizar con cualquiera de los kits definidos en el presente documento.

La divulgación proporciona además un método para preparar una muestra de ácido nucleico para secuenciar, en donde el método comprende las etapas de: (a) poner en contacto la muestra de ácido nucleico con un primer y segundo oligonucleótido adaptador tal como se define en el presente documento; (b) hibridar o ligar el primer oligonucleótido adaptador a una primera subsecuencia de un ácido nucleico diana, e hibridar o ligar el segundo oligonucleótido adaptador a una segunda subsecuencia del ácido nucleico diana; (c) poner en contacto la muestra de ácido nucleico con un reactivo de código de barras multimérico tal como se define en el presente documento; (d) hibridar la región adaptadora del primer oligonucleótido adaptador con la región adaptadora de la primera molécula de código de barras e hibridar la región adaptadora del segundo oligonucleótido adaptador con la región adaptadora de la segunda molécula de código de barras; y (e) ligar el extremo 3' del primer oligonucleótido con código de barras al extremo 5' del primer oligonucleótido adaptador para producir un primer oligonucleótido adaptador con código de barras y ligar el extremo 3' del segundo oligonucleótido con código de barras al extremo 5' del segundo oligonucleótido adaptador para producir un segundo oligonucleótido adaptador con código de barras.

La divulgación proporciona además un método para preparar una muestra de ácido nucleico para secuenciar, en donde el método comprende las etapas de: (a) poner en contacto la muestra de ácido nucleico con un primer y segundo oligonucleótido adaptador tal como se define en el presente documento; (b) el primer oligonucleótido adaptador a una primera subsecuencia de un ácido nucleico diana, y ligar el segundo oligonucleótido adaptador a una segunda subsecuencia del ácido nucleico diana; (c) poner en contacto la muestra de ácido nucleico con un reactivo de código de barras multimérico tal como se define en el presente documento; (d) hibridar la región adaptadora del primer oligonucleótido adaptador con la región adaptadora de la primera molécula de código de barras e hibridar la región adaptadora del segundo oligonucleótido adaptador con la región adaptadora de la segunda molécula de código de barras; y (e) extender el primer oligonucleótido adaptador usando la región de código de barras de la primera molécula de código de barras como molde para producir una primera molécula de ácido nucleico diana con código de barras, y extender el segundo oligonucleótido adaptador usando la región de código de barras de la segunda molécula de código de barras como molde para producir una segunda molécula de ácido nucleico diana con código de barras, en donde la primera molécula de ácido nucleico diana con código de barras comprende una secuencia complementaria a la región de código de barras de la primera molécula de código de barras y la segunda molécula de ácido nucleico diana con código de barras comprende una secuencia complementaria a la región de código de barras de la segunda molécula de código de barras.

La divulgación proporciona además un método para preparar una muestra de ácido nucleico para secuenciar, en donde el método comprende las etapas de: (a) poner en contacto la muestra de ácido nucleico con un primer y segundo oligonucleótido adaptador tal como se define en el presente documento; (b) hibridar la región diana del primer oligonucleótido adaptador con una primera subsecuencia de un ácido nucleico diana e hibridar la región diana del segundo oligonucleótido adaptador con una segunda subsecuencia del ácido nucleico diana; (c) poner en contacto la muestra de ácido nucleico con un reactivo de código de barras multimérico tal como se define en el presente documento; (d) hibridar la región adaptadora del primer oligonucleótido adaptador con la región adaptadora de la primera molécula de código de barras e hibridar la región adaptadora del segundo oligonucleótido adaptador con la región adaptadora de la segunda molécula de código de barras; y (e) ligar el extremo 3' del primer oligonucleótido con código de barras al extremo 5' del primer oligonucleótido adaptador para producir un primer oligonucleótido adaptador

con código de barras y ligar el extremo 3' del segundo oligonucleótido con código de barras al extremo 5' del segundo oligonucleótido adaptador para producir un segundo oligonucleótido adaptador con código de barras.

En el método, los oligonucleótidos adaptadores de código de barras primero y segundo pueden extenderse para producir moléculas de ácido nucleico diana con código de barras diferentes, cada una de las cuales comprende al menos un nucleótido sintetizado a partir del ácido nucleico diana como molde.

Como alternativa, los oligonucleótidos adaptadores primero y segundo pueden extenderse para producir las moléculas de ácido nucleico diana primera y segunda diferentes, cada una de las cuales comprende al menos un nucleótido sintetizado a partir del ácido nucleico diana como molde. En este método, la etapa (f) produce una primera molécula de ácido nucleico diana con código de barras (es decir, el primer oligonucleótido con código de barras ligado al primer oligonucleótido adaptador extendido) y una segunda molécula de ácido nucleico diana con código de barras (es decir, el segundo oligonucleótido con código de barras ligado al segundo oligonucleótido adaptador extendido).

La etapa de extender los oligonucleótidos adaptadores se puede realizar antes de etapa (c), antes de la etapa (d) y/o antes de la etapa (e), y el primer y segundo oligonucleótidos adaptadores pueden permanecer hibridados con la primera y segunda moléculas de código de barras hasta después de la etapa (e).

El método se puede realizar usando una biblioteca de reactivos de códigos de barras multiméricos tal como se define en el presente documento y un oligonucleótido adaptador tal como se define en el presente documento para cada uno de los reactivos de códigos de barras multiméricos. Preferentemente, los oligonucleótidos adaptadores con código de barras del primer reactivo de código de barras multimérico hibridan con subsecuencias de un primer ácido nucleico diana y se producen una primera y una segunda moléculas de ácido nucleico diana con códigos de barras diferentes, en donde cada molécula de ácido nucleico diana con código de barras comprende al menos un nucleótido sintetizado a partir del primer ácido nucleico diana como molde; y los oligonucleótidos adaptadores con código de barras del segundo reactivo de código de barras multimérico hibridan con subsecuencias de un segundo ácido nucleico diana y se producen una primera y segunda moléculas de ácido nucleico diana con códigos de barras diferentes, en donde cada molécula de ácido nucleico diana con código de barras comprende al menos un nucleótido sintetizado a partir del segundo ácido nucleico diana como molde.

El método se puede realizar usando una biblioteca de reactivos de códigos de barras multiméricos tal como se define en el presente documento y un oligonucleótido adaptador tal como se define en el presente documento para cada uno de los reactivos de códigos de barras multiméricos. Preferentemente, los oligonucleótidos adaptadores del primer reactivo de código de barras multimérico hibridan con subsecuencias de un primer ácido nucleico diana y se producen una primera y una segunda moléculas de ácido nucleico diana diferentes, en donde cada molécula de ácido nucleico diana comprende al menos un nucleótido sintetizado a partir del primer ácido nucleico diana como molde; y los oligonucleótidos adaptadores del segundo reactivo de código de barras multimérico hibridan con subsecuencias de un segundo ácido nucleico diana y se producen una primera y segunda moléculas de ácido nucleico diana diferentes, en donde cada molécula de ácido nucleico diana comprende al menos un nucleótido sintetizado a partir del segundo ácido nucleico diana como molde.

Los oligonucleótidos adaptadores de códigos de barras se pueden aislar de la muestra de ácido nucleico después de la hibridación con las subsecuencias del ácido nucleico diana y antes de que se produzcan las moléculas de ácido nucleico diana con código de barras. Opcionalmente, los oligonucleótidos adaptadores con código de barras se aíslan por captura en un soporte sólido a través de una interacción de estreptavidina-biotina.

Las moléculas de ácido nucleico diana con código de barras se pueden aislar de la muestra de ácido nucleico. Opcionalmente, las moléculas de ácido nucleico diana con código de barras se aíslan mediante captura en un soporte sólido a través de una interacción de estreptavidina-biotina.

La Figura 5 muestra un método para preparar una muestra de ácido nucleico para secuenciar usando un reactivo de código de barras multimérico. En el método, los oligonucleótidos adaptadores primero (C1 y G1) y segundo (C2 y G2) hibridan con un ácido nucleico diana en la muestra de ácido nucleico y luego se usan en una reacción de extensión del cebador. Cada oligonucleótido adaptador está compuesto por una región adaptadora (C1 y C2) que es complementaria y, por lo tanto, capaz de hibridar con, la región adaptadora de 5' de una molécula de código de barras (F1 y F2). Cada oligonucleótido adaptador también está compuesto por una región diana (G1 y G2), que puede usarse para hibridar los oligonucleótidos con código de barras con ácidos nucleicos diana, y luego puede usarse como cebadores para una reacción de extensión del cebador o una reacción en cadena de la polimerasa. Estos oligonucleótidos adaptadores se pueden sintetizar para incluir un grupo fosfato 5' terminal.

Los oligonucleótidos adaptadores, cada uno de los cuales se ha extendido para incluir la secuencia del ácido nucleico diana, luego se ponen en contacto con un reactivo de código de barras multimérico que comprende una primera (D1, E1 y F1) y una segunda (D2, E2 y F2) molécula de código de barras, así como el primer (A1 y B1) y el segundo (A2 y B2) oligonucleótidos con código de barras, cada uno de los cuales comprende una región de código de barras (B1 y B2), así como regiones 5' (A1 y A2). La primera y segunda moléculas de código de barras, cada una comprende una región de código de barras (E1 y E2), una región adaptadora (F1 y F2) y una región 3' (D1 y D2), y están unidas entre sí, en esta realización mediante una secuencia de ácido nucleico de conexión (S). Después de poner en contacto la

muestra de ácido nucleico extendido con cebador con un reactivo de código de barras multimérico, las regiones adaptadoras 5' (C1 y C2) de cada oligonucleótido adaptador son capaces de hibridar con una "unión de ligadura" adyacente al extremo 3' de cada oligonucleótido con código de barras (J1 y J2). El extremo 5' de los oligonucleótidos adaptadores extendidos se ligan luego al extremo 3' de los oligonucleótidos con código de barras dentro del reactivo de código de barras multimérico, creando un par de bases ligadas (K1 y K2) en donde anteriormente se encontraba la unión de ligadura. Posteriormente, la solución puede procesarse adicionalmente o amplificarse y usarse en una reacción de secuenciación.

Este método, como los métodos ilustrados en las Figuras 3 y 4, crea moléculas de ácido nucleico diana con código de barras, en el que dos o más subsecuencias de la muestra de ácido nucleico están etiquetadas por un oligonucleótido con código de barras. En este método, no es necesario que esté presente un reactivo de código de barras multimérico para la etapa de hibridación de regiones diana con subsecuencias del ácido nucleico diana, o la etapa de extender las regiones diana hibridadas usando una polimerasa. Esta característica puede tener ventajas en ciertas aplicaciones, por ejemplo, en donde un gran número de secuencias diana son de interés y las regiones diana pueden hibridar más rápidamente con los ácidos nucleicos diana cuando no están restringidas molecularmente por un reactivo de código de barras multimérico.

14. MÉTODOS DE PREPARACIÓN DE UNA MUESTRA DE ÁCIDO NUCLEICO PARA SECUENCIAR UTILIZANDO REACTIVOS DE CÓDIGO DE BARRAS MULTIMÉRICOS, OLIGONUCLEÓTIDOS ADAPTADORES Y CEBADORES DE EXTENSIÓN

La divulgación proporciona además un método para preparar una muestra de ácido nucleico para secuenciar, en donde el método comprende las etapas de: (a) poner en contacto la muestra de ácido nucleico con el primer y segundo oligonucleótidos adaptadores tal como se define en el presente documento; (b) hibridar la región diana del primer oligonucleótido adaptador con una primera subsecuencia de un ácido nucleico diana e hibridar la región diana del segundo oligonucleótido adaptador con una segunda subsecuencia del oligonucleótido diana; (c) poner en contacto la muestra de ácido nucleico con una biblioteca de moléculas de código de barras multiméricas tal como se define en el presente documento y cebadores de extensión primero y segundo tal como se define en el presente documento; (d) hibridar la región adaptadora del primer oligonucleótido adaptador con la región adaptadora de la primera molécula de código de barras e hibridar la región adaptadora del segundo oligonucleótido adaptador con la región adaptadora de la segunda molécula de código de barras; (e) extender el primer cebador de extensión usando la región de código de barras de la primera molécula de código de barras como molde para producir un primer oligonucleótido con código de barras, y extender el segundo cebador de extensión usando la región de código de barras de la segunda molécula de código de barras como molde para producir un segundo oligonucleótido con código de barras, en donde el primer oligonucleótido con código de barras comprende una secuencia complementaria con la región de código de barras de la primera molécula de código de barras y el segundo oligonucleótido con código de barras comprende una secuencia complementaria con la región de código de barras de la segunda molécula de código de barras; y (f) ligar el extremo 3' del primer oligonucleótido con código de barras al extremo 5' del primer oligonucleótido adaptador para producir un primer oligonucleótido adaptador con código de barras y ligar el extremo 3' del segundo oligonucleótido con código de barras al extremo 5' del segundo oligonucleótido adaptador para producir un segundo oligonucleótido adaptador con código de barras.

En el método, los oligonucleótidos adaptadores de código de barras primero y segundo pueden extenderse para producir moléculas de ácido nucleico diana con código de barras diferentes, cada una de las cuales comprende al menos un nucleótido sintetizado a partir del ácido nucleico diana como molde.

Como alternativa, los oligonucleótidos adaptadores primero y segundo pueden extenderse para producir las moléculas de ácido nucleico diana primera y segunda diferentes, cada una de las cuales comprende al menos un nucleótido sintetizado a partir del ácido nucleico diana como molde. En este método, la etapa (f) produce una primera molécula de ácido nucleico diana con código de barras (es decir, el primer oligonucleótido con código de barras ligado al primer oligonucleótido adaptador extendido) y una segunda molécula de ácido nucleico diana con código de barras (es decir, el segundo oligonucleótido con código de barras ligado al segundo oligonucleótido adaptador extendido).

La etapa de extender los oligonucleótidos adaptadores se puede realizar antes de etapa (c), antes de la etapa (d), antes de la etapa (e) y/o antes de la etapa (f), y el primer y segundo oligonucleótidos adaptadores pueden permanecer hibridados con la primera y segunda moléculas de código de barras hasta después de la etapa (f).

Los cebadores de extensión pueden hibridar con las moléculas de código de barras multiméricas antes de la etapa (c). Como alternativa, la muestra de ácido nucleico se puede poner en contacto con una biblioteca de moléculas de código de barras multiméricas tal como se define en el presente documento y cebadores de extensión separados tal como se define en el presente documento. A continuación, los cebadores de extensión pueden hibridar con las moléculas de código de barras multiméricas en la muestra de ácido nucleico. Los cebadores de extensión pueden hibridar con las moléculas de código de barras multiméricas durante la etapa (d).

Los métodos pueden usar una biblioteca de cebadores de extensión primero y segundo, por ejemplo, la biblioteca puede comprender cebadores de extensión primero y segundo para cada molécula de código de barras multimérica. Opcionalmente, cada cebador de extensión en la biblioteca de cebadores de extensión puede comprender una región

de código de barras secundaria, en donde dicha región de código de barras secundaria es diferente a las regiones de código de barras secundarias dentro de los otros cebadores de extensión dentro de la biblioteca. Opcionalmente, tal biblioteca puede comprender 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 500, al menos 1000, al menos 5.000, o al menos 10.000 cebadores de extensión diferentes.

5 15. MÉTODOS DE PREPARACIÓN DE UNA MUESTRA DE ÁCIDO NUCLEICO PARA SECUENCIAR UTILIZANDO REACTIVOS DE CÓDIGO DE BARRAS MULTIMÉRICOS, OLIGONUCLEÓTIDOS ADAPTADORES Y CEBADORES DIANA

La divulgación proporciona además un método para preparar una muestra de ácido nucleico para secuenciar, en donde el método comprende las etapas de: (a) poner en contacto la muestra de ácido nucleico con el primer y segundo oligonucleótidos adaptadores, en donde cada oligonucleótido adaptador comprende en la dirección 5' a 3', una región diana y una región adaptadora, y un primer y segundo cebadores diana; (b) hibridar la región diana del primer oligonucleótido adaptador con una primera subsecuencia de un ácido nucleico diana e hibridar la región diana del segundo oligonucleótido adaptador con una segunda subsecuencia del ácido nucleico diana; (c) hibridar el primer cebador diana con una tercera subsecuencia del ácido nucleico diana, en donde la tercera subsecuencia es 3' de la primera subsecuencia, e hibridar el segundo cebador diana con una cuarta subsecuencia del ácido nucleico diana, en donde la cuarta subsecuencia es 3' de la segunda subsecuencia; (d) extender el primer cebador diana usando el ácido nucleico diana como molde hasta que alcance la primera subsecuencia para producir un primer cebador diana extendido, y extender el segundo cebador diana usando el ácido nucleico diana como molde hasta que alcance la segunda subsecuencia para producir un segundo cebador diana extendido; (e) ligar el extremo 3' del primer cebador diana extendido al extremo 5' del primer oligonucleótido adaptador, y ligar el extremo 3' del segundo cebador diana extendido al extremo 5' del segundo oligonucleótido adaptador; (f) poner en contacto la muestra de ácido nucleico con una biblioteca de moléculas de código de barras multiméricas tal como se define en el presente documento; (g) hibridar la región adaptadora del primer oligonucleótido adaptador con la región adaptadora de la primera molécula de código de barras e hibridar la región adaptadora del segundo oligonucleótido adaptador con la región adaptadora de la segunda molécula de código de barras; y (h) extender el primer oligonucleótido adaptador usando la región de código de barras de la primera molécula de código de barras como molde para producir un primer oligonucleótido con código de barras, y extender el segundo oligonucleótido adaptador usando la región de código de barras de la segunda molécula de código de barras como molde para producir un segundo oligonucleótido con código de barras, en donde el primer oligonucleótido con código de barras comprende una secuencia complementaria con la región de código de barras de la primera molécula de código de barras y el segundo oligonucleótido con código de barras comprende una secuencia complementaria con la región de código de barras de la segunda molécula de código de barras.

En el método, las etapas (b) y (c) se pueden realizar al mismo tiempo.

En el método, las etapas (f)-(h) se pueden realizar antes que las etapas (d) y (e). En este método, primera y segunda moléculas de ácido nucleico diana con códigos de barras diferentes, cada uno de los cuales comprende al menos un nucleótido sintetizado a partir del ácido nucleico diana como molde, se producen al completar la etapa (e).

En el método, las etapas (f)-(h) se pueden realizar después de las etapas (d) y (e). En este método, primera y segunda moléculas de ácido nucleico diana con códigos de barras diferentes, cada uno de los cuales comprende al menos un nucleótido sintetizado a partir del ácido nucleico diana como molde, se producen al completar la etapa (h).

La Figura 6 ilustra una forma en la que se puede realizar este método. En este método, el ácido nucleico diana es el ADN genómico. Se apreciará que el ácido nucleico diana puede ser otro tipo de ácido nucleico, por ejemplo, una molécula de ARN tal como una molécula de ARNm.

16. MÉTODOS PARA PREPARAR UNA MUESTRA DE ÁCIDO NUCLEICO PARA SECUENCIAR UTILIZANDO REACTIVOS DE CÓDIGO DE BARRAS MULTIMÉRICOS Y CEBADORES DIANA

La divulgación proporciona además un método para preparar una muestra de ácido nucleico para secuenciar, en donde el método comprende las etapas de: (a) poner en contacto la muestra de ácido nucleico con el primer y segundo oligonucleótidos de código de barras unidos entre sí, en donde cada oligonucleótido con código de barras comprende en la dirección 5' a 3' una región diana y una región de código de barras, y un primer y segundo cebadores diana; (b) hibridar la región diana del primer oligonucleótido con código de barras con una primera subsecuencia de un ácido nucleico diana, e hibridar la región diana del segundo oligonucleótido con código de barras con una segunda subsecuencia del ácido nucleico diana; (c) hibridar el primer cebador diana con una tercera subsecuencia del ácido nucleico diana, en donde la tercera subsecuencia es 3' de la primera subsecuencia, e hibridar el segundo cebador diana con una cuarta subsecuencia del ácido nucleico diana, en donde la cuarta subsecuencia es 3' de la segunda subsecuencia; (d) extender el primer cebador diana usando el ácido nucleico diana como molde hasta que alcance la primera subsecuencia para producir un primer cebador diana extendido, y extender el segundo cebador diana usando el ácido nucleico diana como molde hasta que alcance la segunda subsecuencia para producir un segundo cebador diana extendido; (e) ligar el extremo 3' del primer cebador diana extendido al extremo 5' del primer oligonucleótido con código de barras para producir una primera molécula de ácido nucleico diana con código de barras, y ligar el extremo 3' del segundo cebador diana extendido al extremo 5' del segundo oligonucleótido con código de barras para producir una segunda molécula de ácido nucleico diana con código de barras, en donde la primera y la segunda moléculas de

ácido nucleico diana con código de barras son diferentes y cada una comprende al menos un nucleótido sintetizado a partir del ácido nucleico diana como molde.

17. MÉTODOS DE ENSAMBLAJE DE MOLÉCULAS DE CÓDIGOS DE BARRAS MULTIMÉRICAS MEDIANTE AMPLIFICACIÓN POR CÍRCULO RODANTE

- 5 La divulgación proporciona además un método para ensamblar una biblioteca de moléculas de códigos de barras multiméricas a partir de una biblioteca de moléculas de ácidos nucleicos con códigos de barras, en donde dichas moléculas de ácido nucleico con código de barras se amplifican mediante uno o más procesos de amplificación por círculo rodante (ACR). En este método, cada una de las moléculas de código de barras de ácido nucleico puede comprender, opcionalmente en la dirección de 5' a 3', una región de código de barras y una región adaptadora.
- 10 Opcionalmente, las moléculas de ácido nucleico con código de barras pueden comprender un extremo 5' fosforilado capaz de ligarse a un extremo 3' de una molécula de ácido nucleico.

- En este método, las moléculas de ácido nucleico con código de barras dentro de la biblioteca se convierten en una forma circular, de tal manera que la región del código de barras y la región adaptadora de una molécula de código de barras estén comprendidas dentro de una molécula de ácido nucleico circular contigua. Opcionalmente, tal etapa de
- 15 convertir moléculas de ácido nucleico con código de barras en forma circular se puede realizar mediante una reacción de ligadura intramolecular monocatenaria. Por ejemplo, las moléculas de ácido nucleico con código de barras que comprenden un extremo 5' fosforilado se pueden circularizar mediante incubación con una ligasa de ácido nucleico monocatenario, como ARN Ligasa 1 de T4, o por incubación con una ligasa de ácido nucleico monocatenario termoestable, tal como la ligasa de ácido nucleico monocatenario termoestable CircLigase (de Epicentre Bio).
 - 20 Opcionalmente, se puede realizar una etapa de exonucleasa para agotar o degradar moléculas no circularizadas y/o no ligadas; opcionalmente, en donde la etapa de exonucleasa la realiza la exonucleasa I de E. coli o la exonucleasa lambda de E. coli.

- Opcionalmente, se puede realizar una etapa de conversión de moléculas de códigos de barras de ácido nucleico en forma circular usando un cebador de circularización. En la presente realización, las moléculas de ácido nucleico con
- 25 código de barras comprenden un extremo 5' fosforilado. Además, en la presente realización, un cebador de circularización que comprende una región 5' complementaria a la región 3' de una molécula de código de barras, y una región 3' complementaria a la región 5' de una molécula de código de barras, está hibridado con una molécula de código de barras, de manera que el extremo 5' y el extremo 3' de la molécula de código de barras estén inmediatamente adyacentes entre sí mientras hibridan a lo largo del cebador de circularización. Tras la etapa de hibridación, las
 - 30 moléculas de código de barras alineadas se ligan con una enzima ligasa, tal como la ADN ligasa de T4, que liga el extremo 3' de la molécula del código de barras al extremo 5' de la molécula del código de barras. Opcionalmente, se puede realizar una etapa de exonucleasa para agotar o degradar moléculas no circularizadas y/o no ligadas; opcionalmente, en donde la etapa de exonucleasa la realiza la exonucleasa I de E. coli o la exonucleasa lambda de E. coli.

- 35 Tras una etapa de circularización, las moléculas de códigos de barras circularizadas se pueden amplificar con una etapa de amplificación por círculo rodante. En este proceso, un cebador hibrida con una hebra de ácido nucleico circularizada que comprende una molécula de código de barras, y el extremo 3' de dicho cebador se extiende con una polimerasa que muestra un comportamiento de desplazamiento de la hebra. Para cada molécula de código de barras circularizada original, este proceso puede formar una molécula de código de barras multimérica lineal (no circular) que
- 40 comprende copias de la molécula de código de barras circularizada original, tal como se ilustra en la Figura 7. En una realización, un cebador de circularización que ha hibridado con una molécula de código de barras puede servir como cebador para una etapa de amplificación por círculo rodante. Opcionalmente, tras la circularización, un cebador de amplificación separado que sea al menos parcialmente complementario a la molécula de código de barras circularizada, puede hibridar con la molécula de código de barras circularizada para cebar una etapa de amplificación
- 45 por círculo rodante.

- Durante dicha etapa de amplificación por círculo rodante, el cebador puede ser extendido por la polimerasa, en donde la polimerasa se extiende a lo largo del molde circularizado hasta que encuentra el extremo 5' del cebador de amplificación y/o cebador de circularización, después de lo cual continúa la amplificación a lo largo del molde circularizado mientras desplaza el extremo 5' del cebador, y luego desplaza la hebra previamente amplificada, en un
- 50 proceso de amplificación por círculo rodante. Después de cualquier etapa de amplificación, se puede realizar una etapa de purificación y/o limpieza para aislar los productos de dicha amplificación por círculo rodante. Opcionalmente, una etapa de purificación y/o limpieza puede comprender un proceso de selección de tamaño, tal como un proceso de selección de tamaño basado en gel, o un proceso de selección de tamaño de inmovilización reversible en fase sólida, tal como un proceso de selección de tamaño de inmovilización reversible en fase sólida basado en perlas magnéticas.
 - 55 Opcionalmente, los productos de amplificación de al menos 100 nucleótidos de longitud, al menos 500 nucleótidos de longitud, al menos 1000 nucleótidos de longitud, al menos 2000 nucleótidos de longitud, al menos 5000 nucleótidos de longitud, al menos 10.000 nucleótidos de longitud, al menos 20.000 nucleótidos de longitud, al menos 50.000 nucleótidos de longitud o al menos 100.000 nucleótidos de longitud se pueden purificar. Opcionalmente, antes y/o durante cualquier etapa de amplificación por círculo rodante, una proteína de unión a ADN monocatenario (tal como la proteína del gen 32 de T4) puede incluirse en una mezcla de reacción, tal como para prevenir la formación de
 - 60 estructuras secundarias por moldes circularizados y/o productos de amplificación. Durante o después de cualquier

etapa de amplificación por círculo rodante, dicha proteína de unión a ADN monocatenario se puede eliminar y/o inactivar, tal como por una etapa de inactivación por calor.

Opcionalmente, dicho proceso de amplificación por círculo rodante se puede realizar mediante la ADN polimerasa phi29. Opcionalmente, dicho proceso de amplificación por círculo rodante se puede realizar mediante una ADN polimerasa Bst o Bsm. Opcionalmente, tal proceso de amplificación por círculo rodante se puede realizar de manera que la polimerasa produzca al menos una copia completa del molde circularizado. Opcionalmente, tal proceso de amplificación por círculo rodante se puede realizar de manera que al menos 2, al menos 3, al menos 5, al menos 10, al menos 50, al menos 100, al menos 200, al menos 500, al menos 1000, al menos 2000, al menos 5000, o al menos 10000 copias completas del molde circularizado se producen por la polimerasa.

En la Figura 7 se proporciona un ejemplo de este método. En la figura, se circulariza una molécula de código de barras que comprende una región adaptadora y una región de código de barras (por ejemplo, usando una reacción de ligadura monocatenaria). A continuación, se hibrida un cebador con el producto circularizado resultante y, a continuación, dicho cebador se extiende utilizando una polimerasa de desplazamiento de cadena (como la ADN polimerasa phi29). Al sintetizar el producto de extensión, la polimerasa luego procesa una circunferencia alrededor del producto circularizado y luego desplaza el cebador original en una reacción de desplazamiento de cadena. El proceso de amplificación por círculo rodante puede proceder a crear una molécula de ácido nucleico contigua larga que comprende muchas copias en tándem de la secuencia circularizada, es decir, muchas copias en tándem de un código de barras y una secuencia adaptadora (y/o secuencias complementarias a un código de barras y una secuencia adaptadora) de una molécula de código de barras.

Las moléculas de códigos de barras multiméricas también se pueden amplificar mediante amplificación por círculo rodante.

18. MÉTODOS DE AMPLIFICACIÓN DE MOLÉCULAS DE CÓDIGOS DE BARRAS MULTIMÉRICAS MEDIANTE AMPLIFICACIÓN POR CÍRCULO RODANTE

A) Propiedades de las moléculas de códigos de barras multiméricas

La divulgación proporciona además un método para amplificar moléculas de códigos de barras multiméricas a partir de una biblioteca de moléculas de códigos de barras multiméricas, en donde dichas moléculas de código de barras multiméricas se amplifican mediante uno o más procesos de amplificación por círculo rodante (ACR). En este método, una molécula de código de barras multimérica comprende al menos dos moléculas de código de barras unidas entre sí dentro de una (única) molécula de ácido nucleico. Opcionalmente, cada región de código de barras de una molécula de código de barras puede ser adyacente a una o más regiones adaptadoras; opcionalmente, dicha región adaptadora puede estar en el extremo 5' de la región del código de barras asociada, o puede estar en el extremo 3' de la región del código de barras asociada. Opcionalmente, cada región de código de barras está asociada con una región adaptadora de 3' y una región adaptadora de 5'; opcionalmente, la región adaptadora de 3' y una región adaptadora de 5' pueden comprender diferentes secuencias adaptadoras. Opcionalmente, una o más regiones adaptadoras pueden comprender una secuencia complementaria o idéntica a una región adaptadora de un oligonucleótido adaptador. Opcionalmente, una o más regiones adaptadoras pueden comprender una secuencia complementaria o idéntica a todo o parte de un cebador de extensión. Una molécula de código de barras multimérica puede adoptar cualquiera de las formas descritas en el presente documento.

Cada molécula de código de barras multimérica puede comprender además, opcionalmente dentro del extremo 5' de la molécula de código de barras multimérica, una secuencia de amplificación de reactivo directo, que puede comprender una secuencia complementaria o idéntica a un cebador de amplificación de reactivo directo. Cada molécula de código de barras multimérica puede comprender además, opcionalmente dentro del extremo 3' de la molécula de código de barras multimérica, una secuencia de amplificación de reactivo inverso, que puede comprender una secuencia complementaria o idéntica a un cebador de amplificación de reactivo inverso.

Una molécula de código de barras multimérica puede comprender al menos 2, al menos 3, al menos 4, al menos 5, al menos 10, al menos 20, al menos 25, al menos 50, al menos 75, al menos 100, al menos 200, al menos 500, al menos 1000, al menos 5000, al menos 10^4 , al menos 10^5 o al menos 10^6 moléculas de códigos de barras diferentes. Cualquier biblioteca de moléculas de códigos de barras multiméricas puede comprender al menos 5, al menos 10, al menos 20, al menos 25, al menos 50, al menos 75, al menos 100, al menos 250, al menos 500, al menos 10^3 , al menos 10^4 , al menos 10^5 , al menos 10^6 , al menos 10^7 , al menos 10^8 o al menos 10^9 moléculas de códigos de barras multiméricas diferentes.

B) Métodos de circularización de moléculas de código de barras multiméricas y/o bibliotecas de las mismas

En un método de amplificación de moléculas de códigos de barras multiméricas, las moléculas de códigos de barras multiméricas (y/o una biblioteca de las mismas) se convierten en una forma circular, tal manera que las 2 o más regiones de código de barras (y, opcionalmente, 2 o más regiones adaptadoras) de una molécula de código de barras multimérica están comprendidas dentro de una molécula de ácido nucleico circular contigua. Opcionalmente, se puede realizar una etapa de conversión de moléculas de código de barras multiméricas en forma circular mediante una reacción de ligadura bicatenaria intramolecular. Por ejemplo, las moléculas de código de barras multiméricas que

comprenden secuencias bicatenarias y extremos 5' fosforilados pueden comprender extremos romos, u opcionalmente pueden tener sus extremos convertidos en una forma roma con una reacción de embotamiento. Dichas moléculas de código de barras multiméricas pueden luego convertirse en forma circular mediante una reacción de ligadura bicatenaria intramolecular con una enzima ADN ligasa T4, de manera que un extremo de una molécula de código de barras multimérica se liga en una o ambas hebras al otro extremo de la misma molécula de código de barras multimérica.

En una realización alternativa, se puede realizar una etapa de conversión de moléculas de código de barras multiméricas en forma circular mediante una reacción de ligadura bicatenaria intramolecular en la que los extremos de moléculas de código de barras multiméricas comprenden extremos generados por una etapa de digestión de restricción. En una realización similar, las moléculas de código de barras multiméricas que comprenden secuencias bicatenarias comprenden sitios de reconocimiento para una o más enzimas endonucleasas de restricción dentro de sus regiones 5' y 3'. En una reacción de digestión, dichas moléculas de código de barras multiméricas se digieren con una o más enzimas endonucleasas de restricción para crear moléculas de código de barras multiméricas digeridas que comprenden extremos con los productos de digestión de restricción. Estas moléculas de código de barras multiméricas digeridas pueden opcionalmente purificarse a continuación, por ejemplo, con una etapa de selección de tamaño basada en gel o basada en perlas. Las moléculas de código de barras multiméricas digeridas se pueden convertir en forma circular mediante una reacción de ligadura bicatenaria intramolecular con una enzima ADN ligasa T4, de manera que el sitio digerido por restricción en un extremo de una molécula de código de barras multimérico se liga al sitio digerido por restricción en el otro extremo de la misma molécula de código de barras multimérico. Opcionalmente, los extremos producidos por la(s) enzima(s) de restricción pueden ser romos, o pueden comprender un saliente 3' de 1 o más nucleótidos, o pueden comprender un saliente 5' de 1 o más nucleótidos.

Durante cualquier etapa del ensamblaje, amplificación, ligadura y/o circularización de moléculas de códigos de barras y/o moléculas de códigos de barras multiméricas, y/o bibliotecas o constituyentes de las mismas, la concentración de tales moléculas dentro de la solución puede mantenerse dentro de ciertos intervalos. Por ejemplo, la concentración de moléculas de códigos de barras y/o moléculas de códigos de barras multiméricas puede ser inferior a 100 nanomolar, menos de 10 nanomolar, menos de 1 nanomolar, menos de 100 picomolar, menos de 10 picomolar, menos de 1 picomolar, menos de 100 femtomolar, menos de 10 femtomolar o menos de 1 femtomolar. Opcionalmente, durante cualquier etapa del ensamblaje, amplificación, ligadura y/o circularización de moléculas de códigos de barras y/o moléculas de códigos de barras multiméricas, y/o bibliotecas o constituyentes de las mismas, la concentración de tales moléculas dentro de la solución puede permitir que dos o más moléculas de códigos de barras diferentes y/o moléculas de códigos de barras multiméricas se agreguen, concatenen o ligen entre sí dentro de la solución, opcionalmente, en donde dichos productos adjuntos, concatenados o ligados se amplifican posteriormente durante una etapa de amplificación.

C) Métodos de amplificación de moléculas de código de barras multiméricas circularizadas con amplificación por círculo rodante

Tras una etapa de circularización, las moléculas de código de barras multiméricas circularizadas se amplifican con una etapa de amplificación por círculo rodante. En este proceso, un cebador hibrida con una hebra de ácido nucleico circularizada que comprende una molécula de código de barras multimérica, y el extremo 3' de dicho cebador se extiende con una polimerasa que muestra un comportamiento de desplazamiento de la hebra. En una realización, un cebador de circularización que ha hibridado con una molécula de código de barras multimérica puede servir como cebador para una etapa de amplificación por círculo rodante. Opcionalmente, tras la circularización, uno o más cebadores de amplificación separados que son al menos parcialmente complementarios a una molécula de código de barras multimérica circularizada, puede hibridar con la molécula de código de barras circularizada para cebar una etapa de amplificación por círculo rodante. Opcionalmente, como cebadores de amplificación se pueden emplear oligonucleótidos al menos parcialmente complementarios a una o más regiones adaptadoras comprendidas dentro de una molécula de código de barras multimérica. Opcionalmente, después de cualquier etapa de hibridación de uno o más cebadores de amplificación con moléculas de código de barras multiméricas circularizadas, se puede realizar una etapa de limpieza para agotar los cebadores no hibridados de la solución y/o para aislar moléculas de código de barras multiméricas hibridadas con cebador. Opcionalmente, tal etapa de limpieza puede comprender una etapa de selección por tamaño, tal como una etapa de selección por tamaño basada en gel o una etapa de selección por tamaño basada en perlas, tal como una etapa de inmovilización reversible en fase sólida. Opcionalmente, antes y/o durante cualquier etapa de amplificación por círculo rodante, una proteína de unión a ADN monocatenario (tal como la proteína del gen 32 de T4) puede incluirse en una mezcla de reacción, tal como para prevenir la formación de estructuras secundarias por moldes circularizados y/o productos de amplificación. Durante o después de cualquier etapa de amplificación por círculo rodante, dicha proteína de unión a ADN monocatenario se puede eliminar y/o inactivar, tal como por una etapa de inactivación por calor.

Durante dicha etapa de amplificación por círculo rodante, cada cebador puede ser extendido por la polimerasa, en donde la polimerasa se extiende a lo largo del molde circularizado hasta que encuentra el extremo 5' de un cebador de amplificación y/o un cebador de circularización, después de lo cual continúa la amplificación a lo largo del molde circularizado mientras desplaza el extremo 5' del cebador, y luego desplaza la hebra previamente amplificada, en un proceso de amplificación por círculo rodante. Después de cualquier etapa de amplificación, se puede realizar una etapa de purificación y/o limpieza para aislar los productos de dicha amplificación por círculo rodante. Opcionalmente,

una etapa de purificación y/o una etapa de limpieza puede comprender un proceso de selección de tamaño, tal como un proceso de selección de tamaño basado en gel, o un proceso de selección de tamaño de inmovilización reversible en fase sólida, tal como un proceso de selección de tamaño de inmovilización reversible en fase sólida basado en perlas magnéticas. Opcionalmente, los productos de amplificación de al menos 100 nucleótidos de longitud, al menos 500 nucleótidos de longitud, al menos 1000 nucleótidos de longitud, al menos 2000 nucleótidos de longitud, al menos 5000 nucleótidos de longitud, al menos 10.000 nucleótidos de longitud, al menos 20.000 nucleótidos de longitud, al menos 50.000 nucleótidos de longitud o al menos 100.000 nucleótidos de longitud se pueden purificar.

Opcionalmente, dicho proceso de amplificación por círculo rodante se puede realizar mediante la ADN polimerasa phi29. Opcionalmente, dicho proceso de amplificación por círculo rodante se puede realizar mediante una ADN polimerasa Bst o Bsm. Opcionalmente, tal proceso de amplificación por círculo rodante se puede realizar de manera que la polimerasa produzca al menos una copia completa del molde circularizado. Opcionalmente, tal proceso de amplificación por círculo rodante se puede realizar de manera que al menos 2, al menos 3, al menos 5, al menos 10, al menos 50, al menos 100, al menos 200, al menos 500, al menos 1000, al menos 2000, al menos 5000 o al menos 10000 copias completas del molde de la molécula de código de barras multimérica circularizada son producidas por la polimerasa a partir de cada cebador que ha hibridado con la molécula de código de barras multimérica circularizada.

Opcionalmente, al menos 2, al menos 3, al menos 5, al menos 10, al menos 20, al menos 50, al menos 100 o al menos 500 cebadores pueden hibridar con una molécula de código de barras multimérica circularizada y se puede utilizar para cebar reacciones de amplificación por círculo rodante a lo largo de la molécula de código de barras multimérica con la que hibridan. Las partes de los productos de extensión producidos a partir de estos cebadores pueden permanecer hibridados con la molécula de código de barras multimérica circularizada con la que se hibridaron originalmente, produciendo así un complejo de ácido nucleico macromolecular que comprende una molécula de código de barras multimérica circularizada y dos o más productos de amplificación por círculo rodante al menos parcialmente hibridados con ella.

Las secuencias dentro de una molécula de código de barras multimérica se pueden configurar de manera que los productos de amplificación por círculo rodante comprendan una o más regiones adaptadoras y/o secuencias adaptadoras, de manera que dichas regiones adaptadoras y/o secuencias adaptadoras puedan hibridar con secuencias complementarias, por ejemplo, secuencias complementarias comprendidas en oligonucleótidos de acoplamiento, oligonucleótidos adaptadores y/o cebadores de extensión. Parte o la totalidad de cualquiera de dichos productos de amplificación por círculo rodante y/o complejo de ácido nucleico macromolecular se puede usar para sintetizar un reactivo de código de barras multimérico, por ejemplo, sirviendo como moléculas de códigos de barras para sintetizar oligonucleótidos con códigos de barras. Parte o la totalidad de cualquiera de dichos productos de amplificación por círculo rodante y/o complejo de ácido nucleico macromolecular puede servir como moléculas de código de barras para su uso para codificar moléculas de ácido nucleico dentro de una muestra de ácido nucleico.

D) Métodos de procesamiento de moléculas de código de barras multiméricas amplificadas por círculo rodante con un proceso de extensión de cebador

Después de cualquier proceso de amplificación por círculo rodante de una molécula de código de barras multimérica y/o biblioteca de la misma, se pueden realizar una o más etapas de extensión del cebador en los productos resultantes.

Los productos de extensión de cebadores resultantes pueden comprender moléculas de ácido nucleico monocatenarias que comprenden la totalidad o parte de moléculas de códigos de barras multiméricas y/o partes de dos o más moléculas de códigos de barras multiméricas. En algunas realizaciones, tales productos de extensión de cebadores pueden comprender una biblioteca de moléculas de ácido nucleico monocatenarias, en donde cada hebra de ácido nucleico comprende una molécula de código de barras multimérica. En otras realizaciones, tales productos de extensión de cebadores pueden hibridar o hibridar parcialmente con las moléculas molde a partir de las cuales se sintetizan. Opcionalmente, cualquier molécula de código de barras multimérica resultante de cualquier proceso de extensión de cebador de este tipo se puede usar para crear un reactivo de código de barras multimérico y/o biblioteca del mismo. Opcionalmente, cualquier molécula de código de barras multimérica que resulte de dicho proceso de extensión primaria se puede usar para codificar moléculas de ácido nucleico dentro de una muestra de ácido nucleico; opcionalmente, las secuencias de códigos de barras que comprenden dichas moléculas de códigos de barras multiméricas pueden adjuntarse a moléculas de ácido nucleico dentro de una muestra de ácido nucleico.

En una de estas formas de realización de un proceso de extensión de cebadores, se puede usar un cebador complementario o idéntico en secuencia a, toda o parte de una secuencia de amplificación de reactivo directo y/o toda o parte de una secuencia de amplificación de reactivo inverso. En una realización similar, se puede usar un cebador al menos parcialmente complementario a una secuencia o secuencias de amplificación de reactivo comprendidas dentro de los productos de extensión de la polimerasa de la reacción de amplificación por círculo rodante para realizar una o más reacciones y/o ciclos de extensión del cebador. En una realización de un proceso de extensión de cebadores, se utiliza una biblioteca de cebadores aleatorios para dicho proceso de extensión de cebadores, por ejemplo, cebadores de hexámeros aleatorios, cebadores de octámeros aleatorios, o cebadores de decámeros aleatorios. Opcionalmente, cualquier cebador utilizado en un proceso de extensión de cebadores puede comprender una o más modificaciones, tales como enlaces fosforotioato, y específicamente tales como enlaces fosforotioato dentro de los enlaces 3' más uno o dos nucleótidos dentro del cebador. Tales enlaces fosforotioato 3' pueden prevenir la

degradación de dichos cebadores por polimerasas que presentan comportamiento de exonucleasa.

Opcionalmente, tal etapa de extensión del cebador se puede realizar por una polimerasa que presenta comportamiento de exonucleasa 5'-3' (tal como la ADN polimerasa I de *E. coli*) y/o comportamiento de endonucleasa flap (como la Taq polimerasa de *Thermus aquaticus*), de manera que las secuencias de ácido nucleico hibridadas inmediatamente aguas abajo de una polimerasa de procesamiento se degradan o se degradan parcialmente durante el proceso de extensión del cebador por dicha polimerasa.

Opcionalmente, tal etapa de extensión del cebador se puede realizar por una polimerasa que presenta un comportamiento de desplazamiento de la hebra, tal como la ADN polimerasa phi29, la polimerasa Vent, la polimerasa Deep Vent, o sus derivados deficientes en exonucleasa (por ejemplo, de New England Biolabs), o ADN polimerasa Bst o Bsm, de tal manera que las secuencias de ácido nucleico hibridadas inmediatamente aguas abajo de una polimerasa de procesamiento son desplazadas durante el proceso de extensión del cebador por dicha polimerasa. Opcionalmente, dichas secuencias de ácido nucleico desplazadas pueden comprender otros productos de extensión de cebadores producidos durante el proceso de extensión de cebadores. Opcionalmente, tal etapa de extensión del cebador puede ser realizado por la ADN polimerasa phi29, en donde los cebadores usados para dicha etapa de extensión del cebador comprenden cebadores aleatorios.

Cualquier etapa de extensión del cebador realizada por una polimerasa que presenta un comportamiento de desplazamiento de cadena puede tener el efecto de desplazar regiones de moléculas de código de barras multiméricas (y/o cadenas de ácido nucleico que comprenden secuencias de moléculas de código de barras multiméricas, por ejemplo, los que se producen mediante un proceso de extensión de cebadores de este tipo) que comprenden una o más regiones adaptadoras y/o secuencias adaptadoras, de manera que dichas regiones adaptadoras y/o secuencias adaptadoras se conviertan en una forma monocatenaria, de modo que las regiones adaptadoras monocatenarias resultantes sean capaces de hibridar con secuencias complementarias, por ejemplo, secuencias complementarias comprendidas en oligonucleótidos de acoplamiento, oligonucleótidos adaptadores y/o cebadores de extensión. Partes de tales moléculas de cadena desplazada pueden permanecer hibridadas con las moléculas molde a partir de las cuales se sintetizaron. Parte o la totalidad de cualquier molécula de ácido nucleico de cadena desplazada dada sintetizada mediante tal proceso de extensión de cebador puede usarse para sintetizar un reactivo de código de barras multimérico. Parte o la totalidad de cualquier molécula de ácido nucleico de cadena desplazada dada sintetizada mediante un proceso de extensión de cebador de este tipo puede usarse para poner códigos de barras de moléculas de ácido nucleico dentro de una muestra de ácido nucleico.

Opcionalmente, tal etapa de extensión del cebador se puede realizar por una polimerasa que no presenta exonucleasa 5'-3', o comportamiento de endonucleasa flap, o comportamiento de desplazamiento de cadena (como las polimerasas Pfu y/o Phusion o derivados de las mismas (New England Biolabs), o ADN polimerasa T4), de manera que las secuencias de ácido nucleico hibridadas inmediatamente aguas abajo de una polimerasa de procesamiento detienen la extensión de la polimerasa cuando las encuentra allí.

Opcionalmente, cualquier etapa de extensión del cebador puede comprender al menos 1, al menos 5, al menos 10, al menos 15, al menos 20, al menos 30, al menos 50 o al menos 100 ciclos de extensión del cebador. Opcionalmente, tales ciclos de extensión de cebadores se pueden realizar dentro de ciclos repetidos de extensión de cebadores, desnaturalización del molde e hibridación del cebador. Opcionalmente, cualquier etapa de extensión del cebador se puede realizar en un tampón que comprende uno o más agentes de agrupación macromolecular, tales como reactivos de polietilenglicol (PEG), por ejemplo PEG 8000.

Opcionalmente, productos de extensión de cebadores de al menos 100 nucleótidos de longitud, al menos 500 nucleótidos de longitud, al menos 1000 nucleótidos de longitud, al menos 2000 nucleótidos de longitud, al menos 5000 nucleótidos de longitud, al menos 10.000 nucleótidos de longitud, al menos 20.000 nucleótidos de longitud, al menos 50.000 nucleótidos de longitud, o al menos 100.000 nucleótidos de longitud se pueden producir mediante cualquier proceso de extensión de cebadores anterior. Opcionalmente, tal proceso de extensión inicial se puede realizar de modo que la polimerasa produzca al menos una copia completa del molde circularizado. Opcionalmente, tal proceso de extensión del cebador se puede realizar de manera que al menos 2, al menos 3, al menos 5, al menos 10, al menos 50, al menos 100, al menos 200, al menos 500, al menos 1000, al menos 2000, al menos 5000 o al menos 10.000 copias del molde de la molécula de código de barras multimérica se producen por la polimerasa durante cada etapa de extensión del cebador. Opcionalmente, la duración en el tiempo (por ejemplo, segundos o minutos) de una reacción de extensión de cebador se puede configurar de modo que cada producto de extensión de cebador tenga aproximadamente la misma longitud que un único reactivo de código de barras multimérico dentro de la biblioteca. Por ejemplo, si una polimerasa utilizada para procesos de extensión de cebadores a una velocidad de 1000 nucleótidos por minuto, y la longitud media de un reactivo de código de barras multimérico dentro de una biblioteca de reactivos de código de barras multimérico es de 1000 nucleótidos, entonces, el ciclo de extensión del cebador se puede configurar para que tenga una duración de 1 minuto.

Opcionalmente, después de una o más etapas de extensión del cebador, los productos de extensión del cebador resultantes se pueden aislar o purificar mediante una reacción de limpieza. Opcionalmente, tal reacción de limpieza puede comprender una etapa de selección de tamaño, tal como una etapa de selección por tamaño basada en gel o una etapa de selección por tamaño basada en perlas, tal como una etapa de inmovilización reversible en fase sólida.

Opcionalmente, productos de extensión de cebadores de al menos 100 nucleótidos de longitud, al menos 500 nucleótidos de longitud, al menos 1000 nucleótidos de longitud, al menos 2000 nucleótidos de longitud, al menos 5000 nucleótidos de longitud, al menos 10.000 nucleótidos de longitud, al menos 20.000 nucleótidos de longitud, al menos 50.000 nucleótidos de longitud o al menos 100.000 nucleótidos de longitud se pueden purificar.

- 5 E) Métodos de procesamiento de moléculas de código de barras multiméricas amplificadas por círculo rodante y/o extendidas por cebador con un proceso de desnaturalización

Antes o después de cualquier etapa de purificación y/o etapa de selección de tamaño, y/o antes de su uso para sintetizar reactivos de códigos de barras multiméricos, y/o antes de su uso para códigos de barras de ácidos nucleicos dentro de una muestra de ácidos nucleicos, cualquier producto de amplificación por círculo rodante o producto de extensión de cebador producido como anteriormente se puede desnaturalizar con una etapa de desnaturalización. Tal etapa de desnaturalización puede ser una etapa de desnaturalización térmica, en donde los productos se incuban a alta temperatura para fusionar secuencias hibridadas y/o estructura secundaria. Dicha etapa de desnaturalización se puede realizar a una temperatura de al menos 60 grados Celsius, al menos 70 grados Celsius, al menos 80 grados Celsius, al menos 90 grados Celsius, o al menos 95 grados Celsius. Dicha etapa de desnaturalización puede tener el efecto de desnaturalizar regiones de moléculas de códigos de barras multiméricas que comprenden una o más regiones adaptadoras y/o secuencias adaptadoras en forma monocatenaria, de modo que las regiones adaptadoras monocatenarias resultantes sean capaces de hibridar con secuencias complementarias, por ejemplo, secuencias complementarias comprendidas en oligonucleótidos de acoplamiento, oligonucleótidos adaptadores y/o cebadores de extensión.

- 20 En realizaciones alternativas, dicha etapa de desnaturalización se puede realizar antes o después de cualquier etapa de purificación y/o etapa de selección de tamaño, y/o antes de su uso para sintetizar reactivos de códigos de barras multiméricos, y/o antes de su uso para códigos de barras de ácidos nucleicos dentro de una muestra de ácidos nucleicos. Por ejemplo, las cadenas de ácido nucleico que comprenden productos de extensión del cebador producidos durante una etapa de extensión del cebador pueden permanecer hibridadas o parcialmente hibridadas con las moléculas molde a partir de las cuales se sintetizaron. Las macromoléculas de ácido nucleico resultantes pueden comprender un total de al menos 2 cadenas de ácido nucleico individuales, al menos 3 cadenas de ácido nucleico individuales, al menos 5 cadenas de ácido nucleico individuales, al menos 10 cadenas de ácido nucleico individuales, al menos 50 cadenas de ácido nucleico individuales, al menos 100 cadenas de ácido nucleico individuales, al menos 500 cadenas de ácido nucleico individuales, al menos 1000 cadenas de ácido nucleico individuales, al menos 5000 cadenas de ácido nucleico individuales o al menos 10.000 cadenas de ácido nucleico individuales. Opcionalmente, las cadenas de ácido nucleico individuales pueden comprender todo o partes de una o más moléculas de códigos de barras multiméricas. Tales macromoléculas de ácido nucleico y/o bibliotecas de las mismas se pueden usar para sintetizar reactivos de códigos de barras multiméricos y/o para codificar ácidos nucleicos dentro de una muestra de ácidos nucleicos.

35 19. MÉTODOS DE SINTETIZAR UN REACTIVO DE CÓDIGO DE BARRAS MULTIMÉRICO

La divulgación proporciona además un método para sintetizar un reactivo de código de barras multimérico para marcar un ácido nucleico diana que comprende: (a) poner en contacto la primera y segunda moléculas de código de barras con el primer y segundo cebadores de extensión, en donde cada una de las moléculas de código de barras comprende un ácido nucleico monocatenario que comprende, en la dirección 5' a 3', una región adaptadora, una región de código de barras y una región diana; (b) hibridar el primer cebador de extensión con la región de cebador de la primera molécula de código de barras e hibridar el segundo cebador de extensión con la región de cebador de la segunda molécula de código de barras; y (c) sintetizar un primer producto de extensión con código de barras extendiendo el primer cebador de extensión y sintetizar un segundo producto de extensión con código de barras extendiendo el segundo cebador de extensión, en donde el primer producto de extensión con código de barras comprende una secuencia complementaria a la región de código de barras de la primera molécula de código de barras y el segundo producto de extensión con código de barras comprende una secuencia complementaria a la región de código de barras de la segunda molécula de código de barras, y en donde el primer producto de extensión con código de barras no comprende una secuencia complementaria a la región adaptadora de la primera molécula de código de barras y el segundo producto de extensión con código de barras no comprende una secuencia complementaria a la región adaptadora de la segunda molécula de código de barras; y en donde la primera y la segunda moléculas de código de barras están unidas entre sí.

El método puede comprender, además, las siguientes etapas antes de la etapa de sintetizar el primer y segundo producto de extensión con código de barras: (a) poner en contacto la primera y segunda moléculas de código de barras con el primer y segundo cebadores de bloqueo; y (b) hibridar el primer cebador de bloqueo con la región adaptadora de la primera molécula de código de barras e hibridar el segundo cebador de bloqueo con la región adaptadora de la segunda molécula de código de barras; y en donde el método comprende además la etapa de disociar los cebadores de bloqueo de las moléculas del código de barras después de la etapa de sintetizar los productos de extensión del código de barras.

En el método, se puede realizar la etapa de extensión, o una segunda etapa de extensión realizada después de la síntesis de un producto de extensión, en donde uno o más de los cuatro desoxirribonucleótidos canónicos se excluyen

de la reacción de extensión, de tal manera que la segunda etapa de extensión termine en una posición antes de la secuencia de la región del adaptador, en donde la posición comprende un nucleótido complementario al desoxirribonucleótido excluido. Esta etapa de extensión se puede realizar con una polimerasa que carece de actividad exonucleasa 3' a 5'.

- 5 Las moléculas de código de barras pueden ser proporcionadas por una molécula de código de barras multimérica monocatenaria tal como se define en el presente documento.

Las moléculas de código de barras se pueden sintetizar mediante cualquiera de los métodos definidos en el presente documento. Las regiones del código de barras pueden identificar de forma única cada una de las moléculas de código de barras. Las moléculas de código de barras se pueden unir a una molécula de ácido nucleico. Las moléculas de

10 código de barras se pueden unir en una reacción de ligadura. Las moléculas de código de barras se pueden unir entre sí mediante una etapa adicional que comprende unir las moléculas de código de barras a un soporte sólido.

- La primera y la segunda moléculas de código de barras se pueden ensamblar como una molécula de código de barras multimérica bicatenaria mediante cualquiera de los métodos definidos en el presente documento antes de la etapa (a) definida anteriormente (es decir, poniendo en contacto la primera y la segunda moléculas de código de barras con el
- 15 primer y segundo cebadores de extensión). La molécula de código de barras multimérica bicatenaria se puede disociar para producir moléculas de código de barras multiméricas monocatenarias para su uso en la etapa (a) definido anteriormente (es decir, poner en contacto la primera y segunda moléculas de código de barras con el primer y segundo cebadores de extensión).

- El método puede comprender además las etapas de: (a) hibridar una región adaptadora de un primer oligonucleótido adaptador con la región adaptadora de la primera molécula de código de barras e hibridar una región adaptadora del
- 20 segundo oligonucleótido adaptador con la región adaptadora de la segunda molécula de código de barras, en donde el primer oligonucleótido adaptador comprende además una región diana capaz de hibridar con una primera subsecuencia del ácido nucleico diana y el segundo oligonucleótido adaptador comprende además una región diana capaz de hibridar con una segunda subsecuencia del ácido nucleico diana; y (b) ligar el extremo 3' del primer producto
- 25 de extensión con código de barras al extremo 5' del primer oligonucleótido adaptador para producir un primer oligonucleótido con código de barras y ligar el extremo 3' del segundo producto de extensión con código de barras al extremo 5' del segundo oligonucleótido adaptador para producir un segundo oligonucleótido con código de barras. Opcionalmente, la etapa de hibridación (a) se puede realizar antes de la etapa de sintetizar el primer y segundo productos de extensión con código de barras y en donde la etapa de sintetizar el primer y segundo productos de
- 30 extensión con código de barras se lleva a cabo en presencia de una enzima ligasa que realiza la etapa de ligadura (b). La ligasa puede ser una ligasa termoestable. La reacción de extensión y ligadura puede proceder a más de 37 grados Celsius, más de 45 grados Celsius o más de 50 grados Celsius.

- Las regiones diana pueden comprender diferentes secuencias. Cada región diana puede comprender una secuencia capaz de hibridar solo con una única subsecuencia de un ácido nucleico diana dentro de una muestra de ácidos
- 35 nucleicos. Cada región diana puede comprender uno o más secuencias aleatorias o una o más degeneradas, para permitir que la región diana hibride con más de una subsecuencia de un ácido nucleico diana. Cada región diana puede comprender al menos 5, al menos 10, al menos 15, al menos 20, al menos 25, al menos 50 o al menos 100 nucleótidos. Preferentemente, cada región diana comprende al menos 5 nucleótidos. Cada región diana puede comprender de 5 a 100 nucleótidos, de 5 a 10 nucleótidos, de 10 a 20 nucleótidos, de 20 a 30 nucleótidos, de 30 a 50
- 40 nucleótidos, de 50 a 100 nucleótidos, de 10 a 90 nucleótidos, de 20 a 80 nucleótidos, de 30 a 70 nucleótidos o de 50 a 60 nucleótidos. Preferentemente, cada región diana comprende de 30 a 70 nucleótidos. Preferentemente, cada región diana comprende desoxirribonucleótidos, opcionalmente, todos los nucleótidos en una región diana son desoxirribonucleótidos. Uno o más de los desoxirribonucleótidos puede ser un desoxirribonucleótido modificado (por ejemplo, un desoxirribonucleótido modificado con un resto de biotina o un nucleótido de desoxiuracilo). Cada región
- 45 diana puede comprender una o más bases universales (por ejemplo, inosina), uno o más nucleótidos modificados y/o uno o más análogos de nucleótidos.

- La región adaptadora de cada oligonucleótido adaptador puede comprender una región constante. Opcionalmente, todas las regiones adaptadoras de oligonucleótidos adaptadores que hibridan con un único reactivo de código de
- 50 barras multimérico son sustancialmente idénticas. La región adaptadora puede comprender al menos 4, al menos 5, al menos 6, al menos 8, al menos 10, al menos 15, al menos 20, al menos 25, al menos 50, al menos 100 o al menos 250 nucleótidos. Preferentemente, la región adaptadora comprende al menos 4 nucleótidos. Preferentemente, cada región adaptadora comprende desoxirribonucleótidos, opcionalmente, todos los nucleótidos en una región adaptadora son desoxirribonucleótidos. Uno o más de los desoxirribonucleótidos puede ser un desoxirribonucleótido modificado (por ejemplo, un desoxirribonucleótido modificado con un resto de biotina o un nucleótido de desoxiuracilo). Cada
- 55 región adaptadora puede comprender una o más bases universales (por ejemplo, inosina), uno o más nucleótidos modificados y/o uno o más análogos de nucleótidos.

- Para cualquiera de los métodos que involucran oligonucleótidos adaptadores, el extremo 3' del oligonucleótido adaptador puede incluir un resto terminador reversible o un nucleótido terminador reversible (por ejemplo, un
- 60 nucleótido O-bloqueado en 3'), por ejemplo, en el nucleótido 3' terminal de la región diana. Cuando se utiliza en una reacción de extensión y/o extensión y ligadura, se puede evitar que los extremos 3' de estos oligonucleótidos

adaptadores ceban cualquier evento de extensión. Esto puede minimizar el cebado incorrecto u otros eventos de extensión espurios durante la producción de oligonucleótidos con códigos de barras. Antes de usar los reactivos de códigos de barras multiméricos ensamblados, el resto terminador del terminador reversible se puede eliminar por medios químicos o de otro medio, permitiendo así que la región diana se extienda a lo largo de un molde de ácido nucleico diana con el que hibrida.

De manera similar, para cualquiera de los métodos que involucran oligonucleótidos adaptadores, se pueden emplear uno o más oligonucleótidos de bloqueo complementarios a una o más secuencias dentro de la región o regiones diana durante las reacciones de extensión y/o extensión y ligadura. Los oligonucleótidos de bloqueo pueden comprender un terminador y/u otro resto en sus extremos 3' y/o 5' de manera que no puedan extenderse mediante polimerasas. Los oligonucleótidos de bloqueo pueden diseñarse de modo que hibriden con secuencias total o parcialmente complementarias con una o más regiones diana, e hibridan con dichas regiones diana antes de una reacción de extensión y/o extensión y ligadura. El uso de cebadores de bloqueo puede evitar que las regiones diana hibriden con secuencias dentro de la solución y potencialmente se ceban erróneamente a lo largo de las secuencias para las que no se desea tal hibridación (por ejemplo, características de secuencia dentro de las propias moléculas de códigos de barras). Los oligonucleótidos de bloqueo se pueden diseñar para lograr temperaturas de hibridación y/o fusión particulares. Antes de usar los reactivos de códigos de barras multiméricos ensamblados, los oligonucleótidos de bloqueo se pueden eliminar mediante, por ejemplo, desnaturalización por calor y luego limpieza selectiva por tamaño, u otros medios. La eliminación del oligonucleótido o los oligonucleótidos de bloqueo puede permitir que la región diana se extienda a lo largo de un molde de ácido nucleico diana con el que hibrida.

El método puede comprender sintetizar un reactivo de código de barras multimérico que comprende al menos 5, al menos 10, al menos 20, al menos 25, al menos 50, al menos 75, o al menos 100 moléculas de código de barras y en donde: (a) cada molécula de código de barras es tal como se define en el presente documento; y (b) se sintetiza un producto de extensión con código de barras a partir de cada molécula de código de barras de acuerdo con cualquier método definido en el presente documento; y, opcionalmente, (c) se liga un oligonucleótido adaptador a cada uno de los productos de extensión con código de barras para producir oligonucleótidos con código de barras de acuerdo con cualquiera de los métodos definidos en el presente documento.

La divulgación proporciona además un método para sintetizar una biblioteca de reactivos de códigos de barras multiméricos, en donde el método comprende repetir las etapas de cualquiera de los métodos definidos en el presente documento para sintetizar dos o más reactivos de códigos de barras multiméricos. Opcionalmente, el método comprende sintetizar una biblioteca de al menos 5, al menos 10, al menos 20, al menos 25, al menos 50, al menos 75, al menos 100, al menos 250, al menos 500, al menos 10^3 , al menos 10^4 , al menos 10^5 , al menos 10^6 , al menos 10^7 , al menos 10^8 , al menos 10^9 o al menos 10^{10} reactivos de códigos de barras multiméricos tal como se define en el presente documento. Preferentemente, la biblioteca comprende al menos 5 reactivos de códigos de barras multiméricos tal como se define en el presente documento. Preferentemente, las regiones de códigos de barras de cada uno de los reactivos de códigos de barras multiméricos pueden ser diferentes a las regiones de códigos de barras de los otros reactivos de códigos de barras multiméricos.

La Figura 8 ilustra un método para sintetizar un reactivo de códigos de barras multimérico para etiquetar un ácido nucleico diana. En este método, las moléculas de código de barras primera (D1, E1 y F1) y segunda (D2, E2 y F2), cada una de las cuales incluye una secuencia de ácido nucleico que comprende una región de código de barras (E1 y E2), y que están unidas por una secuencia de ácido nucleico de conexión (S), se desnaturalizan en forma monocatenaria. A estas moléculas de código de barras monocatenarias, un primer y segundo cebador de extensión (A1 y A2) hibrida con la región 3' de la primera y la segunda moléculas de código de barras (D1 y D2), y un primer y segundo cebador de bloqueo (R1 y R2) hibrida con la región adaptadora de 5' (F1 y F2) de la primera y segunda moléculas de código de barras. Estos cebadores de bloqueo (R1 y R2) se pueden modificar en el extremo 3' de manera que no puedan servir como sitio de cebado para una polimerasa.

Luego se usa una polimerasa para realizar una reacción de extensión del cebador, en donde los cebadores de extensión se extienden para hacer una copia (B1 y B2) de la región del código de barras de las moléculas del código de barras (E1 y E2). Esta reacción de extensión del cebador se realiza de manera que el producto de extensión termine inmediatamente adyacente a la secuencia del cebador de bloqueo, por ejemplo mediante el uso de una polimerasa que carece de desplazamiento de cadena o actividad exonucleasa 5'-3'. A continuación, se eliminan los cebadores de bloqueo (R1 y R2), por ejemplo, mediante desnaturalización a alta temperatura.

Este método crea así un reactivo de código de barras multimérico que contiene una primera y una segunda unión de ligadura (J1 y J2) adyacentes a una región adaptadora monocatenaria (F1 y F2). Este reactivo de códigos de barras multimérico se puede utilizar en el método ilustrado en la Figura 5.

El método puede comprender además la etapa de ligar el extremo 3' del primer y segundo oligonucleótidos con código de barras creado por la etapa de extensión del cebador (el extremo 3' de B1 y B2) al primero (C1 y G1) y al segundo (C2 y G2) oligonucleótidos adaptadores, en donde cada oligonucleótido adaptador comprende una región adaptadora (C1 y C2) que es complementaria y, por tanto, capaz de hibridar con, la región adaptadora de una molécula de código de barras (F1 y F2). Los oligonucleótidos adaptadores se pueden sintetizar para incluir un grupo fosfato 5' terminal.

Cada oligonucleótido adaptador también puede comprender una región diana (G1 y G2), que se puede usar para hibridar los oligonucleótidos con código de barras con ácidos nucleicos diana, y se puede usar por separado o posteriormente como cebadores para una reacción de extensión de cebador o una reacción en cadena de la polimerasa. La etapa de ligar el primer y segundo oligonucleótidos con código de barras a los oligonucleótidos adaptadores produce un reactivo de código de barras multimérico tal como se ilustra en la Figura 1 que se puede usar en los métodos ilustrados en la Figura 3 y/o la Figura 4.

La Figura 9 muestra un método para sintetizar reactivos de códigos de barras multiméricos (tal como se ilustra en la Figura 1) para etiquetar un ácido nucleico diana. En este método, las moléculas de código de barras primera (D1, E1 y F1) y segunda (D2, E2 y F2), cada una de las cuales incluye una secuencia de ácido nucleico que comprende una región de código de barras (E1 y E2), y que están unidas por una secuencia de ácido nucleico de conexión (S), se desnaturalizan en forma monocatenaria. A estas moléculas de código de barras monocatenarias, un primer y segundo cebador de extensión (A1 y A2) hibrida con la región 3' de la primera y segunda moléculas de código de barras (D1 y D2), y las regiones adaptadoras (C1 y C2) del primer (C1 y G1) y el segundo (C2 y G2) oligonucleótidos adaptadores hibridan con las regiones adaptadoras 5' (F1 y F2) de la primera y segunda moléculas de código de barras. Estos oligonucleótidos adaptadores se pueden sintetizar para incluir un grupo fosfato 5' terminal.

Luego se usa una polimerasa para realizar una reacción de extensión del cebador, en donde los cebadores de extensión se extienden para hacer una copia (B1 y B2) de la región del código de barras de las moléculas del código de barras (E1 y E2). Esta reacción de extensión del cebador se realiza de manera que el producto de extensión termine inmediatamente adyacente a la secuencia de la región adaptadora (C1 y C2), por ejemplo mediante el uso de una polimerasa que carece de desplazamiento de cadena o actividad exonucleasa 5'-3'.

A continuación, se usa una enzima ligasa para ligar el extremo 5' de los oligonucleótidos adaptadores al extremo 3' adyacente del producto de extensión correspondiente. En una realización alternativa, se puede incluir una enzima ligasa con la enzima polimerasa en una reacción que afecta simultáneamente tanto la extensión del cebador como la unión del producto resultante al oligonucleótido adaptador. A través de este método, los oligonucleótidos con código de barras resultantes se pueden usar posteriormente como cebadores para una reacción de extensión de cebador o una reacción en cadena de la polimerasa, por ejemplo, tal como en el método mostrado en la Figura 3 y/o la Figura 4.

La divulgación proporciona además un método para sintetizar un reactivo de código de barras multimérico que comprende añadir uno o más reactivos de código de barras multiméricos (donadores) a un soporte. Las moléculas de hibridación multiméricas (por ejemplo, moléculas de código de barras multiméricas) se pueden añadir a un soporte. Adicionalmente o como alternativa, los oligonucleótidos con código de barras, que puede haber sido sintetizado a partir de una molécula de código de barras multimérica, puede adjuntarse a un soporte. El soporte puede ser cualquier soporte descrito en el presente documento, por ejemplo, una macromolécula, soporte sólido o soporte semisólido. El soporte se puede seleccionar basándose en las propiedades estructurales y/o funcionales deseadas del reactivo de códigos de barras multimérico. Por ejemplo: se pueden añadir oligonucleótidos con código de barras a perlas magnéticas. Esto puede permitir que un científico de laboratorio manipule fácilmente los oligonucleótidos con código de barras, por ejemplo para realizar etapas de lavado o etapas de purificación. Además, las propiedades funcionales de la perla pueden permitir a un científico aislar o purificar ácidos nucleicos de una muestra de ácido nucleico que puede hibridar y/o poner códigos de barras con los oligonucleótidos con código de barras. Además, la adición de oligonucleótidos con código de barras a un soporte puede cambiar la naturaleza estructural general de los oligonucleótidos con código de barras. Por ejemplo, la adición de oligonucleótidos con código de barras a un tetrámero de estreptavidina puede cambiar la estructura tridimensional de los oligonucleótidos con código de barras, de modo que se reduce la hibridación cruzada entre las regiones diana de diferentes oligonucleótidos con código de barras, reduciendo así la cantidad de posible cebado erróneo entre oligonucleótidos con código de barras y/o mejorando la accesibilidad de las regiones diana a los posibles ácidos nucleicos diana dentro de una muestra.

20. MÉTODOS DE SECUENCIACIÓN Y/O PROCESAMIENTO DE DATOS DE SECUENCIACIÓN

La divulgación proporciona además un método para secuenciar una muestra, en donde la muestra se ha preparado mediante cualquiera de los métodos de preparación de una muestra de ácido nucleico para secuenciar tal como se define en el presente documento. El método de secuenciación de la muestra comprende las etapas de: aislar las moléculas de ácido nucleico diana con código de barras y producir una secuencia leída de cada molécula de ácido nucleico diana con código de barras que comprende la región del código de barras, la región diana y al menos un nucleótido adicional del ácido nucleico diana. Cada secuencia leída puede comprender al menos 5, al menos 10, al menos 25, al menos 50, al menos 100, al menos 250, al menos 500, al menos 1000, al menos 2000, al menos 5000 o al menos 10000 nucleótidos del ácido nucleico diana. Preferentemente, cada secuencia leída comprende al menos 5 nucleótidos del ácido nucleico diana.

Los métodos pueden producir una secuencia leída de una o más moléculas de ácido nucleico diana con código de barras producidas a partir de al menos 10, al menos 100 o al menos 10^3 , al menos 10^4 , al menos 10^5 , al menos 10^6 , al menos 10^7 , al menos 10^8 o al menos 10^9 ácidos nucleicos diana diferentes.

La secuenciación se puede realizar mediante cualquier método conocido en la técnica. Por ejemplo, por terminación de cadena o secuenciación de Sanger. Preferentemente, la secuenciación se realiza mediante un método de

secuenciación de próxima generación, tal como la secuenciación por síntesis, la secuenciación por síntesis utilizando terminadores reversibles (por ejemplo, secuenciación de Illumina), pirosecuenciación (por ejemplo, secuenciación 454), secuenciación por ligadura (por ejemplo, secuenciación SOLiD), secuenciación de una sola molécula (por ejemplo, secuenciación de una sola molécula, tiempo real (SMRT, por sus siglas en inglés), Pacific Biosciences), o mediante secuenciación de nanoporos (por ejemplo, en las plataformas Minion o Promethion, Oxford Nanopore Technologies).

La divulgación proporciona además un método para procesar datos de secuenciación obtenidos mediante cualquiera de los métodos definidos en el presente documento. El método para procesar datos de secuencia comprende las etapas de: (a) identificar, para cada lectura de secuencia, la secuencia de la región del código de barras y la secuencia del ácido nucleico diana; y (b) usar la información de la etapa (a) para determinar un grupo de secuencias del ácido nucleico diana que se etiquetaron con regiones de código de barras del mismo reactivo de código de barras multimérico.

El método puede comprender además la etapa de determinar una secuencia de un ácido nucleico diana analizando el grupo de secuencias para identificar secuencias contiguas, en donde la secuencia del ácido nucleico diana comprende nucleótidos de al menos dos lecturas de secuencia.

El ácido nucleico diana puede ser una molécula de ácido nucleico intacta, fragmentos colocalizados de una molécula de ácido nucleico, o moléculas de ácido nucleico de una sola célula. Preferentemente, el ácido nucleico diana es una única molécula de ácido nucleico intacta, dos o más fragmentos colocalizados de una única molécula de ácido nucleico, o dos o más moléculas de ácido nucleico de una única célula.

La divulgación proporciona además un algoritmo para procesar (o analizar) los datos de secuenciación obtenidos mediante cualquiera de los métodos definidos en el presente documento. El algoritmo se puede configurar para realizar cualquiera de los métodos para procesar datos de secuencia definidos en el presente documento. El algoritmo se puede usar para detectar la secuencia de una región de código de barras dentro de cada lectura de secuencia, y también para detectar la secuencia dentro de una lectura de secuencia que se obtiene de un ácido nucleico diana, y para separarlos en dos conjuntos de datos asociados.

La divulgación proporciona además un método para generar una lectura larga sintética a partir de un ácido nucleico diana que comprende las etapas de: (a) preparar una muestra de ácido nucleico para secuenciar de acuerdo con cualquiera de los métodos definidos en el presente documento; (b) secuenciar la muestra, en donde, opcionalmente, la muestra se secuenciaría mediante cualquiera de los métodos definidos en el presente documento; y (c) procesar los datos de secuencia obtenidos en la etapa (b), opcionalmente, en donde los datos de secuencia se procesan de acuerdo con cualquiera de los métodos definidos en el presente documento; en donde la etapa (c) genera una lectura larga sintética que comprende al menos un nucleótido de cada una de las al menos dos lecturas de secuencia.

El método puede permitir la fase de una secuencia diana de una molécula de ácido nucleico diana, es decir, puede permitir la determinación de qué copia de un cromosoma (es decir, paterno o materno) está localizada la secuencia. La secuencia diana puede comprender una mutación, translocación, delección o amplificación diana específica y el método puede usarse para asignar la mutación, translocación, delección o amplificación a un cromosoma específico. La puesta en fase de dos o más secuencias diana también puede permitir la detección de aneuploidía.

La lectura larga sintética puede comprender al menos 50, al menos 100, al menos 250, al menos 500, al menos 750, al menos 1000, al menos 2000, al menos 10^4 , al menos 10^5 , al menos 10^6 , al menos 10^7 o al menos 10^8 nucleótidos. Preferentemente, la lectura larga sintética comprende al menos 50 nucleótidos.

La divulgación proporciona además un método de secuenciación de dos o más ácidos nucleicos diana colocalizados que comprende las etapas de: (a) preparar una muestra de ácido nucleico para secuenciar de acuerdo con cualquiera de los métodos definidos en el presente documento; (b) secuenciar la muestra, en donde, opcionalmente, la muestra se secuenciaría mediante cualquiera de los métodos definidos en el presente documento; y (c) procesar los datos de secuencia obtenidos en la etapa (b), opcionalmente, en donde los datos de secuencia se procesan de acuerdo con cualquiera de los métodos definidos en el presente documento; en donde la etapa (c) identifica al menos dos lecturas de secuencia que comprenden nucleótidos de al menos dos ácidos nucleicos diana colocalizados en la muestra.

La divulgación proporciona además un método de secuenciación de ácidos nucleicos diana de una célula individual que comprende las etapas de: (a) preparar una muestra de ácido nucleico para secuenciar de acuerdo con cualquiera de los métodos definidos en el presente documento, en donde el o los reactivos de códigos de barras multiméricos, o las moléculas de códigos de barras multiméricos y/o los oligonucleótidos adaptadores se introducen en la célula; (b) secuenciar la muestra, en donde, opcionalmente, la muestra se secuenciaría mediante cualquiera de los métodos definidos en el presente documento; y (c) procesar los datos de secuencia obtenidos en la etapa (b), opcionalmente, en donde los datos de secuencia se procesan de acuerdo con cualquiera de los métodos definidos en el presente documento; en donde la etapa (c) identifica al menos dos lecturas de secuencia que comprenden nucleótidos de al menos dos ácidos nucleicos diana de la célula.

El o los reactivos de códigos de barras multiméricos y/o los oligonucleótidos adaptadores se pueden introducir en la célula mediante complejación química con un reactivo de transfección de lípidos y luego transfección en la célula.

Los reactivos de códigos de barras multiméricos y/o los oligonucleótidos adaptadores se pueden introducir en la célula mediante las etapas de: (a) permeabilizar la membrana celular poniéndola en contacto con un tensioactivo químico; y luego (b) poner en contacto la célula con el o los reactivos de códigos de barras multiméricos y/o los oligonucleótidos adaptadores. El tensioactivo químico puede ser un tensioactivo no iónico. El tensioactivo químico puede ser Tritón X-100 ($C_{14}H_{22}O(C_2H_4O)_n$ ($n=9-10$)). El tensioactivo químico puede estar en solución a una concentración de menos de 200 micromolar, o menos de 500 micromolar, o menos de 1 milimolar.

En el método, tras la etapa de introducir el o los reactivos de códigos de barras multiméricos y/o los oligonucleótidos adaptadores en la célula, la célula se puede incubar durante un período de tiempo para permitir que las regiones diana de los reactivos de códigos de barras multiméricos u oligonucleótidos adaptadores hibriden con subsecuencias de los ácidos nucleicos diana dentro de la célula. El período de incubación puede ser de al menos 1 minuto, o al menos 5 minutos, o al menos 15 minutos, o al menos 30 minutos, o al menos 60 minutos. Preferentemente, el período de incubación es de al menos 1 minuto. La incubación puede tener lugar dentro de una solución que contiene un desnaturalizante de ácido nucleico, por ejemplo, dimetilsulfóxido (DMSO) o betaína. La incubación puede tener lugar a una temperatura de al menos 20 grados Celsius, al menos 37 grados Celsius, al menos 45 grados Celsius, o al menos 50 grados Celsius. Preferentemente, la incubación tiene lugar a una temperatura de al menos 20 grados Celsius.

En los métodos que implican el uso de reactivos de códigos de barras multiméricos, la etapa de incubación puede disociar sustancialmente los oligonucleótidos con código de barras de las moléculas de código de barras (o molécula de código de barras multimérica). Esto puede permitir que los oligonucleótidos con código de barras se difundan más fácilmente a través de la célula mejorando la eficacia con la que las regiones diana de los oligonucleótidos con código de barras son capaces de hibridar con subsecuencias de los ácidos nucleicos diana.

En el método, después de la introducción de los reactivos de códigos de barras multiméricos y/u oligonucleótidos adaptadores en la célula y, opcionalmente, después de la etapa de incubación, la célula puede ponerse en contacto con una solución de oligonucleótidos complementaria a las regiones diana de los reactivos de códigos de barras multiméricos.

En el método, después de la introducción de los reactivos de códigos de barras multiméricos y/u oligonucleótidos adaptadores en la célula y, opcionalmente, después de la etapa de incubación, la célula se puede aislar de una mezcla de reacción, por ejemplo, por centrifugación.

En el método, después de la introducción de los reactivos de códigos de barras multiméricos y/u oligonucleótidos adaptadores en la célula y, opcionalmente, después de la etapa de incubación, los oligonucleótidos con código de barras y/o moléculas de ácido nucleico diana con código de barras y/o reactivo(s) de código de barras multimérico se pueden aislar de la célula.

Los reactivos de códigos de barras multiméricos, los oligonucleótidos de código de barras y/o los oligonucleótidos adaptadores pueden comprender uno o más restos de biotina.

En el método, después de la introducción de los reactivos de códigos de barras multiméricos y/u oligonucleótidos adaptadores en la célula y, opcionalmente, después de la etapa de incubación, los oligonucleótidos con código de barras y/o moléculas de ácido nucleico diana con código de barras y/o reactivo(s) de código de barras multimérico se pueden aislar por un proceso de: (a) opcionalmente disolver las membranas celulares, por ejemplo, usando un tensioactivo químico o mediante incubación a alta temperatura; (b) poner en contacto la mezcla resultante con un soporte sólido, opcionalmente en donde el soporte sólido comprende restos de estreptavidina; y (c) capturar los oligonucleótidos con código de barras y/o las moléculas de ácido nucleico diana con código de barras y/o los reactivos de códigos de barras multiméricos en el soporte sólido, opcionalmente a través de la interacción estreptavidina-biotina. El soporte sólido puede ser una o más perlas magnéticas, opcionalmente en donde una o más perlas magnéticas comprenden moléculas de estreptavidina en su superficie. Las perlas magnéticas pueden aislarse de una mezcla de reacción con un imán.

Los ácidos nucleicos diana pueden ser moléculas de ADN (por ejemplo, moléculas de ADN genómico) o moléculas de ARN (por ejemplo, moléculas de ARNm).

Preferentemente, cada molécula de ácido nucleico diana con código de barras se produce después del aislamiento del oligonucleótido con código de barras hibridado con una molécula de ARNm diana extendiendo el oligonucleótido con código de barras usando una transcriptasa inversa y la molécula de ARNm diana como molde.

Las moléculas de ARNm pueden ser moléculas de ARNm correspondientes a cadenas alfa y/o beta de una secuencia de receptor de linfocitos T, opcionalmente, en donde se determinan las secuencias de cadenas alfa y beta emparejadas dentro de una célula individual.

Las moléculas de ARNm pueden ser moléculas de ARNm correspondientes a cadenas ligeras y/o pesadas de una secuencia de inmunoglobulina, en donde opcionalmente se determinan las secuencias de cadenas ligeras y pesadas emparejadas dentro de una célula individual.

El método se puede usar para secuenciar ácidos nucleicos diana en al menos 10, al menos 100 o al menos 10^3 , al menos 10^4 , al menos 10^5 , al menos 10^6 , al menos 10^7 , al menos 10^8 o al menos 10^9 células. Preferentemente, el método se puede usar para secuenciar ácidos nucleicos diana en al menos 10 células. Preferentemente, las células son linfocitos T y/o linfocitos B.

- 5 Cualquier método de análisis de moléculas de ácido nucleico con código de barras mediante secuenciación (por ejemplo, para generar lecturas largas sintéticas o para analizar secuencias de ácido nucleico de células individuales) puede comprender una reacción de secuenciación redundante, en donde las moléculas de ácido nucleico diana a las que se han puesto código de barras en una reacción de código de barras se secuencian dos o más veces dentro de una reacción de secuenciación. Opcionalmente, cada molécula con código de barras de una muestra puede ser
10 secuenciada, en promedio, al menos dos veces, al menos 3 veces, al menos 5 veces, al menos 10 veces, al menos 20 veces, al menos 50 veces o al menos 100 veces.

- En cualquier método de análisis de moléculas de ácido nucleico con código de barras mediante secuenciación (por ejemplo, para generar lecturas largas sintéticas o para analizar secuencias de ácido nucleico de células individuales), se puede emplear un proceso de corrección de errores. Este proceso comprende las etapas de: (i) determinar dos o
15 más lecturas de secuencia de un conjunto de datos de secuenciación que comprende la misma secuencia de código de barras, y (ii) alinear las secuencias de dichas dos o más lecturas de secuencia entre sí. Opcionalmente, este proceso de corrección de errores puede comprender además una etapa de (iii) determinar un nucleótido mayoritario y/o más común y/o más probable en cada posición dentro de la lectura de secuencia y/o en cada posición dentro de la secuencia de la molécula de ácido nucleico diana. Esta etapa puede comprender opcionalmente establecer una
20 secuencia consenso de cada secuencia de ácido nucleico diana mediante cualquier proceso de corrección de errores, eliminación de errores, detección de errores, recuento de errores o eliminación de errores estadísticos. Esta etapa puede comprender además la etapa de colapsar múltiples lecturas de secuencia que comprenden la misma secuencia de código de barras en una representación que comprende una sola lectura de errores corregidos. Opcionalmente, cualquier etapa para determinar dos o más lecturas de secuencia de un conjunto de datos de secuenciación que
25 comprende la misma secuencia de código de barras, puede comprender determinar lecturas de secuencia que comprenden secuencias de códigos de barras con al menos un cierto grado de nucleótidos idénticos y/o similitud de secuencia, por ejemplo, al menos el 70 %, al menos el 80 %, al menos el 90 %, o al menos el 95 % de similitud de secuencia (por ejemplo, permitiendo desajustes y/o inserciones o deleciones en cualquier punto entre las secuencias de códigos de barras).

- 30 En cualquier método de análisis de moléculas de ácido nucleico con código de barras mediante secuenciación (por ejemplo, para generar lecturas largas sintéticas o para analizar secuencias de ácido nucleico de células individuales), se puede emplear un proceso alternativo de corrección de errores, que comprende las etapas de: (i) determinar dos o más lecturas de secuencia de un conjunto de datos de secuenciación que comprenden la misma secuencia de ácido nucleico diana, en donde dichas dos o más lecturas de secuencia comprenden además dos o más secuencias de
35 códigos de barras diferentes, en donde las secuencias de código de barras son de la misma molécula de código de barras multimérica y/o reactivo de código de barras multimérico, y (ii) alinear las secuencias de dichas dos o más lecturas de secuencia entre sí. Opcionalmente, este proceso de corrección de errores puede comprender además una etapa de (iii) determinar un nucleótido mayoritario y/o más común y/o más probable en cada posición dentro de la secuencia de la molécula de ácido nucleico diana. Esta etapa puede comprender opcionalmente establecer una
40 secuencia consenso de la molécula de ácido nucleico diana mediante cualquier proceso de corrección de errores, eliminación de errores, detección de errores, recuento de errores o eliminación de errores estadísticos. Esta etapa puede comprender además la etapa de colapsar múltiples lecturas de secuencia que comprenden la misma molécula de ácido nucleico diana en una representación que comprende una sola lectura con corrección de errores. La molécula de ácido nucleico diana puede comprender, por ejemplo, una secuencia de ADN genómico; como alternativa, la
45 molécula de ácido nucleico diana puede comprender la totalidad o parte de una secuencia de ARN mensajero tal como un gen expresado o una cadena de receptor inmunitario adaptativo expresada. Opcionalmente, cualquier etapa de comparar dos secuencias de códigos de barras y/o comparar una secuencia de códigos de barras secuenciada y una secuencia de códigos de barras de referencia, puede comprender la determinación de secuencias que comprenden al menos una cierta extensión de nucleótidos idénticos y/o similitud de secuencia, por ejemplo, al menos el 70 %, al
50 menos el 80 %, al menos el 90 %, o al menos el 95 % de similitud de secuencia (por ejemplo, permitiendo desajustes y/o inserciones o deleciones en cualquier punto entre las secuencias de códigos de barras).

- En cualquier método de análisis de moléculas de ácido nucleico con código de barras mediante secuenciación, el número de secuencias de códigos de barras agregadas a dianas de ácidos nucleicos específicas por cualquier reactivo de códigos de barras multimérico dado, y/o en un grupo de dos o más reactivos de códigos de barras multiméricos
55 diferentes, se puede cuantificar. Por ejemplo, se puede determinar el número de secuencias de códigos de barras diferentes de un reactivo de códigos de barras multimérico adjunto a un transcrito de ARN mensajero particular (o cualquier otro ácido nucleico diana específico) de una sola célula. Se puede cuantificar cualquier tipo de diana de ácido nucleico específico, tal como cualquier transcrito, cualquier secuencia de ADN genómico, cualquier secuencia de código de barras sintética, cualquier cadena de receptor inmune adaptativo y/o secuencia de receptor inmune, o
60 cualquier secuencia de mutación específica. Cualquier proceso de cuantificación de este tipo se puede repetir para cualquier número de dianas de ácido nucleico específicas y/o grupos de las mismas.

21. USOS DE UN REACTIVO DE CÓDIGO DE BARRAS MULTIMÉRICO, BIBLIOTECA O KIT

La divulgación proporciona además el uso de un reactivo de códigos de barras multimérico tal como se define en el presente documento, una biblioteca de reactivos de códigos de barras multiméricos tal como se define en el presente documento, o un kit tal como se define en el presente documento, para producir dos o más lecturas de secuencia a partir de un ácido nucleico diana, en donde dos o más lecturas de secuencia se pueden identificar como que provienen del mismo ácido nucleico diana y se pueden combinar para producir una lectura larga sintética.

5

La divulgación proporciona además el uso de un reactivo de códigos de barras multimérico tal como se define en el presente documento, una biblioteca de reactivos de códigos de barras multiméricos tal como se define en el presente documento, o un kit tal como se define en el presente documento, para etiquetar una muestra de ácido nucleico embebida en parafina y fijada en formalina (EPFF), en donde el reactivo de código de barras multimérico o los componentes del kit se introducen en la muestra y se usan para marcar un conjunto de dos o más ácidos nucleicos diana colocalizados para secuenciación.

10

Los reactivos de códigos de barras multiméricos para su uso en el etiquetado de una muestra de ácido nucleico EPFF pueden ser inferiores a 10 kb, menos de 5 kb, menos de 2 kb, menos de 1 kb de longitud o menos de 500 pb. Preferentemente, los reactivos de códigos de barras multiméricos tienen menos de 1 kb de longitud.

15

La divulgación proporciona además el uso de un reactivo de códigos de barras multimérico tal como se define en el presente documento, una biblioteca de reactivos de códigos de barras multiméricos tal como se define en el presente documento, o un kit tal como se define en el presente documento, para etiquetar los ácidos nucleicos diana en una célula individual, en donde el reactivo de código de barras multimérico o los componentes del kit se introducen en una célula y se utilizan para etiquetar un conjunto de dos o más ácidos nucleicos diana en la célula para su secuenciación.

20

La divulgación proporciona además el uso de un reactivo de códigos de barras multimérico tal como se define en el presente documento, una biblioteca de reactivos de códigos de barras multiméricos tal como se define en el presente documento, o un kit tal como se define en el presente documento, para etiquetar los ácidos nucleicos diana en una muestra de plasma o suero humano, en donde el reactivo de código de barras multimérico o los componentes del kit se utilizan para etiquetar un conjunto de dos o más ácidos nucleicos diana en el plasma o suero para secuenciación.

25

Breve descripción de los dibujos

La invención, junto con otros objetos y ventajas de la misma, puede entenderse mejor haciendo referencia a la descripción tomada junto con los dibujos adjuntos, en donde:

La Figura 1 ilustra un reactivo de código de barras multimérico que se puede usar en el método ilustrado en la Figura 3 o la Figura 4.

30

La Figura 2 ilustra un kit que comprende un reactivo de código de barras multimérico y oligonucleótidos adaptadores para etiquetar un ácido nucleico diana.

La Figura 3 ilustra un método para preparar una muestra de ácido nucleico para secuenciar usando un reactivo de código de barras multimérico.

35

La Figura 4 ilustra un segundo método para preparar una muestra de ácido nucleico para secuenciar usando un reactivo de código de barras multimérico.

La Figura 5 ilustra un método para preparar una muestra de ácido nucleico para secuenciar usando un reactivo de código de barras multimérico y oligonucleótidos adaptadores.

La Figura 6 ilustra un método para preparar una muestra de ácido nucleico para secuenciar usando un reactivo de código de barras multimérico, oligonucleótidos adaptadores y oligonucleótidos diana.

40

La Figura 7 ilustra un método para ensamblar una molécula de código de barras multimérica usando un proceso de amplificación por círculo rodante.

La Figura 8 ilustra un método para sintetizar reactivos de códigos de barras multiméricos para etiquetar un ácido nucleico diana que se puede usar en los métodos ilustrados en la Figura 3, la Figura 4 y/o la Figura 5.

45

La Figura 9 ilustra un método alternativo de sintetizar reactivos de códigos de barras multiméricos (tal como se ilustra en la Figura 1) para etiquetar un ácido nucleico diana que se puede usar en el método ilustrado en la Figura 3 y/o la Figura 4.

La Figura 10 es un gráfico que muestra el número total de nucleótidos dentro de cada secuencia de código de barras.

La Figura 11 es un gráfico que muestra el número total de moléculas de código de barras únicas en cada molécula de código de barras multimérica secuenciada.

50

La Figura 12 muestra moléculas de código de barras multiméricas representativas que fueron detectadas por el comando de análisis.

La Figura 13 es un gráfico que muestra el número de códigos de barras únicos por identificador de secuencia molecular frente al número de identificadores de secuencia molecular siguiendo el código de barras de los moldes de ADN sintético de secuencia conocida con reactivos de códigos de barras multiméricos que contienen oligonucleótidos con código de barras.

- 5 La Figura 14 es un gráfico que muestra el número de códigos de barras únicos por identificador de secuencia molecular frente al número de identificadores de secuencia molecular siguiendo el código de barras de los moldes de ADN sintético de secuencia conocida con reactivos de códigos de barras multiméricos y oligonucleótidos adaptadores separados.

- 10 La Figura 15 es una tabla que muestra los resultados del código de barras de los loci de ADN genómico de tres genes humanos (BRCA1, HLA-A y DQB1) con reactivos de códigos de barras multiméricos que contienen oligonucleótidos con códigos de barras.

La Figura 16 es una ilustración esquemática de una lectura de secuencia obtenida a partir de loci de ADN genómico de códigos de barras con reactivos de códigos de barras multiméricos que contienen oligonucleótidos con códigos de barras.

- 15 La Figura 17 es un gráfico que muestra el número de códigos de barras del mismo reactivo de código de barras multimérico que etiquetó secuencias en la misma molécula de molde sintético frente al número de moléculas de molde sintéticas.

La Figura 18 ilustra ejemplos de reactivos de códigos de barras multiméricos que comprenden restos de unión a células.

- 20 La Figura 19 ilustra un método de transferencia de reactivos de códigos de barras multiméricos a células mediante restos de unión a células.

La Figura 20 ilustra un método para transferir reactivos de códigos de barras multiméricos a células mediante administración liposomal.

- 25 La Figura 21 ilustra un método para transferir reactivos de códigos de barras multiméricos a células mediante transfección.

La Figura 22 ilustra un método de transferencia de reactivos de códigos de barras multiméricos a células mediante un proceso de permeabilización.

La Figura 23 ilustra un método de colocación de códigos de barras de ácidos nucleicos celulares con una etapa de permeabilización de la membrana.

- 30 La Figura 24 ilustra un método de colocación de códigos de barras de ácidos nucleicos celulares con una etapa de permeabilización de la membrana y liberación de oligonucleótidos con código de barras.

- 35 La Figura 18 ilustra ejemplos de reactivos de códigos de barras multiméricos que comprenden restos de unión a células. La figura muestra dos variantes esquemáticas diferentes de un reactivo de código de barras multimérico que comprende restos de unión a células. En una primera realización de este tipo (izquierda), una serie de restos de unión a células se unen a un soporte (tal como una perla o una molécula de ácido nucleico), y también se unen al soporte varios oligonucleótidos con códigos de barras. Los restos de unión a células pueden comprender cualquier tipo de molécula o compuesto capaz de interactuar preferentemente con las superficies celulares, tales como anticuerpos o aptámeros que tienen afinidad por proteínas específicas en la superficie de las células, o moléculas de carga tales como restos de polilisina que tienen afinidad electrostática por la membrana celular cargada. La unión de tales restos de unión a células y oligonucleótidos con códigos de barras al soporte puede ser directa (por ejemplo, mediante complejación química covalente directa), puede ser no covalente (por ejemplo, a través de interacciones proteína-proteína) y/o puede ser indirecto, tal como la participación de moléculas de unión secundarias.

- 45 En una segunda realización (a la derecha), una serie de restos de unión a células se añaden a un soporte, al igual que varias moléculas enlazadoras que comprenden una secuencia de ácido nucleico. Estas moléculas enlazadoras se pueden unir directamente al soporte (por ejemplo, mediante complejación química) o mediante cualquier otra unión indirecta y/o no covalente. Un oligonucleótido con código de barras hibrida con la secuencia de ácido nucleico de cada molécula enlazadora, formando así una unión indirecta de cada oligonucleótido con código de barras al soporte dentro del reactivo de código de barras multimérico global. La región de hibridación formada entre las moléculas enlazadoras y los oligonucleótidos con código de barras puede permitir además la manipulación de la interacción entre los oligonucleótidos con código de barras y el soporte; por ejemplo, se puede usar un proceso de incubación a alta temperatura para desnaturalizar la región de hibridación y así permitir que los oligonucleótidos con código de barras se difundan en solución desde el propio soporte.

La Figura 19 ilustra un ejemplo de un método de transferencia de reactivos de códigos de barras multiméricos a células mediante restos de unión a células. En el método, los reactivos de códigos de barras multiméricos se transfieren a las

células mediante un proceso de transferencia que implica restos de unión a células. Estos restos de unión a células pueden comprender cualquier tipo de resto molecular, macromolecular y/o sólido que sea capaz de interactuar preferentemente con una célula. Por ejemplo, esto puede comprender un anticuerpo capaz de unirse a una proteína específica en la superficie celular; como alternativa, por ejemplo, esto puede comprender una macromolécula catiónica tal como un resto de polilisina que interactúa preferentemente con la superficie celular por atracción electrostática.

En la primera etapa, una biblioteca de dos o más reactivos de códigos de barras multiméricos que cada uno comprende uno o más restos de unión a células, se incuban con una muestra de células durante un período de tiempo, durante cuyo tiempo los reactivos de códigos de barras multiméricos migran para entrar en contacto con una membrana celular y se unen a dicha membrana celular a través de uno o más restos de unión a células asociados.

En una segunda etapa después de esta etapa de unión celular, la muestra de células unidas a reactivos de códigos de barras multiméricos se incuba durante un período de tiempo, durante ese tiempo, los reactivos de códigos de barras multiméricos se transfieren a las células. Este proceso de transferencia se puede efectuar mediante uno o más procesos conocidos de células que internalizan constituyentes unidos a o dentro de su membrana celular, tal como la endocitosis, pinocitosis y/o fagocitosis. En esta ilustración, un primer complejo de lípido-reactivo de código de barras multimérico se transfiere a una primera célula, y un segundo complejo de lípido-reactivo de código de barras multimérico se transfiere a una segunda célula; en realizaciones reales, se puede transferir una gran biblioteca de reactivos de códigos de barras multiméricos a una gran muestra de células.

Siguiendo esta etapa de transferencia, se realiza una etapa de incubación, durante cuyo tiempo se permite que las moléculas de ARN mensajero complementarias a las regiones diana de los oligonucleótidos con código de barras comprendidas dentro de los reactivos de códigos de barras multiméricos transferidos hibriden con dichas regiones diana. Esta incubación se puede realizar a una temperatura propicia para dicho proceso de hibridación, y/o se puede realizar en presencia de un tampón de hibridación modificado que puede conducir a dicho proceso de hibridación (tal como un tampón que contiene un desnaturizante de ácido nucleico, tal como betaína o DMSO).

Tras la etapa de hibridación, por tanto, las moléculas de ARN mensajero de células individuales hibridan con oligonucleótidos con códigos de barras del reactivo de códigos de barras multimérico que se transfirió a esa célula. En las etapas de procesamiento posteriores (por ejemplo, después de una etapa de aislamiento de las moléculas de ARN mensajero hibridado y los oligonucleótidos con códigos de barras), el ARN mensajero se puede transcribir de forma inversa con una transcriptasa inversa y luego, opcionalmente, amplificar, como con un proceso de PCR, antes de realizar una reacción de secuenciación. La transcripción inversa puede incluir una y/o ambas de transcripción inversa de la primera cadena (por ejemplo, síntesis de ADNc de la primera cadena) y también síntesis de la segunda cadena. Además, cualquier etapa de transcripción inversa y/o síntesis de ADNc puede incluir cualquier etapa estándar adicional de procesamiento de ADNc, tal como la fragmentación (por ejemplo, fragmentación acústica como ultrasonidos Covaris, o por ejemplo, fragmentación enzimática tal como con una enzima fragmentasa, una enzima de restricción, y/o una enzima transposasa *in vitro*) y adaptador (por ejemplo, adaptador de PCR y/o adaptador de secuenciación) ligadura y/o transposición de adaptador *in vitro* en cualquier etapa antes y/o después de la transcripción inversa y/o síntesis de la segunda cadena y/o PCR.

La Figura 20 ilustra un método para transferir reactivos de códigos de barras multiméricos a células mediante administración liposomal. En el método, los reactivos de códigos de barras multiméricos se transfieren a las células mediante un proceso de transferencia que implica oligonucleótidos con código de barras que se encuentran comprendidos dentro de compuestos liposomales, y luego se transfieren dichos oligonucleótidos con código de barras mediante administración liposómica. En la presente realización, los oligonucleótidos con código de barras se encapsulan dentro de los liposomas. Estos oligonucleótidos con códigos de barras se pueden asociar opcionalmente con otros restos moleculares.

En la primera etapa, la biblioteca de liposomas se incuba con una muestra de dos o más células, y se permite que los liposomas interactúen con las membranas celulares de las células dentro de la muestra. Al igual que con los métodos estándar de administración de liposomas, el liposoma se puede fusionar con la membrana celular y/o ser internalizado en la célula y liberar sus oligonucleótidos con código de barras constituyentes en el citoplasma, logrando así la administración liposómica de oligonucleótidos con código de barras en las células de la muestra.

Después de esta etapa de administración de liposomas, se realiza una etapa de incubación, durante cuyo tiempo se permite que las moléculas de ARN mensajero complementarias a las regiones diana de los oligonucleótidos con código de barras administrados por los liposomas hibriden con dichas regiones diana. Esta incubación se puede realizar a una temperatura propicia para dicho proceso de hibridación, y/o se puede realizar en presencia de un tampón de hibridación modificado que puede conducir a dicho proceso de hibridación (tal como un tampón que contiene un desnaturizante de ácido nucleico, tal como betaína o DMSO).

Tras la etapa de hibridación, por lo tanto, las moléculas de ARN mensajero de las células individuales hibridan con oligonucleótidos con códigos de barras que se han administrado por un liposoma. En las etapas de procesamiento posteriores (por ejemplo, después de una etapa de aislamiento de las moléculas de ARN mensajero hibridado y los oligonucleótidos con códigos de barras), el ARN mensajero se puede transcribir de forma inversa con una transcriptasa inversa y luego, opcionalmente, amplificar, como con un proceso de PCR, antes de realizar una reacción de

secuenciación.

La Figura 21 ilustra un ejemplo de un método de transferencia de reactivos de códigos de barras multiméricos a células mediante transfección. En el método, los reactivos de códigos de barras multiméricos se transfieren a las células mediante un proceso de transfección. En la primera etapa, los reactivos de códigos de barras multiméricos (por ejemplo, oligonucleótidos con códigos de barras hibridados a lo largo de una molécula de código de barras multimérico) forman complejo con un reactivo de transfección de lípidos. Estos complejos, análogos a los plásmidos complejos de lípidos, tendrán un carácter biofísico y electrostático propicio para la interacción con una membrana celular y luego la transfección en las células.

Los complejos reactivo-lípido de códigos de barras multiméricos resultantes se incuban luego con una muestra de células durante un período de tiempo, durante ese tiempo los complejos migran para entrar en contacto con una membrana celular y se transfectan a las células. En esta ilustración, un primer complejo de lípido-reactivo de código de barras multimérico se transfecta a una primera célula, y un segundo complejo de lípido-reactivo de código de barras multimérico se transfecta a una segunda célula; en realizaciones reales, se puede transfectar una gran biblioteca de reactivos de códigos de barras multiméricos a una gran muestra de células.

Tras esta etapa de transfección, se realiza una etapa de incubación, durante cuyo tiempo se permite que las moléculas de ARN mensajero complementarias a las regiones diana de los oligonucleótidos con código de barras comprendidas dentro de los reactivos de códigos de barras multiméricos transfectados hibriden con dichas regiones diana.

Esta incubación se puede realizar a una temperatura propicia para dicho proceso de hibridación, y/o se puede realizar en presencia de un tampón de hibridación modificado que puede conducir a dicho proceso de hibridación (tal como un tampón que contiene un desnaturante de ácido nucleico, tal como betaína o DMSO).

Tras la etapa de hibridación, por tanto, las moléculas de ARN mensajero de células individuales hibridan con oligonucleótidos con códigos de barras del reactivo de códigos de barras multimérico que se transfirió a esa célula. En las etapas de procesamiento posteriores (por ejemplo, después de una etapa de aislamiento de las moléculas de ARN mensajero hibridado y los oligonucleótidos con códigos de barras), el ARN mensajero se puede transcribir de forma inversa con una transcriptasa inversa y luego, opcionalmente, amplificar, como con un proceso de PCR, antes de realizar una reacción de secuenciación.

La Figura 22 ilustra un ejemplo de un método de transferencia de reactivos de códigos de barras multiméricos a células mediante un proceso de permeabilización. En el método, los reactivos de códigos de barras multiméricos se transfieren a las células mediante un proceso de permeabilización. En la primera etapa, las membranas de las células se permeabilizan con un proceso de permeabilización. Esto puede, en una realización, realizarse mediante exposición a un tensioactivo químico, tal como un detergente no iónico. Después de este proceso de permeabilización, la membrana de cada célula tendrá un carácter biofísico propicio para la difusión de especies macromoleculares tales como reactivos de códigos de barras multiméricos a través de ella.

Las células permeabilizadas resultantes se incuban luego con una biblioteca de dos o más reactivos de códigos de barras multiméricos durante un período de tiempo, durante el cual, los reactivos de códigos de barras multiméricos migran para entrar en contacto con una membrana celular y se transfieren a las células mediante un proceso de difusión. En esta ilustración, un primer reactivo de código de barras multimérico se difunde en una primera célula, y un segundo reactivo de código de barras multimérico se difunde en una segunda célula; en realizaciones reales, se puede transferir una gran biblioteca de reactivos de códigos de barras multiméricos a una gran muestra de células mediante este método.

Después de esta etapa de difusión, se realiza una etapa de incubación, durante cuyo tiempo se permite que las moléculas de ARN mensajero complementarias a las regiones diana de los oligonucleótidos con código de barras comprendidas dentro de los reactivos de códigos de barras multiméricos transferidos hibriden con dichas regiones diana. Esta incubación se puede realizar a una temperatura propicia para dicho proceso de hibridación, y/o se puede realizar en presencia de un tampón de hibridación modificado que puede conducir a dicho proceso de hibridación (tal como un tampón que contiene un desnaturante de ácido nucleico, tal como betaína o DMSO).

Tras la etapa de hibridación, por tanto, las moléculas de ARN mensajero de células individuales hibridan con oligonucleótidos con códigos de barras del reactivo de códigos de barras multimérico que se transfirió a esa célula. En las etapas de procesamiento posteriores (por ejemplo, después de una etapa de aislamiento de las moléculas de ARN mensajero hibridado y los oligonucleótidos con códigos de barras), el ARN mensajero se puede transcribir de forma inversa con una transcriptasa inversa y luego, opcionalmente, amplificar, como con un proceso de PCR, antes de realizar una reacción de secuenciación.

La Figura 23 ilustra un ejemplo de un método de colocación de códigos de barras de ácidos nucleicos celulares con una etapa de permeabilización de la membrana. En el método, las moléculas de ARN mensajero se liberan de las células, después de lo cual se le coloca el código de barras por oligonucleótidos con código de barras que se encuentran dentro de la proximidad espacial de la propia célula. En la primera etapa, se mezcla una biblioteca de dos o más reactivos de códigos de barras multiméricos con una muestra de dos o más células. Opcionalmente, tal como se muestra, dichos reactivos de códigos de barras multiméricos pueden comprender restos de unión a células que los

impulsan a interactuar preferentemente con las membranas de las células dentro de las muestras; se realiza una etapa de incubación para permitir que los reactivos de códigos de barras multiméricos se unan a las superficies de las células.

En una segunda etapa, se realiza un proceso de permeabilización de membranas y/o lisis celular, en donde la membrana celular se vuelve permeable a las macromoléculas de modo que las moléculas de ARN mensajero y/o los oligonucleótidos pueden difundirse a través del espacio de la membrana. Esta etapa se puede realizar por varios medios, tal como mediante una etapa de incubación a alta temperatura tal como se ilustra en el presente documento. Esta etapa de permeabilización y/o lisis permite la interacción molecular entre los oligonucleótidos con códigos de barras y sus ácidos nucleicos diana.

Después de esta etapa de permeabilización y/o lisis de la membrana, se realiza una etapa de incubación, durante cuyo tiempo se permite que las moléculas de ARN mensajero complementarias a las regiones diana de los oligonucleótidos con código de barras comprendidas dentro de los reactivos de códigos de barras multiméricos hibriden con dichas regiones diana. Esta incubación se puede realizar a una temperatura propicia para dicho proceso de hibridación, y/o se puede realizar en presencia de un tampón de hibridación modificado que puede conducir a dicho proceso de hibridación (tal como un tampón que contiene un desnaturante de ácido nucleico, tal como betaína o DMSO). Esta incubación se puede realizar además en presencia de un agente espesante, tal como poli(etilen)glicol (PEG), para retardar la difusión de oligonucleótidos con código de barras y/o moléculas de ácido nucleico diana dentro de la solución.

Tras la etapa de hibridación, por tanto, las moléculas de ARN mensajero de células individuales hibridan con oligonucleótidos con códigos de barras del reactivo de códigos de barras multimérico que estaba dentro de la proximidad espacial a esa célula. En las etapas de procesamiento posteriores (por ejemplo, después de una etapa de aislamiento de las moléculas de ARN mensajero hibridado y los oligonucleótidos con códigos de barras), el ARN mensajero se puede transcribir de forma inversa con una transcriptasa inversa y luego, opcionalmente, amplificar, como con un proceso de PCR, antes de realizar una reacción de secuenciación.

La Figura 24 ilustra un método de colocación de códigos de barras de ácidos nucleicos celulares con una etapa de permeabilización de la membrana y liberación de oligonucleótidos con código de barras. En el método, las moléculas de ARN mensajero se pueden liberar de las células, después de lo cual se les coloca el código de barras mediante oligonucleótidos con códigos de barras que se liberan de reactivos de códigos de barras multiméricos que estaban dentro de la proximidad espacial de dicha misma célula. En la primera etapa, se mezcla una biblioteca de dos o más reactivos de códigos de barras multiméricos con una muestra de dos o más células. Opcionalmente, tal como se muestra, dichos reactivos de códigos de barras multiméricos pueden comprender restos de unión a células que los impulsan a interactuar preferentemente con las membranas de las células dentro de las muestras; se realiza una etapa de incubación para permitir que los reactivos de códigos de barras multiméricos se unan a las superficies de las células.

En una segunda etapa, se realiza un proceso de permeabilización de membranas y/o lisis celular, en donde la membrana celular se vuelve permeable a las macromoléculas de modo que las moléculas de ARN mensajero y/o los oligonucleótidos pueden difundirse a través del espacio de la membrana. Esta etapa se puede realizar por varios medios, tal como mediante una etapa de incubación a alta temperatura tal como se ilustra en el presente documento. Esta etapa de permeabilización y/o lisis permite la interacción molecular entre los oligonucleótidos con códigos de barras y sus dianas de ácidos nucleicos.

En la presente realización, esta etapa de incubación a alta temperatura disocia aún más los oligonucleótidos de códigos de barras de sus respectivos reactivos de códigos de barras multiméricos, específicamente en esta realización, dichos oligonucleótidos con código de barras hibridan con moléculas enlazadoras que, a su vez, están unidas al soporte sólido/molecular de cada reactivo multimérico de códigos de barras. Esta etapa de incubación a alta temperatura se realiza a una temperatura por encima de la temperatura de fusión de la región de hibridación oligonucleótido-enlazador con código de barras y, por lo tanto, los oligonucleótidos con código de barras quedan libres para difundirse en solución.

Después de esta etapa de permeabilización y/o lisis de la membrana, se realiza una etapa de incubación, durante cuyo tiempo se permite que las moléculas de ARN mensajero complementarias a las regiones diana de los oligonucleótidos con código de barras liberadas de los reactivos de códigos de barras multiméricos hibriden con dichas regiones diana. Esta incubación se puede realizar a una temperatura propicia para dicho proceso de hibridación, y/o se puede realizar en presencia de un tampón de hibridación modificado que puede conducir a dicho proceso de hibridación (tal como un tampón que contiene un desnaturante de ácido nucleico, tal como betaína o DMSO). Esta incubación se puede realizar además en presencia de un agente espesante, tal como poli(etilen)glicol (PEG), para retardar la difusión de oligonucleótidos con código de barras y/o moléculas de ácido nucleico diana dentro de la solución.

Tras la etapa de hibridación, por tanto, las moléculas de ARN mensajero de células individuales hibridan con oligonucleótidos con códigos de barras liberadas del reactivo de códigos de barras multimérico que estaba dentro de la proximidad espacial a esa célula. En las etapas de procesamiento posteriores (por ejemplo, después de una etapa de aislamiento de las moléculas de ARN mensajero hibridado y los oligonucleótidos con códigos de barras), el ARN mensajero se puede transcribir de forma inversa con una transcriptasa inversa y luego, opcionalmente, amplificar,

como con un proceso de PCR, antes de realizar una reacción de secuenciación.

Ejemplos

MATERIALES Y MÉTODOS

Método 1 - Síntesis de una biblioteca de moléculas de códigos de barras de ácidos nucleicos

5 Síntesis de la biblioteca de moléculas de subcódigo de barras bicatenarias

En un tubo de PCR, se añadieron 10 microlitros de BC_MX3 a 10 micromolar (una mezcla equimolar de todas las secuencias en la SEQ ID NO: 18 a 269) a 10 microlitros de BC_ADD_TP1 a 10 micromolar (SEQ ID NO: 1), más 10 microlitros de tampón CutSmart a 10X (New England Biolabs) más 1,0 microlitros de mezcla de nucleótidos de desoxinucleótido trifosfato a 10 milimolar (Invitrogen) más 68 microlitros de H₂O, hasta un volumen final de 99 microlitros. El tubo de PCR se colocó en un termociclador y se incubó a 75 °C durante 5 minutos, luego hibridó lentamente a 4 °C, luego se mantuvo a 4 °C, luego se colocó en hielo. a la solución se añadió 1,0 microlitro de fragmento de polimerasa Klenow (New England Biolabs; a 5 U/ul) y se mezcló. El tubo de PCR se colocó nuevamente en un termociclador y se incubó a 25 °C durante 15 minutos, luego se mantuvo a 4 °C. A continuación, la solución se purificó con una columna de purificación (Nucleotide Removal Kit; Qiagen), se eluyó en 50 microlitros de H₂O, y se cuantificó espectrofotométricamente.

Síntesis de la molécula adaptadora bicatenaria aguas abajo

En un tubo de PCR, se añadieron 0,5 microlitros de BC_ANC_TP1 100 micromolar (SEQ ID NO: 2) a 0,5 microlitros de BC_ANC_BT1 100 micromolar (SEQ ID NO: 3), más 20 microlitros de tampón CutSmart a 10X (New England Biolabs) más 178 microlitros de H₂O, hasta un volumen final de 200 microlitros. El tubo de PCR se colocó en un termociclador y se incubó a 95 °C durante 5 minutos, luego hibridó lentamente a 4 °C, luego se mantuvo a 4 °C, luego se colocó en hielo, luego se almacenó a -20 °C.

Ligadura de la biblioteca de moléculas de subcódigo de barras bicatenarias a una molécula adaptadora bicatenaria aguas abajo

En un tubo Eppendorf de 1,5 mililitros, se añadió 1,0 microlitros de solución de molécula de adaptador bicatenario aguas abajo a 2,5 microlitros de biblioteca de moléculas de subcódigo de barras bicatenarias, más 2,0 microlitros de tampón de ADN ligasa T4 a 10X y 13,5 microlitros de H₂O hasta un volumen final de 19 microlitros. Se añadió 1,0 microlitro de ADN ligasa T4 (New England Biolabs; alta concentración) a la solución y se mezcló. El tubo se incubó a temperatura ambiente durante 60 minutos, luego se purificó con 1,8X de volumen (34 microlitros) de perlas Ampure XP (Agencourt; según las instrucciones del fabricante), y se eluyó en 40 microlitros de H₂O.

30 Amplificación por PCR de la biblioteca ligada

En un tubo de PCR, se añaden 2,0 microlitros de la biblioteca ligada a 2,0 microlitros de BC_FWD_PR1 50 micromolar (SEQ ID NO: 4), más 2,0 microlitros de BC_REV_PR1 50 micromolar (SEQ ID NO: 5), más 10 microlitros de tampón de PCR Taq a 10X (Qiagen) más 2,0 microlitros de mezcla de nucleótidos de desoxinucleótido trifosfato 10 milimolar (Invitrogen) más 81,5 microlitros de H₂O, más 0,5 microlitros de Taq Polimerasa de Qiagen (a 5U/ul) hasta un volumen final de 100 microlitros. El tubo de PCR se colocó en un termociclador y se amplificó durante 15 ciclos de: 95 °C durante 30 segundos, luego 59 °C durante 30 segundos, luego 72 °C durante 30 segundos; luego se mantuvo a 4 °C. Después, la solución se purificó con un volumen de 1,8X (180 microlitros) de perlas Ampure XP (Agencourt; según las instrucciones del fabricante), y se eluyó en 50 microlitros de H₂O.

Digestión de la enzima uracil glucosilasa

A un tubo eppendorf se añadieron 15 microlitros de la amplificación por PCR eluida, 1,0 microlitros de H₂O, más 2,0 microlitros de tampón CutSmart a 10X (New England Biolabs), más 2,0 microlitros de solución de enzima USER (New England Biolabs) y mezcló. El tubo se incubó a 37 °C durante 60 minutos, luego, la solución se purificó con un volumen de 1,8X (34 microlitros) de perlas Ampure XP (Agencourt; según las instrucciones del fabricante), y se eluyó en 34 microlitros de H₂O.

45 Escisión de enzimas de restricción de MlyI

Al eluato de la etapa anterior (digestión con glicosilasa), se añadieron 4,0 microlitros de tampón CutSmart a 10X (New England Biolabs), más 2,0 microlitros de enzima MlyI (New England Biolabs, a 5 U/ul) y se mezclaron. El tubo se incubó a 37 °C durante 60 minutos, luego, la solución se purificó con un volumen de 1,8X (72 microlitros) de perlas Ampure XP (Agencourt; según las instrucciones del fabricante), y se eluyó en 40 microlitros de H₂O.

50 Ligadura de la biblioteca de subcódigos de barras a una solución de MlyI escindida

En un tubo Eppendorf de 1,5 mililitros, se agregaron 10 microlitros de solución de MlyI escindida a 2,5 microlitros de biblioteca de moléculas de subcódigos de barras bicatenarias, más 2,0 microlitros de tampón de ADN ligasa T4 a 10X

y 4,5 microlitros de H₂O hasta un volumen final de 19 microlitros. 1,0 microlitro de ADN ligasa T4 (New England Biolabs; alta concentración) y se mezcló. El tubo se incubó a temperatura ambiente durante 60 minutos, luego se purificó con 1,8X de volumen (34 microlitros) de perlas Ampure XP (Agencourt; según las instrucciones del fabricante), y se eluyó en 40 microlitros de H₂O.

5 Ciclos repetidos de adición de subcódigos de barras

Las etapas experimentales de: 1) ligadura de la biblioteca de códigos de barras secundarios a la solución escindida de MlyI, 2) Amplificación por PCR de la biblioteca ligada, 3) Digestión de la enzima de uracilo glicosilasa y 4) la escisión de la enzima de restricción de MlyI se repitieron, en secuencia, por un total de cinco ciclos.

Síntesis de la molécula adaptadora bicatenaria aguas arriba

- 10 En un tubo de PCR, se añadieron 1,0 microlitros de BC_USO_TP1 100 micromolar (SEQ ID NO: 6) a 1,0 microlitros de BC_USO_BT1 100 micromolar (SEQ ID NO: 7), más 20 microlitros de tampón CutSmart a 10X (New England Biolabs) más 178 microlitros de H₂O, hasta un volumen final de 200 microlitros. El tubo de PCR se colocó en un termociclador y se incubó a 95 °C durante 60 segundos, luego hibridó lentamente a 4 °C, luego se mantuvo a 4 °C, luego se colocó en hielo, luego se almacenó a -20 °C.

15 Ligadura de la molécula adaptadora bicatenaria aguas arriba

- En un tubo Eppendorf de 1,5 mililitros, se añadieron 3,0 microlitros de solución de adaptador aguas arriba a 10,0 microlitros de solución final escindida de MlyI (después del quinto ciclo), más 2,0 microlitros de tampón de ADN ligasa T4 a 10X y 5,0 microlitros de H₂O hasta un volumen final de 19 microlitros. Se añadió a la solución 1,0 microlitro de ADN ligasa T4 (New England Biolabs; alta concentración) y se mezcló. El tubo se incubó a temperatura ambiente durante 60 minutos, luego se purificó con 1,8X de volumen (34 microlitros) de perlas Ampure XP (Agencourt; según las instrucciones del fabricante), y se eluyó en 40 microlitros de H₂O.

Amplificación por PCR de la biblioteca ligada con adaptador aguas arriba

- En un tubo de PCR, se añadieron 6,0 microlitros de biblioteca ligada con adaptador aguas arriba a 1,0 microlitros de BC_CS_PCR_FWD1 100 micromolar (SEQ ID NO: 8), más 1,0 microlitros de BC_CS_PCR_REV1 100 micromolar (SEQ ID NO: 9), más 10 microlitros de tampón de PCR Taq a 10X (Qiagen) más 2,0 microlitros de mezcla de nucleótidos de desoxinucleótido trifosfato 10 milimolar (Invitrogen) más 73,5 microlitros de H₂O, más 0,5 microlitros de Taq Polimerasa de Qiagen (a 5U/ul) hasta un volumen final de 100 microlitros. El tubo de PCR se colocó en un termociclador y se amplificó durante 15 ciclos de: 95 °C durante 30 segundos, luego 61 °C durante 30 segundos, luego 72 °C durante 30 segundos; luego se mantuvo a 4 °C. La solución, que contiene una biblioteca de moléculas de códigos de barras de ácido nucleico amplificadas, luego se purificó con un volumen de 1,8X (180 microlitros) de perlas de Ampure XP (Agencourt; según las instrucciones del fabricante). La biblioteca de moléculas de códigos de barras de ácido nucleico amplificadas se eluyó luego en 40 microlitros de H₂O.

- La biblioteca de moléculas de código de barras de ácido nucleico amplificadas sintetizadas por el método descrito anteriormente se usó luego para ensamblar una biblioteca de moléculas de código de barras multiméricas tal como se describe a continuación.

Método 2- Ensamblaje de una biblioteca de moléculas de código de barras multiméricas

Se ensambló una biblioteca de moléculas de código de barras multiméricas utilizando la biblioteca de moléculas de código de barras de ácido nucleico sintetizadas de acuerdo con los métodos del Método 1.

Extensión de cebador con cebador de terminación directo y cebador de ferulización directo

- 40 En un tubo de PCR, se añadieron 5,0 microlitros de la biblioteca de moléculas de código de barras de ácido nucleico amplificado a 1,0 microlitros de CS_SPLT_FWD1 100 micromolar (SEQ ID NO: 10), más 1,0 microlitros de CS_TERM_FWD1 5 micromolar (SEQ ID NO: 11), más 10 microlitros de tampón Thermopol a 10X (NEB) más 2,0 microlitros de mezcla de nucleótidos de desoxinucleótido trifosfato 10 milimolar (Invitrogen) más 80,0 microlitros de H₂O, más 1,0 microlitros polimerasa Vent Exo-Minus (New England Biolabs, a 2U/ul) hasta un volumen final de 100 microlitros. El tubo de PCR se colocó en un termociclador y se amplificó durante 1 ciclo de: 95 °C durante 30 segundos, luego 53 °C durante 30 segundos, luego 72 °C durante 60 segundos, luego 1 ciclo de: 95 °C durante 30 segundos, luego 50 °C durante 30 segundos, luego 72 °C durante 60 segundos, luego se mantuvo a 4 °C. A continuación, la solución se purificó en una columna de purificación por PCR (Qiagen) y se eluyó en 85,0 microlitros de H₂O.

Extensión de cebador con cebador de terminación inverso y cebador de ferulización inverso

- 50 En un tubo de PCR, los 85,0 microlitros de productos de extensión del cebador de extensión directa se añadieron a 1,0 microlitros de CS_SPLT_REV1 100 micromolar (SEQ ID NO: 12), más 1,0 microlitros de CS_TERM_REV1 5 micromolar (SEQ ID NO: 13), más 10 microlitros de tampón Thermopol a 10X (NEB) más 2,0 microlitros de mezcla de nucleótidos de desoxinucleótido trifosfato 10 milimolar (Invitrogen), más 1,0 microlitros polimerasa Vent Exo-Minus (New England Biolabs, a 2U/ul) hasta un volumen final de 100 microlitros. El tubo de PCR se colocó en un termociclador

y se amplificó durante 1 ciclo de: 95 °C durante 30 segundos, luego 53 °C durante 30 segundos, luego 72 °C durante 60 segundos, luego 1 ciclo de: 95 °C durante 30 segundos, luego 50 °C durante 30 segundos, luego 72 °C durante 60 segundos, luego se mantuvo a 4 °C. A continuación, la solución se purificó en una columna de purificación por PCR (Qiagen) y se eluyó en 43,0 microlitros de H₂O.

5 Unión de productos de extensión de cebador con PCR superposición-extensión

En un tubo de PCR se añadieron los 43,0 microlitros de productos de extensión de cebadores de extensión inversa, más 5,0 microlitros de tampón Thermopol a 10X (NEB) más 1,0 microlitros de mezcla de nucleótidos de desoxinucleótido trifosfato 10 milimolar (Invitrogen), más 1,0 microlitros polimerasa Vent Exo-Minus (New England Biolabs, a 2U/ul) hasta un volumen final de 50 microlitros. El tubo de PCR se colocó en un termociclador y se amplificó durante 5 ciclos de: 95 °C durante 30 segundos, luego 60 °C durante 60 segundos, luego 72 °C durante 2 minutos; luego 5 ciclos de: 95 °C durante 30 segundos, luego 60 °C durante 60 segundos, luego 72 °C durante 5 minutos; luego 5 ciclos de: 95 °C durante 30 segundos, luego 60 °C durante 60 segundos, luego 72 °C durante 10 minutos; luego se mantuvo a 4 °C. Después, la solución se purificó con un volumen de 0,8X (80 microlitros) de perlas Ampure XP (Agencourt; según las instrucciones del fabricante), y se eluyó en 40 microlitros de H₂O.

15 Amplificación de productos de superposición-extensión

En un tubo de PCR se agregaron 2,0 microlitros de solución de PCR de superposición-extensión, más 1,0 microlitros de CS_PCR_FWD1 100 micromolar (SEQ ID NO: 14), más 1,0 microlitros de CS_PCR_REV1 100 micromolar (SEQ ID NO: 15), más 10 microlitros de tampón Thermopol a 10X (NEB) más 2,0 microlitros de mezcla de nucleótidos de desoxinucleótido trifosfato 10 milimolar (Invitrogen), más 1,0 microlitros de polimerasa Vent Exo-Minus (New England Biolabs, a 2U/ul), más 83,0 microlitros de H₂O hasta un volumen final de 100 microlitros. El tubo de PCR se colocó en un termociclador y se amplificó durante 15 ciclos de: 95 °C durante 30 segundos, luego 58 °C durante 30 segundos, luego 72 °C durante 10 minutos; luego se mantuvo a 4 °C. Después, la solución se purificó con un volumen de 0,8X (80 microlitros) de perlas Ampure XP (Agencourt; según las instrucciones del fabricante), y se eluyó en 50 microlitros de H₂O, y se cuantificó espectrofotométricamente.

25 Selección de tamaño basado en gel de productos de superposición-extensión amplificados

Se cargaron aproximadamente 250 nanogramos de productos de superposición-extensión amplificados y se procesaron en un gel de agarosa al 0,9 %, y luego se tiñeron y visualizaron con bromuro de etidio. Se cortó una banda correspondiente a un tamaño de 1000 nucleótidos (más y menos 100 nucleótidos) y se purificó con una columna de extracción de gel (Gel Extraction Kit, Qiagen) y se eluyó en 50 microlitros de H₂O.

30 Amplificación de productos de superposición-extensión

En un tubo de PCR se agregaron 10,0 microlitros de solución seleccionada por tamaño de gel, más 1,0 microlitros de CS_PCR_FWD1 100 micromolar (SEQ ID NO: 14), más 1,0 microlitros de CS_PCR_REV1 100 micromolar (SEQ ID NO: 15), más 10 microlitros de tampón Thermopol a 10X (NEB) más 2,0 microlitros de mezcla de nucleótidos de desoxinucleótido trifosfato 10 milimolar (Invitrogen), más 1,0 microlitros de polimerasa Vent Exo-Minus (New England Biolabs, a 2U/ul) más 75,0 microlitros de H₂O hasta un volumen final de 100 microlitros. El tubo de PCR se colocó en un termociclador y se amplificó durante 15 ciclos de: 95 °C durante 30 segundos, luego 58 °C durante 30 segundos, luego 72 °C durante 4 minutos; luego se mantuvo a 4 °C. Después, la solución se purificó con un volumen de 0,8X (80 microlitros) de perlas Ampure XP (Agencourt; según las instrucciones del fabricante), y se eluyó en 50 microlitros de H₂O, y se cuantificó espectrofotométricamente.

40 Selección y amplificación de un número cuantitativamente conocido de moléculas de código de barras multiméricas

La solución extraída en gel amplificada se diluyó a una concentración de 1 picogramo por microlitro, y luego a un tubo de PCR se agregaron 2,0 microlitros de esta solución diluida (aproximadamente 2 millones de moléculas individuales), más 0,1 microlitros de CS_PCR_FWD1 100 micromolar (SEQ ID NO: 14), más 0,1 microlitros de CS_PCR_REV1 100 micromolar (SEQ ID NO: 15), más 1,0 microlitros de tampón Thermopol a 10X (NEB) más 0,2 microlitros de mezcla de desoxinucleótido trifosfato de 10 milimolar (Invitrogen), más 0,1 microlitros de polimerasa Vent Exo-Minus (New England Biolabs, a 2U/ul) más 6,5 microlitros de H₂O hasta un volumen final de 10 microlitros. El tubo de PCR se colocó en un termociclador y se amplificó durante 11 ciclos de: 95 °C durante 30 segundos, luego 57 °C durante 30 segundos, luego 72 °C durante 4 minutos; luego se mantuvo a 4 °C.

Al tubo de PCR se añadieron 1,0 microlitros de CS_PCR_FWD1 100 micromolar (SEQ ID NO: 14), más 1,0 microlitros de CS_PCR_REV1 100 micromolar (SEQ ID NO: 15), más 9,0 microlitros de tampón Thermopol a 10X (NEB) más 2,0 microlitros de mezcla de nucleótidos de desoxinucleótido trifosfato 10 milimolar (Invitrogen), más 1,0 microlitros de polimerasa Vent Exo-Minus (New England Biolabs, a 2U/ul) más 76,0 microlitros de H₂O hasta un volumen final de 100 microlitros. El tubo de PCR se colocó en un termociclador y se amplificó durante 10 ciclos de: 95 °C durante 30 segundos, luego 57 °C durante 30 segundos, luego 72 °C durante 4 minutos; luego se mantuvo a 4 °C. Después, la solución se purificó con un volumen de 0,8X (80 microlitros) de perlas Ampure XP (Agencourt; según las instrucciones del fabricante), y se eluyó en 50 microlitros de H₂O, y se cuantificó espectrofotométricamente.

Método 3: Producción de moléculas de código de barras multiméricas monocatenarias por transcripción *in vitro* y síntesis de ADNc

Este método describe una serie de etapas para producir cadenas de ADN monocatenario, con las que pueden hibridar los oligonucleótidos y luego colocarles código de barras. Este método comienza con cuatro reacciones idénticas realizadas en paralelo, en las que un sitio promotor para la ARN polimerasa de T7 se adjunta al extremo 5' de una biblioteca de moléculas de código de barras multiméricas utilizando una reacción de amplificación por PCR de superposición-extensión. Se realizan cuatro reacciones idénticas en paralelo y luego se combinan para aumentar la cantidad cuantitativa y la concentración de este producto disponible. En cada uno de los cuatro tubos de PCR idénticos, se mezclaron aproximadamente 500 picogramos de moléculas de código de barras multiméricas de tamaño seleccionado y amplificadas por PCR (como se produce en la etapa 'Selección y amplificación del número cuantitativamente conocido de moléculas de código de barras multiméricas' del Método 2) con 2,0 microlitros de CS_PCR_FWD1_T7 100 micromolar (SEQ ID NO 270) y 2,0 microlitros de CS_PCR_REV4 100 micromolar (SEQ ID NO. 271), más 20,0 microlitros de tampón de PCR Thermopol a 10X, más 4,0 microlitros de mezcla de nucleótidos de desoxinucleótido trifosfato 10 milimolar y 2,0 microlitros de polimerasa Vent Exo Minus (a 5 unidades por microlitro) más agua hasta un volumen total de 200 microlitros. El tubo de PCR se colocó en un termociclador y se amplificó durante 22 ciclos de: 95 °C durante 60 segundos, luego 60 °C durante 30 segundos, luego 72 °C durante 3 minutos; luego se mantuvo a 4 °C. La solución de las cuatro reacciones se purificó luego con una columna de extracción de gel (Gel Extraction Kit, Qiagen) y se eluyó en 52 microlitros de H₂O.

Se mezclaron cincuenta (50) microlitros del eluato con 10 microlitros de NEBuffer 2 a 10X (NEB), más 0,5 microlitros de mezcla de nucleótidos de desoxinucleótido trifosfato 10 milimolar y 1,0 microlitros de polimerasa Vent Exo Minus (a 5 unidades por microlitro) más agua hasta un volumen total de 100 microlitros. La reacción se incubó durante 15 minutos a temperatura ambiente, luego se purificó con un volumen de 0,8X (80 microlitros) de perlas Ampure XP (Agencourt; según las instrucciones del fabricante), y se eluyó en 40 microlitros de H₂O, y se cuantificó espectrofotométricamente.

Luego se realiza una etapa de transcripción, en donde la biblioteca de moldes amplificados por PCR que contienen el sitio promotor de la ARN polimerasa de T7 (como se produjo en la etapa anterior) se usa como molde para la ARN polimerasa de T7. Esto comprende una etapa de amplificación para producir una gran cantidad de ácido nucleico basado en ARN correspondiente a la biblioteca de moléculas de código de barras multiméricas (dado que cada molécula de PCR de entrada puede servir como molde para producir una gran cantidad de moléculas de ARN afines).

En la etapa posterior, estas moléculas de ARN luego se transcriben de forma inversa para crear las moléculas de código de barras multiméricas monocatenarias deseadas. Se mezclaron diez (10) microlitros del eluato con 20 microlitros de tampón de transcripción a 5X (Promega), más 2,0 microlitros de mezcla de nucleótidos de desoxinucleótido trifosfato 10 milimolar, más 10 microlitros de DTT 0,1 milimolar, más 4,0 microlitros de SuperAseln (Ambion) y 4,0 microlitros de ARN polimerasa T7 de Promega (a 20 unidades por microlitro) más agua hasta un volumen total de 100 microlitros. La reacción se incubó 4 horas a 37 °C, luego se purificó con un RNEasy Mini Kit (Qiagen) y se eluyó en 50 microlitros de H₂O, y se añadió a 6,0 microlitros de SuperAseln (Ambion).

La solución de ARN producida en la anterior etapa de transcripción *in vitro* luego se transcribe de forma inversa (usando un cebador específico para los extremos 3' de las moléculas de ARN) y luego se digiere con ribonucleasa H para crear moléculas de ADN monocatenarias correspondientes a moléculas de códigos de barras multiméricas, con las que pueden hibridar oligonucleótidos y luego colocarles códigos de barras. En dos tubos de réplica idénticos, se mezclaron 23,5 microlitros del eluato con 5,0 microlitros de mezcla de nucleótidos de desoxinucleótido trifosfato 10 milimolar, más 3,0 microlitros de SuperAseln (Ambion) y 10,0 microlitros de CS_PCR_REV1 2,0 micromolar (SEQ ID NO. 272) más agua hasta un volumen final de 73,5 microlitros. La reacción se incubó en un termociclador a 65 °C durante 5 minutos, luego 50 °C durante 60 segundos; luego se mantuvo a 4 °C. Al tubo se le añadieron 20 microlitros de tampón de transcripción inversa a 5X (Invitrogen), más 5,0 microlitros de DTT 0,1 milimolar y 1,75 microlitros de transcriptasa inversa Superscript III (Invitrogen). La reacción se incubó a 55 °C durante 45 minutos, luego 60 °C durante 5 minutos; luego 70 °C durante 15 minutos, luego se mantuvo a 4 °C, luego se purificó con una columna de limpieza de PCR (Qiagen) y se eluyó en 40 microlitros de H₂O.

Se mezclaron sesenta (60) microlitros del eluato con 7,0 microlitros de tampón de ribonucleasa H a 10X (Promega), más 4,0 microlitros de ribonucleasa H (Promega). La reacción se incubó 12 horas a 37 °C, luego 95 °C durante 10 minutos, luego se mantuvo a 4 °C, luego se purificó con un volumen de 0,7X (49 microlitros) de perlas Ampure XP (Agencourt; según las instrucciones del fabricante), y se eluyó en 30 microlitros de H₂O, y se cuantificó espectrofotométricamente.

Método 4: Producción de reactivos de códigos de barras multiméricos que contienen oligonucleótidos con códigos de barras

Este método describe las etapas para producir reactivos de códigos de barras multiméricos a partir de moléculas de código de barras multiméricas monocatenarias (tal como se produce en el Método 3) y cebadores de extensión apropiados y oligonucleótidos adaptadores.

En un tubo de PCR, se mezclaron aproximadamente 45 nanogramos de moléculas de código de barras multiméricas

digeridas con ribonucleasa H monocatenaria (como se produjo en la última etapa del Método 3) con 0,25 microlitros de DS_ST_05 10 micromolar (SEQ ID NO. 273, un oligonucleótido adaptador) y 0,25 microlitros de US_PCR_Prm_Only_03 10 micromolar (SEQ ID NO. 274, un cebador de extensión), más 5,0 microlitros de tampón de extensión/ligadura isotérmica a 5X, más agua hasta un volumen final de 19,7 microlitros.

- 5 Con el fin de hibridar los oligonucleótidos adaptadores y los cebadores de extensión a las moléculas de código de barras multiméricas, en un termociclador, el tubo se incubó a 98 °C durante 60 segundos, luego hibridó lentamente a 55 °C, luego se mantuvo a 55 °C durante 60 segundos, luego hibridó lentamente a 50 °C y luego se mantuvo a 50 °C durante 60 segundos, luego hibridó lentamente a 20 °C a 0,1 °C/seg, luego se mantuvo a 4 °C. Al tubo se le añadieron 0,3 microlitros (0,625 U) de Polimerasa Phusion (NEB; 2 U/ul) 2,5 microlitros (100 U) ADN Ligasa Taq (NEB; 40 U/ul);
- 10 y 2,5 microlitros de DTT 100 milimolar. Para extender el(los) cebador(es) de extensión a través de la(s) región(es) de código de barras adyacente(s) de cada molécula de código de barras multimérica, y luego ligar este producto de extensión al extremo 5' fosforilado del oligonucleótido adaptador hibridado con la corriente abajo del mismo, el tubo se incubó luego a 50 °C durante 3 minutos, luego se mantuvo a 4 °C. Luego, la reacción se purificó con una columna de limpieza de PCR (Qiagen) y se eluyó en 30 microlitros de H₂O, y se cuantificó espectrofotométricamente.

15 Método 5: Producción de moldes de ADN sintético de secuencia conocida

Este método describe una técnica para producir moldes de ADN sintético con un gran número de identificadores de secuencias moleculares colineales repetidos en tándem, circularizando y luego amplificando en tándem (con una polimerasa de desplazamiento de cadena procesiva) oligonucleótidos que contienen dichos identificadores de secuencia molecular. Este reactivo puede usarse luego para evaluar y medir los reactivos de códigos de barras multiméricos descritos en el presente documento.

En una PCR se añadieron 0,4 microlitros de Syn_Temp_01 1,0 micromolar (SEQ ID NO. 275) y 0,4 microlitros de ST_Splint_02 1,0 micromolar (SEQ ID NO. 276) y 10,0 microlitros de tampón NEB CutSmart a 10X. En un termociclador, el tubo se incubó a 95 °C durante 60 segundos, después se mantuvo a 75 °C durante 5 minutos, luego hibridó lentamente a 20 °C y luego se mantuvo a 20 °C durante 60 segundos, luego se mantuvo a 4 °C. Para

25 circularizar las moléculas a través de una reacción de ligadura intramolecular, a continuación, se añadieron al tubo 10,0 microlitros de ribo-ATP y 5,0 microlitros de ADN ligasa de T4 (NEB; Alta concentración). Luego, el tubo se incubó a temperatura ambiente durante 30 minutos, luego a 65 °C durante 10 minutos, luego hibridó lentamente a 20 °C y luego se mantuvo a 20 °C durante 60 segundos, luego se mantuvo a 4 °C. A cada tubo se le añadió tampón NEB CutSmart a 10X, 4,0 microlitros de mezcla de nucleótidos de desoxinucleótido trifosfato 10 milimolar y 1,5 microlitros

30 de ADN polimerasa phi29 diluida (NEB; diluida a 1:20 en tampón CutSmart a 1X) más agua hasta un volumen total de 200 microlitros. La reacción se incubó a 30 °C durante 5 minutos, luego se mantuvo a 4 °C, luego se purificó con un volumen de 0,7X (140 microlitros) de perlas Ampure XP (Agencourt; según las instrucciones del fabricante), y se eluyó en 30 microlitros de H₂O, y se cuantificó espectrofotométricamente.

35 Método 6: Moldes de códigos de barras de ADN sintético de secuencia conocida con reactivos de códigos de barras multiméricos que contienen oligonucleótidos con códigos de barras

En un tubo de PCR se añadieron 10,0 microlitros de tampón Phusion HF a 5X (NEB), más 1,0 microlitros de mezcla de nucleótidos de desoxinucleótido trifosfato 10 milimolar, más 2,0 microlitros (10 nanogramos) 5,0 nanogramos/microlitros de moldes de ADN sintético de secuencia conocida (tal como se produce mediante el Método 5), más agua hasta un volumen final de 42,5 microlitros. Luego, el tubo se incubó a 98 °C durante 60 segundos, luego se mantuvo a 20 °C. Al tubo se le añadieron 5,0 microlitros de 5,0 picogramos/microlitro de reactivos de códigos de barras multiméricos que contienen oligonucleótidos con códigos de barras (tal como se produce mediante el Método 4). Luego, la reacción se incubó a 70 °C durante 60 segundos, luego hibridó lentamente a 60 °C, luego 60 °C durante cinco minutos, luego hibridó lentamente a 55 °C, luego 55 °C durante cinco minutos, luego hibridó lentamente a 50 °C, luego 50 °C durante cinco minutos, luego se mantuvo a 4 °C. A la reacción se le añadieron 0,5 microlitros de Polimerasa Phusion (NEB), más 2,0 microlitros de SynTemp_PE2_B1_Short1 10 uM (SEQ ID NO. 277, un cebador que es complementario a parte de los productos de extensión producidos por hibridación y extensión de los reactivos de códigos de barras multiméricos creados por el Método 4 a lo largo de los moldes de ADN sintético creados por el Método 5, sirve como cebador para la extensión del cebador y luego las reacciones de PCR descritas en este método). De esta reacción, se añadió un volumen de 5,0 microlitros a un nuevo tubo de PCR, que luego se incubó durante 30

50 segundos a 55 °C, 30 segundos a 60 °C y 30 segundos a 72 °C, luego seguido de 10 ciclos de: 98 °C luego 65 °C luego 72 °C durante 30 segundos cada uno, luego se mantuvo a 4 °C. A cada tubo se le añadió 9,0 microlitros de tampón Phusion a 5X, más 1,0 microlitros de mezcla de nucleótidos de desoxinucleótido trifosfato 10 milimolar, más 1,75 microlitros de SynTemp_PE2_B1_Short1 10 uM (SEQ ID NO. 277), más 1,75 microlitros de US_PCR_Prm_Only_02 10 uM (SEQ ID NO. 278, un cebador parcialmente complementario al cebador de extensión empleado para generar los reactivos de códigos de barras multiméricos según el Método 4, y que sirve como el

55 cebador 'directo' en esta reacción de amplificación por PCR), más 0,5 microlitros de Polimerasa Phusion (NEB), más agua hasta un volumen final de 50 microlitros. El tubo de PCR se colocó en un termociclador y se amplificó durante 24 ciclos de: 98 °C durante 30 segundos, luego 72 °C durante 30 segundos; luego se mantuvo a 4 °C, luego se purificó con 1,2X volúmenes (60 microlitros) de perlas Ampure XP (Agencourt; según las instrucciones del fabricante), y se eluyó en 30 microlitros de H₂O, y se cuantificó espectrofotométricamente.

60

A la biblioteca resultante se le colocó el código de barras luego para la identificación de muestras mediante un método basado en PCR, se amplificó y se secuenció por métodos estándar usando un ciclo de 150, celda de flujo NextSeq de salida media (Illumina) y se desmultiplexó informáticamente para un análisis más detallado.

Método 7: Colocación de códigos de barras en moldes de ADN sintético de secuencia conocida con reactivos de códigos de barras multiméricos y oligonucleótidos adaptadores separados

Para hibridar y extender oligonucleótidos adaptadores a lo largo de los moldes de ADN sintético, en un tubo de PCR se añadieron 10,0 microlitros de tampón Phusion HF a 5X (NEB), más 1,0 microlitros de mezcla de nucleótidos de desoxinucleótido trifosfato 10 milimolar, más 5,0 microlitros (25 nanogramos) 5,0 nanogramos/microlitros de moldes de ADN sintético de secuencia conocida (según lo producido por el Método 5), más 0,25 microlitros de DS_ST_05 10 micromolar (SEQ ID NO: 273, un oligonucleótido adaptador), más agua hasta un volumen final de 49,7 microlitros. En un termociclador, el tubo se incubó a 98 °C durante 2 minutos, luego 63 °C durante 1 minuto, luego hibridó lentamente a 60 °C y luego se mantuvo a 60 °C durante 1 minuto, luego hibridó lentamente a 57 °C y luego se mantuvo a 57 °C durante 1 minuto, luego hibridó lentamente a 54 °C y luego se mantuvo a 54 °C durante 1 minuto, luego hibridó lentamente a 50 °C y luego se mantuvo a 50 °C durante 1 minuto, luego hibridó lentamente a 45 °C y luego se mantuvo a 45 °C durante 1 minuto, luego hibridó lentamente a 40 °C y luego se mantuvo a 40 °C durante 1 minuto, luego se mantuvo a 4 °C. Al tubo se le añadió 0,3 microlitros de Polimerasa Phusion (NEB) y la reacción se incubó a 45 °C durante 20 segundos, luego 50 °C durante 20 segundos, luego 55 °C durante 20 segundos, 60 °C durante 20 segundos, luego 72 °C durante 20 segundos, luego se mantuvo a 4 °C; a continuación, la reacción se purificó con un volumen de 0,8X (40 microlitros) de perlas Ampure XP (Agencourt; según las instrucciones del fabricante), y se eluyó en 30 microlitros de H₂O, y se cuantificó espectrofotométricamente.

Para hibridar oligonucleótidos adaptadores (hibridados y extendidos a lo largo de los moldes de ADN sintético como en la etapa anterior) con moléculas de código de barras multiméricas, y luego hibridar y luego extender cebadores de extensión a través de las regiones de códigos de barras adyacentes de cada molécula de código de barras multimérica, y luego para ligar este producto de extensión al extremo 5' fosforilado del oligonucleótido adaptador hibridado con la zona de aguas abajo del mismo, a un tubo de PCR se le añadieron 10 microlitros del eluato de la etapa anterior (que contiene los moldes de ADN sintético a lo largo de los cuales se han hibridado y extendido los oligonucleótidos adaptadores), más 3,0 microlitros de una solución 50,0 nanomolar de moléculas de código de barras multiméricas digeridas con ribonucleasa H (como se produjo en la última etapa del Método 3), más 6,0 microlitros de tampón de extensión/ligadura isotérmica a 5X, más agua hasta un volumen final de 26,6 microlitros. En un termociclador, el tubo se incubó a 70 °C durante 60 segundos, luego hibridó lentamente a 60 °C, después se mantuvo a 60 °C durante 5 minutos, luego hibridó lentamente a 55 °C y luego se mantuvo a 55 °C durante 5 minutos, luego hibridó lentamente a 50 °C a 0,1 °C/seg y luego se mantuvo a 50 °C durante 30 minutos, luego se mantuvo a 4 °C. Al tubo se le añadieron 0,6 microlitros de US_PCR_Prm_Only_02 10 uM (SEQ ID NO: 278, un cebador de extensión), y la reacción se incubó a 50 °C durante 10 minutos, luego se mantuvo a 4 °C. Al tubo se le añadieron 0,3 microlitros (0,625 U) de Polimerasa Phusion (NEB; 2 U/ul) 2,5 microlitros (100 U) ADN Ligasa Taq (NEB; 40 U/ul); y 2,5 microlitros de DTT 100 milimolar. El tubo se incubó luego a 50 °C durante 5 minutos, luego se mantuvo a 4 °C. A continuación, la reacción se purificó con un volumen de 0,7 veces (21 microlitros) de perlas Ampure XP

(Agencourt; según las instrucciones del fabricante), y se eluyó en 30 microlitros de H₂O, y se cuantificó espectrofotométricamente.

A un nuevo tubo de PCR se añadieron 25,0 microlitros del eluato, más 10,0 microlitros de tampón Phusion HF a 5X (NEB), más 1,0 microlitros de mezcla de nucleótidos de desoxinucleótido trifosfato 10 milimolar, más 2,0 microlitros de SynTemp_PE2_B1_Short1 10 uM (SEQ ID NO: 277; un cebador que es complementario a parte de los productos de extensión producidos por las etapas anteriores; sirve como cebador para la extensión del cebador y luego las reacciones de PCR descritas en el presente documento), más 0,5 ul de polimerasa Phusion (NEB), más agua hasta un volumen final de 49,7 microlitros. De esta reacción, se añadió un volumen de 5,0 microlitros a un nuevo tubo de PCR, que luego se incubó durante 30 segundos a 55 °C, 30 segundos a 60 °C y 30 segundos a 72 °C, luego seguido de 10 ciclos de: 98 °C luego 65 °C luego 72 °C durante 30 segundos cada uno, luego se mantuvo a 4 °C. A cada tubo se le añadió 9,0 microlitros de tampón Phusion a 5X, más 1,0 microlitros de mezcla de nucleótidos de desoxinucleótido trifosfato 10 milimolar, más 1,75 microlitros de SynTemp_PE2_B1_Short1 10 uM (SEQ ID NO: 277), más 1,75 microlitros de US_PCR_Prm_Only_02 10 uM (SEQ ID NO: 278), más 0,5 microlitros de polimerasa Phusion (NEB), más agua hasta un volumen final de 50 microlitros. El tubo de PCR se colocó en un termociclador y se amplificó durante 24 ciclos de: 98 °C durante 30 segundos, luego 72 °C durante 30 segundos; luego se mantuvo a 4 °C, luego se purificó con 1,2X volúmenes (60 microlitros) de perlas Ampure XP (Agencourt; según las instrucciones del fabricante), y se eluyó en 30 microlitros de H₂O, y se cuantificó espectrofotométricamente.

A la biblioteca resultante se le colocó el código de barras luego para la identificación de muestras mediante un método basado en PCR, se amplificó y se secuenció por métodos estándar usando un ciclo de 150, celda de flujo NextSeq de salida media (Illumina) y se desmultiplexó informáticamente para un análisis más detallado.

Método 9: Loci de ADN genómico de códigos de barras con reactivos de códigos de barras multiméricos que contienen oligonucleótidos con códigos de barras

Este método describe un marco para las dianas de códigos de barras dentro de loci genómicos específicos (por ejemplo, colocación de códigos de barras de varios exones dentro de un gen específico) utilizando reactivos de códigos de barras multiméricos que contienen oligonucleótidos con códigos de barras. En primer lugar, una solución de moléculas de código de barras multiméricas fue producida por transcripción *in vitro* y síntesis de ADNc (tal como se describe en el Método 3). A continuación, las soluciones de reactivos de códigos de barras multiméricos que contienen oligonucleótidos con códigos de barras se produjeron tal como se describe en el Método 4, con una modificación hecha de tal manera que en lugar de usar un oligonucleótido adaptador que se dirija a un molde de ADN sintético (es decir, DS_ST_05, SEQ ID NO: 273, tal como se usa en el Método 4), en esa etapa se incluyeron oligonucleótidos adaptadores dirigidos a los loci genómicos específicos. De manera específica, se produjo individualmente una solución de reactivos de códigos de barras multiméricos que contenían oligonucleótidos con códigos de barras apropiados para cada uno de los tres genes humanos diferentes: BRCA1 (que contiene 7 oligonucleótidos adaptadores, SEQ ID NO 279-285), HLA-A (que contiene 3 oligonucleótidos adaptadores, SEQ ID NO 286-288) y DQB1 (que contiene 2 oligonucleótidos adaptadores, SEQ ID NO 289-290). El proceso del Método 4 se llevó a cabo para cada una de estas tres soluciones como se describió anteriormente. Estas tres soluciones luego se fusionaron, en igual volumen, y diluido hasta una concentración final de todos los oligonucleótidos con código de barras de aproximadamente 50 nanomolar.

En un tubo de PCR había más 2,0 microlitros de tampón Phusion HF a 5X (NEB), más 1,0 microlitros de 100 nanogramos/microlitros de ADN genómico humano (NA12878 del Coriell Institute) hasta un volumen final de 9,0 microlitros. En ciertas versiones variantes de este protocolo, los reactivos de códigos de barras multiméricos (que contienen oligonucleótidos con códigos de barras) también se agregaron en esta etapa, antes de la incubación a alta temperatura a 98 °C. La reacción se incubó a 98 °C durante 120 segundos, luego se mantuvo a 4 °C. Al tubo se le añadieron 1,0 microlitros de la solución 50 nanomolar anterior de reactivos de códigos de barras multiméricos y luego la reacción se incubó durante 1 hora a 55 °C, luego 1 hora a 50 °C, luego 1 hora a 45 °C, luego se mantuvo a 4 °C. (Téngase en cuenta que para ciertas muestras, este último proceso de hibridación se extendió para que tenga lugar durante la noche, durante un total de aproximadamente 4 horas por etapa de temperatura).

Para añadir una secuencia de cebador universal inversa a cada secuencia de amplicón (y así permitir la posterior amplificación de toda la biblioteca a la vez, usando solo un cebador de amplificación directo y uno inverso), la reacción se diluyó a 1:100 y se añadió 1,0 microlitro de la solución resultante en un nuevo tubo de PCR a 20,0 microlitros de tampón Phusion HF a 5X (NEB), más 2,0 microlitros de mezcla de nucleótidos de desoxinucleótido trifosfato 10 milimolar, más 1,0 microlitros de una mezcla de cebador inverso (concentración equimolar de SEQ ID NO 291-303, cada cebador a una concentración de 5 micromolar), más 1,0 ul de polimerasa Phusion (NEB), más agua hasta un volumen final de 100 microlitros. La reacción se incubó a 53 °C durante 30 segundos, 72 °C durante 45 segundos, 98 °C durante 90 segundos, luego 68 °C durante 30 segundos, luego 64 °C durante 30 segundos, luego 72 °C durante 30 segundos; luego se mantuvo a 4 °C. Después, la reacción se purificó con un volumen de 0,8X (80 microlitros) de perlas Ampure XP (Agencourt; según las instrucciones del fabricante), y se eluyó en 30 microlitros de H₂O, y se cuantificó espectrofotométricamente.

A la biblioteca resultante se le colocó el código de barras luego para la identificación de muestras mediante un método basado en PCR, se amplificó y se secuenció por métodos estándar usando un ciclo de 150, celda de flujo NextSeq de salida media (Illumina) y se desmultiplexó informáticamente para un análisis más detallado.

Método 10 - Secuenciación de la biblioteca de moléculas de código de barras multiméricas

Preparación de moléculas seleccionadas amplificadas para evaluación con secuenciación de alto rendimiento

A un tubo de PCR se le añadieron 1,0 microlitros de la solución de la molécula seleccionada amplificada, más 1,0 microlitros de CS_SQ_AMP_REV1 100 micromolar (SEQ ID NO: 16), más 1,0 microlitros de US_PCR_Prm_Only_02 100 micromolar (SEQ ID NO: 17), más 10 microlitros de tampón Thermopol a 10X (NEB) más 2,0 microlitros de mezcla de nucleótidos de desoxinucleótido trifosfato 10 milimolar (Invitrogen), más 1,0 microlitros de polimerasa Vent Exo-Minus (New England Biolabs, a 2U/ul) más 84,0 microlitros de H₂O hasta un volumen final de 100 microlitros. El tubo de PCR se colocó en un termociclador y se amplificó durante 3 ciclos de: 95 °C durante 30 segundos, luego 56 °C durante 30 segundos, luego 72 °C durante 3 minutos; luego se mantuvo a 4 °C. Después, la solución se purificó con un volumen de 0,8X (80 microlitros) de perlas Ampure XP (Agencourt; según las instrucciones del fabricante), y se eluyó en 85 microlitros de H₂O.

Luego, esta solución se añadió a un nuevo tubo de PCR, más 1,0 microlitros de Illumina_PE1 100 micromolar, más 1,0 microlitros de Illumina_PE2 100 micromolar, más 10 microlitros de tampón Thermopol a 10X (NEB) más 2,0 microlitros de mezcla de nucleótidos de desoxinucleótido trifosfato 10 milimolar (Invitrogen), más 1,0 microlitros polimerasa Vent Exo-Minus (New England Biolabs, a 2U/ul) hasta un volumen final de 100 microlitros. El tubo de PCR se colocó en un termociclador y se amplificó durante 4 ciclos de: 95 °C durante 30 segundos, luego 64 °C durante 30 segundos, luego 72 °C durante 3 minutos; luego 18 ciclos de: 95 °C durante 30 segundos, luego 67 °C durante 30 segundos, luego 72 °C durante 3 minutos; luego se mantuvo a 4 °C. Después, la solución se purificó con un volumen de 0,8X (80 microlitros) de perlas Ampure XP (Agencourt; según las instrucciones del fabricante), y se eluyó en 40 microlitros de H₂O.

A continuación, se realizó una secuenciación de Illumina de alto rendimiento en esta muestra utilizando un

secuenciador MiSeq con extremos emparejados, química de secuenciación V2 de 250 ciclos.

Método 11 - Evaluación de la naturaleza multimérica de códigos de barras hibridados y extendidos a lo largo de moléculas de ADN molde sintético únicas

- 5 Se creó una biblioteca de moldes de ADN sintético con códigos de barras usando una solución de reactivos de códigos de barras multiméricos producidos de acuerdo con un protocolo tal como se describe generalmente en el Método 3 y el Método 4, y usando una solución de moldes de ADN sintético tal como se describe en el Método 5, y usando un protocolo de laboratorio tal como se describe en el Método 6; a la biblioteca resultante se le colocó el código de barras luego para la identificación de muestras mediante un método basado en PCR, se amplificó y se secuenció por métodos estándar usando un ciclo de 150, celda de flujo NextSeq de salida media (Illumina) y se desmultiplexó informáticamente para un análisis más detallado. Los resultados de la secuenciación de ADN de este método se compararon luego de forma informática con los datos producidos a partir del Método 10 para evaluar el grado de superposición entre el código de barras multimérico de los moldes de ADN sintético y la disposición de dichos códigos de barras en reactivos de códigos de barras multiméricos individuales (los resultados se muestran en la Figura 17).

RESULTADOS

- 15 Estructura y contenido de secuencia esperado de cada secuencia de molécula de reactivo de código de barras multimérico

- 20 La biblioteca de moléculas de códigos de barras multiméricas sintetizadas tal como se describe en los Métodos 1 a 3 se preparó para una secuenciación de alto rendimiento, en donde cada molécula secuenciada incluye un tramo contiguo de una molécula de código de barras multimérica específica (que incluye una o más secuencias de códigos de barras y una o más secuencias adaptadoras asociadas aguas arriba y/o secuencias adaptadoras aguas abajo), todas colineales dentro de la molécula secuenciada. A continuación, esta biblioteca se secuenció con lecturas de 250 nucleótidos de extremos emparejados en un secuenciador MiSeq (Illumina) tal como se describe. Esto produjo aproximadamente 13,5 millones de moléculas totales secuenciadas de la biblioteca, secuenciadas una vez desde cada extremo, para un total de aproximadamente 27 millones de lecturas de secuencia.

- 25 Se espera que cada lectura directa comience con una secuencia de seis nucleótidos, correspondiente al extremo 3' del adaptador aguas arriba: TGACCT

Esta lectura directa es seguida por la primera secuencia de código de barras dentro de la molécula (se espera que tenga 20 nt de longitud).

- 30 Luego, este código de barras es seguido por una 'secuencia de código de barras interna' (en este caso se secuencian en la dirección 'directa' (que es de 82 nucleótidos, incluida la secuencia del adaptador descendente y la secuencia del adaptador ascendente en serie):

ATACCTGACTGCTCGTCAGTTGAGCGAATTCCGTATGGTGGTACACACCTACACTACTCGGA
CGCTCTTCCGATCTTGACCT

- 35 Dentro de la lectura directa de 250 nucleótidos, esto será seguido por un segundo código de barras, otra secuencia de código de barras interna, y luego un tercer código de barras, y luego una fracción de otra secuencia de código de barras interna.

Se espera que cada lectura inversa comience con una secuencia correspondiente a la secuencia del adaptador de aguas abajo:

GCTCAACTGACGAGCAGTCAGGTAT

- 40 Esta lectura inversa es seguida por el primer código de barras que proviene del extremo opuesto de la molécula (también de 20 nucleótidos de longitud, pero secuenciado desde la hebra opuesta de la molécula y, por lo tanto, de orientación inversa a los secuenciados por la lectura directa)

Este código de barras es seguido por la 'secuencia de código de barras interna' pero en la orientación inversa (como está en la cadena opuesta):

AGGTCAAGATCGGAAGAGCGTCCGAGTAGTGTAGGTGTGTACCACCATACGGAATTCGCTC
AACTGACGAGCAGTCAGGTAT

- 45 Asimismo, esta lectura inversa de 250 nucleótidos será seguida por un segundo código de barras, otra secuencia de código de barras interna, y luego un tercer código de barras, y luego una fracción de otra secuencia de código de barras interna.

Extracción y análisis de secuencias

Con secuencias de comandos en Python, se aislaron cada par asociado de código de barras y la secuencia de adaptador aguas arriba y adaptador aguas abajo que flanqueaban, con cada secuencia de código de barras individual de cada molécula de código de barras luego aislada, y cada secuencia de código de barras que fue secuenciada dentro de la misma molécula se anotó como perteneciente a la misma molécula de código de barras multimérica en la biblioteca de moléculas de código de barras multiméricas. Se empleó un comando de análisis simple (Networkx; Python) para determinar los grupos de códigos de barras de moléculas de códigos de barras multiméricas generales, examinando la superposición de pares de códigos de barras en diferentes moléculas secuenciadas. Se realizaron varias métricas de estos datos, incluida la longitud del código de barras, el contenido de la secuencia y el tamaño y la complejidad de las moléculas de códigos de barras multiméricas en la biblioteca de moléculas de códigos de barras multiméricas.

Número de nucleótidos dentro de cada secuencia de código de barras

Cada secuencia de código de barras individual de cada molécula de código de barras, contenida dentro de cada molécula secuenciada por Illumina, y la longitud total de cada código de barras se determinó contando el número de nucleótidos entre la secuencia de la molécula adaptadora aguas arriba y la secuencia de la molécula adaptadora aguas abajo. En la Figura 10 se muestran los resultados.

La inmensa mayoría de los códigos de barras tienen 20 nucleótidos de longitud, que corresponde a cinco adiciones de las moléculas de subcódigo de barras de cuatro nucleótidos de la biblioteca de subcódigos de barras de doble cadena de los presentes inventores. Este es, por tanto, el resultado esperado y deseado, e indica que cada 'ciclo' de: ligadura de la biblioteca de subcódigos de barras a la solución de MlyI escindida, amplificación por PCR de la biblioteca ligada, digestión de la enzima uracil-glicosilasa y escisión de la enzima de restricción de MlyI, tuvo éxito y pudo agregar de manera eficaz nuevas moléculas de código de barras de cuatro nucleótidos en cada ciclo, y luego fue capaz de amplificar y llevar estas moléculas a través del protocolo para un procesamiento posterior continuo, incluso a través de los cinco ciclos totales de adición de códigos de barras secundarios, para hacer las bibliotecas finales ligadas al adaptador de aguas arriba.

Los presentes inventores también utilizaron este método de análisis de secuencia para cuantificar el número total de códigos de barras únicos en total, en todas las moléculas de código de barras multiméricas secuenciadas: esto ascendió a un total de 19.953.626 códigos de barras únicos, que es esencialmente idéntico a los 20 millones de códigos de barras que cabría esperar, dado que los presentes inventores sintetizaron 2 millones de moléculas de códigos de barras multiméricas, cada uno con aproximadamente 10 moléculas de código de barras individuales.

Conjuntamente, estos datos y análisis muestran que los métodos para crear complejos, códigos de barras combinatorios de secuencias de códigos de barras secundarios son eficaces y útiles para sintetizar moléculas de códigos de barras multiméricas.

Número total de moléculas de código de barras únicas en cada molécula de código de barras multimérica

La Figura 11 muestra los resultados de la cuantificación del número total de moléculas de código de barras únicas (según lo determinado por sus respectivas secuencias de código de barras) en cada molécula de código de barras multimérica secuenciada. Como se ha descrito anteriormente, para hacer esto, los presentes inventores examinaron, en el primer caso, secuencias de códigos de barras que estaban presentes y detectadas dentro de las mismas moléculas individuales secuenciadas en el secuenciador. Luego, los presentes inventores emplearon una etapa adicional para agrupar aún más las secuencias de códigos de barras, en donde los presentes inventores emplearon un comando de análisis de red simple (Networkx) que puede determinar enlaces entre secuencias de códigos de barras individuales basados en el conocimiento explícito de los enlaces (donde los códigos de barras se encuentran dentro de la misma molécula secuenciada contigua), y también puede determinar enlaces 'implícitos', en donde dos o más códigos de barras, que no están secuenciados dentro de la misma molécula secuenciada, en cambio, ambos comparten un enlace directo a una tercera secuencia de código de barras común (este enlace común compartido dicta que las dos primeras secuencias de código de barras están ubicadas de hecho en la misma molécula de código de barras multimérica).

Esta figura muestra que la mayoría de las moléculas de códigos de barras multiméricas secuenciadas dentro de la reacción de los presentes inventores tienen dos o más códigos de barras únicos contenidos en la misma, mostrando así que, a través del proceso de enlace de PCR por solapamiento-extensión, los presentes inventores pueden unir varias moléculas de códigos de barras en moléculas de códigos de barras multiméricas. Si bien los presentes inventores esperarían ver más moléculas de código de barras multiméricas que estuviesen más cerca del número esperado de moléculas de código de barras (10), los presentes inventores esperan que este efecto observado se deba a una profundidad de secuenciación insuficientemente alta, y que con un mayor número de moléculas secuenciadas, fuesen capaces de observar una fracción mayor de los verdaderos enlaces entre moléculas de códigos de barras individuales. No obstante, estos datos sugieren que el procedimiento de síntesis fundamental que describen los presentes inventores en el presente documento es eficaz para el propósito previsto.

Moléculas de código de barras multiméricas representativas

La Figura 12 muestra moléculas de código de barras multiméricas representativas que han sido detectadas por el comando de análisis de los presentes inventores. En esta figura, cada 'nodo' es una sola molécula de código de barras (de su secuencia de código de barras asociada), cada línea es un 'enlace directo' entre dos moléculas de código de barras que se han secuenciado al menos una vez en la misma molécula secuenciada, y cada grupo de nodos es una molécula de código de barras multimérica individual, que contiene tanto códigos de barras con enlaces directos como aquellos implícitos, enlaces indirectos según lo determinado por el comando de análisis de los presentes inventores. La figura insertada incluye una sola molécula de código de barras multimérica y las secuencias de las moléculas de código de barras que la constituyen.

Esta figura ilustra el procedimiento de síntesis de la molécula de código de barras multimérica: que los presentes inventores pueden construir moléculas de códigos de barras a partir de bibliotecas de moléculas de códigos de barras secundarias, que los presentes inventores pueden enlazar múltiples moléculas de códigos de barras con una reacción de PCR de superposición-extensión, que los presentes inventores pueden aislar un número cuantitativamente conocido de moléculas de códigos de barras multiméricas individuales, y que pueden amplificarlas y someterlas a análisis y uso posteriores.

Colocación de códigos de barras en moldes de ADN sintéticos de secuencia conocida con (i) reactivos de códigos de barras multiméricos que contienen oligonucleótidos con códigos de barras y (ii) reactivos de códigos de barras multiméricos y oligonucleótidos adaptadores separados

Extracción y análisis de secuencias

Con comandos en Python e implementado en un marco de Amazon Web Services (AWS), para cada secuencia leída después de la desmultiplexación de muestra, cada región de código de barras del reactivo de código de barras multimérico dado se aisló de su secuencia de adaptador aguas arriba y adaptador aguas abajo flanqueantes. Análogamente, cada región de identificación de secuencia molecular de la molécula de molde de ADN sintético dada se aisló de sus secuencias flanqueantes aguas arriba y aguas abajo. Este proceso se repitió para cada molécula en la biblioteca de muestras; se realizó una sola etapa de filtrado en donde los códigos de barras individuales y los identificadores de secuencia molecular que estaban presentes en una sola lectura (por lo tanto, probablemente representarían un error de secuenciación o un error del proceso de preparación de la muestra enzimática) se censuraron de los datos. Para cada identificador de secuencia molecular, se cuantificó el número total de regiones de código de barras únicas (es decir, con secuencias diferentes) que se encuentran asociadas con las mismas dentro de las lecturas de secuencia única. A continuación, se creó un gráfico de histograma para visualizar la distribución de este número en todos los identificadores de secuencia molecular que se encuentran en la biblioteca.

Discusión

La Figura 13 muestra los resultados de este análisis para el Método 6 (Colocación de códigos de barras en moldes de ADN sintético de secuencia conocida con reactivos de códigos de barras multiméricos que contienen oligonucleótidos con códigos de barras). Esta figura deja en claro que la mayoría de los reactivos de códigos de barras multiméricos son capaces de etiquetar con éxito dos o más de las copias repetidas en tándem de cada identificador de secuencia molecular con el que están asociados. Se observa una distribución de 1 a aproximadamente 5 o 6 'eventos de etiquetado', lo que indica que puede haber un grado de interacciones estocásticas que se dan con este sistema, quizás debido a reacciones enzimáticas incompletas, o impedimentos estéricos en la interfaz de reactivo de código de barras/molde sintético, u otros factores.

La Figura 14 muestra los resultados de este mismo análisis realizado usando el Método 7 (Colocación de código de barras en moldes de ADN sintético de oligonucleótidos de código de barras de secuencia conocida con moléculas de código de barras multiméricas y oligonucleótidos adaptadores separados). Esta figura también muestra claramente que la mayoría de los reactivos de códigos de barras multiméricos pueden etiquetar con éxito dos o más de las copias repetidas en tándem de cada identificador de secuencia molecular con el que están asociados, con una distribución similar a la observada para el análisis anterior.

Conjuntamente, estas dos figuras muestran que este marco para el código de barras molecular multimérico es eficaz y, además, que el marco puede configurarse de diferentes formas metodológicas. La Figura 13 muestra los resultados basados en un método en el que el marco está configurado de tal manera que los reactivos de códigos de barras multiméricos ya contienen oligonucleótidos con códigos de barras, antes de que se pongan en contacto con un molde de ADN diana (sintético). Por el contrario, la Figura 14 muestra resultados basados en un método alternativo en donde los oligonucleótidos adaptadores primero se ponen en contacto con el molde de ADN sintético, y luego, en una etapa posterior, a los oligonucleótidos adaptadores se les colocan códigos de barras a través del contacto con un reactivo de código de barras multimérico. En conjunto, estas cifras demuestran tanto la capacidad de colocación de códigos de barras multiméricos de estos reactivos como su versatilidad en diferentes protocolos de laboratorio clave.

Para analizar si, y en qué medida, los reactivos de códigos de barras multiméricos individuales etiquetan con éxito dos o más subsecuencias del mismo molde de ADN sintético, los grupos de códigos de barras diferentes en cada reactivo de códigos de barras multimérico individual en la biblioteca (como se predijo a partir del análisis Networkx descrito en el párrafo anterior y como se ilustra en la Figura 12) se compararon con los códigos de barras hibridados y extendidos

a lo largo de moldes de ADN sintético individuales (tal como se describe en el Método 11). A cada grupo de códigos de barras que se encuentran en los reactivos de códigos de barras multiméricos individuales se le asignó una "etiqueta de identificación de reactivo" numérica. Para cada identificador de secuencia molecular de molde de ADN sintético (es decir, para cada molécula de molde de ADN sintético individual) que se representó en los datos de secuenciación del Método 11 por dos o más códigos de barras (es decir, donde dos o más subsecuencias de la molécula de molde sintético hibridada y extendida por un oligonucleótido con código de barras), se determinó la correspondiente "etiqueta de identificación del reactivo". Para cada molécula de molde sintético, se calculó el número total de códigos de barras multiméricos provenientes del mismo reactivo de código de barras multimérico único (es decir, el número de subsecuencias diferentes en la molécula de molde sintético que fueron etiquetadas por un oligonucleótido con código de barras *diferente* pero del que se calculó el *mismo* reactivo de código de barras multimérico único). Este análisis luego se repitió y se comparó con una condición de 'control negativo', en el que los códigos de barras asignados a cada 'etiqueta de identificación de reactivo' fueron aleatorios (es decir, las mismas secuencias de códigos de barras permanecen presentes en los datos, pero ya no corresponden al enlace molecular real de diferentes secuencias de códigos de barras en la biblioteca de reactivos de códigos de barras multiméricos).

Los datos de este análisis se muestran en la Figura 17, tanto para los datos experimentales reales como para los datos de control con asignaciones de códigos de barras aleatorias (obsérvese la escala logarítmica del eje vertical). Tal como muestra esta figura, aunque el número de eventos de códigos de barras únicos por molécula de molde de ADN sintético diana es pequeño, se superponen casi a la perfección con el contenido de código de barras conocido de los reactivos de códigos de barras multiméricos individuales. Es decir, en comparación con los datos de códigos de barras aleatorizados (que esencialmente no contienen moléculas de molde que parezcan tener un 'código de barras multivalente'), la abrumadora mayoría (más del 99,9 %) de las moléculas de molde en el experimento real que parecen estar etiquetadas por múltiples oligonucleótidos con códigos de barras del mismo reactivo de código de barras multimérico individual, de hecho están etiquetados como multiplicar por los mismos reactivos individuales en solución. Por el contrario, si no hubiera una asociación no aleatoria entre los diferentes códigos de barras que etiquetaban los moldes individuales de ADN sintético (es decir, si la Figura 17 no mostraba ninguna diferencia entre los datos experimentales reales y los datos aleatorizados), entonces esto habría indicado que el código de barras no se había producido de una manera espacialmente restringida según las indicaciones de los reactivos de códigos de barras multiméricos. Sin embargo, tal como se ha explicado anteriormente, los datos indican de manera convincente que se produjeron las reacciones de códigos de barras deseadas, en las que las subsecuencias encontradas en moldes de ADN sintético individuales interactuaron con (y luego se les puso un código de barras mediante) solo con reactivos de códigos de barras multiméricos individuales únicos.

Loci de ADN genómico de códigos de barras con reactivos de códigos de barras multiméricos que contienen oligonucleótidos con códigos de barras

Extracción y análisis de secuencias

Al igual que con otros análisis, la secuencia de comandos se compuso en Python y se implementó en un marco de Amazon Web Services (AWS). Para cada secuencia leída después de la desmultiplexación de muestra, cada región de código de barras del reactivo de código de barras multimérico dado se aisló de su secuencia de adaptador aguas arriba y adaptador aguas abajo flanqueantes y se registró de forma independiente para un análisis adicional. Análogamente, cada secuencia hasta el extremo 3' de la región aguas abajo (que representa la secuencia que contiene el oligonucleótido con código de barras y cualquier secuencia que el oligonucleótido haya cebado durante el protocolo experimental) se aisló para su análisis adicional. Se analizó cada secuencia aguas abajo de cada lectura para la presencia de secuencias de oligonucleótidos adaptadores esperadas (es decir, de los cebadores correspondientes a uno de los tres genes a los que se dirigieron los oligonucleótidos) y secuencias aguas abajo adicionales relevantes. Luego, cada lectura se registró como 'en diana' (con la secuencia correspondiente a una de las secuencias diana esperadas) o 'fuera de diana'. Además, para cada una de las regiones diana, se calculó el número total de códigos de barras multiméricos únicos (es decir, con códigos de barras idénticos pero duplicados fusionados en una representación de copia única). Un esquema de cada secuencia esperada leída, y los componentes constituyentes de la misma, se muestra en la Figura 16.

Discusión

La Figura 15 muestra los resultados de este análisis para este método, para cuatro muestras independientes diferentes. Estas cuatro muestras representan un método en el que el proceso de hibridación de los reactivos de códigos de barras multiméricos tuvo lugar durante 3 horas o durante la noche (aproximadamente 12 horas). Además, para cada una de estas dos condiciones, el método se realizó con los reactivos de códigos de barras multiméricos conservados intactos como se sintetizaron originalmente, o con un protocolo modificado en el que los oligonucleótidos con códigos de barras se desnaturalizan primero de las moléculas del código de barras (mediante una etapa de fusión a alta temperatura). Cada fila representa una diana de amplicón diferente como se indica, y cada célula representa el número total de códigos de barras únicos encontrados asociados con cada amplicón en cada una de las cuatro muestras. También se enumera la proporción total de lecturas en la diana, en todas las dianas sumadas, para cada muestra.

Como se ve en la figura, la mayoría de las lecturas de todas las muestras están en la diana; sin embargo, se observa

un amplio intervalo en el número de moléculas de códigos de barras únicas observadas para cada diana de amplicón. Estas tendencias en diferentes amplicones parecen ser consistentes en las diferentes condiciones experimentales y podrían deberse a diferentes eficacias de cebado (o mal cebado) de los diferentes oligonucleótidos, o diferentes eficacias de amplificación, o diferentes eficacias de mapeo, más otros posibles factores que actúan de forma independiente o en combinación. Además, está claro que las muestras que hibridaron durante más tiempo tienen un mayor número de códigos de barras observados, probablemente debido a una hibridación general más completa de los reactivos multiméricos con sus dianas genómicas afines. Y además, las muestras en las que los oligonucleótidos con códigos de barras se desnaturalizaron por primera vez de las moléculas del código de barras muestran un menor número total de códigos de barras únicos, quizás debido a un efecto de avidez en el que las moléculas de código de barras completamente ensambladas pueden hibridar de manera más eficaz grupos de cebadores con dianas genómicas cercanas en el mismo locus. En cualquier caso, en conjunto, esta figura ilustra la capacidad de los reactivos multiméricos para etiquetar moléculas de ADN genómico, a través de un gran número de moléculas simultáneamente, y para hacerlo si los oligonucleótidos con código de barras permanecen unidos a los reactivos de códigos de barras multiméricos o si se han desnaturalizado a partir de ellos y, por tanto, posiblemente capaces de difundirse más fácilmente en solución.

REIVINDICACIONES

1. Un método para preparar una muestra de ácido nucleico para secuenciación, en donde la muestra comprende al menos 10 células, en donde la muestra de células se ha incubado con un conjunto de dos o más sondas de afinidad con código de barras diferentes, en donde el grupo comprende una primera sonda de afinidad con código de barras que comprende un primer resto de afinidad y un primer oligonucleótido con código de barras de sonda, en donde el primer resto de afinidad es capaz de unirse a una primera proteína diana, y una segunda sonda de afinidad con código de barras que comprende un segundo resto de afinidad y un segundo oligonucleótido con código de barras de sonda, en donde el segundo resto de afinidad es capaz de unirse a una segunda proteína diana, en donde cada oligonucleótido con código de barras de sonda comprende una secuencia de código de barras asociada con y/o identificación del resto de afinidad al que está unido, y en donde el método comprende las etapas de:

(a) poner en contacto la muestra con una biblioteca que comprende al menos dos reactivos de códigos de barras multiméricos, en donde cada reactivo de códigos de barras multimérico comprende un primer y un segundo oligonucleótido con código de barras unidos entre sí y un resto de unión a células, en donde los oligonucleótidos con código de barras comprenden, cada uno, una región de código de barras y en donde las regiones de código de barras del primer y segundo oligonucleótidos con código de barras de un primer reactivo de código de barras multimérico de la biblioteca son diferentes a las regiones de código de barras del primer y segundo oligonucleótidos con código de barras de un segundo reactivo de código de barras multimérico de la biblioteca, en donde el resto de unión celular del primer reactivo de código de barras multimérico se une a la membrana celular de una primera célula antes de la etapa (b), y en donde el resto de unión celular del segundo reactivo de código de barras multimérico se une a la membrana celular de una segunda célula antes de la etapa (b);

(b) lisar las células o permeabilizar las membranas celulares de las células; y

(c) hibridar o ligar el primer y segundo oligonucleótidos con código de barras del primer reactivo de código de barras multimérico a la primera y segunda subsecuencia de un ácido nucleico diana de la primera célula para producir una primera y segunda moléculas de ácido nucleico diana con código de barras, e hibridar o ligar el primer y segundo oligonucleótidos con código de barras del segundo reactivo de código de barras multimérico a la primera y segunda subsecuencias de un ácido nucleico diana de la segunda célula para producir una primera y segunda moléculas de ácido nucleico diana con código de barras,

en donde las células están comprendidas dentro de un único volumen acuoso contiguo durante las etapas (a), (b) y (c), y en donde los ácidos nucleicos diana comprenden oligonucleótidos con código de barras de sonda dentro de las sondas de afinidad con código de barras.

2. El método de la reivindicación 1, en donde la etapa (c) comprende:

(i) hibridar el primer y segundo oligonucleótidos con código de barras del primer reactivo de código de barras multimérico con la primera y segunda subsecuencia de un ácido nucleico diana de la primera célula, e hibridar el primer y segundo oligonucleótidos con código de barras del segundo reactivo de código de barras multimérico con la primera y segunda subsecuencias de un ácido nucleico diana de la segunda célula; y

(ii) extender el primer y segundo oligonucleótidos con código de barras del primer reactivo de código de barras multimérico para producir una primera y segunda moléculas de ácido nucleico diana con código de barras diferentes y extender el primer y segundo oligonucleótidos con código de barras del segundo reactivo de código de barras multimérico para producir una primera y segunda molécula de ácido nucleico diana con código de barras diferente, en donde cada una de las moléculas de ácido nucleico diana con código de barras comprende al menos un nucleótido sintetizado a partir del ácido nucleico diana como molde.

3. El método de la reivindicación 1 o la reivindicación 2, en donde los reactivos de códigos de barras multiméricos comprenden cada uno:

(i) la primera y segunda moléculas de hibridación unidas entre sí, en donde cada una de las moléculas de hibridación comprende una secuencia de ácido nucleico que comprende una región de hibridación; y

(ii) el primer y segundo oligonucleótidos con código de barras, en donde el primer oligonucleótido con código de barras hibrida con la región de hibridación de la primera molécula de hibridación y en donde el segundo oligonucleótido con código de barras hibrida con la región de hibridación de la segunda molécula de hibridación.

4. El método de la reivindicación 3, en donde los reactivos de códigos de barras multiméricos comprenden cada uno:

(i) la primera y segunda moléculas de código de barras unidas entre sí, en donde cada una de las moléculas de código de barras comprende una secuencia de ácido nucleico que comprende una región de código de barras; y

(ii) el primer y segundo oligonucleótidos con código de barras, en donde el primer oligonucleótido con código de barras comprende una región de código de barras hibridada con la región de código de barras de la primera molécula de código de barras, y

en donde el segundo oligonucleótido con código de barras comprende una región de código de barras hibridada con la región de código de barras de la segunda molécula de código de barras.

5. El método de una cualquiera de las reivindicaciones 1-4, en donde la etapa (b) se realiza aumentando la temperatura de la muestra.
- 5 6. El método de una cualquiera de las reivindicaciones 1-5, en donde la etapa (b) se realiza en presencia de un tensioactivo químico.
7. El método de una cualquiera de las reivindicaciones 1-6, en donde la etapa (b) se realiza en condiciones hipotónicas o hipertónicas.
8. El método de una cualquiera de las reivindicaciones 1-7, en donde los ácidos nucleicos diana comprenden además moléculas de ARNm comprendidas dentro de las células.
- 10 9. El método de una cualquiera de las reivindicaciones 1-8, en donde un resto de unión a células se une a cada uno de los oligonucleótidos con código de barras.
10. El método de una cualquiera de las reivindicaciones 1-9, en donde cada uno de los reactivos de códigos de barras multiméricos comprenden un soporte sólido o un soporte semisólido, y en donde un resto de unión a células está unido al soporte sólido o al soporte semisólido.
- 15 11. El método de una cualquiera de las reivindicaciones 1-10, en donde la muestra de células se ha reticulado químicamente antes de la incubación de la muestra de células con el conjunto de sondas de afinidad con código de barras.
12. El método de una cualquiera de las reivindicaciones 1-11, en donde la muestra comprende células que han sido reticuladas con formalina y embebidas en parafina.
- 20

FIGURA 1

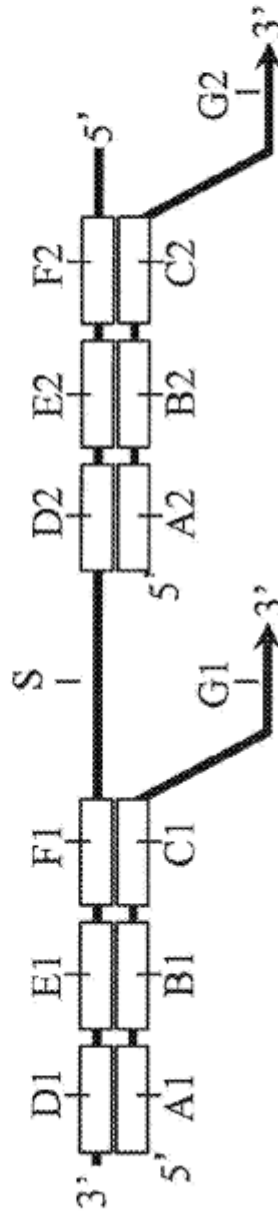


FIGURA 2

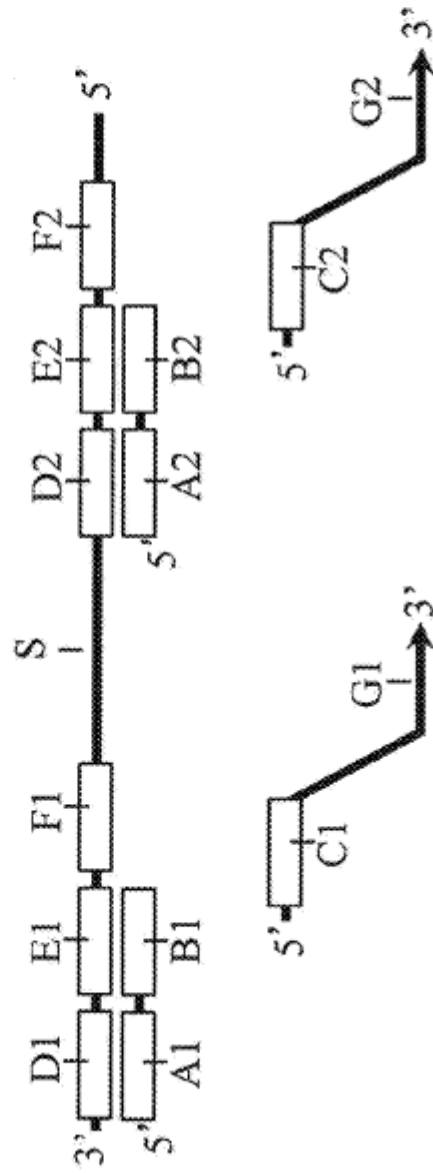


FIGURA 3

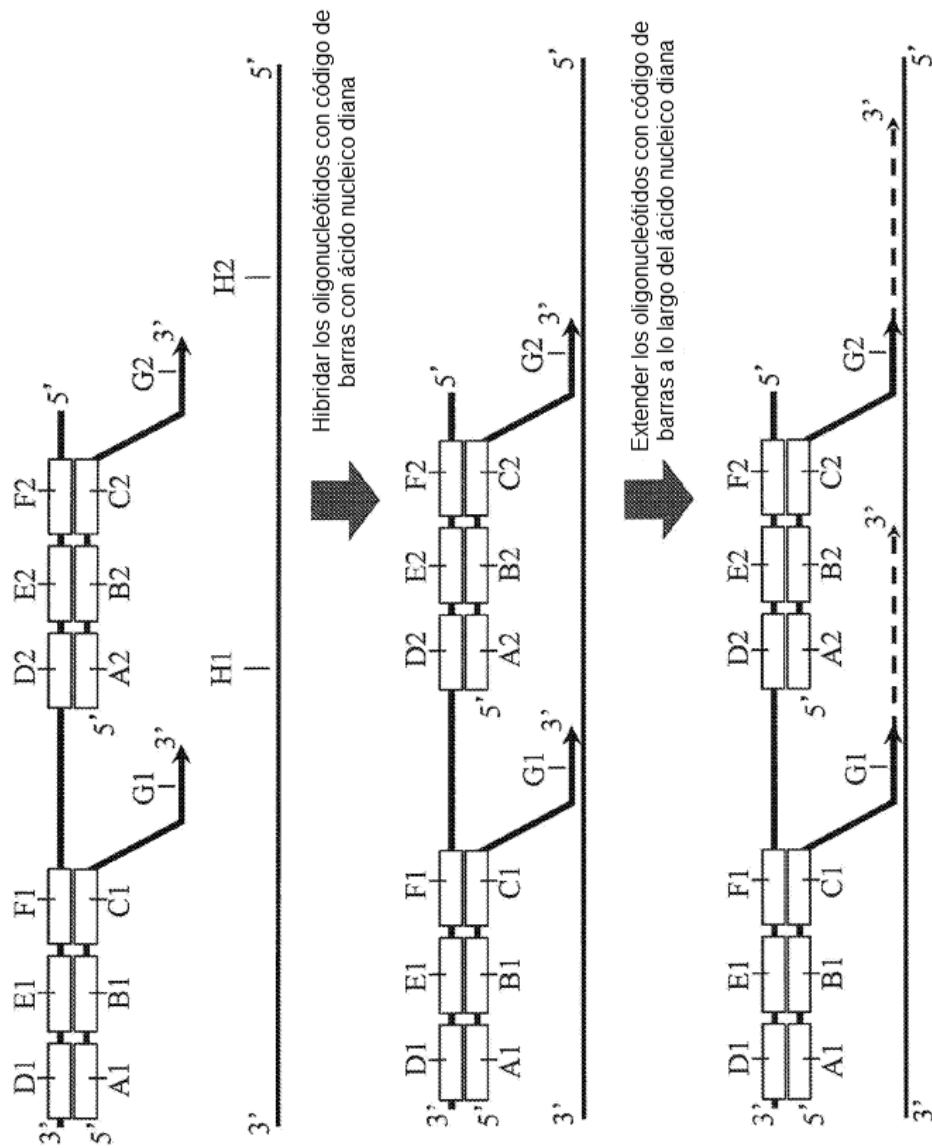


FIGURA 4

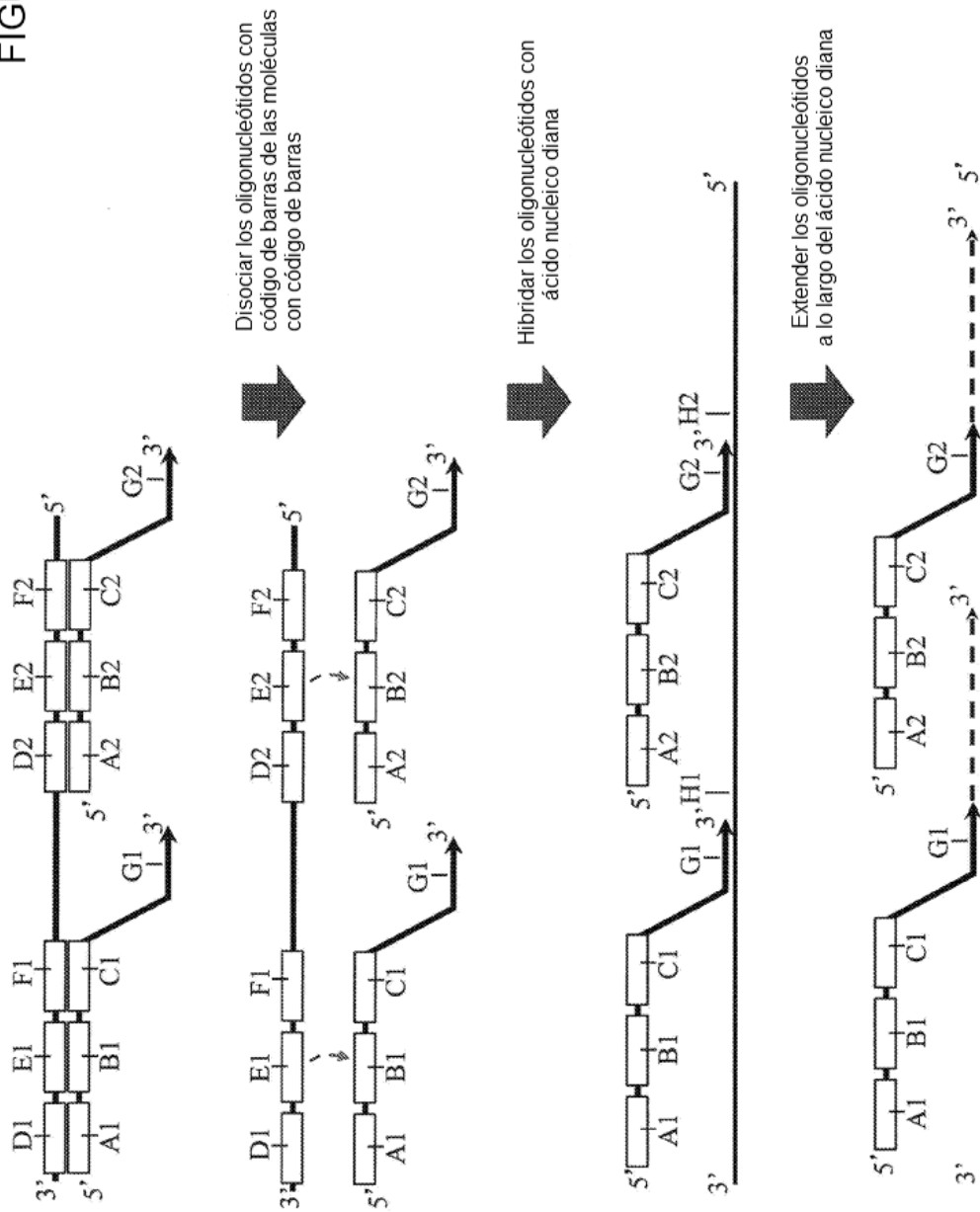


FIGURA 5

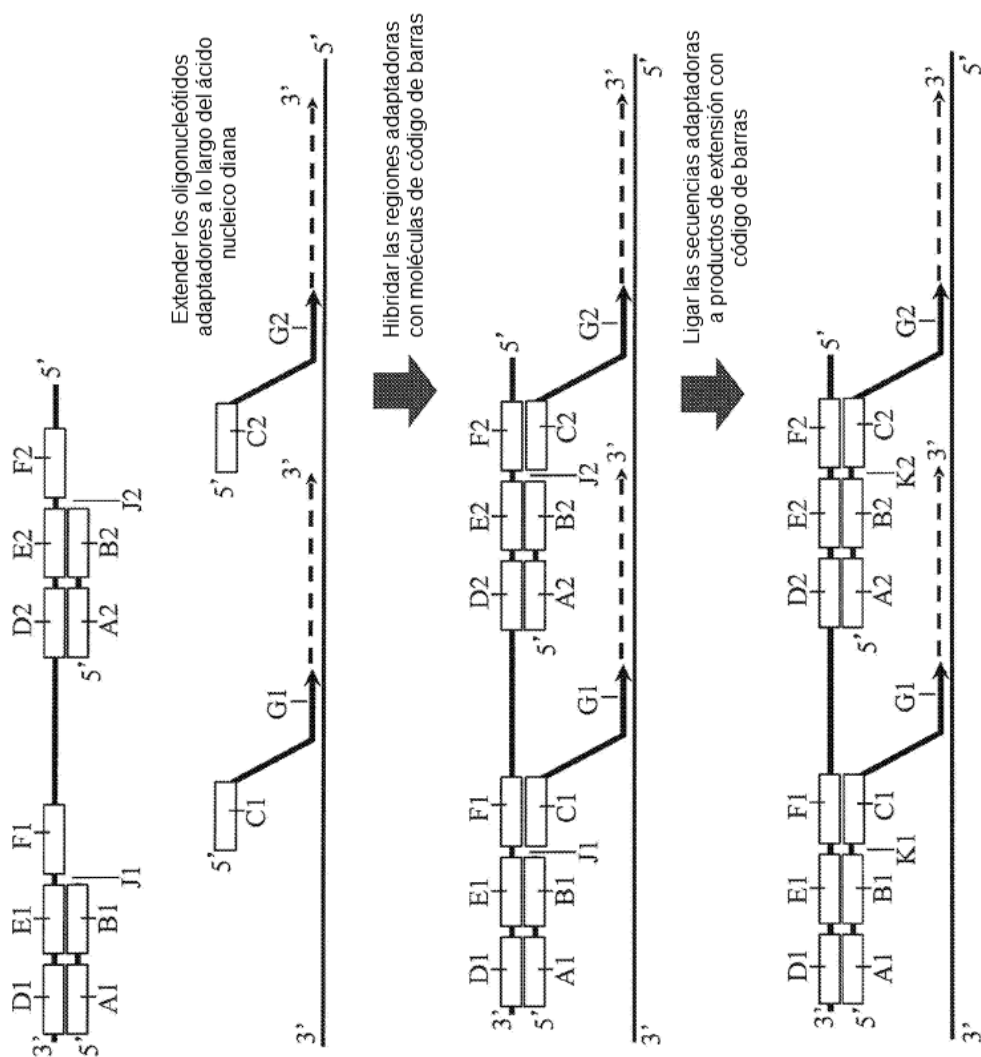


FIGURA 6

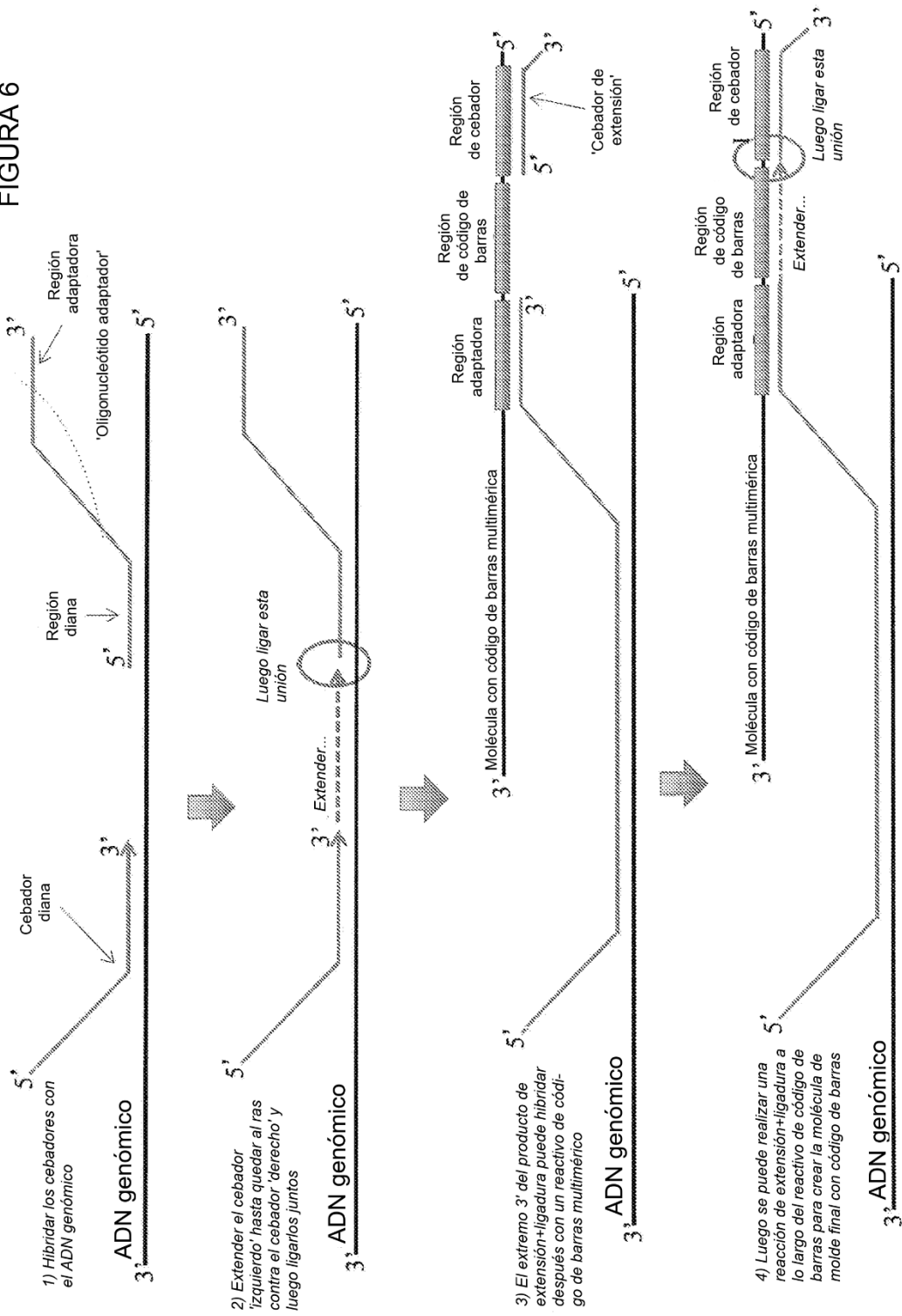


FIGURA 7

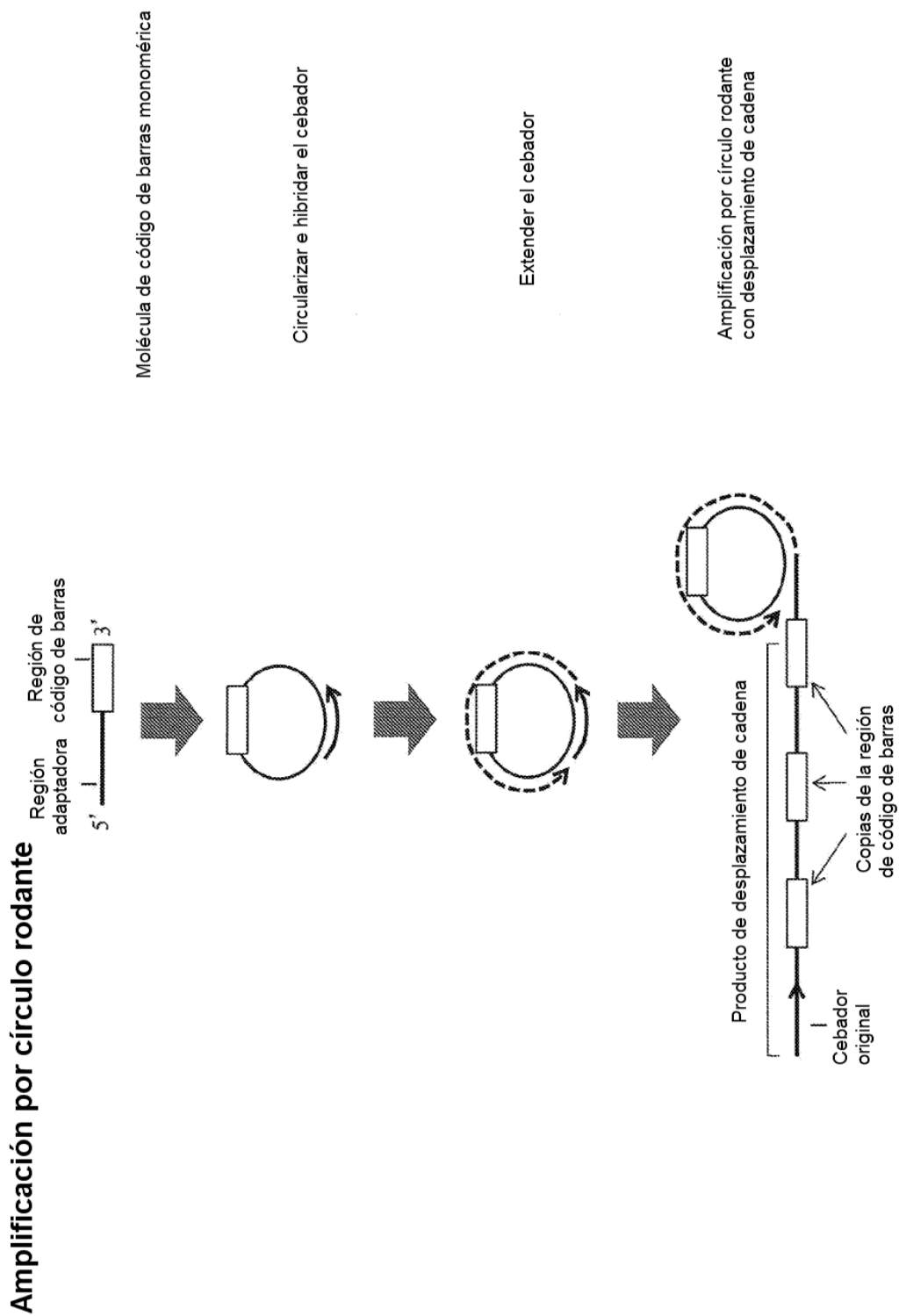


FIGURA 8

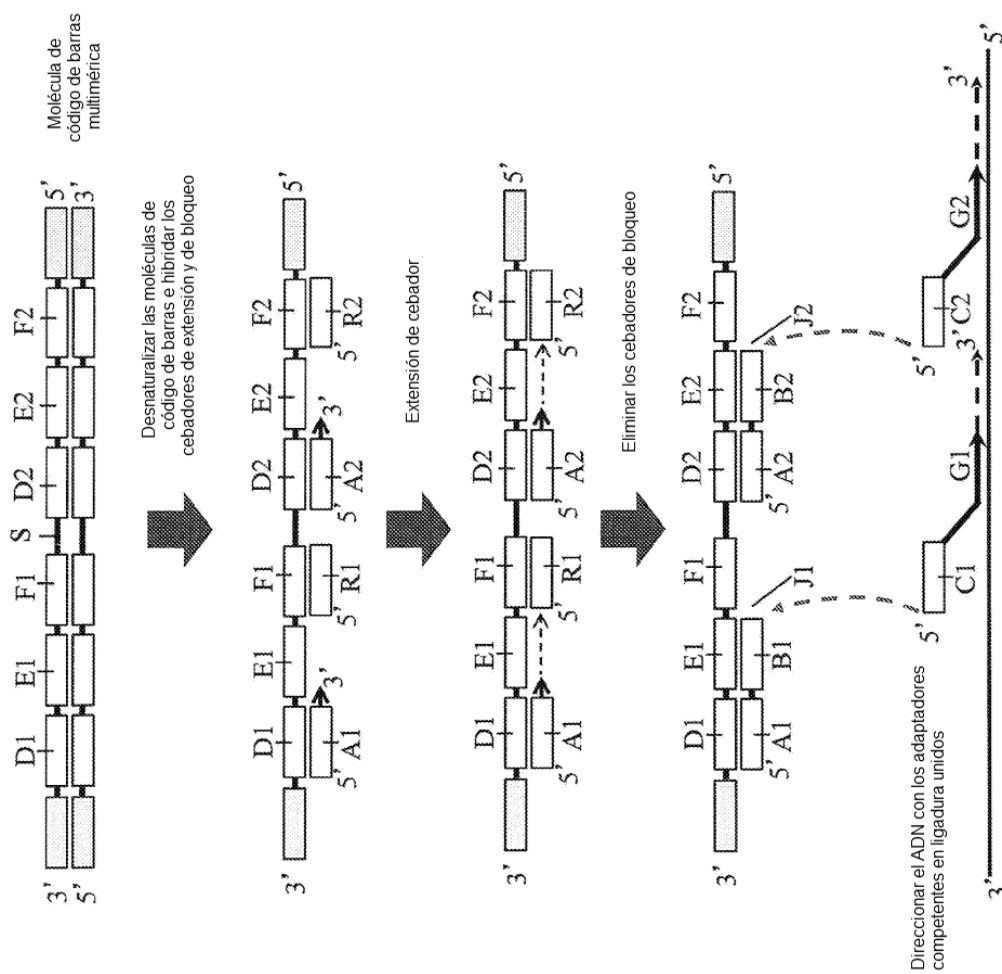


FIGURA 9

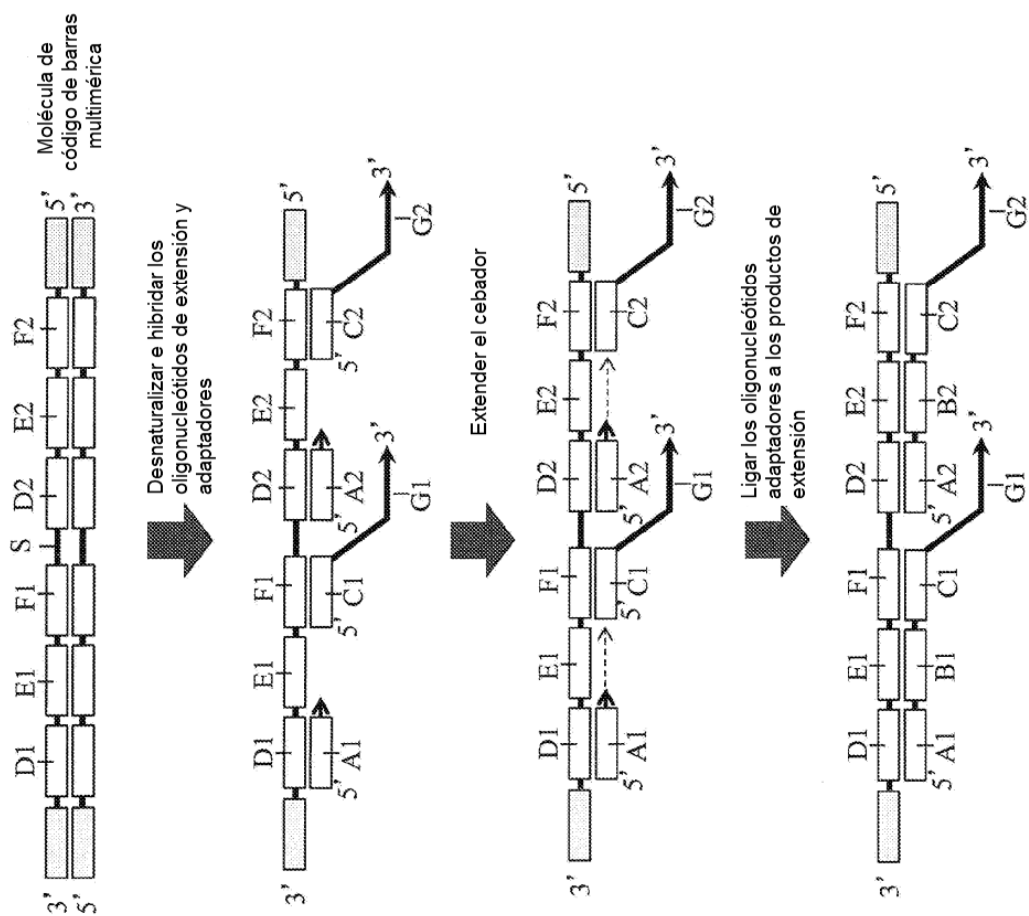


FIGURA 10

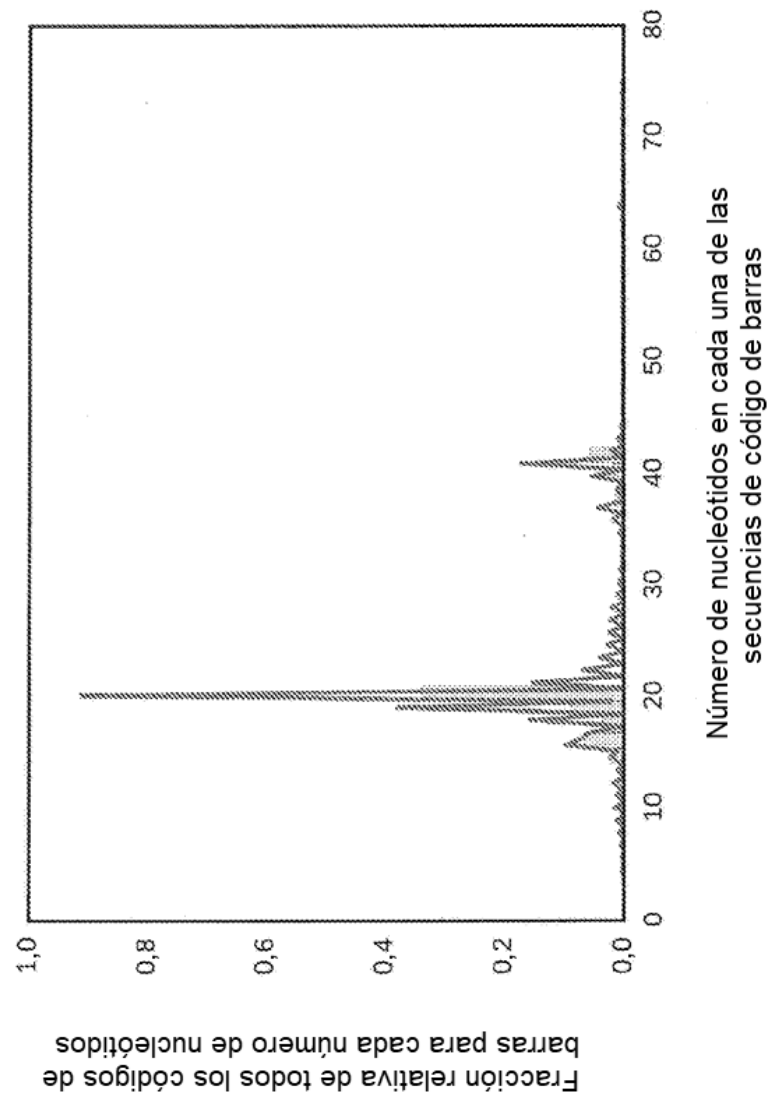


FIGURA 11

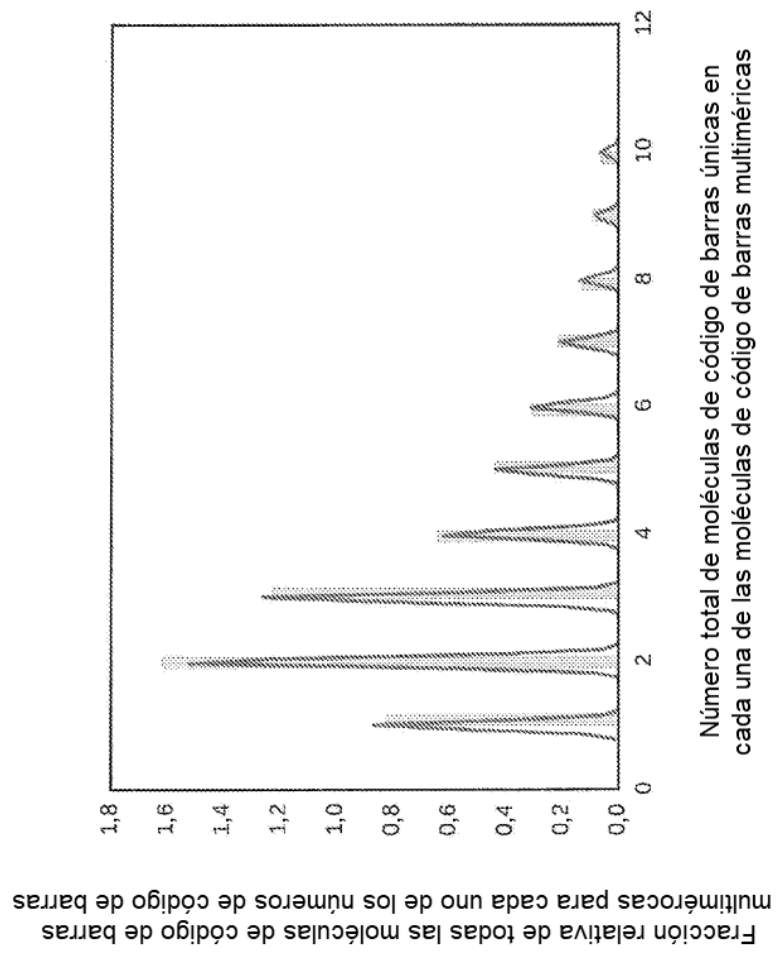


FIGURA 12

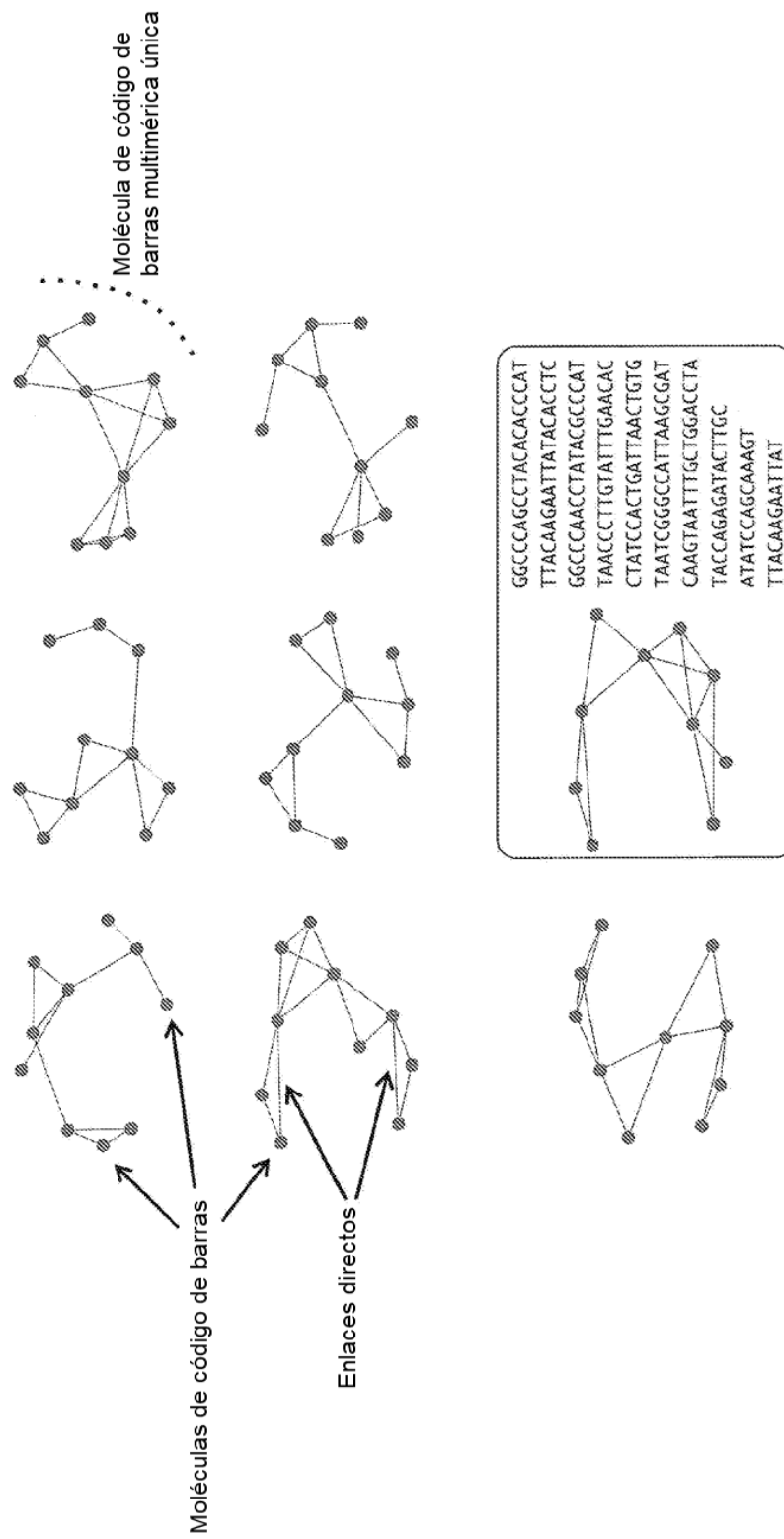


FIGURA 13

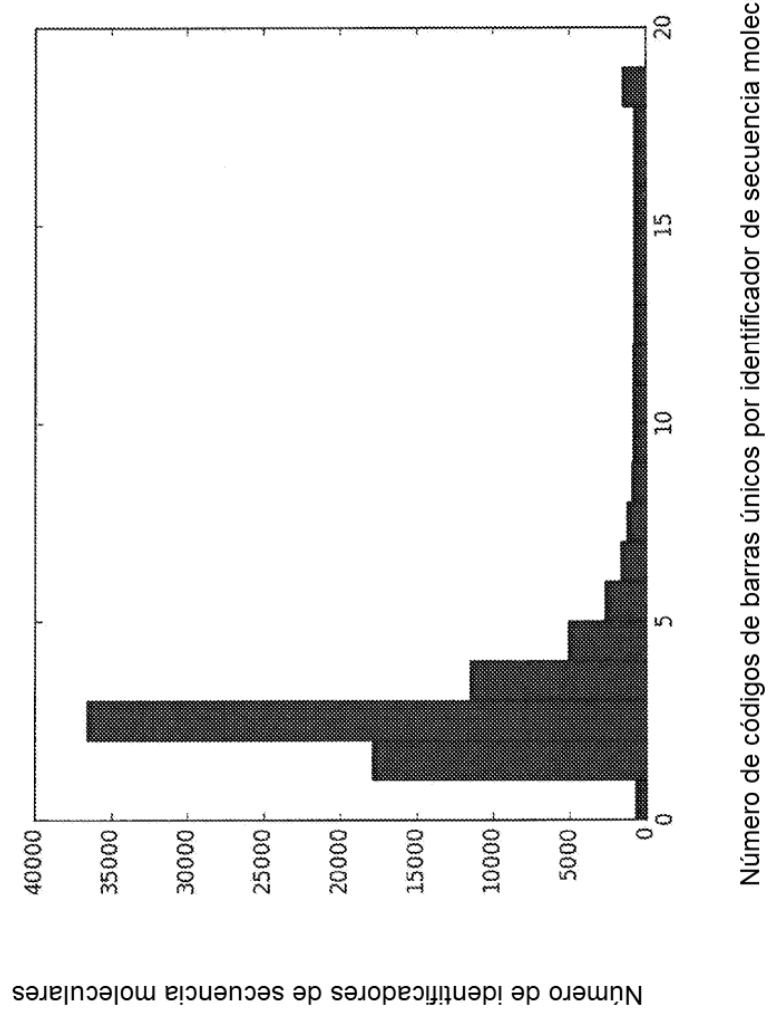


FIGURA 14

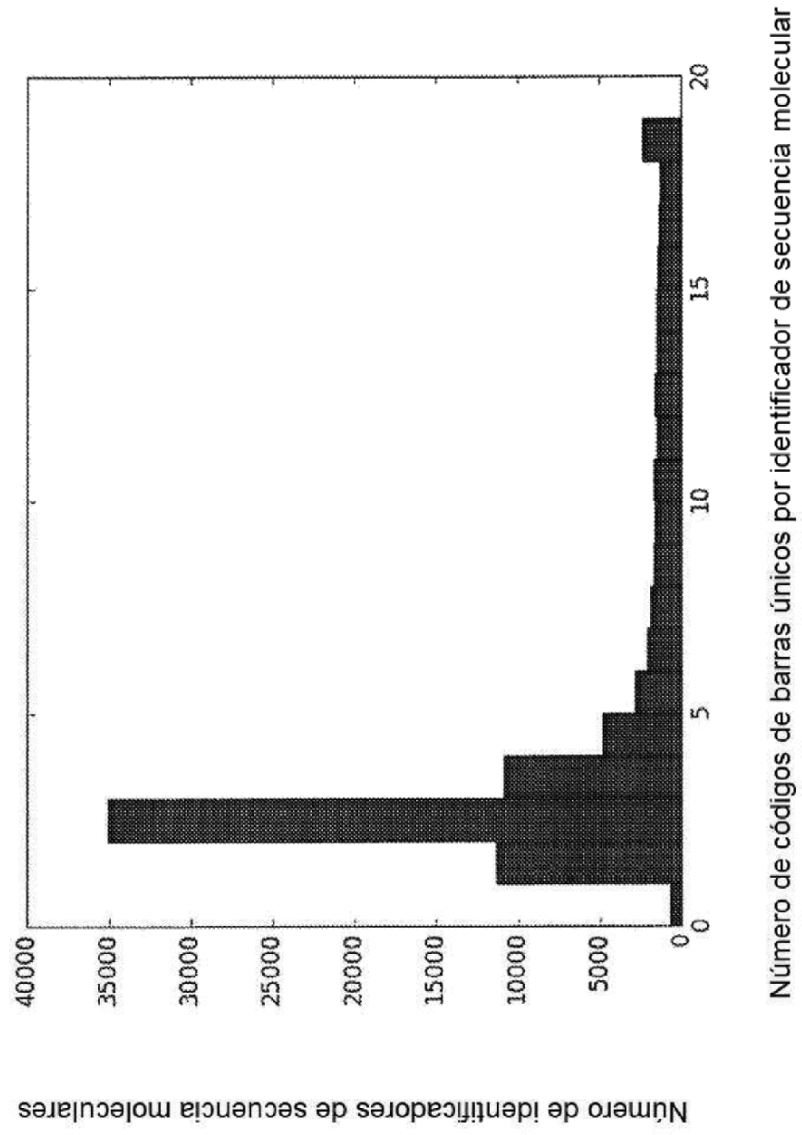
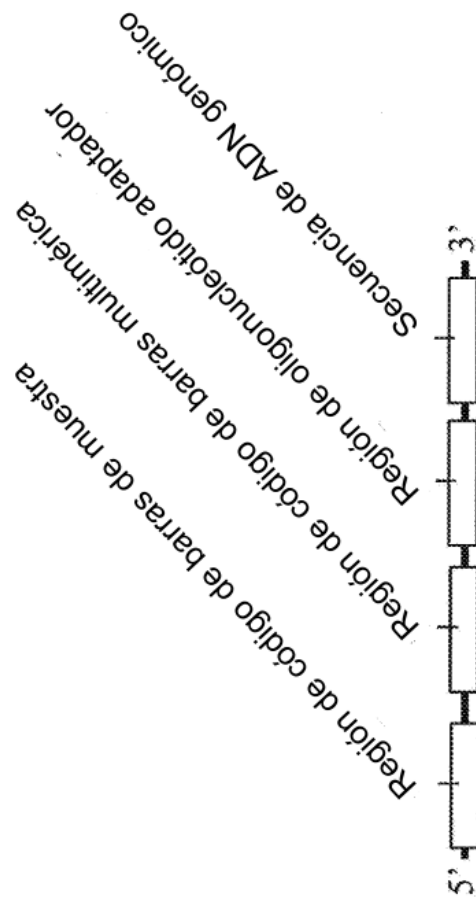


FIGURA 15

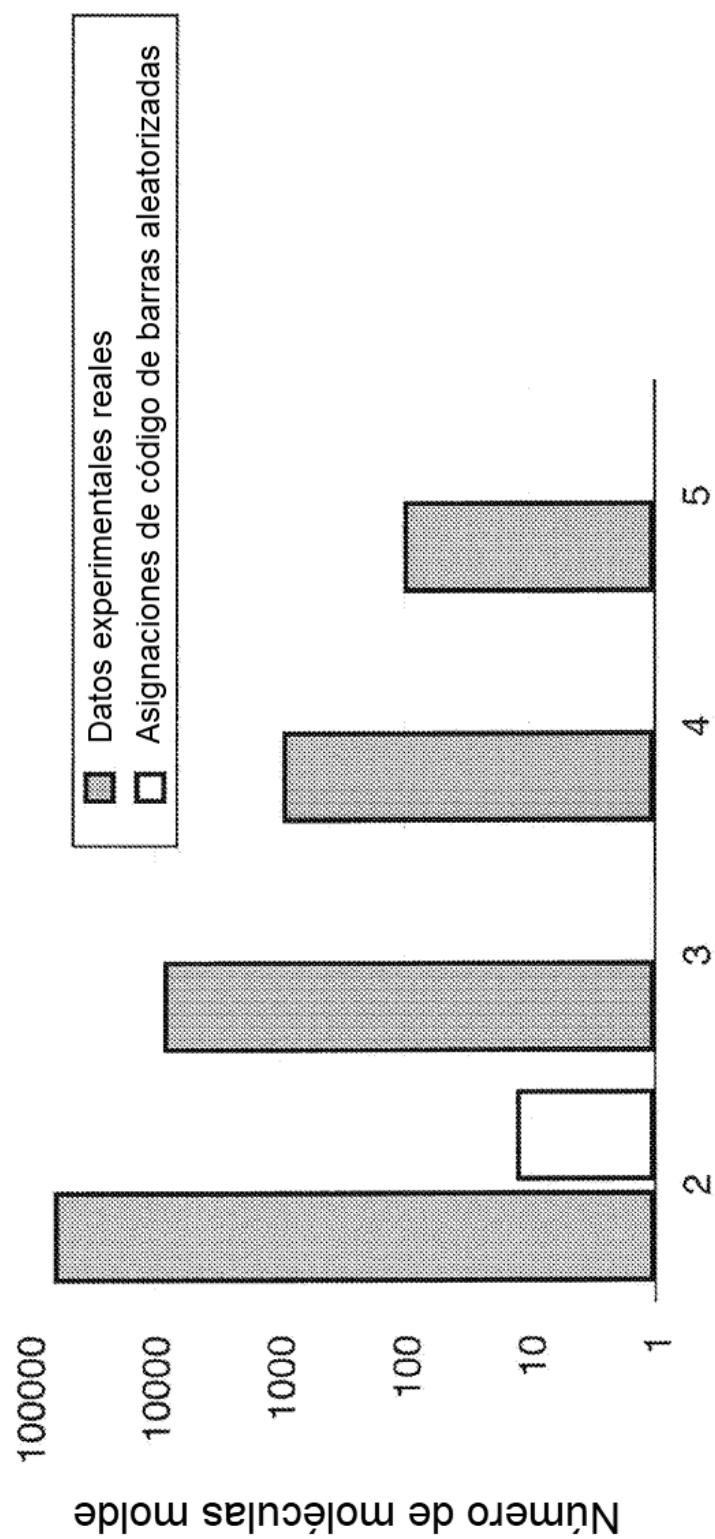
	Hibridado con ADN genómico toda la noche	Hibridado con ADN genómico durante 3 horas; reactivos de código de barras desnaturalizados primero	Hibridado con ADN genómico durante 3 horas	Hibridado con ADN genómico durante 3 horas; reactivos de código de barras desnaturalizados primero
HLA-A_Exon2	153	62	103	50
HLA-A_Exon3	2548	341	818	106
HLA-A_Exon4	489	235	400	108
DQB1_Exon2	557	627	285	635
DQB1_Exon3	2186	1614	1085	221
BRCA1_US1	1771	941	1047	544
BRCA1_SNP7	2524	1261	90	43
BRCA1_SNP6	731	537	245	339
BRCA1_SNP1	10033	5787	2905	1813
BRCA1_DS1	4	103	5	0
BRCA_AMP6	81	69	37	24
BRCA_AMP7	731	391	40	472
Fracción en diana:				
	81 %	76 %	82 %	70 %

FIGURA 16



Esquema de lectura de secuencia de colocación
de código de barras de loci de ADN genómico
(En dirección 5' a 3')

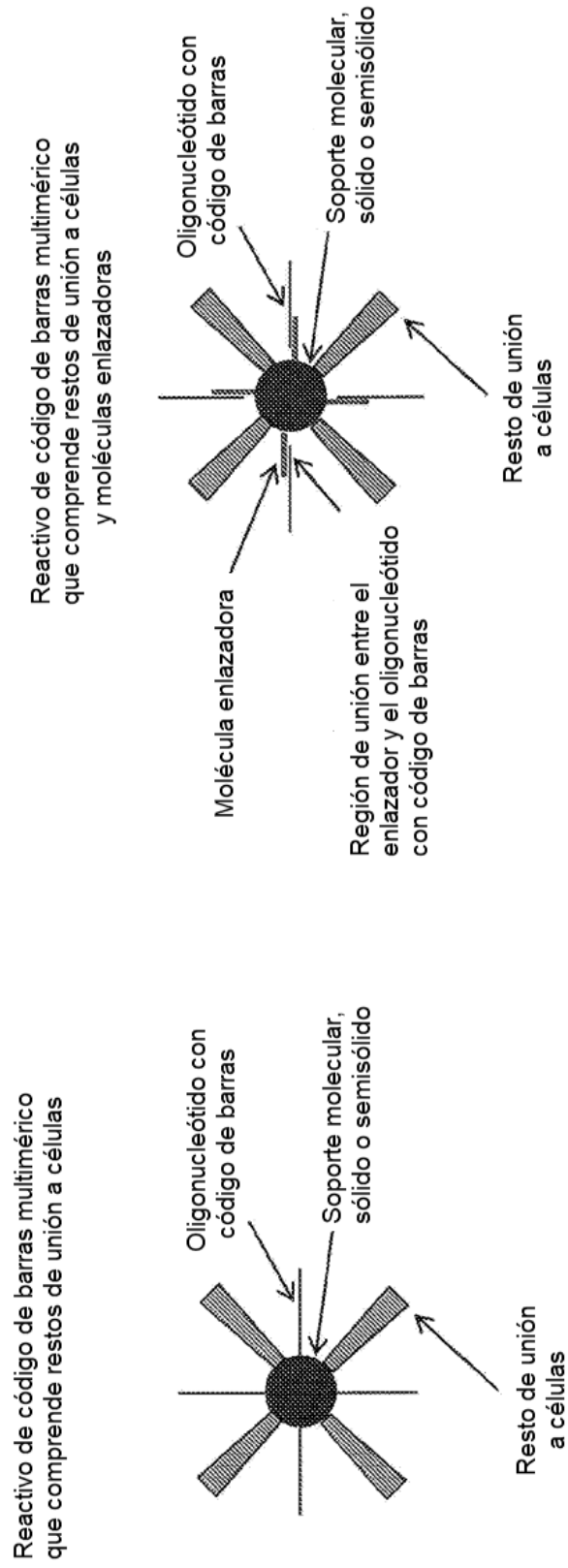
FIGURA 17



Número de códigos de barras del mismo reactivo de código de barras multimérico por molécula molde

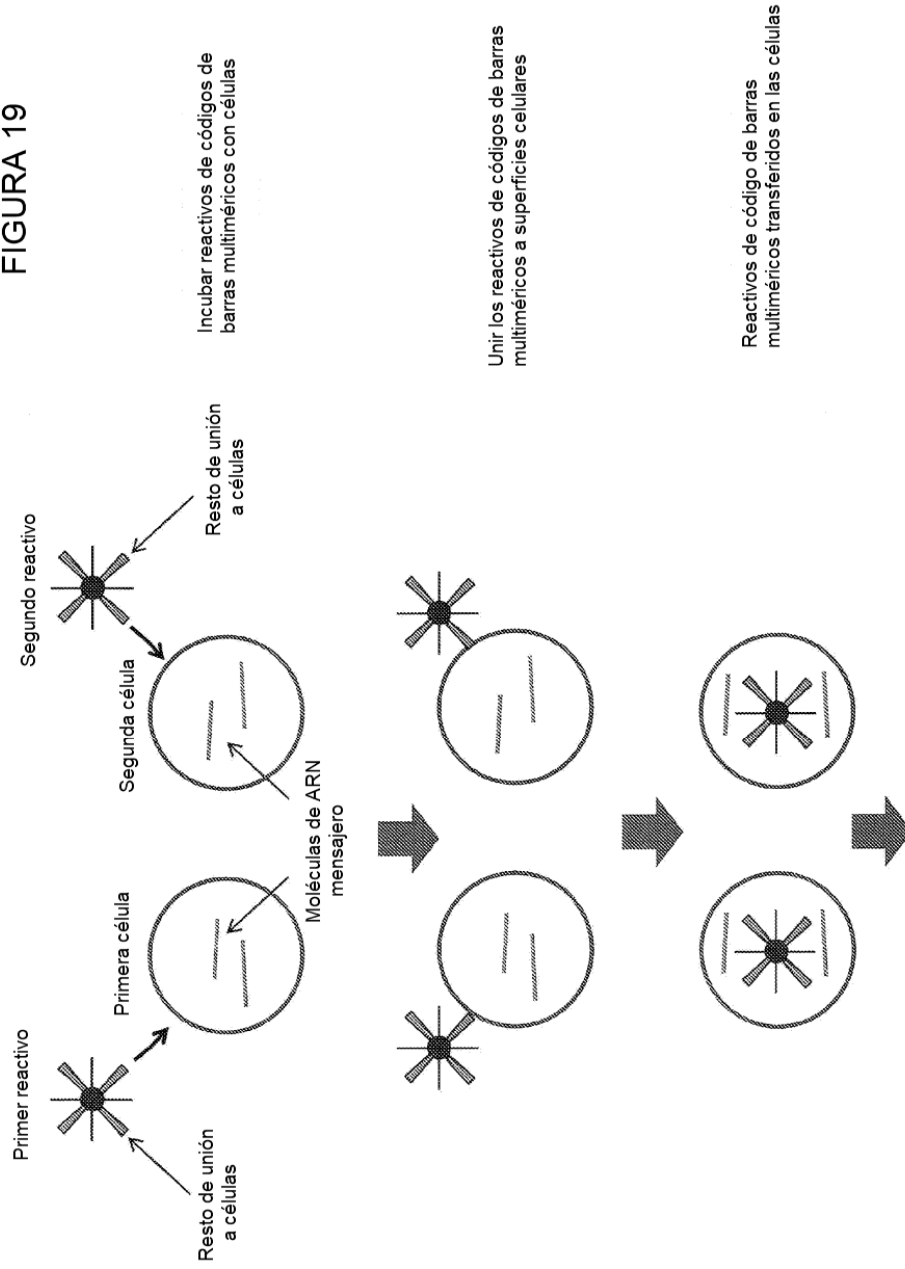
FIGURA 18

Reactivos de códigos de barras multiméricos que comprenden restos de unión a células



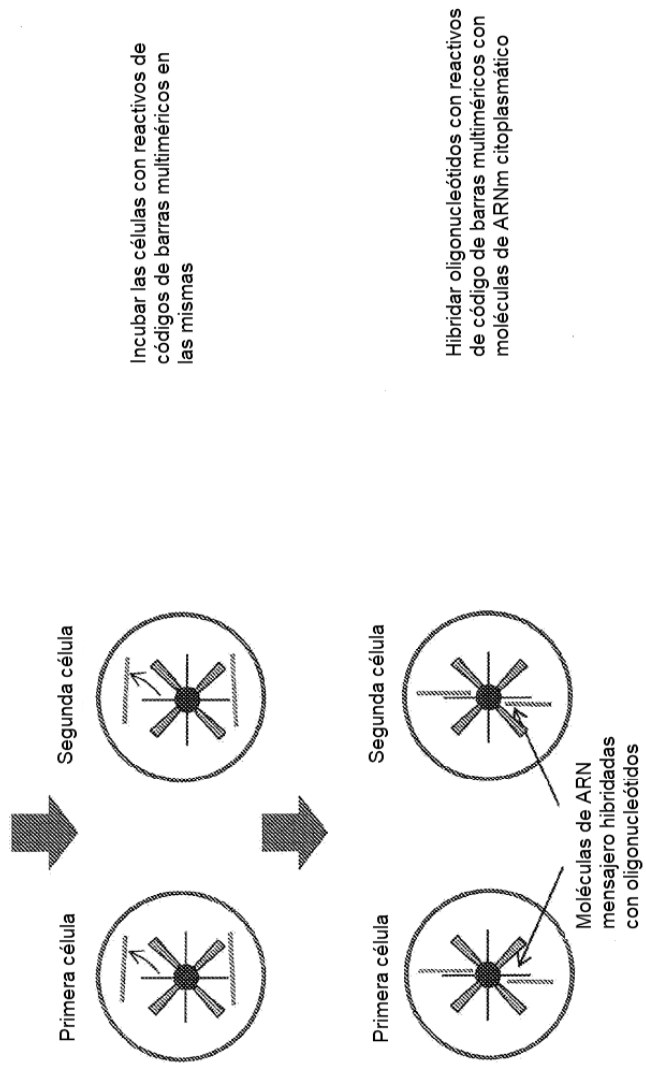
Transferencia de reactivos de códigos de barras multiméricos en células por restos de unión a células

FIGURA 19



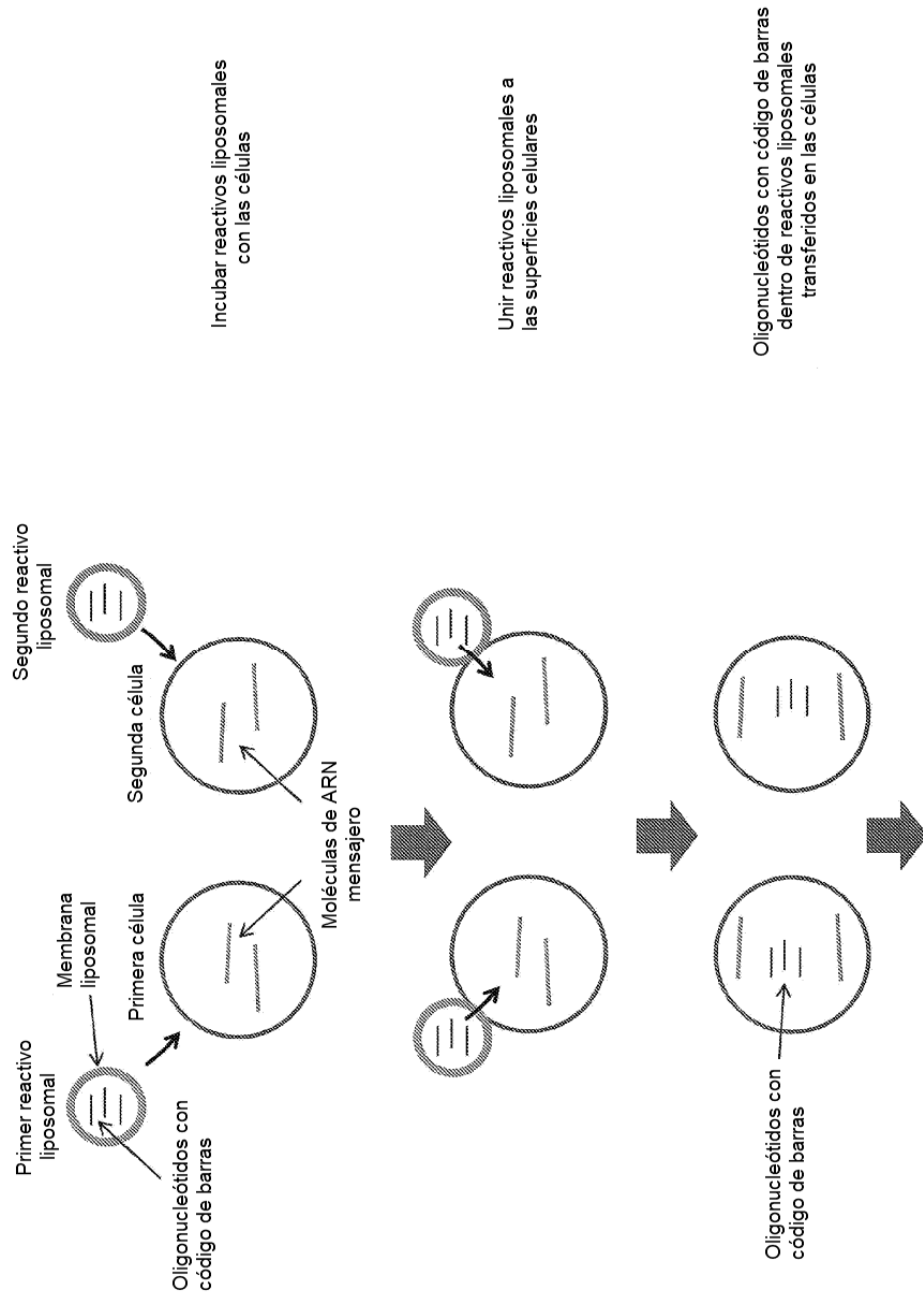
Transferencia de reactivos de códigos de barras multiméricos en células por restos de unión a células

FIGURA 19



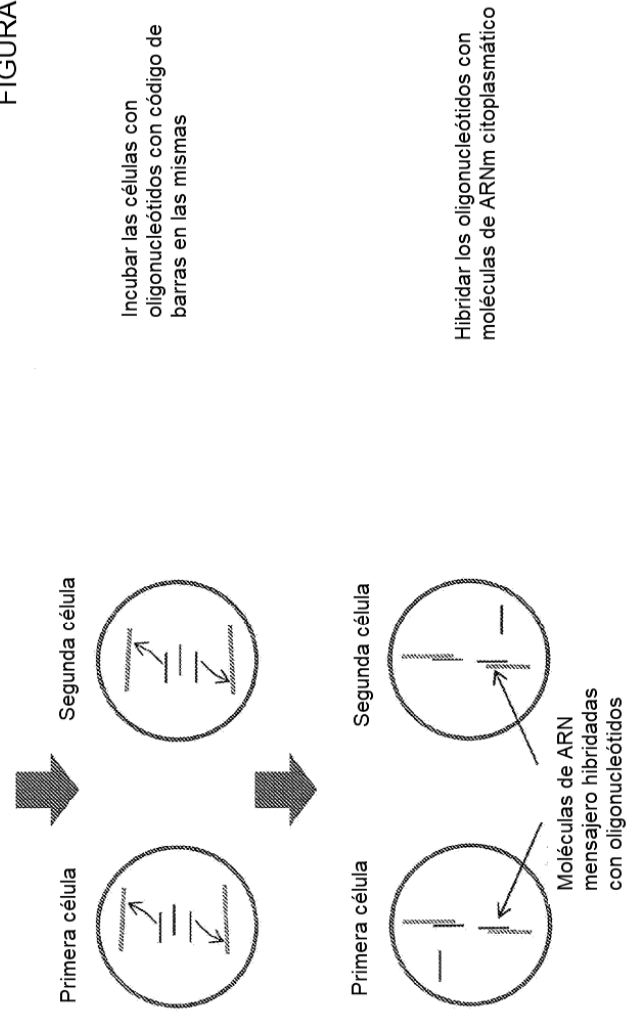
Transferencia de reactivos de código de barras multiméricos en células mediante administración liposomal

FIGURA 20



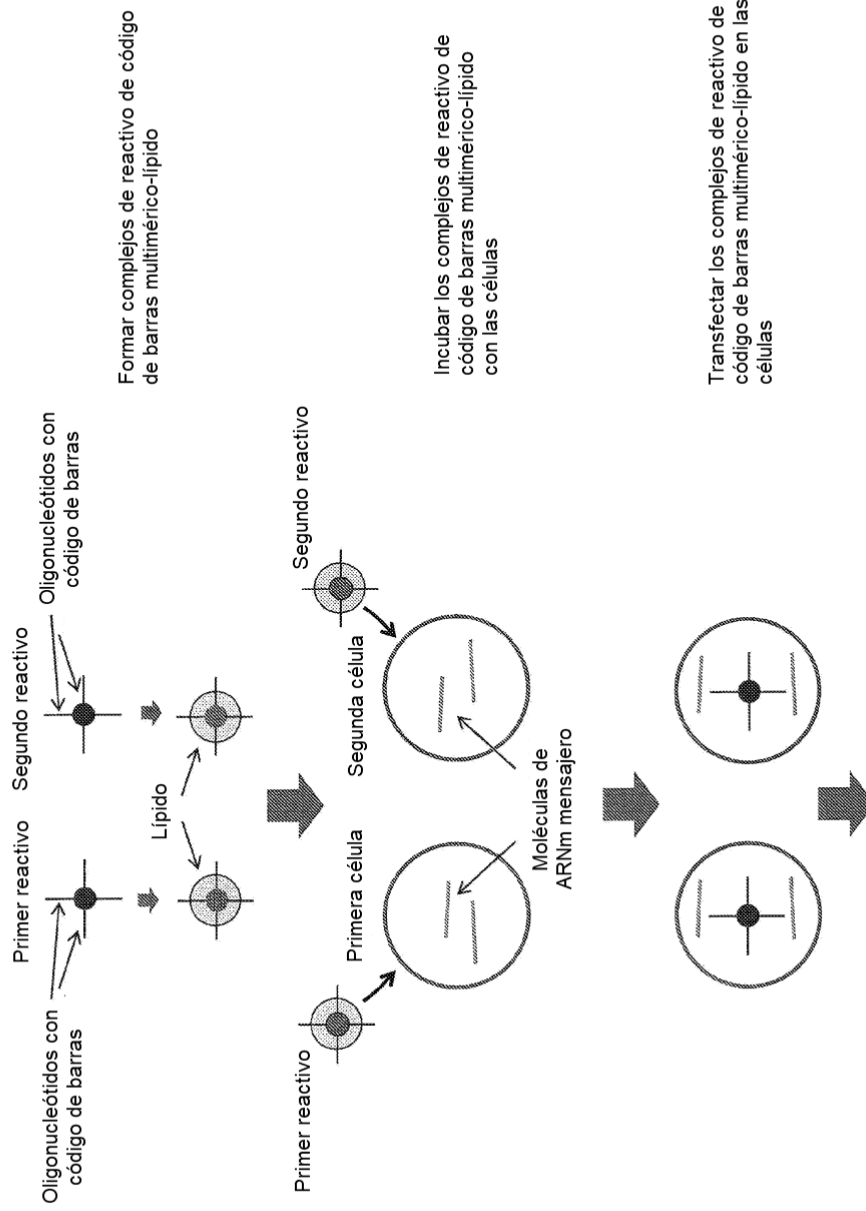
Transferencia de reactivos de códigos de barras multiméricos en células mediante administración liposomal

FIGURA 20



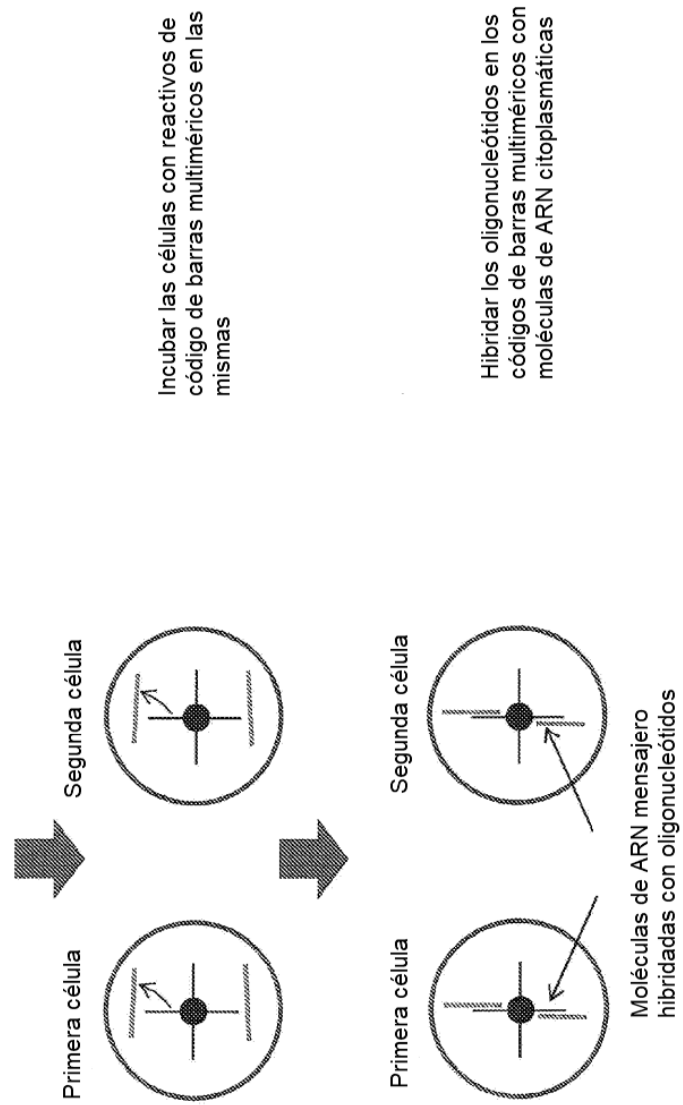
Transferencia de reactivos de códigos de barras multiméricos en células mediante transfección

FIGURA 21



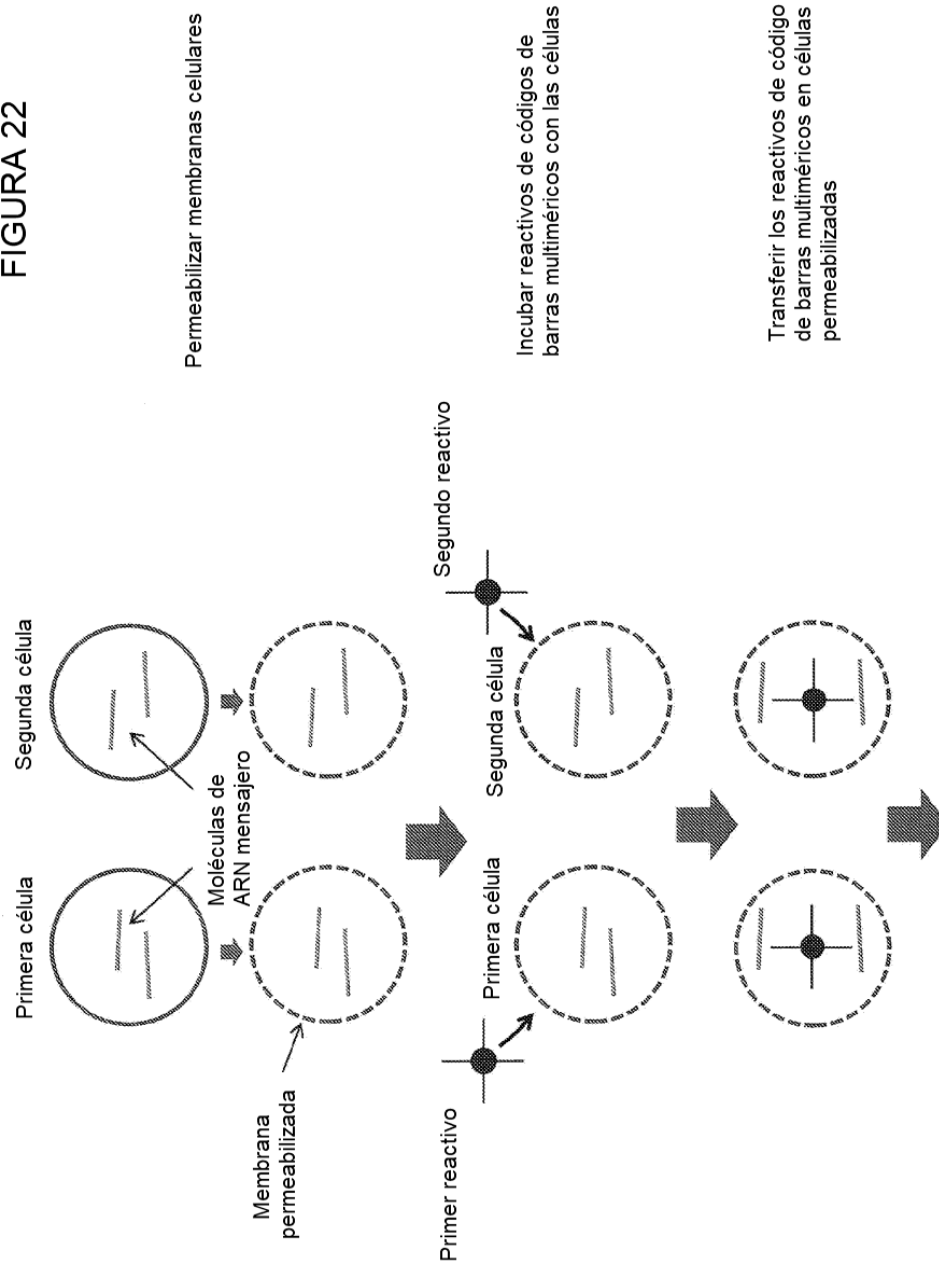
Transferencia de reactivos de códigos de barras multiméricos en células mediante transfección

FIGURA 21



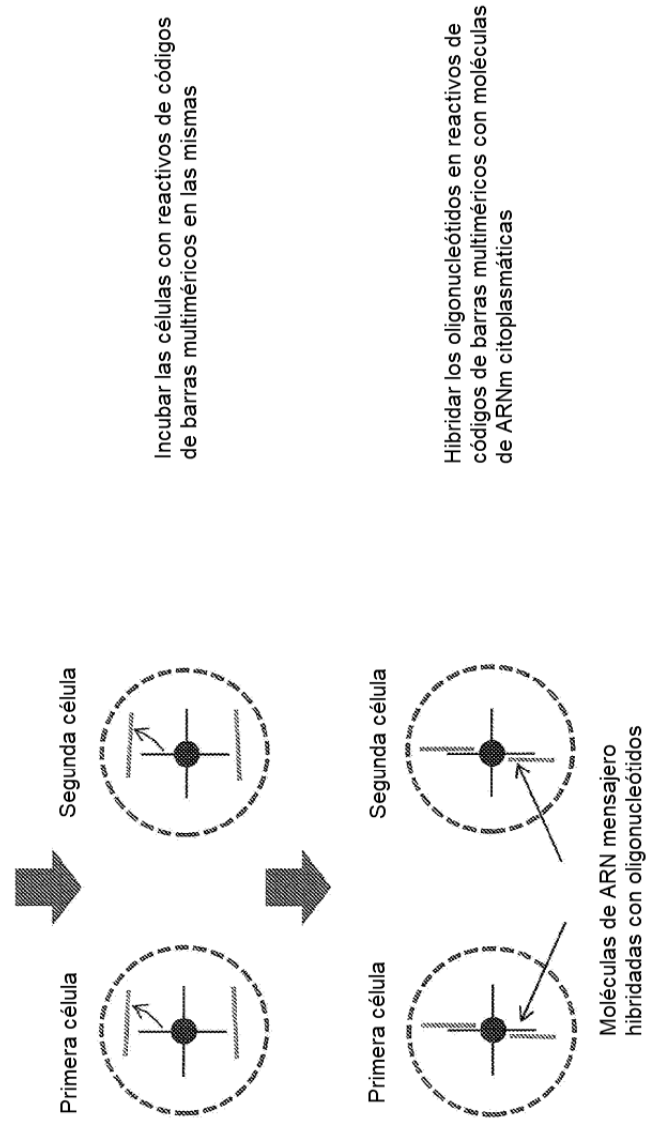
Transferencia de reactivos de código de barras multiméricos en células mediante permeabilización celular

FIGURA 22



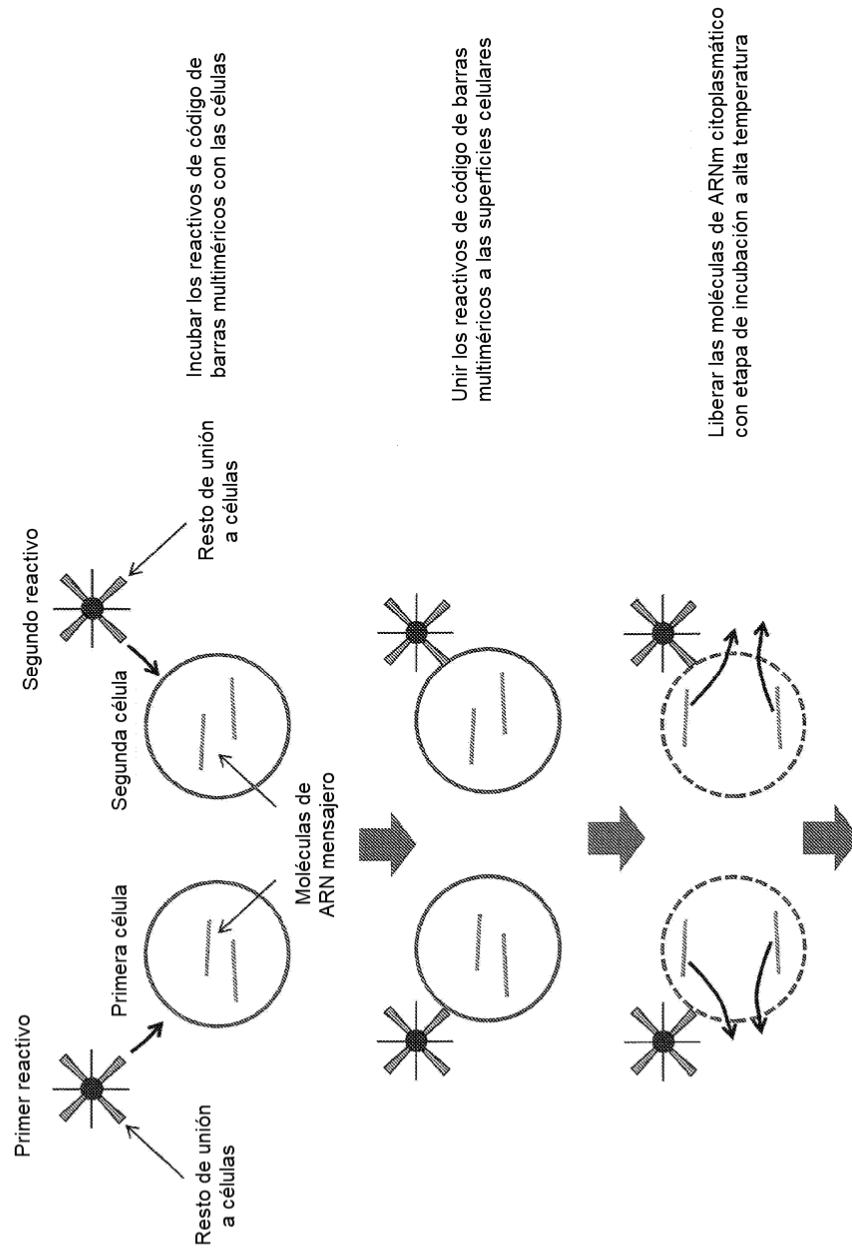
Transferencia de reactivos de código de barras multiméricos en células mediante permeabilización celular

FIGURA 22



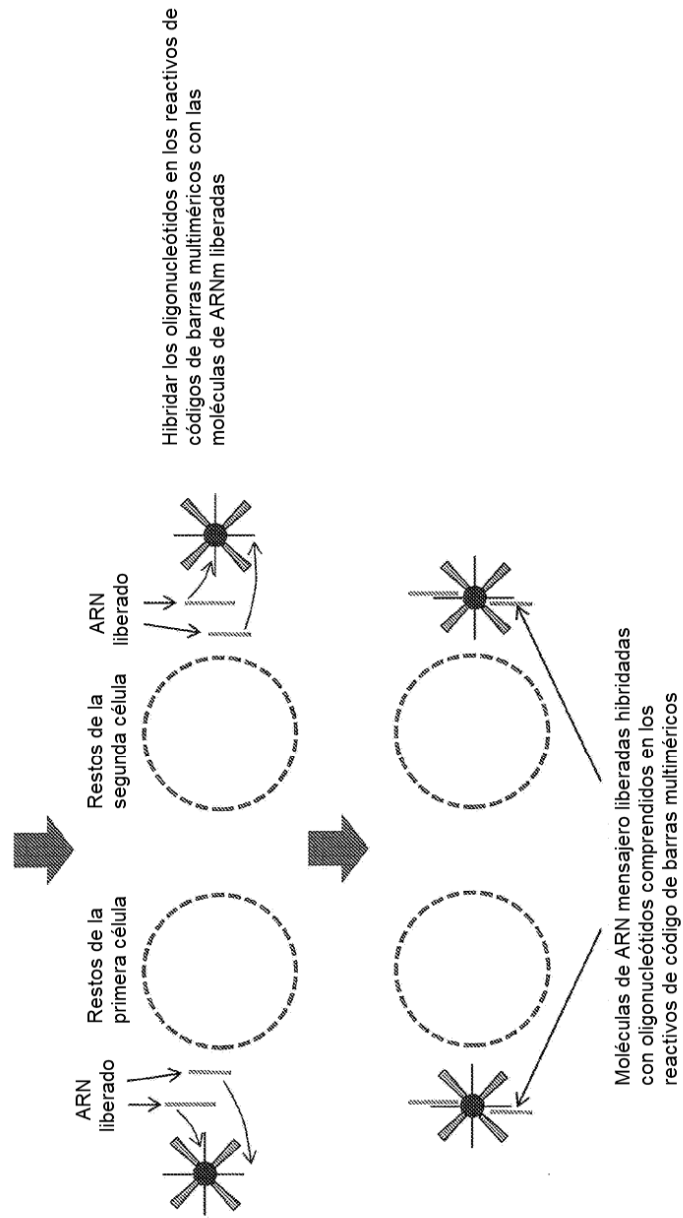
Colocación de códigos de barras en ácidos nucleicos celulares con una etapa de permeabilización de membrana

FIGURA 23



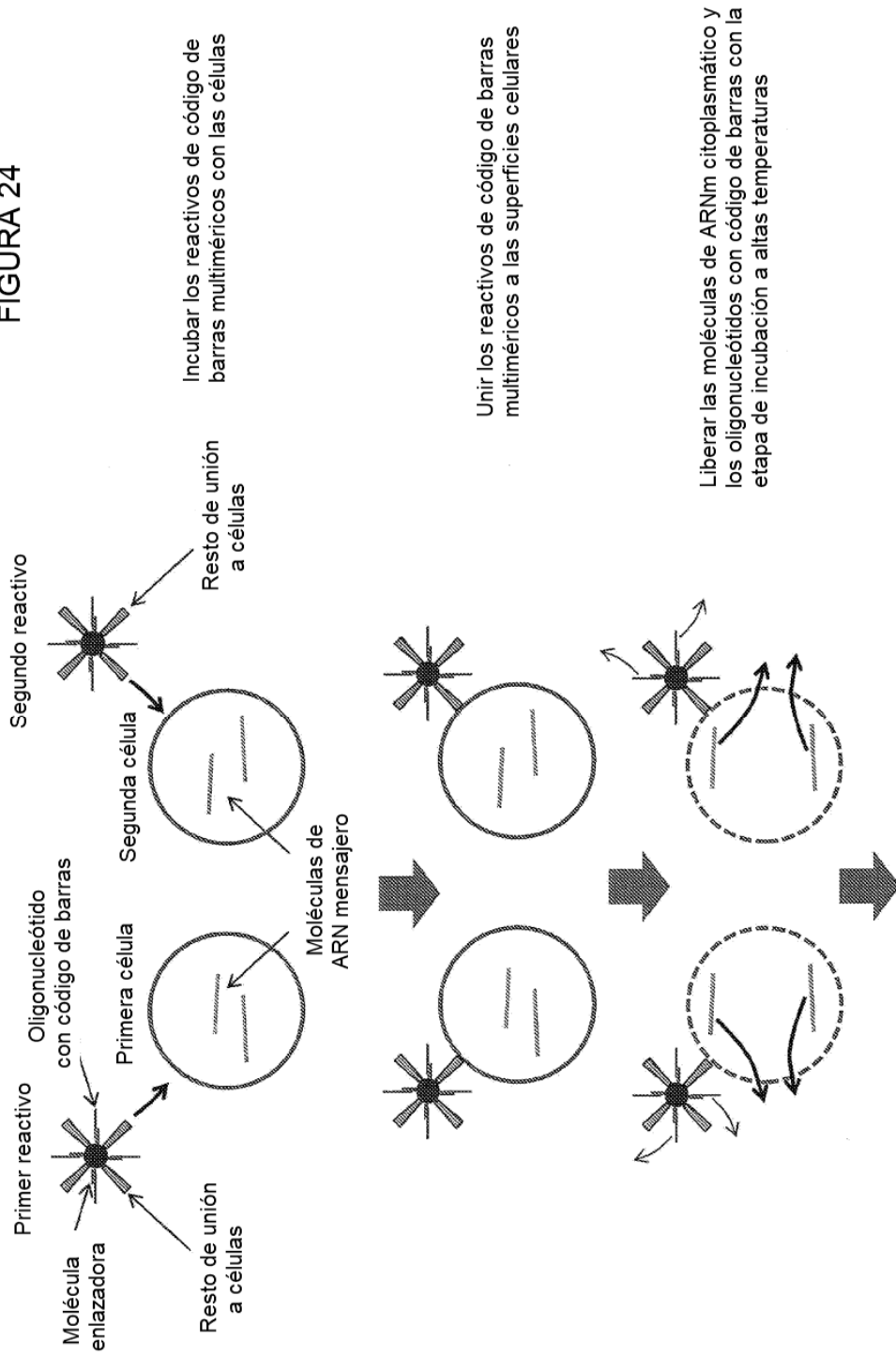
Colocación de códigos de barras en ácidos nucleicos celulares con una etapa de permeabilización de membrana

FIGURA 23



Colocación de códigos de barras en ácidos nucleicos celulares con una etapa de permeabilización de membrana y de liberación de oligonucleótidos con código de barras

FIGURA 24



Colocación de códigos de barras en ácidos nucleicos celulares con una etapa de permeabilización de membrana y de liberación de oligonucleótidos con código de barras

FIGURA 24

