

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 727 656**

51 Int. Cl.:

**C12N 15/10** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **15.09.2015 PCT/US2015/050119**

87 Fecha y número de publicación internacional: **24.03.2016 WO16044227**

96 Fecha de presentación y número de la solicitud europea: **15.09.2015 E 15778415 (8)**

97 Fecha y número de publicación de la concesión europea: **06.02.2019 EP 3194593**

54 Título: **Secuenciación de alto rendimiento de banco de nucleótidos**

30 Prioridad:

**15.09.2014 US 201462050549 P**  
**17.09.2014 US 201462051832 P**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:  
**17.10.2019**

73 Titular/es:

**ABVITRO LLC (100.0%)**  
**27 Drydock Avenue 6th Floor**  
**Boston, Massachusetts 02210, US**

72 Inventor/es:

**VIGNEAULT, FRANCOIS;**  
**WRANGHAM BRIGGS, ADRIAN;**  
**CLOUSER, CHRISTOPHER RYAN;**  
**GOLDFLESS, STEPHEN JACOB y**  
**TIMBERLAKE, SONIA**

74 Agente/Representante:

**INGENIAS CREACIONES, SIGNOS E**  
**INVENCIONES, SLP**

**ES 2 727 656 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

Secuenciación de alto rendimiento de banco de nucleótidos

5 **Antecedentes**

Las tecnologías actuales de presentación de anticuerpos (fagos, levaduras, ribosomas, mamíferos, etc.) están limitadas, porque la calidad de los candidatos a anticuerpo seleccionados está limitada por el banco de partida a partir de la que se generan. Los enfoques, tales como los enfoques de diseño de anticuerpos combinatorios e "inteligentes" y los enfoques de descubrimiento de hibridomas, suelen producir anticuerpos sintéticos que presentan complicaciones cadena abajo, incluyendo dificultades de expresión a gran escala, alto riesgo de inmunogenicidad en pacientes y falta de función inmunitaria suficiente, aparte de las afinidades de unión altas. Pocos anticuerpos derivados de tecnologías de visualización han superado con éxito los ensayos clínicos en la última década, incluso cuando demuestran características preclínicas positivas. Actualmente, la capacidad de predecir o comprender el mecanismo mediante el que una determinada secuencia de anticuerpos reconoce y activa la respuesta inmunitaria contra una diana foránea ha sido difícil de alcanzar. Por lo tanto, existe la necesidad en la técnica de procedimientos para descubrir y generar anticuerpos que tengan afinidades de unión elevadas, que puedan generarse a gran escala y que tengan suficiente función inmunitaria. Los procedimientos descritos en el presente documento pretenden utilizar los millones de años de evolución del repertorio inmunitario para satisfacer estas necesidades, y para comprender mejor estos conceptos y cómo se relacionan con la generación de anticuerpos. Los procedimientos descritos en el presente documento pueden usarse para producir un banco de secuencias de anticuerpos y/o anticuerpos para la selección de candidatos a anticuerpos de alta calidad.

El repertorio de anticuerpos humanos es casi ilimitado en su complejidad y tamaño. Como resultado de ello, se ha demostrado estadísticamente que los bancos combinatorios raramente proporcionan un emparejamiento correcto de las cadenas pesadas ( $V_H$ ) o ligeras ( $V_L$ ). Otros se han centrado en barajar la única de las familias marco expresadas más frecuentemente de regiones determinantes de la complementariedad (CDR) (tales como V3-23, V1-69, o frecuencias  $V_H$  y  $V_L$  coincidentes), y por lo tanto, limitan la diversidad del repertorio a un tamaño manejable. Se esperaba que la familia expresada con mayor frecuencia se seleccionara y evolucionara con mayor frecuencia durante una respuesta inmunitaria. Sorprendentemente, a través del uso de la secuenciación inmunitaria de los repertorios de anticuerpos humanos, se ha descubierto que no existe una relación entre las frecuencias de expresión marco del anticuerpo y el potencial de activación de un anticuerpo en respuesta a un desafío inmunitario. Los procedimientos descritos en el presente documento se pueden usar para diseñar y/o generar un banco de anticuerpos no limitante para superar estos desafíos para el descubrimiento y la selección de anticuerpos. Se pueden generar bancos de donantes autoinmunes, con cáncer, infecciosos y normales/sanos para la medicina personalizada a fin de abordar las necesidades biológicas fundamentales no cubiertas.

El documento WO 2012/048340 desvela procedimientos y composiciones para determinar y/o controlar el estado inmunitario de un individuo. El documento WO 2012/048341 desvela procedimientos y composiciones para análisis de células individuales de alto rendimiento. El documento WO2014/071361 desvela procedimientos para codificar con códigos de barras ácidos nucleicos, tales como el ADN genómico. El documento WO2005/084134 desvela la secuenciación de genes expresados que pertenecen a la superfamilia de genes de inmunoglobulina, en particular, inmunoglobulinas y receptores de linfocitos T. Dekosky *et al* (2013) *Nature* 31, 166-9 desvelan una secuenciación de alto rendimiento del repertorio de cadenas pesadas y ligeras de inmunoglobulina humana emparejadas. Alon *et al* (2011) "Genome Research", Cold Spring Harbor Laboratory Press 21, 1506-11, desvelan el sesgo de la codificación de barras en la secuenciación multiplex de alto rendimiento de microARN.

**Sumario**

50 La presente divulgación proporciona un procedimiento que comprende:

(a) formar una pluralidad de vasos comprendiendo cada uno

- 55 (i) una célula individual de una muestra que comprende una pluralidad de células,  
 (ii) una pluralidad de polinucleótidos con código de barras molecular y  
 (iii) un polinucleótido con código de barras del vaso;

(b) producir:

- 60 (i) un primer polinucleótido complementario que es complementario a un primer polinucleótido celular de la célula individual y  
 (ii) un segundo polinucleótido complementario que es complementario a un segundo polinucleótido celular de la célula individual;

65 (c) unir:

- (i) un primer polinucleótido con código de barras molecular de la pluralidad de polinucleótidos con código de barras molecular con el primer polinucleótido complementario y
- (ii) un segundo polinucleótido con código de barras molecular de la pluralidad de polinucleótidos con código de barras molecular con el segundo polinucleótido complementario,

5 formando así un primer polinucleótido con un solo código de barras de una sola célula y un segundo polinucleótido con un solo código de barras de una sola célula; y

(d) unir el polinucleótido con código de barras del vaso, o un producto amplificado del mismo, a

- 10 (i) el primer polinucleótido con un solo código de barras de una sola célula o un producto amplificado del mismo e
- (ii) el segundo polinucleótido con un solo código de barras de una sola célula o un producto amplificado del mismo, formando así una primera secuencia de código de barras doble de una sola célula y una segunda secuencia de código de barras doble de una sola célula.

15 **Descripción detallada**

La presente divulgación proporciona un procedimiento que comprende: formar una pluralidad de vasos, comprendiendo cada uno una sola célula de una muestra que comprende una pluralidad de células, una pluralidad de polinucleótidos con código de barras molecular y un polinucleótido con código de barras del vaso; producir: un primer polinucleótido complementario que es complementario a un primer polinucleótido celular de la célula individual, y un segundo polinucleótido complementario que es complementario a un segundo polinucleótido celular de la célula individual; unir: un primer polinucleótido con código de barras molecular de la pluralidad con el primer polinucleótido complementario, y un segundo polinucleótido con código de barras molecular con el segundo polinucleótido complementario, formando así un primer y un segundo polinucleótido con un solo código de barras de una sola célula; y unir el polinucleótido con código de barras del vaso, o un producto amplificado del mismo, con el primer polinucleótido con un solo código de barras de una sola célula y el segundo polinucleótido con un solo código de barras de una sola célula, formando así una primera y una segunda secuencia de código de barras doble de una sola célula.

30 La presente divulgación proporciona una composición que comprende: una pluralidad de vasos, comprendiendo cada uno una sola célula de una muestra que comprende una pluralidad de células, una pluralidad de polinucleótidos con código de barras molecular, un polinucleótido con código de barras del vaso; un primer polinucleótido complementario que es complementario a un primer polinucleótido celular de la célula individual, y un segundo polinucleótido complementario que es complementario a un segundo polinucleótido celular de la célula individual; en la que el primer polinucleótido complementario comprende un primer código de barras molecular de la pluralidad de polinucleótidos con código de barras molecular y el código de barras del vaso del polinucleótido con código de barras del vaso o un producto amplificado del polinucleótido con código de barras del vaso, y en la que el segundo polinucleótido complementario comprende un segundo código de barras molecular de la pluralidad de polinucleótidos con código de barras molecular y el código de barras del vaso del polinucleótido con código de barras del vaso o un producto amplificado del polinucleótido con código de barras del vaso.

La presente divulgación proporciona un procedimiento que comprende: (a) formar una pluralidad de vasos, comprendiendo cada uno una sola célula de una muestra que comprende una pluralidad de células, una pluralidad de polinucleótidos con código de barras molecular y un polinucleótido con código de barras del vaso; (b) producir: un primer polinucleótido complementario que es complementario a un primer polinucleótido celular de la célula individual, y un segundo polinucleótido complementario que es complementario a un segundo polinucleótido celular de la célula individual; (c) unir: un primer polinucleótido con código de barras molecular de la pluralidad con el primer polinucleótido complementario, y un segundo polinucleótido con código de barras molecular con el segundo polinucleótido complementario, formando así un primer y un segundo polinucleótido con un solo código de barras de una sola célula; y (d) unir el polinucleótido con código de barras del vaso, o un producto amplificado del mismo, con el primer polinucleótido con un solo código de barras de una sola célula, o un producto amplificado del mismo, formando así una primera y una segunda secuencia de código de barras doble de una sola célula.

55 La presente divulgación proporciona un procedimiento que comprende: (a) producir un primer polinucleótido complementario a partir de un polinucleótido de inmunoglobulina (IgH) de cadena pesada, y un segundo polinucleótido complementario a partir de un polinucleótido de inmunoglobulina (IgL) de cadena ligera de una pluralidad de células inmunitarias de una muestra con: un primer cebador diana que comprende una región complementaria a una misma región de los polinucleótidos de IgH de la pluralidad de células inmunitarias; un segundo cebador diana que comprende una región complementaria a una misma región de los polinucleótidos de IgL de la pluralidad de células inmunitarias; una transcriptasa inversa que comprende una actividad de transferasa terminal no molde, en el que se añaden 3 o más nucleótidos idénticos que no son de molde al extremo 3' del primer y segundo polinucleótido complementario; una pluralidad de polinucleótidos con código de barras molecular, comprendiendo cada uno: un código de barras molecular, una región del extremo 5' complementaria a una región de un polinucleótido con código de barras del vaso, y una región del extremo 3' complementaria a los 3 o más nucleótidos que no son de molde; y un polinucleótido con código de barras del vaso, formando así un primer y un segundo polinucleótido con un solo código

de barras de una sola célula; (b) amplificar el polinucleótido con código de barras del vaso, formando así un primer y un segundo polinucleótido con código de barras doble de una sola célula; (c) amplificar el primer y el segundo polinucleótido con código de barras doble de una sola célula, formando así un banco de secuencias que comprenden una región variable de los polinucleótidos de IgH o IgL, o una combinación de los mismos; y (d) secuenciar una o más de las secuencias del banco, en el que (a) se realiza en un vaso de una pluralidad de vasos, en el que el vaso comprende una sola célula inmunitaria de la pluralidad de células inmunitarias.

La presente divulgación proporciona un procedimiento que comprende: (a) producir un primer polinucleótido complementario a partir de un polinucleótido receptor de linfocitos T alfa ( $TCR\alpha$ ), y un segundo polinucleótido complementario a partir de un polinucleótido receptor de linfocitos T beta ( $TCR\beta$ ) de una pluralidad de células inmunitarias de una muestra con: un primer cebador diana que comprende una región complementaria a una misma región de los polinucleótidos  $TCR\alpha$  de la pluralidad de células inmunitarias; un segundo cebador diana que comprende una región complementaria a una misma región de los polinucleótidos  $TCR\beta$  de la pluralidad de células inmunitarias; una transcriptasa inversa que comprende una actividad de transferasa terminal no molde, en el que se añaden 3 o más nucleótidos idénticos no molde al extremo 3' del primer y segundo polinucleótido complementario; una pluralidad de polinucleótidos con código de barras molecular, comprendiendo cada uno: un código de barras molecular, una región del extremo 5' complementaria a una región de un polinucleótido con código de barras del vaso, y una región del extremo 3' complementaria a los 3 o más nucleótidos que no son de molde; y un polinucleótido con código de barras del vaso, formando así un primer y un segundo polinucleótido con un solo código de barras de una sola célula; (b) amplificar el polinucleótido con código de barras del vaso, formando así un primer y un segundo polinucleótido con código de barras doble de una sola célula; (c) amplificar el primer y el segundo polinucleótido con código de barras doble de una sola célula, formando así un banco de secuencias que comprenden una región variable de los polinucleótidos de  $TCR\alpha$  o  $TCR\beta$ , o una combinación de los mismos; y (d) secuenciar una o más de las secuencias del banco, en el que (a) se realiza en un vaso de una pluralidad de vasos, en el que el vaso comprende una sola célula inmunitaria de la pluralidad de células inmunitarias.

La presente divulgación proporciona un procedimiento que comprende: (a) producir un primer polinucleótido complementario a partir de un polinucleótido receptor de linfocitos T gamma ( $TCR\gamma$ ), y un segundo polinucleótido complementario a partir de un polinucleótido receptor de linfocitos T delta ( $TCR\delta$ ) de una pluralidad de células inmunitarias de una muestra con: un primer cebador diana que comprende una región complementaria a una misma región de los polinucleótidos  $TCR\gamma$  de la pluralidad de células inmunitarias; un segundo cebador diana que comprende una región complementaria a una misma región de los polinucleótidos  $TCR\delta$  de la pluralidad de células inmunitarias; una transcriptasa inversa que comprende una actividad de transferasa terminal no molde, en el que se añaden 3 o más nucleótidos idénticos no molde al extremo 3' del primer y segundo polinucleótido complementario; una pluralidad de polinucleótidos con código de barras molecular, comprendiendo cada uno: un código de barras molecular, una región del extremo 5' complementaria a una región de un polinucleótido con código de barras del vaso, y una región del extremo 3' complementaria a los 3 o más nucleótidos que no son de molde; y un polinucleótido con código de barras del vaso, formando así un primer y un segundo polinucleótido con un solo código de barras de una sola célula; (b) amplificar el polinucleótido con código de barras del vaso, formando así un primer y un segundo polinucleótido con código de barras doble de una sola célula; (c) amplificar el primer y el segundo polinucleótido con código de barras doble de una sola célula, formando así un banco de secuencias que comprenden una región variable de los polinucleótidos de  $TCR\gamma$  o  $TCR\delta$ , o una combinación de los mismos; y (d) secuenciar una o más de las secuencias del banco, en el que (a) se realiza en un vaso de una pluralidad de vasos, en el que el vaso comprende una sola célula inmunitaria de la pluralidad de células inmunitarias.

En algunas realizaciones, el banco representa un estado inmunitario de la muestra. En algunas realizaciones, la primera y la segunda secuencia de código de barras doble de una sola célula son un banco de primera y segunda secuencia de código de barras doble de una sola célula. En algunas realizaciones, los códigos de barras moleculares del primer y segundo polinucleótido con código de barras molecular son diferentes. En algunas realizaciones, el primer y el segundo polinucleótido con un solo código de barras de una sola célula comprenden un código de barras molecular diferente. En algunas realizaciones, la primera y la segunda secuencia con código de barras doble de una sola célula comprenden un código de barras molecular diferente. En algunas realizaciones, la primera y la segunda secuencia de código de barras doble de una sola célula comprenden el mismo código de barras del vaso. En algunas realizaciones, la pluralidad de polinucleótidos con código de barras molecular no son productos amplificados. En algunas realizaciones, el código de barras molecular de un polinucleótido con código de barras molecular de un primer vaso es diferente del código de barras molecular de un polinucleótido con código de barras molecular de un segundo vaso. En algunas realizaciones, el código de barras molecular de cada polinucleótido con código de barras molecular de un primer vaso de la pluralidad de vasos es único. En algunas realizaciones, el código de barras molecular de cada polinucleótido con código de barras molecular de un segundo vaso de la pluralidad de vasos es único. En algunas realizaciones, el código de barras molecular de cada polinucleótido con código de barras molecular de un primer vaso y un segundo vaso son únicos. En algunas realizaciones, el código de barras molecular de cada polinucleótido con código de barras molecular de un tercer vaso de la pluralidad de vasos es único. En algunas realizaciones, el código de barras molecular de cada polinucleótido con código de barras molecular del primer vaso, segundo vaso y tercer vaso son únicos. En algunas realizaciones, el código de barras molecular de cada polinucleótido con código de barras molecular de cualquier vaso individual de la pluralidad de vasos es único. En algunas realizaciones, el código de barras molecular de cada polinucleótido con código de barras molecular en cualquier vaso de la pluralidad de vasos es diferente al código de barras molecular de



pluralidad de vasos. En algunas realizaciones, una segunda secuencia molecular común del primer polinucleótido con código de barras molecular es la misma que una segunda secuencia molecular común del segundo polinucleótido con código de barras molecular. En algunas realizaciones, cada polinucleótido con código de barras molecular de un vaso cualquiera de la pluralidad de vasos comprende una primera secuencia molecular común que comprende la misma secuencia que una primera secuencia molecular común de un polinucleótido con código de barras molecular de cualquier otro vaso de la pluralidad de vasos. En algunas realizaciones, cada polinucleótido con código de barras molecular de un vaso cualquiera de la pluralidad de vasos comprende una segunda secuencia molecular común que comprende la misma secuencia que una segunda secuencia molecular común de un polinucleótido con código de barras molecular de cualquier otro vaso de la pluralidad de vasos. En algunas realizaciones, la primera secuencia de vaso común comprende una secuencia que comprende la misma secuencia que la primera secuencia molecular común. En algunas realizaciones, la primera secuencia de vaso común comprende una secuencia complementaria a la primera secuencia molecular común o un complemento de la misma. En algunas realizaciones, la segunda secuencia molecular común comprende una región complementaria a tres o más nucleótidos no molde añadidos al extremo 3' del primer polinucleótido complementario. En algunas realizaciones, la región complementaria a tres o más nucleótidos no molde añadido al extremo 3' del primer polinucleótido complementario es una región terminal.

En algunas realizaciones, un primer y un segundo polinucleótido con código de barras molecular no se fusionan entre sí. En algunas realizaciones, el primer y segundo polinucleótido con un solo código de barras de una sola célula no se fusionan entre sí. En algunas realizaciones, la primera y la segunda secuencia de código de barras doble de una sola célula no se fusionan entre sí.

En algunas realizaciones, el primer polinucleótido celular es ADN. En algunas realizaciones, el segundo polinucleótido celular es ADN. En algunas realizaciones, el primer polinucleótido celular es ARN. En algunas realizaciones, el segundo polinucleótido celular es ARN. En algunas realizaciones, el ARN es ARNm. En algunas realizaciones, el primer polinucleótido complementario de (b) es ADNc. En algunas realizaciones, el segundo polinucleótido complementario de (b) es ADNc.

En algunas realizaciones, (b) comprende extender un primer cebador diana hibridado con el primer polinucleótido celular, y extender un segundo cebador diana hibridado con el segundo polinucleótido celular. En algunas realizaciones, la extensión comprende la transcripción inversa del primer polinucleótido celular con un primer cebador diana, y la transcripción inversa del segundo polinucleótido celular con un segundo cebador diana. En algunas realizaciones, el primer cebador diana comprende una secuencia complementaria a una secuencia diana del primer polinucleótido celular. En algunas realizaciones, el segundo cebador diana comprende una secuencia complementaria a una secuencia diana del segundo polinucleótido celular. En algunas realizaciones, el primer cebador diana comprende una secuencia poli (T). En algunas realizaciones, el segundo cebador diana comprende una secuencia poli (T). En algunas realizaciones, la secuencia diana del primer polinucleótido celular es una secuencia de inmunoglobulina de cadena pesada (IgH), una secuencia de TCR $\alpha$ , una secuencia de TCR $\gamma$  o una combinación de las mismas. En algunas realizaciones, la secuencia diana del primer polinucleótido celular es una secuencia de región constante de cadena pesada (C<sub>H</sub>), una secuencia de región constante de TCR $\alpha$  (C $\alpha$ ), una secuencia de región constante de TCR $\gamma$  (C $\gamma$ ) o una combinación de las mismas. En algunas realizaciones, la secuencia diana del segundo polinucleótido celular es una secuencia de inmunoglobulina de cadena ligera (IgL), una secuencia de TCR $\beta$ , una secuencia de TCR $\delta$  o una combinación de las mismas. En algunas realizaciones, la secuencia diana del segundo polinucleótido celular es una secuencia de región constante de cadena ligera (C<sub>L</sub>), una secuencia de región constante de TCR $\beta$  (C $\beta$ ), una secuencia de región constante de TCR $\delta$  (C $\delta$ ) o una combinación de las mismas. En algunas realizaciones, el primer cebador diana comprende una pluralidad de primeros cebadores diana. En algunas realizaciones, el segundo cebador diana comprende una pluralidad de segundos cebadores diana. En algunas realizaciones, la pluralidad de primeros cebadores diana comprende una pluralidad de secuencias complementarias a una pluralidad de secuencias de inmunoglobulina de cadena pesada (IgH), secuencias de TCR $\alpha$ , secuencias de TCR $\gamma$  o una combinación de las mismas. En algunas realizaciones, la pluralidad de secuencias de inmunoglobulina de cadena pesada (IgH), secuencias de TCR $\alpha$  o secuencias de TCR $\gamma$  comprende una pluralidad de secuencias de región constante de cadena pesada (C<sub>H</sub>), secuencias de región constante de TCR $\alpha$  (C $\alpha$ ), secuencias de región constante de TCR $\gamma$  (C $\gamma$ ) o una combinación de las mismas. En algunas realizaciones, la pluralidad de secuencias de región constante de cadena pesada (C<sub>H</sub>) comprende dos o más secuencias seleccionadas del grupo que consiste en secuencias de región constante de cadena pesada (C<sub>H</sub>) de IgM, IgD, IgA, IgE, IgG y combinaciones de las mismas. En algunas realizaciones, la pluralidad de segundos cebadores diana comprende una pluralidad de secuencias complementarias a una pluralidad de secuencias de inmunoglobulina de cadena ligera (IgL), secuencias de TCR $\beta$ , secuencias de TCR $\delta$  o una combinación de las mismas. En algunas realizaciones, la pluralidad de secuencias de inmunoglobulina de cadena ligera (IgL), secuencias de TCR $\beta$  o secuencias de TCR $\delta$  comprende una pluralidad de secuencias de región constante de cadena ligera (C<sub>L</sub>), secuencias de región constante de TCR $\beta$  (C $\beta$ ), secuencias de región constante de TCR $\delta$  (C $\delta$ ) o una combinación de las mismas. En algunas realizaciones, la pluralidad de secuencias de región constante de cadena ligera (C<sub>L</sub>) comprende dos o más secuencias seleccionadas del grupo que consiste en secuencias de región constante de cadena ligera (C<sub>L</sub>) de Ig $\kappa$ , Ig $\lambda$  y combinaciones de las mismas. En algunas realizaciones, en (b), la extensión comprende el uso de una transferasa terminal no molde, en la que se añaden tres o más nucleótidos no molde al extremo 3' del primer polinucleótido complementario. En algunas realizaciones, la transferasa terminal no molde es una transcriptasa inversa o una polimerasa. En algunas realizaciones, la transferasa terminal no molde es una transcriptasa inversa, y en las que la transcriptasa inversa se

selecciona del grupo que consiste en la transcriptasa inversa Superscript II, la transcriptasa inversa máxima, la transcriptasa inversa Protoscript II, la transcriptasa inversa del virus de la leucemia murina de Moloney (MMLV-RT), la transcriptasa inversa HighScriber, la transcriptasa inversa del virus de la mieloblastosis aviar (AMV), cualquier transcriptasa inversa que comprenda la actividad de la desoxinucleotidil transferasa terminal y sus combinaciones. En algunas realizaciones, se añaden tres o más nucleótidos no molde al extremo 3' del segundo polinucleótido complementario.

En algunas realizaciones, en (c), la unión comprende hibridar una región de un primer polinucleótido con código de barras molecular con los tres o más nucleótidos no molde añadidos al extremo 3' del primer polinucleótido complementario. En algunas realizaciones, en (c), la unión comprende hibridar una región de un segundo polinucleótido con código de barras molecular con los tres o más nucleótidos no molde añadidos al extremo 3' del segundo polinucleótido complementario. En algunas realizaciones, en (c), un primer polinucleótido con código de barras molecular unido con el primer polinucleótido complementario comprende una región complementaria a los tres o más nucleótidos no molde en el extremo 3' del primer polinucleótido complementario. En algunas realizaciones, en (c), un segundo polinucleótido con código de barras molecular unido con el segundo polinucleótido complementario comprende una región complementaria a tres o más nucleótidos no molde en el extremo 3' del segundo polinucleótido complementario. En algunas realizaciones, los tres o más nucleótidos no molde son idénticos. En algunas realizaciones, al menos uno de los tres o más nucleótidos no molde no es idéntico a otro nucleótido de los tres o más nucleótidos no molde. En algunas realizaciones, al menos un nucleótido de la región hibridada del primer polinucleótido con código de barras molecular no es idéntico a otro ácido nucleico de la región hibridada del primer polinucleótido con código de barras molecular. En algunas realizaciones, al menos un nucleótido de la región hibridada del segundo polinucleótido con código de barras molecular no es idéntico a otro ácido nucleico de la región hibridada del segundo polinucleótido con código de barras molecular. En algunas realizaciones, el al menos un nucleótido no idéntico es un desoxirribonucleótido o un análogo del mismo. En algunas realizaciones, el al menos un nucleótido no idéntico no es un ribonucleótido o un análogo del mismo. En algunas realizaciones, el al menos un nucleótido no idéntico es una desoxirriboguanosina. En algunas realizaciones, el al menos un nucleótido no idéntico es un análogo de desoxirriboguanosina. En algunas realizaciones, el al menos un nucleótido no idéntico es un nucleótido terminal del primer o segundo polinucleótido con código de barras molecular. En algunas realizaciones, el al menos un nucleótido no idéntico es un ribonucleótido o un análogo del mismo. En algunas realizaciones, un nucleótido terminal de la región hibridada del primer o segundo polinucleótido con código de barras molecular es un desoxirribonucleótido o un análogo del mismo. En algunas realizaciones, un nucleótido terminal de la región hibridada del primer o segundo polinucleótido con código de barras molecular no es un ribonucleótido o análogo del mismo. En algunas realizaciones, un nucleótido terminal de la región hibridada del primer o segundo polinucleótido con código de barras molecular es una desoxirriboguanosina. En algunas realizaciones, un nucleótido terminal de la región hibridada del primer o segundo polinucleótido con código de barras molecular es un análogo de desoxirriboguanosina. En algunas realizaciones, un nucleótido terminal de la región hibridada del primer o segundo polinucleótido con código de barras molecular es un ribonucleótido o análogo del mismo. En algunas realizaciones, al menos dos nucleótidos no terminales de la región hibridada del primer o segundo polinucleótido con código de barras molecular son ribonucleótidos o análogos de los mismos. En algunas realizaciones, al menos dos nucleótidos no terminales de la región hibridada del primer o segundo polinucleótido con código de barras molecular no son desoxirribonucleótidos o análogos de los mismos. En algunas realizaciones, al menos dos nucleótidos no terminales de la región hibridada del primer o segundo polinucleótido con código de barras molecular son desoxirribonucleótidos o análogos de los mismos. En algunas realizaciones, (c) comprende además extender el primer polinucleótido complementario y el segundo polinucleótido complementario después de la unión. En algunas realizaciones, el primer polinucleótido complementario comprende una región complementaria a un primer polinucleótido con código de barras molecular. En algunas realizaciones, el segundo polinucleótido complementario comprende una región complementaria a un segundo polinucleótido con código de barras molecular. En algunas realizaciones, el primer polinucleótido complementario comprende una región complementaria a un segundo polinucleótido con código de barras molecular. En algunas realizaciones, la región del primer polinucleótido complementario que es complementario al primer o segundo polinucleótido con código de barras molecular no es complementaria a una secuencia de código de barras molecular. En algunas realizaciones, la región del primer polinucleótido complementario que es complementario al primer o segundo polinucleótido con código de barras molecular no es complementaria a una región del polinucleótido con código de barras del vaso o un producto amplificado del mismo. En algunas realizaciones, la región del primer polinucleótido complementario al primer o segundo polinucleótido con código de barras molecular comprende tres o más nucleótidos no molde añadidos al extremo 3' del primer polinucleótido complementario. En algunas realizaciones, la región del segundo polinucleótido complementario que es complementario al segundo polinucleótido con código de barras molecular comprende tres o más nucleótidos no molde añadidos al extremo 3' del segundo polinucleótido complementario. En algunas realizaciones, el primer polinucleótido complementario no es complementario al polinucleótido con código de barras del vaso. En algunas realizaciones, el segundo polinucleótido complementario no es complementario al polinucleótido con código de barras del vaso. En algunas realizaciones, una región de un complemento de un primer polinucleótido con código de barras molecular es complementaria a una región del polinucleótido con código de barras del vaso. En algunas realizaciones, una región de un complemento de un segundo polinucleótido con código de barras molecular es complementaria a una región del polinucleótido con código de barras del vaso. En algunas realizaciones, una región del primer polinucleótido con código de barras de una sola célula es complementaria a una región del polinucleótido con código de barras del vaso. En algunas realizaciones, una región del segundo polinucleótido con código de barras de una sola célula es complementaria a una región del polinucleótido con código de barras del vaso. En algunas

realizaciones, una región del primer polinucleótido de un solo código de barras de una sola célula es complementaria a la región del polinucleótido con código de barras del vaso al que es complementario el segundo polinucleótido de un solo código de barras de una sola célula. En algunas realizaciones, el procedimiento comprende además amplificar el polinucleótido con código de barras del vaso con un primer conjunto de cebadores, en el que la amplificación se realiza antes de unir el polinucleótido con código de barras del vaso o simultáneamente con la unión del polinucleótido con código de barras del vaso. En algunas realizaciones, el polinucleótido con código de barras del vaso comprende un primer y un segundo polinucleótido con código de barras del vaso seleccionado del grupo que consiste en el polinucleótido con código de barras del vaso, un complemento del polinucleótido con código de barras del vaso, un producto amplificado del polinucleótido con código de barras del vaso y cualquier combinación de los mismos. En algunas realizaciones, la unión del polinucleótido con código de barras del vaso comprende: hibridar una región del polinucleótido con código de barras del vaso o un producto amplificado del mismo a una región del primer polinucleótido con un solo código de barras de una sola célula, e hibridar una región del polinucleótido con código de barras del vaso o producto amplificado del mismo a una región del segundo polinucleótido con un solo código de barras de una sola célula.

En algunas realizaciones, el procedimiento comprende además extender el primer polinucleótido de secuencia con un solo código de barras de una sola célula y el segundo polinucleótido de secuencia de un solo código de barras de una sola célula después de unir el polinucleótido de código de barras del vaso, formando así la primera y segunda secuencia de código de barras doble de una sola célula. En algunas realizaciones, la primera secuencia con código de barras doble de una sola célula comprende una región complementaria al polinucleótido con código de barras del vaso. En algunas realizaciones, la segunda secuencia de código de barras doble de una sola célula comprende una región complementaria al polinucleótido con código de barras del vaso. En algunas realizaciones, las regiones de la primera y la segunda secuencia con código de barras doble de una sola célula que son complementarias al polinucleótido con código de barras del vaso son la misma secuencia. En algunas realizaciones, la región del primer polinucleótido con un solo código de barras de una sola célula que es complementaria al primer o segundo polinucleótido con código de barras molecular no es complementaria a una región del polinucleótido con código de barras del vaso o un producto amplificado del mismo. En algunas realizaciones, un primer cebador del primer conjunto de cebadores es complementario a una región de un primer polinucleótido con código de barras molecular, un complemento del primer polinucleótido con un solo código de barras de una sola célula, un complemento de la primera secuencia con código de barras doble de una sola célula o cualquier combinación de los mismos. En algunas realizaciones, el primer cebador del primer conjunto de cebadores es complementario a una región de un segundo polinucleótido con código de barras molecular, un complemento del segundo polinucleótido con un solo código de barras de una sola célula, un complemento de la segunda secuencia con código de barras doble de una sola célula o cualquier combinación de los mismos. En algunas realizaciones, un primer cebador del primer conjunto de cebadores no es complementario al primer polinucleótido celular o un complemento del mismo. En algunas realizaciones, el primer cebador del primer conjunto de cebadores no es complementario al segundo polinucleótido celular o un complemento del mismo. En algunas realizaciones, un primer cebador del primer conjunto de cebadores es complementario a una región de un complemento de la primera secuencia con un solo código de barras de una sola célula que está cadena abajo del código de barras molecular. En algunas realizaciones, el primero de los mismos del primer conjunto de cebadores es complementario a una región de un complemento del segundo polinucleótido con un solo código de barras de una sola célula que está cadena abajo del código de barras molecular. En algunas realizaciones, un primer cebador del primer conjunto de cebadores es complementario a una región de un complemento de la primera secuencia con código de barras doble de una sola célula que está cadena arriba del código de barras del vaso. En algunas realizaciones, el primer cebador del primer conjunto de cebadores es complementario a una región de un complemento del segundo polinucleótido con código de barras doble de una sola célula que está cadena arriba del código de barras del vaso. En algunas realizaciones, un segundo cebador del primer conjunto de cebadores no es complementario a una región del primer polinucleótido celular o un complemento del mismo, el primer polinucleótido complementario o un complemento del mismo, un primer polinucleótido con código de barras molecular o complemento del mismo, el primer polinucleótido con un solo código de barras de una sola célula o un complemento del mismo, o cualquier combinación de los mismos. En algunas realizaciones, el segundo cebador del primer conjunto de cebadores no es complementario a una región del segundo polinucleótido celular o un complemento del mismo, el segundo polinucleótido complementario o un complemento del mismo, un segundo polinucleótido con código de barras molecular o complemento del mismo, el segundo polinucleótido con un solo código de barras de una sola célula o un complemento del mismo, o cualquier combinación de los mismos. En algunas realizaciones, un segundo cebador del primer conjunto de cebadores es complementario a una región de la primera secuencia de código de barras doble de una sola célula. En algunas realizaciones, un segundo cebador del primer conjunto de cebadores es complementario a una región de la segunda secuencia con código de barras doble de una sola célula. En algunas realizaciones, un segundo cebador del primer conjunto de cebadores es complementario a una región de la primera secuencia con código de barras doble de una sola célula que está cadena arriba del código de barras molecular. En algunas realizaciones, el segundo cebador del primer conjunto de cebadores es complementario a una región del segundo polinucleótido con código de barras molecular que está cadena arriba del código de barras molecular. En algunas realizaciones, un segundo del mismo del primer conjunto de cebadores es complementario a una región de la primera secuencia de código de barras doble de una sola célula que está cadena arriba del código de barras del vaso. En algunas realizaciones, el segundo del mismo del primer conjunto de cebadores es complementario a una región del segundo polinucleótido con código de barras molecular que está cadena arriba del código de barras del vaso.

En algunas realizaciones, el procedimiento comprende además romper dos o más vasos de la pluralidad de vasos. En algunas realizaciones, el procedimiento comprende además agrupar la primera y la segunda secuencia con código de barras doble de una sola célula de los dos o más vasos rotos.

5 En algunas realizaciones, el procedimiento comprende además amplificar la primera y la segunda secuencia con código de barras doble de una sola célula. En algunas realizaciones, la amplificación de la primera y la segunda secuencia con código de barras doble de una sola célula se realiza fuera de un vaso de la pluralidad de vasos. En algunas realizaciones, el procedimiento comprende además amplificar la primera y la segunda secuencia con código de barras doble de una sola célula con un segundo conjunto de cebadores. En algunas realizaciones, un primer cebador del segundo conjunto de cebadores no es complementario a una región del primer polinucleótido celular o un complemento del mismo, el primer polinucleótido complementario o un complemento del mismo, un primer polinucleótido con código de barras molecular o complemento del mismo, el primer polinucleótido con un solo código de barras de una sola célula o un complemento del mismo, o cualquier combinación de los mismos. En algunas realizaciones, el primer cebador del segundo conjunto de cebadores no es complementario a una región del segundo polinucleótido celular o un complemento del mismo, el segundo polinucleótido complementario o un complemento del mismo, un segundo polinucleótido con código de barras molecular o complemento del mismo, el segundo polinucleótido con un solo código de barras de una sola célula o un complemento del mismo, o cualquier combinación de los mismos. En algunas realizaciones, un primer cebador del segundo conjunto de cebadores es complementario a una región de la primera secuencia de código de barras doble de una sola célula. En algunas realizaciones, el primer cebador del segundo conjunto de cebadores es complementario a una región de la segunda secuencia con código de barras doble de una sola célula. En algunas realizaciones, un primer cebador del segundo conjunto de cebadores es complementario a una región de la primera secuencia con código de barras doble de una sola célula que está cadena arriba del código de barras molecular. En algunas realizaciones, el primer cebador del segundo conjunto de cebadores es complementario a una región de la segunda secuencia con código de barras doble de una sola célula que está cadena arriba del código de barras molecular. En algunas realizaciones, un primer cebador del segundo conjunto de cebadores es complementario a una región de la primera secuencia de código de barras doble de una sola célula que está cadena arriba del código de barras del vaso. En algunas realizaciones, el primer cebador del segundo conjunto de cebadores es complementario a una región de la segunda secuencia con código de barras doble de una sola célula que está cadena arriba del código de barras del vaso. En algunas realizaciones, el segundo cebador del primer conjunto de cebadores es el primer cebador del segundo conjunto de cebadores. En algunas realizaciones, un segundo cebador del segundo conjunto de cebadores es complementario a una región del primer y segundo polinucleótido celular, un complemento del primer y segundo polinucleótido complementario, un complemento del primer y segundo polinucleótido con un solo código de barras de una sola célula, un complemento de la primera y segunda secuencia con código de barras doble de una sola célula o cualquier combinación de los mismos. En algunas realizaciones, el segundo cebador del segundo conjunto de cebadores comprende una secuencia poli (T). En algunas realizaciones, un segundo cebador del segundo conjunto de cebadores es complementario a una región del primer o segundo polinucleótido celular, un complemento del primer o segundo polinucleótido complementario, un complemento del primer o segundo polinucleótido con un solo código de barras de una sola célula, un complemento de la primera o segunda secuencia con código de barras doble de una sola célula o cualquier combinación de los mismos. En algunas realizaciones, el segundo cebador del segundo conjunto de cebadores no es complementario a un primer o segundo polinucleótido con código de barras molecular o complemento del mismo, el polinucleótido con código de barras del vaso o complemento del mismo, o cualquier combinación de los mismos. En algunas realizaciones, un tercer cebador del segundo conjunto de cebadores es complementario a una región del segundo polinucleótido celular, un complemento del segundo polinucleótido complementario, un complemento del segundo polinucleótido con un solo código de barras de una sola célula, un complemento de la segunda secuencia con código de barras doble de una sola célula o cualquier combinación de los mismos. En algunas realizaciones, el segundo cebador del segundo conjunto de cebadores es complementario a una región del primer polinucleótido celular, un complemento del primer polinucleótido complementario, un complemento del primer polinucleótido con un solo código de barras de una sola célula, un complemento de la primera secuencia con código de barras doble de una sola célula o cualquier combinación de los mismos. En algunas realizaciones, el tercer cebador del segundo conjunto de cebadores es complementario a una región del primer polinucleótido celular, un complemento del primer polinucleótido complementario, un complemento del primer polinucleótido con un solo código de barras de una sola célula, un complemento de la primera secuencia con código de barras doble de una sola célula o cualquier combinación de los mismos. En algunas realizaciones, el tercer cebador del segundo conjunto de cebadores no es complementario a un primer o segundo polinucleótido con código de barras molecular o complemento del mismo, el polinucleótido con código de barras del vaso o complemento del mismo, o cualquier combinación de los mismos. En algunas realizaciones, el segundo cebador del segundo conjunto de cebadores comprende una secuencia específica de la diana. En algunas realizaciones, el tercer cebador del segundo conjunto de cebadores comprende una secuencia específica de la diana. En algunas realizaciones, la secuencia específica de la diana del segundo cebador del segundo conjunto de cebadores se dirige a una secuencia de inmunoglobulina de cadena pesada (IgH), secuencia de TCR $\alpha$ , secuencia de TCR $\gamma$  o una combinación de las mismas. En algunas realizaciones, la secuencia específica de la diana del segundo cebador del segundo conjunto de cebadores se dirige a una secuencia de región constante de cadena pesada (C<sub>H</sub>), secuencia de región constante de TCR $\alpha$  (C $\alpha$ ), secuencias de región constante de TCR $\gamma$  (C $\gamma$ ) o una combinación de las mismas. En algunas realizaciones, la secuencia específica de la diana del segundo cebador se selecciona del grupo que consiste en GGGTTGGGGCGGATGCAC, CATCCGGAGCCTTGGTGG, CCTTGGGGCTGGTCGGGG, CGGATGGGCTCTGTGTGG, CCGATGGGCCCTTGGTGG, GGATTTAGAGTCTCTCAGCTG,

CACGGCAGGGTCAGGGTTC y GGGGAAACATCTGCATCAAGT. En algunas realizaciones, la secuencia específica de la diana del tercer cebador del segundo conjunto de cebadores se dirige a una secuencia de inmunoglobulina de cadena ligera (IgL), secuencia de TCR $\beta$ , secuencia de TCR $\delta$  o una combinación de las mismas. En algunas realizaciones, la secuencia específica de la diana del tercer cebador del segundo conjunto de cebadores se dirige a una secuencia de región constante de cadena ligera (CL), una secuencia de región constante de TCR $\beta$  (C $\beta$ ), una secuencia de región constante de TCR $\delta$  (C $\delta$ ) o una combinación de las mismas. En algunas realizaciones, la secuencia específica de la diana del tercer cebador se selecciona del grupo que consiste en TTTGATCTCCACCTTGGTCCCCTCCGC, TTTGATCTCCAGCTTGGTCCCCTGG, TTTGATATCCACTTTGGTCCCAGGGC, TTTGATTTCCACCTTGGTCCCCTGGC, TTTAATCTCCAGTCGTGTCCTTGGC, GAGGACGGTCACCTTGGTGCCA, TAGGACGGTCAGCTTGGTCCCCTCC, GAGGACGGTCAGCTGGGTGCC, TAAAATGATCAGCTGGGTTCCCTCCAC, TAGGACGGTGACCTTGGTCCCAG, GGGAGATCTCTGCTTCTGATG, CGACCTCGGGTGGGAACAC y CGGATGGTTTGGTATGAGGC. En algunas realizaciones, el segundo cebador del segundo conjunto de cebadores comprende una pluralidad de segundos cebadores. En algunas realizaciones, el tercer cebador del segundo conjunto de cebadores comprende una pluralidad de terceros cebadores. En algunas realizaciones, las secuencias específicas diana de la pluralidad de segundos cebadores se dirigen a una pluralidad de secuencias de inmunoglobulina de cadena pesada (IgH), secuencias de TCR $\alpha$ , secuencias de TCR $\gamma$  o una combinación de las mismas. En algunas realizaciones, la pluralidad de secuencias de inmunoglobulina de cadena pesada (IgH), secuencias de TCR $\alpha$  o secuencias de TCR $\gamma$  comprende una pluralidad de secuencias de región constante de cadena pesada (C<sub>H</sub>), secuencias de región constante de TCR $\alpha$  (C $\alpha$ ), secuencias de región constante de TCR $\gamma$  (C $\gamma$ ) o una combinación de las mismas. En algunas realizaciones, la pluralidad de secuencias de región constante de cadena pesada (C<sub>H</sub>) comprende dos o más secuencias seleccionadas del grupo que consiste en secuencias de región constante de cadena pesada (C<sub>H</sub>) de IgM, IgD, IgA, IgE, IgG y combinaciones de las mismas. En algunas realizaciones, las secuencias específicas diana de la pluralidad de terceros cebadores se dirigen a una pluralidad de secuencias de inmunoglobulina de cadena ligera (IgL), secuencias de TCR $\beta$ , secuencias de TCR $\delta$  o una combinación de las mismas. En algunas realizaciones, la pluralidad de secuencias de inmunoglobulina de cadena ligera (IgL), secuencias de TCR $\beta$  o secuencias de TCR $\delta$ , comprende una pluralidad de secuencias de región constante de cadena ligera (C<sub>L</sub>), secuencias de región constante de TCR $\beta$  (C $\beta$ ), secuencias de región constante de TCR $\delta$  (C $\delta$ ) o una combinación de las mismas. En algunas realizaciones, la pluralidad de secuencias de región constante de cadena ligera (C<sub>L</sub>) comprende dos o más secuencias seleccionadas del grupo que consiste en secuencias de región constante de cadena ligera (C<sub>L</sub>) de Ig $\kappa$ , Ig $\lambda$  y combinaciones de las mismas.

En algunas realizaciones, un primer cebador diana, un segundo cebador diana, el polinucleótido con código de barras del vaso, un polinucleótido con código de barras molecular o cualquier combinación de los mismos no está unido a un soporte sólido. En algunas realizaciones, un primer cebador diana, un segundo cebador diana, un cebador del primer conjunto de cebadores, un cebador del segundo conjunto de cebadores o cualquier combinación de los mismos, no comprende un código de barras molecular, un código de barras del vaso, un código de barras o cualquier combinación de los mismos. En algunas realizaciones, un primer cebador diana, un segundo cebador diana, un cebador del primer conjunto de cebadores, un cebador del segundo conjunto de cebadores o cualquier combinación de los mismos, no comprende una región saliente. En algunas realizaciones, cada vaso de la pluralidad de vasos no comprende un soporte sólido. En algunas realizaciones, el polinucleótido con código de barras del vaso está unido a un soporte sólido. En algunas realizaciones, el polinucleótido con código de barras del vaso está unido a una perla. En algunas realizaciones, el polinucleótido con código de barras del vaso, un polinucleótido con código de barras molecular o cualquier combinación de los mismos no es un cebador. En algunas realizaciones, el polinucleótido con código de barras del vaso, un polinucleótido con código de barras molecular o cualquier combinación de los mismos no se extiende.

En algunas realizaciones, (a)-(d) se realizan en el vaso individual.

50 En algunas realizaciones, (a)-(d) se realizan en una sola reacción.

En algunas realizaciones, el procedimiento comprende además la lisis de la célula individual. En algunas realizaciones, la lisis libera el primer y el segundo polinucleótido celular de la célula individual. En algunas realizaciones, la célula individual se lisa después de (a). En algunas realizaciones, la célula individual se lisa antes de (b). En algunas realizaciones, la célula individual se lisa en el vaso. En algunas realizaciones, la lisis comprende lisis química. En algunas realizaciones, la lisis comprende congelación-descongelación.

En algunas realizaciones, el código de barras del vaso se amplifica antes de (d). En algunas realizaciones, el código de barras del vaso se amplifica simultáneamente con (d). En algunas realizaciones, el código de barras del vaso y el primer polinucleótido con código de barras de una sola célula se amplifican o extienden simultáneamente. En algunas realizaciones, el código de barras del vaso, el primer polinucleótido con un solo código de barras de una sola célula y el segundo polinucleótido con un solo código de barras de una sola célula se amplifican o extienden simultáneamente. En algunas realizaciones, el primer polinucleótido con código de barras de una sola célula y el segundo polinucleótido con código de barras de una sola célula se amplifican o extienden simultáneamente. En algunas realizaciones, el primer polinucleótido con código de barras de una sola célula y el segundo polinucleótido con código de barras doble de una sola célula se amplifican o extienden simultáneamente. En algunas realizaciones, la pluralidad de vasos

comprende una pluralidad de pocillos. En algunas realizaciones, la pluralidad de vasos comprende una pluralidad de emulsiones. En algunas realizaciones, cada emulsión de la pluralidad de emulsiones es de aproximadamente 0,01 picolitros a 10 microlitros en volumen. En algunas realizaciones, la pluralidad de vasos comprende una pluralidad de recipientes. En algunas realizaciones, el primer cebador diana, el segundo cebador diana, un cebador del primer conjunto de cebadores o un cebador del segundo conjunto de cebadores comprende un código de barras de muestra. En algunas realizaciones, el procedimiento comprende además recuperar el primer polinucleótido con código de barras doble de una sola célula, el segundo polinucleótido con código de barras doble de una sola célula y sus productos amplificados del vaso. En algunas realizaciones, el procedimiento comprende además la secuenciación del primer polinucleótido con código de barras doble de una sola célula, el segundo polinucleótido con código de barras doble de una sola célula, sus productos amplificados o cualquier combinación de los mismos. En algunas realizaciones, el primer polinucleótido con código de barras doble de una sola célula, el segundo polinucleótido con código de barras doble de una sola célula, sus productos amplificados o cualquier combinación de los mismos, se secuencian simultáneamente. En algunas realizaciones, el primer polinucleótido con código de barras doble de una sola célula, el segundo polinucleótido con código de barras doble de una sola célula, sus productos amplificados o cualquier combinación de los mismos se secuencian en la misma reacción.

En algunas realizaciones, el procedimiento comprende además determinar el origen celular del primer polinucleótido celular y el segundo polinucleótido celular para que sea el mismo basándose en el código de barras del vaso. En algunas realizaciones, la determinación comprende hacer coincidir la secuencia del código de barras del primer polinucleótido con código de barras doble de una sola célula o su producto amplificado con la secuencia del código de barras del vaso del segundo polinucleótido con código de barras doble de una sola célula o su producto amplificado. En algunas realizaciones, el procedimiento comprende además determinar un número de moléculas de partida con una secuencia del primer polinucleótido celular, el segundo polinucleótido celular, o ambos, basándose en el código de barras molecular. En algunas realizaciones, la determinación comprende determinar el número de secuencias con un mismo primer código de barras molecular, un mismo segundo código de barras molecular o ambos. En algunas realizaciones, cuando una primera secuencia de un polinucleótido con código de barras doble de una sola célula o un producto amplificado del mismo y una segunda secuencia de un polinucleótido con un código de barras doble de una sola célula o un producto amplificado del mismo contienen un mismo código de barras del vaso o complemento del mismo, son del mismo vaso o célula individual. En algunas realizaciones, cuando la primera secuencia de un polinucleótido con código de barras doble de una sola célula o su producto amplificado y la segunda secuencia de un polinucleótido con código de barras doble de una sola célula o su producto amplificado contienen un código de barras molecular diferente o un complemento del mismo, son de una molécula de polinucleótido celular diferente. En algunas realizaciones, cuando la primera secuencia de un polinucleótido con código de barras doble de una sola célula o un producto amplificado del mismo y la segunda secuencia de un polinucleótido con un código de barras doble de una sola célula o un producto amplificado del mismo contienen un mismo código de barras molecular o complemento del mismo, son de una misma molécula de polinucleótido celular. En algunas realizaciones, cuando la primera secuencia de un polinucleótido con código de barras doble de una sola célula o su producto amplificado y la segunda secuencia de un polinucleótido con código de barras doble de una sola célula o su producto amplificado contienen un código de barras del vaso diferente o un complemento del mismo, son de un solo vaso o una sola célula diferente.

En algunas realizaciones, la célula individual comprende una célula inmunitaria. En algunas realizaciones, la pluralidad de células comprende una pluralidad de células inmunitarias. En algunas realizaciones, la célula inmunitaria es un linfocito o subtipo del mismo, un linfocito B o subtipo del mismo, un linfocito T o subtipo del mismo o una combinación de los mismos. En algunas realizaciones, la pluralidad de células está enriquecida en linfocitos B de memoria, linfocitos B vírgenes, linfocitos B plasmablasticos, linfocitos T vírgenes, linfocitos T plasmablasticos, cualquier subtipo de linfocito B, cualquier subtipo de linfocito T o cualquier combinación de los mismos. En algunas realizaciones, la célula individual comprende una célula cancerosa. En algunas realizaciones, la pluralidad de células comprende una pluralidad de células cancerosas. En algunas realizaciones, la célula cancerosa es una célula de carcinoma de células escamosas, una célula de adenocarcinoma, una célula de carcinoma de células de transición, una célula de sarcoma de hueso, una célula de sarcoma de cartílago, una célula de sarcoma muscular, una célula de leucemia, una célula de linfoma, una célula de glioma o cualquier combinación de las mismas. En algunas realizaciones, la pluralidad de células cancerosas está enriquecida en células cancerosas circulantes, células cancerosas endoteliales, células cancerosas epiteliales, células cancerosas raras, o cualquier tipo o subtipo de célula cancerosa. En algunas realizaciones, la muestra es una muestra biológica. En algunas realizaciones, la muestra biológica es de un sujeto. En algunas realizaciones, el procedimiento comprende además diagnosticar que el sujeto tiene una enfermedad o afección. En algunas realizaciones, el sujeto es un animal. En algunas realizaciones, el animal es un ser humano. En algunas realizaciones, el procedimiento comprende además determinar si un sujeto es homocigoto o heterocigoto para un alelo. En algunas realizaciones, el procedimiento comprende además diagnosticar, pronosticar o tratar a un sujeto con una enfermedad o afección. En algunas realizaciones, la muestra es una muestra de sangre. En algunas realizaciones, el primer o segundo polinucleótido celular se aísla de la muestra. En algunas realizaciones, el primer o segundo polinucleótido celular no está aislado de la muestra.

En algunas realizaciones, la muestra comprende una pluralidad de muestras que comprenden una primera muestra y una segunda muestra. En algunas realizaciones, la pluralidad de muestras comprende al menos 3, 4, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90 o 100, o más muestras. En algunas realizaciones, la pluralidad de muestras comprende al menos aproximadamente 100, 200, 300, 400, 500, 600, 700, 800, 900 o 1.000, o más muestras. En algunas realizaciones, la

pluralidad de muestras comprende al menos aproximadamente 1.000, 2.000, 3.000, 4.000, 5.000, 6.000, 7.000, 8.000 muestras, 9.000 o 10.000 muestras, o 100.000 muestras, o 1.000.000 o más muestras. En algunas realizaciones, la pluralidad de muestras comprende al menos aproximadamente 10.000 muestras. En algunas realizaciones, la primera muestra es de un primer sujeto y la segunda muestra es de un segundo sujeto. En algunas realizaciones, el primer sujeto es un sujeto con una enfermedad o afección. En algunas realizaciones, el segundo sujeto es un sujeto sin una enfermedad o afección. En algunas realizaciones, el primer o segundo polinucleótido celular comprende una secuencia variante. En algunas realizaciones, la secuencia variante comprende una mutación, polimorfismo, eliminación o inserción. En algunas realizaciones, el polimorfismo es un polimorfismo de un solo nucleótido. En algunas realizaciones, el primer o segundo polinucleótido celular es un biomarcador para una enfermedad o afección. En algunas realizaciones, el primer o segundo polinucleótido celular es de un patógeno. En algunas realizaciones, el patógeno es un virus, una bacteria o un hongo.

En algunas realizaciones, el procedimiento comprende además comparar las secuencias de un banco del primer y segundo polinucleótido con código de barras doble de una sola célula de un sujeto con un banco del primer y segundo polinucleótido con código de barras doble de una sola célula del mismo sujeto en un punto de tiempo diferente. En algunas realizaciones, el procedimiento comprende además comparar las secuencias de un banco del primer y segundo polinucleótido con código de barras doble de una sola célula de un sujeto con una enfermedad o afección con un banco del primer y segundo polinucleótido con código de barras doble de una sola célula de un sujeto sin la enfermedad o afección.

En algunas realizaciones, el procedimiento comprende además determinar una secuencia de línea germinal del primer polinucleótido celular, del segundo polinucleótido celular, o de ambos, en el que el primer polinucleótido celular comprende una secuencia de IgH o V<sub>H</sub>, y en el que el segundo polinucleótido celular comprende una secuencia de IgL o V<sub>L</sub>, o cualquier combinación de las mismas. En algunas realizaciones, el procedimiento comprende además determinar una varianza de la secuencia de IgL IgH, V<sub>H</sub>, V<sub>L</sub> o cualquier combinación de las mismas a partir de una secuencia de las de la línea germinal. En algunas realizaciones, el procedimiento comprende además determinar al menos uno de: el número total de secuencias únicas de IgH; el número total de secuencias únicas de IgL; el número total de secuencias únicas de IgH e IgL; el número total de secuencias únicas emparejadas de IgL e IgH; la frecuencia de una secuencia de IgH o una secuencia de IgL; o la frecuencia de una combinación de una secuencia de IgH y una secuencia de IgL frente a una u otras más. En algunas realizaciones, el procedimiento comprende además determinar una secuencia de línea germinal del primer polinucleótido celular, el segundo polinucleótido celular, o ambos, en el que el primer polinucleótido celular comprende una secuencia de TCRα o Vα, y en el que el segundo polinucleótido celular comprende una secuencia de TCRβ o Vβ, o cualquier combinación de las mismas. En algunas realizaciones, el procedimiento comprende además determinar una varianza de la secuencia del TCRα, TCRβ, Vα, Vβ, o cualquier combinación de las mismas a partir de una secuencia de las de la línea germinal. En algunas realizaciones, el procedimiento comprende además determinar al menos uno de: el número total de secuencias únicas de TCRα; el número total de secuencias únicas de TCRβ; el número total de secuencias únicas de TCRα y TCRβ; el número total de secuencias únicas emparejadas de TCRβ y TCRα; la frecuencia de una secuencia de TCRα o una secuencia de TCRβ; o la frecuencia de una combinación de una secuencia de TCRα y una secuencia de TCRβ frente a una u otras más. En algunas realizaciones, el procedimiento comprende además determinar una secuencia de línea germinal del primer polinucleótido celular, el segundo polinucleótido celular, o ambos, en el que el primer polinucleótido celular comprende una secuencia de TCRγ o Vγ, y en el que el segundo polinucleótido celular comprende una secuencia de TCRδ o Vδ, o cualquier combinación de las mismas. En algunas realizaciones, el procedimiento comprende además determinar una varianza de la secuencia del TCRγ, TCRδ, Vγ, Vδ o cualquier combinación de las mismas a partir de una secuencia de las de la línea germinal. En algunas realizaciones, el procedimiento comprende además determinar al menos uno de: el número total de secuencias únicas de TCRγ; el número total de secuencias únicas de TCRδ; el número total de secuencias únicas de TCRγ y TCRδ; el número total de secuencias únicas emparejadas de TCRδ y TCRγ; la frecuencia de una secuencia de TCRγ o una secuencia de TCRδ; o la frecuencia de una combinación de una secuencia de TCRγ y una secuencia de TCRδ frente a una u otras más. En algunas realizaciones, el procedimiento comprende además determinar al menos uno de: el número total de secuencias de un primer gen; el número total de secuencias de un segundo gen; el número total de secuencias únicas de un primer gen; el número total de secuencias únicas de un segundo gen; o la frecuencia de una secuencia de un primer gen o una secuencia de un segundo gen. En algunas realizaciones, el procedimiento comprende además la selección de un anticuerpo o TCR basándose en una cantidad total de uno o más pares de secuencias de IgL e IgH emparejadas individualmente, o secuencias de TCRα y TCRβ, o secuencias de TCRγ y TCRδ, y una varianza de una línea germinal. En algunas realizaciones, el procedimiento comprende además la selección de un anticuerpo o TCR basándose en una o más secuencias de IgL o de IgH, secuencias de TCRα y de TCRβ, o secuencias de TCRγ y de TCRδ, y una varianza de una línea germinal. En algunas realizaciones, el procedimiento comprende además la selección de un anticuerpo o TCR basándose en uno o más de los patrones de secuencia, análisis de varianza, dinámica o frecuencia. En algunas realizaciones, el procedimiento comprende además seleccionar un anticuerpo o TCR basándose en la frecuencia.

En algunas realizaciones, el anticuerpo o TCR seleccionado se une a un epítipo con una KD inferior a aproximadamente o igual a  $1 \times 10^{-7}$ ,  $1 \times 10^{-8}$ ,  $1 \times 10^{-9}$ ,  $1 \times 10^{-10}$ ,  $1 \times 10^{-11}$  o  $1 \times 10^{-12}$  M.

En algunas realizaciones, el anticuerpo o TCR seleccionado es un anticuerpo o TCR terapéutico humano. En algunas realizaciones, el anticuerpo o TCR seleccionado es un anticuerpo o TCR neutralizante. En algunas realizaciones, una

diana a la que se une el anticuerpo o TCR seleccionado es desconocida. En algunas realizaciones, una diana a la que se une el anticuerpo o TCR seleccionado es desconocida en el momento en el que se selecciona el anticuerpo o TCR seleccionado.

5 En algunas realizaciones, el procedimiento comprende además poner en contacto el anticuerpo o TCR seleccionado con al menos un candidato a biomarcador para descubrir un biomarcador. En algunas realizaciones, el candidato a biomarcador está en un soporte sólido. En algunas realizaciones, el biomarcador está en solución. En algunas realizaciones, el anticuerpo o TCR está en un soporte sólido. En algunas realizaciones, el anticuerpo o TCR está en solución. En algunas realizaciones, el soporte sólido es una matriz. En algunas realizaciones, el soporte sólido es una perla.

10 En algunas realizaciones, el procedimiento comprende además insertar el primer polinucleótido celular en un vector. En algunas realizaciones, el procedimiento comprende además insertar el segundo polinucleótido celular en el vector. En algunas realizaciones, el vector es un vector de clonación. En algunas realizaciones, el vector es un vector de expresión.

15 En algunas realizaciones, el procedimiento comprende además secuencias de emparejamiento con códigos de barras moleculares idénticos. En algunas realizaciones, el procedimiento comprende además formar secuencias de consenso a partir del banco. En algunas realizaciones, la secuenciación y los errores de PCR se minimizan, se eliminan o son inferiores al 0,01 %, 0,001 %, 0,0001 %, 0,00001 %, 0,000001 % o 0,0000001 %. En algunas realizaciones, el número de ciclos de una reacción de amplificación se limita a cualquiera de 1-40 ciclos.

20 La presente divulgación proporciona un anticuerpo o TCR aislado, purificado, identificado mediante cualquiera de los procedimientos descritos en el presente documento. En un aspecto, en el presente documento, se proporciona un anticuerpo aislado, purificado IgL, TCR $\beta$  o TCR $\delta$  identificado por cualquiera de los procedimientos descritos en el presente documento. En un aspecto, se proporciona en el presente documento un anticuerpo aislado, purificado IgH, TCR $\alpha$  o TCR $\gamma$  mediante cualquiera de los procedimientos descritos en el presente documento. En un aspecto, en el presente documento, se proporciona un fragmento Fab aislado, purificado de un anticuerpo o TCR identificado mediante cualquiera de los procedimientos descritos en el presente documento. En un aspecto, en el presente documento, se proporciona un fragmento Fab2 aislado, purificado de un anticuerpo o TCR identificado mediante cualquiera de los procedimientos descritos en el presente documento. En un aspecto, en el presente documento, se proporciona un fragmento Fv aislado, purificado de un anticuerpo o TCR identificado mediante cualquiera de los procedimientos descritos en el presente documento. En un aspecto, en el presente documento, se proporciona un fragmento ScFv aislado, purificado de un anticuerpo identificado mediante cualquiera de los procedimientos descritos en el presente documento. En un aspecto, en el presente documento, se proporciona un procedimiento para el tratamiento de un sujeto que lo necesita, que comprende administrar el anticuerpo o TCR seleccionado, o un fragmento del mismo, a un sujeto que lo necesite. En algunas realizaciones, el anticuerpo, el TCR o el fragmento del mismo se identifica del sujeto que lo necesita. En algunas realizaciones, el anticuerpo, el TCR o el fragmento del mismo no se identifica del sujeto que lo necesita. En algunas realizaciones, el sujeto que lo necesita muestra uno o más síntomas de una enfermedad. En algunas realizaciones, el sujeto que lo necesita tiene una enfermedad. En algunas realizaciones, la enfermedad es desconocida. En algunas realizaciones, la enfermedad es conocida. En algunas realizaciones, la muestra comprende una primera muestra de un sujeto tomada en un primer punto de tiempo y una segunda muestra del sujeto tomado y un segundo punto de tiempo. En algunas realizaciones, el procedimiento comprende además determinar un aumento o una disminución de la cantidad del primer o segundo polinucleótido celular de las muestras tomadas en el primer y el segundo punto de tiempo. En algunas realizaciones, el aumento o la disminución de la cantidad es un aumento o una disminución que varía de al menos aproximadamente: 0,1 veces, 0,2 veces, 0,3 veces, 0,4 veces, 0,5 veces, 0,6 veces, 0,7 veces, 0,8 veces, 0,9 veces, 1,5 veces, 2 veces, 3 veces, 5 veces, 10 veces, 50 veces, 100 veces, 1.000 veces, 10.000 veces, 100.000 veces, 1.000.000 veces o más. En algunas realizaciones, el tiempo entre el primer y el segundo punto de tiempo es de aproximadamente, o al menos, aproximadamente de: 1 hora, 2 horas, 3 horas, 4 horas, 5 horas, 6 horas, 7 horas, 8 horas, 9 horas, 10 horas, 11 horas, 12 horas, 13 horas, 14 horas, 15 horas, 16 horas, 17 horas, 18 horas, 19 hora, 20 horas, 21 horas, 22 horas, 23 horas, 1 día, 2 días, 3 días, 4 días, 5 días, 6 días, 1 semana, 2 semanas, 3 semanas, 4 semanas, 5 semanas, 6 semanas, 7 semanas, 8 semanas, 9 semanas, 10 semanas, 11 semanas, 12 semanas, 4 meses, 5 meses, 6 meses, 7 meses, 8 meses, 9 meses, 10 meses, 11 meses, 12 meses o más.

55 En algunas realizaciones, la secuenciación es de alto rendimiento. En algunas realizaciones, el procedimiento no comprende un multiplex de cebadores ni/o un multiplex de cebadores unidos a un soporte sólido. En algunas realizaciones, el procedimiento no emplea una multiplicidad de cebadores de segmento V que comprenden una secuencia que es complementaria a un solo segmento V funcional o una pequeña familia de segmentos V. En algunas realizaciones, el procedimiento no emplea una etapa de aislamiento del primer o segundo polinucleótido celular. En algunas realizaciones, la secuenciación se realiza mediante síntesis masiva paralela.

60 En algunas realizaciones, el procedimiento comprende además comparar las lecturas de secuencias con una secuencia de línea germinal y determinar una acumulación somática de hipermutación de las lecturas de secuencias. En algunas realizaciones, el procedimiento comprende además determinar una distribución de isotipos de secuencias de anticuerpos para seleccionar un isotipo específico. En algunas realizaciones, el anticuerpo seleccionado comprende

un isotipo de Ig específico. En algunas realizaciones, el isotipo Ig es IgA, IgG, IgM, IgD o IgE.

En algunas realizaciones, el procedimiento comprende además generar un banco de secuencias de anticuerpos IgH e IgL emparejadas o secuencias de TCR $\alpha$  y TCR $\beta$ . En algunas realizaciones, el banco es una base de datos. En algunas realizaciones, el primer y el segundo polinucleótido con código de barras doble de una sola célula comprenden una CDR1, CDR2, CDR3 y/o región de hipermutación a través de secuencias codificantes de anticuerpos o TCR.

En algunas realizaciones, el procedimiento comprende además la clonación del anticuerpo o TCR seleccionado directamente en la tecnología de visualización de superficies. En algunas realizaciones, el procedimiento comprende además desarrollar el anticuerpo o TCR seleccionado mediante evolución dirigida. En algunas realizaciones, el procedimiento comprende además seleccionar el anticuerpo o TCR seleccionado para determinar su especificidad funcional, afinidad o capacidad de neutralización. En algunas realizaciones, las mutaciones somáticas se determinan con una confianza del 99 % o superior. En algunas realizaciones, se identifica cada segmento V, D y J de cada molécula de polinucleótido.

En algunas realizaciones, el código de barras del vaso comprende al menos 2 nucleótidos. En algunas realizaciones, el código de barras del vaso comprende al menos 3, 4, 5, 6, 7, 8 o 9 nucleótidos. En algunas realizaciones, el código de barras del vaso comprende al menos 10 nucleótidos. En algunas realizaciones, el código de barras del vaso comprende al menos 15 nucleótidos. En algunas realizaciones, el código de barras del vaso comprende como máximo 50 nucleótidos. En algunas realizaciones, el código de barras del vaso comprende de 10-30 nucleótidos. En algunas realizaciones, el código de barras del vaso comprende una secuencia degenerada. En algunas realizaciones, el código de barras del vaso comprende la secuencia NNNNNNNNNNNNNNNN, en la que N es cualquier ácido nucleico. En algunas realizaciones, el código de barras del vaso comprende la secuencia NNNNNWNNNNNNWNNNNN, en la que N es cualquier ácido nucleico y W es adenina o timina. En algunas realizaciones, el código de barras del vaso comprende la secuencia NNNNNXNNNNXNNNNN, en la que N es cualquier ácido nucleico y X es cualquier nucleótido conocido. En algunas realizaciones, el código de barras del vaso comprende la secuencia NNNNNNNNNNNNNNNNNN, en la que N es cualquier ácido nucleico y al menos uno o dos N de la secuencia es W, en la que W es adenina o timina. En algunas realizaciones, el código de barras del vaso comprende la secuencia NNNNNNNNNNNNNNNNNN, en la que N es cualquier ácido nucleico y al menos uno o dos N de la secuencia es X, en la que X es cualquier nucleótido conocido. En algunas realizaciones, el código de barras molecular comprende al menos 2 nucleótidos. En algunas realizaciones, el código de barras molecular comprende al menos 3, 4, 5, 6, 7, 8 o 9 nucleótidos. En algunas realizaciones, el código de barras molecular comprende al menos 10 nucleótidos. En algunas realizaciones, el código de barras molecular comprende al menos 15 nucleótidos. En algunas realizaciones, el código de barras molecular comprende como máximo 50 nucleótidos. En algunas realizaciones, el código de barras molecular comprende de 10-30 nucleótidos. En algunas realizaciones, el código de barras molecular comprende una secuencia degenerada. En algunas realizaciones, el código de barras molecular comprende la secuencia NNNNNNNN, en la que N es cualquier ácido nucleico. En algunas realizaciones, el código de barras molecular comprende la secuencia NNTNNANN, en la que N es cualquier ácido nucleico. En algunas realizaciones, el código de barras molecular comprende la secuencia NNWNNWNN, en la que N es cualquier ácido nucleico y W es adenina o timina. En algunas realizaciones, el código de barras molecular comprende la secuencia NNXNNXNN, en la que N es cualquier ácido nucleico y X es cualquier nucleótido conocido. En algunas realizaciones, el código de barras molecular comprende la secuencia NNNNNNNN, en la que N es cualquier ácido nucleico y al menos uno o dos N de la secuencia es W, en la que W es adenina o timina. En algunas realizaciones, el código de barras molecular comprende la secuencia NNNNNNNN, en la que N es cualquier ácido nucleico y al menos uno o dos N de la secuencia es X, en la que X es cualquier nucleótido conocido.

En algunas realizaciones, el procedimiento comprende además corregir errores de amplificación. En algunas realizaciones, el procedimiento comprende además corregir errores de secuenciación. En algunas realizaciones, el procedimiento comprende además reunir o agrupar secuencias que comprenden el mismo código de barras molecular. En algunas realizaciones, el procedimiento comprende además reunir o agrupar secuencias que comprenden el mismo código de barras molecular usando un ordenador o un algoritmo. En algunas realizaciones, el procedimiento comprende además reunir o agrupar secuencias que comprenden el mismo código de barras del vaso usando un ordenador o un algoritmo. En algunas realizaciones, el procedimiento comprende además secuencias de agrupamiento con al menos aproximadamente el 90 %, el 95 % o el 99 % de homología de secuencia. En algunas realizaciones, el procedimiento comprende además alinear secuencias con al menos aproximadamente el 90 %, el 95 % o el 99 % de homología de secuencia. En algunas realizaciones, el agrupamiento o la alineación se realiza con ayuda de un ordenador o algoritmo. En algunas realizaciones, el procedimiento comprende determinar el número de lecturas de secuencias que contienen el mismo código de barras molecular. En algunas realizaciones, el procedimiento comprende determinar el número de lecturas de secuencias que contienen tanto el mismo código de barras molecular como la misma primera secuencia de polinucleótido celular con al menos aproximadamente el 90 %, el 95 % o el 99 % de homología de secuencia. En algunas realizaciones, el procedimiento comprende determinar el número de lecturas de secuencias que contienen tanto el mismo código de barras molecular como la misma segunda secuencia de polinucleótidos celulares con al menos aproximadamente el 90 %, el 95 % o el 99 % de homología de secuencia. En algunas realizaciones, el procedimiento comprende determinar la cantidad de un primer o un segundo polinucleótido celular en la muestra. En algunas realizaciones, el procedimiento comprende formar una secuencia de consenso a

partir de dos o más secuencias, lecturas de secuencias, secuencias de amplicones, secuencias reunidas, secuencias alineadas, secuencias agrupadas o secuencias de conjuntos de amplicones que comprenden el mismo código de barras molecular o código de barras del vaso, o ambos. En algunas realizaciones, el procedimiento comprende determinar una primera o una segunda secuencia de polinucleótidos celulares con al menos aproximadamente el 80 %, 81 %, 82 %, 83 %, 84 %, 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 %, 99,5 %, 99,6 %, 99,7 %, 99,8 %, 99,9 %, 99,99 % o 100 % de exactitud o confianza. En algunas realizaciones, la secuenciación y los errores de PCR se minimizan, se eliminan o son inferiores al 0,01 %, 0,001 %, 0,0001 %, 0,00001 % o 0,000001 %. En algunas realizaciones, la tasa de error de la secuenciación es inferior o igual al 0,00001 %, 0,0001 %, 0,001 %, 0,01 % o 0 %. En algunas realizaciones, la tasa de error de la secuenciación no es 0. En algunas realizaciones, se secuencian al menos 1.000, 100.000,  $1 \times 10^6$ ,  $1 \times 10^7$ ,  $1 \times 10^8$ ,  $1 \times 10^9$ ,  $1 \times 10^{10}$ ,  $1 \times 10^{11}$ ,  $1 \times 10^{12}$  o  $9 \times 10^{12}$  polinucleótidos. En algunas realizaciones, el procedimiento se realiza en una cantidad positiva de tiempo inferior o igual a 4 semanas, 3 semanas, 2 semanas, 1 semana, 6 días, 5 días, 5 días, 4 días, 3 días, 2 días, 1 día, 18 horas, 12 horas, 9 horas, 6 horas, 3 horas, 2 horas o 1 hora. En algunas realizaciones, el número de lecturas usadas para lograr una determinada confianza o precisión de identificación de bases es de al menos aproximadamente 1,1, 1,5, 2, 2,5, 3, 3,5, 4, 4,5, 5, 5,5, 6, 6,5, 7, 7,5, 8, 8,5, 9, 9,5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900 o 1.000 veces inferior al número de lecturas usadas para lograr la misma, una similar o una mayor confianza o precisión de identificación de bases usando un procedimiento similar sin el uso con códigos de barras moleculares, códigos de barras del vaso o ambos. En algunas realizaciones, el número de lecturas usadas para lograr una determinada confianza o precisión de identificación de bases es de al menos aproximadamente 1.000, 100.000,  $1 \times 10^6$ ,  $1 \times 10^7$ ,  $1 \times 10^8$ ,  $1 \times 10^9$ ,  $1 \times 10^{10}$ ,  $1 \times 10^{11}$ ,  $1 \times 10^{12}$  o  $9 \times 10^{12}$  lecturas menos que el número de lecturas usadas para lograr la misma, una similar o una mayor confianza o precisión de identificación de bases usando un procedimiento similar sin el uso con códigos de barras moleculares, códigos de barras del vaso o ambos. En algunas realizaciones, la pluralidad de vasos comprende al menos 1.000, 100.000,  $1 \times 10^6$ ,  $1 \times 10^7$ ,  $1 \times 10^8$ ,  $1 \times 10^9$ ,  $1 \times 10^{10}$ ,  $1 \times 10^{11}$ ,  $1 \times 10^{12}$  o  $9 \times 10^{12}$  o más vasos. En algunas realizaciones, la pluralidad de polinucleótidos celulares comprende al menos 1.000, 100.000,  $1 \times 10^6$ ,  $1 \times 10^7$ ,  $1 \times 10^8$ ,  $1 \times 10^9$ ,  $1 \times 10^{10}$ ,  $1 \times 10^{11}$ ,  $1 \times 10^{12}$  o  $9 \times 10^{12}$  o más polinucleótidos celulares.

La presente divulgación proporciona una composición que comprende: una pluralidad de vasos, comprendiendo cada uno una sola célula de una muestra que comprende una pluralidad de células, una pluralidad de polinucleótidos con código de barras molecular, un polinucleótido con código de barras del vaso; un primer polinucleótido complementario que es complementario a un primer polinucleótido celular de la célula individual, y un segundo polinucleótido complementario que es complementario a un segundo polinucleótido celular de la célula individual; en el que el primer polinucleótido complementario comprende un primer código de barras molecular de la pluralidad de polinucleótidos con código de barras molecular y el código de barras del vaso del polinucleótido con código de barras del vaso o un producto amplificado del polinucleótido con código de barras del vaso, y en el que el segundo polinucleótido complementario comprende un segundo código de barras molecular de la pluralidad de polinucleótidos con código de barras molecular y el código de barras del vaso del polinucleótido con código de barras del vaso o un producto amplificado del polinucleótido con código de barras del vaso.

En algunas realizaciones, los códigos de barras moleculares del primer y segundo polinucleótido con código de barras molecular son diferentes. En algunas realizaciones, el primer y el segundo polinucleótido complementario comprenden un código de barras molecular diferente. En algunas realizaciones, el primer y el segundo polinucleótido complementario comprenden el mismo código de barras del vaso. En algunas realizaciones, la pluralidad de polinucleótidos con código de barras molecular no son productos amplificados. En algunas realizaciones, el código de barras molecular de un polinucleótido con código de barras molecular de un primer vaso es diferente al código de barras molecular de un polinucleótido con código de barras molecular de un segundo vaso. En algunas realizaciones, el código de barras molecular de cada polinucleótido con código de barras molecular de un primer vaso de la pluralidad de vasos es único. En algunas realizaciones, el código de barras molecular de cada polinucleótido con código de barras molecular de un segundo vaso de la pluralidad de vasos es único. En algunas realizaciones, el código de barras molecular de cada polinucleótido con código de barras molecular de un primer vaso y un segundo vaso son únicos. En algunas realizaciones, el código de barras molecular de cada polinucleótido con código de barras molecular de un tercer vaso de la pluralidad de vasos es único. En algunas realizaciones, el código de barras molecular de cada polinucleótido con código de barras molecular del primer vaso, segundo vaso y tercer vaso son únicos. En algunas realizaciones, el código de barras molecular de cada polinucleótido con código de barras molecular de cualquier vaso individual de la pluralidad de vasos es único. En algunas realizaciones, el código de barras molecular de cada polinucleótido con código de barras molecular en cualquier vaso de la pluralidad de vasos es diferente al código de barras molecular de cada polinucleótido con código de barras molecular de cualquier otro vaso de la pluralidad de vasos. En algunas realizaciones, el código de barras molecular de un polinucleótido con código de barras molecular de un primer vaso es el mismo que el código de barras molecular de un polinucleótido con código de barras molecular de un segundo vaso. En algunas realizaciones, el código de barras molecular de un polinucleótido con código de barras molecular de un primer vaso es el mismo que el código de barras molecular de un polinucleótido con código de barras molecular del primer vaso. En algunas realizaciones, el código de barras molecular de un polinucleótido con código de barras molecular de un segundo vaso es el mismo que el código de barras molecular de un polinucleótido con código de barras molecular del segundo vaso. En algunas realizaciones, el código de barras del vaso de un polinucleótido con código de barras o amplicón del mismo en un primer vaso de la pluralidad de vasos es diferente del código de barras del vaso de un polinucleótido con código de barras o amplicón del mismo de un segundo vaso de la



extremo 3' del primer polinucleótido complementario. En algunas realizaciones, la región complementaria a tres o más nucleótidos no molde añadido al extremo 3' del primer polinucleótido complementario es una región terminal.

5 En algunas realizaciones, un primer y un segundo polinucleótido con código de barras molecular no se fusionan entre sí. En algunas realizaciones, el primer y el segundo polinucleótido complementario no se fusionan entre sí.

10 En algunas realizaciones, el primer polinucleótido celular es ADN. En algunas realizaciones, el segundo polinucleótido celular es ADN. En algunas realizaciones, el primer polinucleótido celular es ARN. En algunas realizaciones, el segundo polinucleótido celular es ARN. En algunas realizaciones, el ARN es ARNm. En algunas realizaciones, el primer polinucleótido complementario es ADNc. En algunas realizaciones, el segundo polinucleótido complementario es ADNc.

15 En algunas realizaciones, la composición comprende además una transferasa terminal no molde, una transcriptasa inversa, una polimerasa o cualquier combinación de las mismas. En algunas realizaciones, el primer y/o el segundo polinucleótido complementario comprenden tres o más nucleótidos no molde añadidos al extremo 3'. En algunas realizaciones, la transferasa terminal no molde es una transcriptasa inversa, y en las que la transcriptasa inversa se selecciona del grupo que consiste en la transcriptasa inversa Superscript II, la transcriptasa inversa máxima, la transcriptasa inversa Protoscript II, la transcriptasa inversa del virus de la leucemia murina de Moloney (MMLV-RT), la transcriptasa inversa HighScriber, la transcriptasa inversa del virus de la mieloblastosis aviar (AMV), cualquier transcriptasa inversa que comprenda la actividad de la desoxinucleotidil transferasa terminal y sus combinaciones.

20 En algunas realizaciones, un primer polinucleótido con código de barras molecular comprende una región complementaria a los tres o más nucleótidos no molde en el extremo 3' del primer polinucleótido complementario. En algunas realizaciones, un segundo polinucleótido con código de barras molecular comprende una región complementaria a los tres o más nucleótidos no molde en el extremo 3' del segundo polinucleótido complementario. En algunas realizaciones, los tres o más nucleótidos no molde son idénticos. En algunas realizaciones, al menos uno de los tres o más nucleótidos no molde no es idéntico a otro nucleótido de los tres o más nucleótidos no molde. En algunas realizaciones, al menos un nucleótido de la región complementaria del primer polinucleótido con código de barras molecular no es idéntico a otro ácido nucleico de la región complementaria del primer polinucleótido con código de barras molecular. En algunas realizaciones, al menos un nucleótido de la región complementaria del segundo polinucleótido con código de barras molecular no es idéntico a otro ácido nucleico de la región complementaria del segundo polinucleótido con código de barras molecular. En algunas realizaciones, el al menos un nucleótido no idéntico es un desoxirribonucleótido o un análogo del mismo. En algunas realizaciones, el al menos un nucleótido no idéntico no es un ribonucleótido o un análogo del mismo. En algunas realizaciones, el al menos un nucleótido no idéntico es una desoxirriboguanosina. En algunas realizaciones, el al menos un nucleótido no idéntico es un análogo de desoxirriboguanosina. En algunas realizaciones, el al menos un nucleótido no idéntico es un nucleótido terminal del primer o segundo polinucleótido con código de barras molecular. En algunas realizaciones, el al menos un nucleótido no idéntico es un ribonucleótido o un análogo del mismo. En algunas realizaciones, un nucleótido terminal de la región complementaria del primer o segundo polinucleótido con código de barras molecular es un desoxirribonucleótido o un análogo del mismo. En algunas realizaciones, un nucleótido terminal de la región hibridada del primer o segundo polinucleótido con código de barras molecular no es un ribonucleótido o análogo del mismo. En algunas realizaciones, un nucleótido terminal de la región hibridada del primer o segundo polinucleótido con código de barras molecular es una desoxirriboguanosina. En algunas realizaciones, un nucleótido terminal de la región hibridada del primer o segundo polinucleótido con código de barras molecular es un análogo de desoxirriboguanosina. En algunas realizaciones, un nucleótido terminal de la región hibridada del primer o segundo polinucleótido con código de barras molecular es un ribonucleótido o análogo del mismo. En algunas realizaciones, al menos dos nucleótidos no terminales de la región hibridada del primer o segundo polinucleótido con código de barras molecular son ribonucleótidos o análogos de los mismos. En algunas realizaciones, al menos dos nucleótidos no terminales de la región hibridada del primer o segundo polinucleótido con código de barras molecular no son desoxirribonucleótidos o análogos de los mismos. En algunas realizaciones, al menos dos nucleótidos no terminales de la región hibridada del primer o segundo polinucleótido con código de barras molecular son desoxirribonucleótidos o análogos de los mismos. En algunas realizaciones, el primer polinucleótido complementario comprende una región complementaria a un primer polinucleótido con código de barras molecular. En algunas realizaciones, el segundo polinucleótido complementario comprende una región complementaria a un segundo polinucleótido con código de barras molecular. En algunas realizaciones, el primer polinucleótido complementario comprende una región complementaria a un segundo polinucleótido con código de barras molecular. En algunas realizaciones, la región del primer polinucleótido complementario que es complementario al primer o segundo polinucleótido con código de barras molecular no es complementaria a una secuencia de código de barras molecular. En algunas realizaciones, la región del primer polinucleótido complementario que es complementario al primer o segundo polinucleótido con código de barras molecular no es complementaria a una región del polinucleótido con código de barras del vaso o un producto amplificado del mismo. En algunas realizaciones, la región del primer polinucleótido complementario al primer o segundo polinucleótido con código de barras molecular comprende tres o más nucleótidos no molde añadidos al extremo 3' del primer polinucleótido complementario. En algunas realizaciones, la región del segundo polinucleótido complementario que es complementario al segundo polinucleótido con código de barras molecular comprende tres o más nucleótidos no molde añadidos al extremo 3' del segundo polinucleótido complementario. En algunas realizaciones, el primer polinucleótido complementario no es complementario al polinucleótido con código de barras del vaso. En algunas realizaciones, el segundo polinucleótido complementario no es complementario al polinucleótido

con código de barras del vaso. En algunas realizaciones, una región de un complemento de un primer polinucleótido con código de barras molecular es complementaria a una región del polinucleótido con código de barras del vaso. En algunas realizaciones, una región de un complemento de un segundo polinucleótido con código de barras molecular es complementaria a una región del polinucleótido con código de barras del vaso. En algunas realizaciones, la composición comprende además uno o más cebadores de los procedimientos anteriores. En algunas realizaciones, cada vaso de la pluralidad de vasos no comprende un soporte sólido. En algunas realizaciones, el polinucleótido con código de barras del vaso está unido a un soporte sólido. En algunas realizaciones, el polinucleótido con código de barras del vaso está unido a una perla. En algunas realizaciones, el polinucleótido con código de barras del vaso, un polinucleótido con código de barras molecular o cualquier combinación de los mismos no es un cebador. En algunas realizaciones, el polinucleótido con código de barras del vaso, un polinucleótido con código de barras molecular o cualquier combinación de los mismos no es un polinucleótido extendido. En algunas realizaciones, la célula está lisada. En algunas realizaciones, la pluralidad de vasos comprende una pluralidad de pocillos. En algunas realizaciones, la pluralidad de vasos comprende una pluralidad de emulsiones. En algunas realizaciones, cada emulsión de la pluralidad de emulsiones es de aproximadamente 0,01 picolitros a 10 microlitros en volumen.

En algunas realizaciones, la célula individual comprende una célula inmunitaria. En algunas realizaciones, la pluralidad de células comprende una pluralidad de células inmunitarias. En algunas realizaciones, la célula inmunitaria es un linfocito o subtipo del mismo, un linfocito B o subtipo del mismo, un linfocito T o subtipo del mismo o una combinación de los mismos. En algunas realizaciones, la pluralidad de células está enriquecida en linfocitos B de memoria, linfocitos B vírgenes, linfocitos B plasmablasticos, linfocitos T vírgenes, linfocitos T plasmablasticos, cualquier subtipo de linfocito B, cualquier subtipo de linfocito T o cualquier combinación de los mismos. En algunas realizaciones, la célula individual comprende una célula cancerosa. En algunas realizaciones, la pluralidad de células comprende una pluralidad de células cancerosas. En algunas realizaciones, la célula cancerosa es una célula de carcinoma de células escamosas, una célula de adenocarcinoma, una célula de carcinoma de células de transición, una célula de sarcoma de hueso, una célula de sarcoma de cartílago, una célula de sarcoma muscular, una célula de leucemia, una célula de linfoma, una célula de glioma o cualquier combinación de las mismas. En algunas realizaciones, la pluralidad de células cancerosas está enriquecida en células cancerosas circulantes, células cancerosas endoteliales, células cancerosas epiteliales, células cancerosas raras, o cualquier tipo o subtipo de célula cancerosa. En algunas realizaciones, el primer o segundo polinucleótido celular comprende una secuencia variante. En algunas realizaciones, la secuencia variante comprende una mutación, polimorfismo, eliminación o inserción. En algunas realizaciones, el polimorfismo es un polimorfismo de un solo nucleótido. En algunas realizaciones, el primer o segundo polinucleótido celular es un biomarcador para una enfermedad o afección. En algunas realizaciones, el primer o segundo polinucleótido celular es de un patógeno. En algunas realizaciones, el primer o segundo polinucleótido complementario comprende una CDR1, CDR2, CDR3 y/o región de hipermutación a través de secuencias codificantes de anticuerpos o TCR.

En algunas realizaciones, el código de barras del vaso comprende al menos 2 nucleótidos. En algunas realizaciones, el código de barras del vaso comprende al menos 3, 4, 5, 6, 7, 8 o 9 nucleótidos. En algunas realizaciones, el código de barras del vaso comprende al menos 10 nucleótidos. En algunas realizaciones, el código de barras del vaso comprende al menos 15 nucleótidos. En algunas realizaciones, el código de barras del vaso comprende como máximo 50 nucleótidos. En algunas realizaciones, el código de barras del vaso comprende de 10-30 nucleótidos. En algunas realizaciones, el código de barras del vaso comprende una secuencia degenerada. En algunas realizaciones, el código de barras del vaso comprende una secuencia degenerada total o parcial. En algunas realizaciones, el código de barras del vaso comprende la secuencia NNNNNNNNNNNNNNNN, en la que N es cualquier ácido nucleico. En algunas realizaciones, el código de barras del vaso comprende la secuencia NNNNNWNNNNNNWNNNNN, en la que N es cualquier ácido nucleico y W es adenina o timina. En algunas realizaciones, el código de barras del vaso comprende la secuencia NNNNNXNNNNNNXNNNNN, en la que N es cualquier ácido nucleico y X es cualquier nucleótido conocido. En algunas realizaciones, el código de barras del vaso comprende la secuencia NNNNNNNNNNNNNNNNNN, en la que N es cualquier ácido nucleico y al menos uno o dos N de la secuencia es W, en la que W es adenina o timina. En algunas realizaciones, el código de barras del vaso comprende la secuencia NNNNNNNNNNNNNNNNNN, en la que N es cualquier ácido nucleico y al menos uno o dos N de la secuencia es X, en la que X es cualquier nucleótido conocido. En algunas realizaciones, el código de barras molecular comprende al menos 2 nucleótidos. En algunas realizaciones, el código de barras molecular comprende al menos 3, 4, 5, 6, 7, 8 o 9 nucleótidos. En algunas realizaciones, el código de barras molecular comprende al menos 10 nucleótidos. En algunas realizaciones, el código de barras molecular comprende al menos 15 nucleótidos. En algunas realizaciones, el código de barras molecular comprende como máximo 50 nucleótidos. En algunas realizaciones, el código de barras molecular comprende de 10-30 nucleótidos. En algunas realizaciones, el código de barras molecular comprende una secuencia degenerada. En algunas realizaciones, el código de barras molecular comprende una secuencia degenerada total o parcial. En algunas realizaciones, el código de barras molecular comprende la secuencia NNNNNNNN, en la que N es cualquier ácido nucleico. En algunas realizaciones, el código de barras molecular comprende la secuencia NNTNNANN, en la que N es cualquier ácido nucleico. En algunas realizaciones, el código de barras molecular comprende la secuencia NNWNNWNN, en la que N es cualquier ácido nucleico y W es adenina o timina. En algunas realizaciones, el código de barras molecular comprende la secuencia NNXNNXNN, en la que N es cualquier ácido nucleico y X es cualquier nucleótido conocido. En algunas realizaciones, el código de barras molecular comprende la secuencia NNNNNNNN, en la que N es cualquier ácido nucleico y al menos uno o dos N de la secuencia es W, en la que W es adenina o timina. En algunas realizaciones, el código de barras molecular comprende la secuencia NNNNNNNN, en la que N es cualquier ácido nucleico y al menos uno o dos N de la secuencia es X, en la que X es cualquier nucleótido conocido.

En algunas realizaciones, la pluralidad de vasos comprende al menos 1.000, 100.000,  $1 \times 10^6$ ,  $1 \times 10^7$ ,  $1 \times 10^8$ ,  $1 \times 10^9$ ,  $1 \times 10^{10}$ ,  $1 \times 10^{11}$ ,  $1 \times 10^{12}$  o  $9 \times 10^{12}$  o más vasos. En algunas realizaciones, la pluralidad de polinucleótidos celulares comprende al menos 1.000, 100.000,  $1 \times 10^6$ ,  $1 \times 10^7$ ,  $1 \times 10^8$ ,  $1 \times 10^9$ ,  $1 \times 10^{10}$ ,  $1 \times 10^{11}$ ,  $1 \times 10^{12}$  o  $9 \times 10^{12}$  o más polinucleótidos celulares.

La presente divulgación proporciona un procedimiento de polinucleótidos con código de barras que comprende (a) hibridar un polinucleótido con código de barras molecular con cada uno de una pluralidad de polinucleótidos de una sola célula, en el que el polinucleótido con código de barras molecular hibridado es de una pluralidad de polinucleótidos con códigos de barras moleculares únicos de dentro de un vaso que comprende esa célula individual; (b) extender un polinucleótido de la célula individual que está hibridado a un polinucleótido con código de barras molecular para formar un polinucleótido celular con código de barras molecular; (c) hibridar un polinucleótido con código de barras del vaso con un polinucleótido celular con código de barras molecular, en el que el polinucleótido con código de barras del vaso es único para un solo vaso de una pluralidad de vasos; (d) extender un polinucleótido celular con código de barras molecular que está hibridado a un polinucleótido con código de barras del vaso para formar un polinucleótido celular con código de barras doble; y (e) secuenciar el polinucleótido celular con código de barras doble. En algunas realizaciones, la hibridación de (a) no se realiza mediante el emparejamiento de bases de una secuencia de origen natural en los polinucleótidos de una sola célula. En algunas realizaciones, el polinucleótido con código de barras del vaso hibridado al polinucleótido celular con código de barras molecular es un producto amplificado. En algunas realizaciones, la hibridación de (c) no se realiza mediante el emparejamiento de bases de un complemento de una secuencia de origen natural en los polinucleótidos de una sola célula. En algunas realizaciones, la hibridación de (c) es mediante el emparejamiento de bases en una región del polinucleótido de la célula individual que se extendió en (b). En algunas realizaciones, (a)-(d) se realizan en el vaso individual. En algunas realizaciones, (a)-(d) se realizan en una sola reacción.

### Breve descripción de los dibujos

Las características novedosas descritas en el presente documento se exponen con particularidad en las reivindicaciones adjuntas. Se obtendrá una mejor comprensión de las características y ventajas de las características descritas en el presente documento haciendo referencia a la siguiente descripción detallada que expone ejemplos ilustrativos, en los que se utilizan los principios de las características descritas en el presente documento, y los dibujos adjuntos de los que:

La **FIG. 1A** representa un esquema de una fase de código de barras de un procedimiento ilustrativo descrito en el presente documento. El esquema representa un procedimiento para amplificar y codificar con código de barras dos o más polinucleótidos, tal como las secuencias de Ig (por ejemplo, ARNm de  $V_H$  y  $V_L$ ) y de TCR (por ejemplo, ARNm de  $V\alpha/V\beta$  y  $V\gamma/V\delta$ ) variables emparejadas, tal como para la preparación del bancos y la secuenciación inmunológica. Código de barras del vaso (DB); Código de barras molecular (MB). (Parte superior) Una sola gotita (de una pluralidad de gotitas) de una emulsión que contiene una sola célula y otros componentes de reacción (por ejemplo, enzimas, tampones, oligonucleótidos). (Medio) Lisis celular y transcripción inversa de ARN de células lisadas. (Parte inferior) Marcaje de código de barras molecular (MB) de moléculas individuales durante la transcripción inversa.

La **FIG. 1A** representa un esquema de una fase de amplificación de un procedimiento ilustrativo descrito en el presente documento. El esquema representa un procedimiento para amplificar y codificar con código de barras dos o más polinucleótidos, tal como las secuencias de Ig (por ejemplo, ARNm de  $V_H$  y  $V_L$ ) y de TCR (por ejemplo, ARNm de  $V\alpha/V\beta$  y  $V\gamma/V\delta$ ) variables emparejadas, tal como para la preparación del bancos y la secuenciación inmunológica. (Parte superior) La amplificación independiente de los códigos de barras de los vasos (VB) genera una pluralidad de copias de VB idénticos en cada gotita. Las moléculas de ADNc-MB se marcan simultáneamente con las VB durante las fases de hibridación y extensión de la amplificación. (Medio) Amplificación simultánea de moléculas de ADNc con código de barras doble durante el ciclo de amplificación. (Parte inferior) Productos de amplificación recuperados de gotitas de la emulsión.

La **FIG. 2** ilustra un esquema que muestra que la identidad de secuencia del código de barras del vaso (DB) permite la identificación de la célula de origen para cada ARN.

La **FIG. 3** ilustra un esquema que muestra que si el mismo código de barras molecular (MB) está unido a las mismas secuencias de ARN idénticas, esta especie de ARN-MB-DB es probablemente el resultado de la duplicación de la PCR. Cuando dos MB diferentes están unidos a las mismas secuencias de ARN idénticas, entonces estos ARN1-MB1-DB y ARN1-MB2-DB son la observación real de dos moléculas de ARN independientes de origen y no de duplicación de PCR.

La **FIG. 4A** representa un esquema de un procedimiento ilustrativo descrito en el presente documento. El esquema representa un procedimiento para amplificar y codificar con código de barras las secuencias de Ig (por ejemplo, secuencias de  $V_H$  y  $V_L$ ) y de TCR (por ejemplo, secuencias de  $V\alpha/V\beta$  y  $V\gamma/V\delta$ ) variables emparejadas, para la preparación del bancos y la secuenciación inmunológica. Código de barras del vaso (DB); Código de barras molecular (MB). Cada una de las reacciones mostradas se puede realizar en una sola fase de emulsión y se muestran por separado para facilitar la representación.

La **FIG. 4B** representa un esquema de un procedimiento ilustrativo descrito en el presente documento. El esquema representa un procedimiento para amplificar y codificar con código de barras el ARNm del anticuerpo  $V_H$  y  $V_L$  para

la preparación de bancos y la secuenciación inmunológica.

La **FIG. 4C** representa un esquema de un procedimiento ilustrativo descrito en el presente documento. El esquema ilustra un procedimiento para amplificar y codificar con código de barras el ARNm del anticuerpo  $V_H$  y  $V_L$  para la preparación de bancos y la secuenciación inmunológica.

La **FIG. 4D** representa un esquema de un procedimiento ilustrativo descrito en el presente documento. El esquema ilustra un procedimiento para amplificar y codificar con código de barras las secuencias de Ig (por ejemplo, secuencias de  $V_H$  y  $V_L$ ) y de TCR (por ejemplo, secuencias de  $V\alpha/V\beta$  y  $V\gamma/V\delta$ ) variables emparejadas para la preparación de bancos y la secuenciación inmunológica. Una etapa opcional de amplificación de ADNc antes de marcar los ADNc con un código de barras del vaso (DB).

La **FIG. 5** ilustra un esquema que muestra que la identidad de secuencia del código de barras del vaso (DB) permite la identificación de la célula de origen para cada ARN. Los procedimientos se pueden usar con una emulsión que contenga una pluralidad de gotitas, cada una de las cuales contiene una sola célula para producir ADNc con códigos de barras dobles en una sola reacción.

La **FIG. 6** ilustra un esquema que muestra que si el mismo código de barras molecular (MB) está unido a las mismas secuencias de ARN idénticas, esta especie de ARN-MB-DB es probablemente el resultado de la duplicación de la PCR. Cuando dos MB diferentes están unidos a las mismas secuencias de ARN idénticas, entonces estos ARN1-MB1-DB y ARN1-MB2-DB son la observación real de dos moléculas de ARN independientes de origen y no de duplicación de PCR.

La **FIG. 7A** representa un esquema de un procedimiento ilustrativo descrito en el presente documento. El esquema representa una leyenda para los términos de las reivindicaciones.

La **FIG. 7B** representa un esquema de un procedimiento ilustrativo descrito en el presente documento. El esquema representa un procedimiento para amplificar y codificar con código de barras dos o más polinucleótidos, tal como las secuencias de Ig (por ejemplo, ARNm de  $V_H$  y  $V_L$ ) y de TCR (por ejemplo, ARNm de  $V\alpha/V\beta$  y  $V\gamma/V\delta$ ) variables emparejadas. tal como para la preparación de los bancos y la secuenciación inmunológica.

## DESCRIPCIÓN DETALLADA ADICIONAL

A continuación, se describen varios aspectos con referencia a las aplicaciones de ejemplo para ilustración. Debe entenderse que se establecen numerosos detalles, relaciones y procedimientos específicos para proporcionar una comprensión completa de las características descritas en el presente documento.

La terminología usada en el presente documento solo tiene el fin de describir casos particulares, y no pretende ser limitante. Como se usan en el presente documento, las formas en singular "un", "una", "el" y "la" también pretenden incluir las formas en plural, a menos que el contexto indique claramente otra cosa. Además, en la medida en que los términos "que incluye/n", "incluye", "que tiene", "tiene", "con" o variantes de los mismos se usan bien en la descripción detallada y/o en las reivindicaciones, dichos términos pretenden ser inclusivos de manera similar al término "que comprende".

El término "aproximadamente" o "de manera aproximada" puede determinar dentro de un intervalo de error aceptable para el valor particular determinado por un experto en la materia, que dependerá en parte de cómo se mide o determina el valor, es decir, de las limitaciones del sistema de medida. Por ejemplo, "aproximadamente" puede significar más/menos 1 o más de 1 desviación típica, por la práctica en la técnica.

Como alternativa, "aproximadamente" puede significar un intervalo de hasta el 20 %, hasta el 10 %, hasta el 5 % o hasta el 1 % de un valor dado. Como alternativa, particularmente con respecto a los sistemas o procesos biológicos, el término puede significar dentro de un orden de magnitud, entre 5 veces y más preferentemente entre 2 veces, un valor. Cuando se describan valores particulares en la solicitud y reivindicaciones, a menos que se indique lo contrario, se debe asumir el término "aproximadamente" dentro de un intervalo de error aceptable para el valor particular.

Tanto los pares de cadenas de receptores de linfocitos T como los pares de cadenas de anticuerpo de inmunoglobulina son tipos de receptores inmunitarios y están relacionados evolutivamente. Es un objeto de la invención generar bancos de polinucleótidos para la secuenciación y el diagnóstico de alto rendimiento. También es un objeto de la invención desarrollar paneles de bancos derivados de ser humano para el descubrimiento de anticuerpos y/o TCR de pacientes o cohortes con atributos comunes específicos. El material de partida puede ser sangre periférica o proceder de una biopsia de tejido, a partir de la que las células inmunitarias se aíslan o se subdividen globalmente para vírgenes, De memoria y ASC si se desea. La invención desvelada se puede aplicar a múltiples tipos diferentes de secuencias variables emparejadas, por ejemplo, pares de cadenas de receptores de linfocitos T y pares de cadenas de anticuerpos de inmunoglobulina.

Las células aisladas, tales como las células inmunitarias, se pueden encapsular en vasos, tales como las emulsiones de agua en aceite (gotitas), de manera que se creen compartimentos individuales de picolitros que contengan una sola célula inmunitaria o menos por gotita. Se pueden procesar millones de células para cada muestra, tal como una muestra biológica de un sujeto, lo que permite un alto rendimiento en la tecnología de secuenciación de una sola célula. El uso de un soporte sólido, tal como una perla, puede evitarse usando los procedimientos descritos en el presente documento. La necesidad de generar poblaciones separadas de vasos también puede evitarse usando los

procedimientos descritos en el presente documento. Por ejemplo, se pueden generar bancos de secuencias en una misma reacción o en una sola reacción, o en una sola pluralidad o población de vasos. Los polinucleótidos complementarios a los polinucleótidos celulares, tales como las cadenas de anticuerpo  $V_H$  y  $V_L$  y/o las cadenas de receptor de linfocitos T (TCR)  $V\alpha/V\beta$  y  $V\gamma/V\delta$  se introducen durante la formación de los vasos. También se puede introducir un polinucleótido que alberga un código de barras del vaso durante la formación de los vasos. Estos polinucleótidos con código de barras del vaso pueden portar códigos de barras degenerados de modo que cada polinucleótido celular que contiene un código de barras del vaso contenga un código de identidad único correspondiente al vaso en el que se encuentra. También se puede introducir una pluralidad de polinucleótidos que albergan un código de barras molecular durante la formación de los vasos. Estos polinucleótidos con código de barras molecular pueden portar códigos de barras degenerados de manera que cada molécula de polinucleótido celular que contiene un código de barras molecular contiene un código de identidad único correspondiente a una molécula de polinucleótido de una sola célula de la que proceden. Los millones de células inmunitarias individuales pueden ser lisadas dentro de la emulsión y las transcripciones celulares, tales como las transcripciones de cadenas de  $V_H$  y  $V_L$  y/o  $V\alpha/V\beta$  y/o  $V\gamma/V\delta$ , se pueden transcribir de forma inversa o copiar usando cebadores, seguido del marcaje con un código de barras del vaso y un código de barras molecular, y la amplificación por PCR de los polinucleótidos con código de barras. Cada cadena de  $V_H$  y  $V_L$  y/o  $V\alpha/V\beta$  y/o  $V\gamma/V\delta$  derivada de una sola célula inmunitaria (por ejemplo, un linfocito B o un linfocito T) puede estar prácticamente unida entre sí con la misma identidad del código de barras del vaso.

Las cadenas de  $V_H$  y  $V_L$  y/o  $V\alpha/V\beta$  y/o  $V\gamma/V\delta$  se pueden recuperar entonces a partir de los vasos y la PCR enriquecida con el fin de añadir marcadores de siguiente generación (NGS). El banco puede secuenciarse usando una plataforma de secuenciación de alto rendimiento seguida de un análisis de la diversidad del repertorio, la frecuencia de los anticuerpos, la caracterización de CDR3, el análisis de la filogenia de hipermutación somática, etc. Se puede generar una base de datos de pares correctamente emparejados de  $V_H$  y  $V_L$  y/o  $V\alpha/V\beta$  y/o  $V\gamma/V\delta$  mediante la deconvolución de las secuencias con códigos de barras moleculares y de vasos. Debido a que cada célula inmunitaria individual se aísla en su respectivo vaso, por cada código de barras observado dos veces en el vaso, las transcripciones secuenciadas se originaron a partir de las mismas gotitas de emulsión y, por lo tanto, de una sola célula individual. Por cada código de barras molecular diferente observado, para secuencias que contienen el mismo código de barras del vaso, las transcripciones secuenciadas se originaron a partir de una molécula de transcripción diferente de una sola célula. Por cada código de barras molecular igual observado, para secuencias que contienen el mismo código de barras del vaso, las transcripciones secuenciadas se originaron a partir de una misma molécula de transcripción de una sola célula (por ejemplo, duplicados de PCR).

En paralelo a la secuenciación, se puede clonar un banco de cadenas de  $V_H$  y  $V_L$  y/o  $V\alpha/V\beta$  y/o  $V\gamma/V\delta$  recuperadas de los vasos en vectores de expresión de anticuerpos y transfectarlas conjuntamente para la detección mediante visualización de levaduras. La clonación de este grupo de bancos idéntico es el procedimiento preferido en comparación con la división de una muestra biológica al principio, ya que solo se capturarían algunas células inmunitarias raras en uno u otro ensayo. El banco de cadenas de  $V_H$  y  $V_L$  y/o  $V\alpha$  y  $V\beta$  y/o  $V\gamma$  y  $V\delta$  derivadas de ser humano puede expresarse independientemente del par correcto o incorrecto que coincida con los ensayos de visualización clásicos. Se puede realizar la visualización de levaduras frente a una o más dianas antigénicas para enriquecer en posibles candidatos a anticuerpo.

Los anticuerpos candidatos positivos que surgen de tecnologías de visualización, tales como una visualización de levaduras, pueden secuenciarse y consultarse frente a la base de datos de códigos de barras de pares emparejados. Cada cadena  $V_H$  y/o  $V\alpha$  y/o  $V\gamma$  mostrada en las levaduras se puede volver a emparejar con su respectiva cadena  $V_L$  o  $V\beta$  o  $V\delta$ , respectivamente, y cada cadena  $V_L$  y/o  $V\beta$  y/o  $V\delta$  mostrada en las levaduras se puede volver a emparejar con su respectiva cadena  $V_H$  o  $V\alpha$  o  $V\gamma$ , respectivamente. Estos candidatos correctamente emparejados se pueden sintetizar y expresar genéticamente en estirpes celulares de mamífero y validar funcionalmente frente a la diana de interés. Estos candidatos pueden ser anticuerpos y/o TCR completamente humanos.

Un "anticuerpo" se refiere a una inmunoglobulina (Ig) ya sea natural, o parcial o totalmente producida sintéticamente. Un "receptor de linfocitos T" ("TCR") se refiere a una molécula, ya sea natural, o parcial o totalmente producida sintéticamente, que se encuentra en la superficie de los linfocitos T (linfocitos T) que reconoce antígenos unidos a las moléculas del complejo principal de histocompatibilidad (MHC). Se incluyen polipéptidos o proteínas que tienen un dominio de unión que es un dominio de unión al antígeno o es homólogo a un dominio de unión al antígeno. La expresión incluye además "fragmentos de unión al antígeno" y otros términos intercambiables para fragmentos de unión similares, tal como se describe a continuación. Estas expresiones también contemplan anticuerpos y TCR injertados de la región determinante de la complementariedad (CDR), y otros anticuerpos humanizados y TCR (incluyendo las modificaciones de la CDR y las modificaciones de la región marco conservada). Cabe señalar que, aunque se puede hacer referencia únicamente a las cadenas de inmunoglobulina (por ejemplo, cadenas pesadas y cadenas ligeras), la invención desvelada se puede aplicar a otros múltiples tipos diferentes de secuencias emparejadas, por ejemplo, pares de cadenas de receptores de linfocitos T (cadenas de  $TCR\alpha$  y  $TCR\beta$ , y cadenas de  $TCR\gamma$  y  $TCR\delta$ , y no se limita a las inmunoglobulinas.

Los anticuerpos nativos y las inmunoglobulinas nativas son generalmente glicoproteínas heterotetraméricas de aproximadamente 150.000 Da, compuestas de dos cadenas ligeras (L) idénticas y dos cadenas pesadas (H) idénticas.

Cada cadena ligera está unida normalmente a una cadena pesada por un enlace disulfuro covalente, mientras que el número de enlaces disulfuro varía entre las cadenas pesadas de diferentes isotipos de inmunoglobulina. Cada cadena pesada y ligera tiene también enlaces disulfuro intracatenarios a distancias regulares. Cada cadena pesada tiene en un extremo un dominio variable ( $V_H$ ) seguido de varios dominios constantes (CH). Cada cadena ligera tiene un dominio variable en un extremo ( $V_L$ ) y un dominio constante (CL) en su otro extremo; el dominio constante de la cadena ligera está alineado con el primer dominio constante de la cadena pesada y el dominio variable de la cadena ligera está alineado con el dominio variable de la cadena pesada. Se cree que restos de aminoácidos particulares forman una interfaz entre los dominios variables de cadena ligera y pesada. Los anticuerpos pueden asignarse a diferentes clases en función de la secuencia de aminoácidos del dominio constante de sus cadenas pesadas, incluyendo IgA, IgD, IgE, IgG e IgM, y varios de estos pueden además dividirse en subclases (isotipos), por ejemplo, IgG1, IgG2, IgG3, IgG4, IgA e IgA2. Las cadenas pesadas (IgH) de los anticuerpos corresponden a diferentes clases de inmunoglobulinas denominadas  $\alpha$ ,  $\delta$ ,  $\epsilon$ ,  $\gamma$  y  $\mu$ , respectivamente, basándose en las secuencias de aminoácidos de sus dominios constantes. Las cadenas ligeras (IgL) de los anticuerpos de cualquier especie de vertebrado pueden asignarse a uno de dos tipos claramente distintos, denominados kappa ( $\kappa$ ) y lambda ( $\lambda$ ), basándose en las secuencias de aminoácidos de sus dominios constantes.

La capacidad de los linfocitos T para reconocer antígenos asociados con diversos cánceres u organismos infecciosos es conferida por su TCR, que se compone de una cadena alfa ( $\alpha$ ) y una cadena beta ( $\beta$ ) o gamma ( $\gamma$ ) y una cadena delta ( $\delta$ ). Las proteínas que forman estas cadenas están codificadas por ADN, que emplea un mecanismo único para generar la tremenda diversidad del TCR. Este receptor de reconocimiento inmunitario de varias subunidades se asocia con el complejo CD3 y se une a los péptidos presentados por las proteínas de clase I y II del MHC en la superficie de las células presentadoras de antígeno (APC). La unión de un TCR al péptido antigénico en la APC es un evento central en la activación de los linfocitos T, que se produce en una sinapsis inmunitaria en el punto de contacto entre el linfocito T y la APC.

Cada TCR contiene regiones determinantes de la complementariedad variables (CDR), así como regiones marco conservadas (FR) y una región constante. La secuencia de aminoácidos de los bucles de la tercera región determinante de la complementariedad (CDR3) de los dominios variables de las cadenas  $\alpha$  y  $\beta$  determina en gran medida la diversidad de la secuencia de los linfocitos T  $\alpha\beta$  que surgen de la recombinación entre la variable ( $V\beta$ ), la diversidad ( $D\beta$ ) y la unión de los segmentos de genes ( $J\beta$ ) en el locus de la cadena  $\beta$ , y entre segmentos de genes  $V\alpha$  y  $J\alpha$  análogos en el locus de la cadena  $\alpha$ , respectivamente. La existencia de dichos múltiples segmentos de genes en los locus de las cadenas  $\alpha$  y  $\beta$  de TCR permite que se codifique un gran número de secuencias CDR3 distintas. La adición y eliminación independientes de nucleótidos en las uniones  $V\beta$ - $D\beta$ ,  $D\beta$ - $J\beta$  y  $V\alpha$ - $J\alpha$  durante el proceso de reordenamiento de los genes de TCR aumenta aún más la diversidad de las secuencias CDR3. En este sentido, la inmunocompetencia se refleja en la diversidad de los TCR.

El TCR  $\gamma\delta$  se distingue del TCR  $\alpha\beta$  en que codifica un receptor que interactúa estrechamente con el sistema inmunitario innato. TCR $\gamma\delta$ , se expresa temprano en el desarrollo, tiene una distribución anatómica especializada, tiene especificidades únicas hacia patógenos y moléculas pequeñas, y tiene un amplio espectro de interacciones celulares innatas y adaptativas. Al comienzo de la ontogenia, como los subconjuntos restringidos de células TCR $\gamma\delta$  pueblan varios tejidos antes del nacimiento, se establece un patrón sesgado de la expresión de los segmentos V y J de TCR $\gamma$ . Por lo tanto, la expansión periférica extensa después de la estimulación por exposición ambiental a patógenos y moléculas tóxicas causa gran parte del repertorio diverso de TCR $\gamma$  en los tejidos adultos.

Las Ig expresadas por linfocitos B son proteínas que consisten en cuatro cadenas polipeptídicas, dos cadenas pesadas (IgH) y dos cadenas ligeras (IgL), que forman una estructura H<sub>2</sub>L<sub>2</sub>. Cada par de cadenas de IgH e IgL contiene un dominio hipervariable, que consiste en una región  $V_L$  y  $V_H$ , y un dominio constante. Las cadenas de IgH de las Ig son de varios tipos,  $\mu$ ,  $\delta$ ,  $\gamma$ ,  $\alpha$  y  $\beta$ . La diversidad de las Ig dentro de un individuo está determinada principalmente por el dominio hipervariable. Al igual que el TCR, el dominio V de las cadenas IgH se crea por la unión combinatoria de los segmentos génicos  $V_H$ ,  $D_H$  y  $J_H$ . La adición y eliminación independientes de nucleótidos en las uniones  $V_H$ - $D_H$ ,  $D_H$ - $J_H$  y  $V_H$ - $J_H$  durante el proceso de reorganización de genes de Ig aumenta aún más la diversidad de la secuencia del dominio hipervariable. En este caso, la inmunocompetencia se refleja en la diversidad de las Ig.

"Variable" con referencia a cadenas de anticuerpos, por ejemplo, cadenas pesadas y ligeras, o cadenas de TCR, por ejemplo, cadenas alfa ( $\alpha$ ) y beta o cadenas gamma ( $\gamma$ ) y delta ( $\delta$ ), se refiere a partes de las cadenas de anticuerpo o de TCR que difieren en la secuencia entre anticuerpos o TCR, y participan en la unión y especificidad de cada anticuerpo o TCR particular por su antígeno en particular. Dicha variabilidad se concentra en tres segmentos denominados regiones hipervariables de los dominios variables tanto de cadena ligera como de cadena pesada o de los dominios variables alfa y beta. Las partes más conservadas de los dominios variables se denominan "regiones marco conservadas" (FR). Los dominios variables de las cadenas pesadas y ligeras nativas comprenden cada uno cuatro FR (FR1, FR2, FR3 y FR4, respectivamente), conectados por tres regiones hipervariables. Las regiones hipervariables en cada cadena se mantienen juntas en proximidad estrecha por las FR y, con las regiones hipervariables de la otra cadena, contribuyen a la formación del sitio de unión a antígeno de los anticuerpos (véase Kabat *et al.*, "Sequences of Proteins of Immunological Interest", 5ª Ed. Public Health Service, National Institutes of Health, Bethesda, Md. (1991), páginas 647-669). Los dominios constantes no están implicados directamente en la unión de un anticuerpo o TCR con un antígeno, pero muestran diversas funciones efectoras, por ejemplo, la

participación del anticuerpo en la toxicidad celular dependiente de anticuerpos.

Una "región hipervariable" se refiere a los restos de aminoácidos de un anticuerpo o TCR que son responsables de la unión del antígeno. La región hipervariable comprende restos de aminoácidos de una "región determinante de la complementariedad" o "CDR". Los restos "marco" o "FR" son aquellos restos de dominio variable distintos de los restos de la región hipervariable según lo definido en el presente documento.

Los "fragmentos de anticuerpo" y los "fragmentos de TCR" comprenden una parte de un anticuerpo de longitud completa o TCR, en general, el dominio de unión al antígeno o dominio variable del mismo. Los ejemplos de anticuerpos y fragmentos de TCR incluyen, aunque sin limitación, fragmentos Fab, Fab', F(ab')<sub>2</sub>, Fv y scFv, anticuerpos lineales o TCR, anticuerpo monocatenario o moléculas de TCR, diacuerpos y anticuerpos multiespecíficos o TCR formados a partir de fragmentos de anticuerpos o de TCR.

Un "anticuerpo monoclonal" se refiere a una molécula de anticuerpo sintetizada por un solo clon de células inmunitarias. El modificador "monoclonal" indica la naturaleza del anticuerpo como obtenido de una población sustancialmente homogénea de anticuerpos, y no debe considerarse que requiere la producción del anticuerpo mediante cualquier procedimiento concreto. Por lo tanto, los anticuerpos monoclonales se pueden producir mediante el procedimiento de hibridoma descrito por primera vez por Kohler y Milstein, *Nature* 256:495 (1975); *Eur. J. Immunol.* 6:511 (1976), mediante técnicas de ADN recombinante, o también pueden aislarse de bancos de anticuerpos de fagos.

Un "anticuerpo policlonal" se refiere a una población de moléculas de anticuerpo sintetizadas por una población de células inmunitarias.

Un "Fv monocatenario" o "scFv" se refiere a fragmentos de anticuerpos o de TCR que comprenden los dominios de cadena pesada variable (V<sub>H</sub>) y de cadena ligera variable (V<sub>L</sub>) de un anticuerpo o los dominios de cadena alfa o gamma variable (V<sub>α</sub> o V<sub>γ</sub>) y los dominios de cadena beta o delta variables (V<sub>β</sub> o V<sub>δ</sub>) de un TCR, en donde estos dominios están presentes en una sola cadena de polipéptido. En general, el polipéptido Fv comprende además un enlazador polipeptídico entre los dominios V<sub>H</sub> y V<sub>L</sub> o los dominios V<sub>α</sub> y V<sub>β</sub> o los dominios V<sub>γ</sub> y V<sub>δ</sub> que permiten que el sFv forme la estructura deseada para la unión al antígeno.

Un "diacuerpo" se refiere a pequeños fragmentos de anticuerpo y/o de TCR con dos sitios de unión al antígeno, fragmentos que comprenden un V<sub>H</sub> conectado a un V<sub>L</sub> en la misma cadena polipeptídica (V<sub>H</sub>-V<sub>L</sub>) o un V<sub>α</sub> conectado a un V<sub>β</sub> en la misma cadena polipeptídica (V<sub>α</sub>-V<sub>β</sub>) o un V<sub>γ</sub> conectado a un V<sub>δ</sub> en la misma cadena polipeptídica (V<sub>γ</sub>-V<sub>δ</sub>). Al usar un enlazador que es demasiado corto para permitir el emparejamiento entre los dos dominios de la misma cadena, los dominios se obligan a emparejarse con los dominios complementarios de otra cadena y crean dos sitios de unión al antígeno. Los ejemplos de diacuerpos se describen más detalladamente en, por ejemplo, los documentos EP404097 y WO93111161.

Un "anticuerpo biespecífico" o "TCR biespecífico" se refiere a un anticuerpo o TCR que muestra especificidades hacia dos tipos diferentes de antígenos. Los términos como se usan en el presente documento específicamente incluyen, sin limitación, anticuerpos y TCR que muestran especificidad de unión hacia un antígeno diana y hacia otra diana que facilita el suministro a un determinado tejido. De manera similar, los anticuerpos y los TCR multiespecíficos tienen dos o más especificidades de unión.

Un "anticuerpo lineal" o "TCR" lineal se refiere a un par de segmentos Fd en tándem (por ejemplo, V<sub>H</sub>-CH1-V<sub>H</sub>-CH1 o V<sub>α</sub>-Ca<sub>1</sub>-V<sub>α</sub>-Ca<sub>1</sub>) que forman un par de regiones de unión al antígeno. Los anticuerpos lineales y los TCR pueden ser biespecíficos o mono-específicos, por ejemplo, según lo descrito por Zapata *et al.*, *Protein Eng.* 8(10):1057-1062 (1995).

Un "dominio de unión al antígeno" se refiere a uno o más fragmentos de un anticuerpo o TCR que conservan la capacidad de unirse específicamente a un antígeno. Los ejemplos no limitantes de fragmentos de anticuerpos incluidos dentro de dichos términos incluyen, aunque sin limitación, (i) un fragmento Fab, un fragmento monovalente que consiste en los dominios V<sub>L</sub>, V<sub>H</sub>, CL y C<sub>H1</sub>; (ii) un fragmento F(ab')<sub>2</sub>, un fragmento bivalente que contiene dos fragmentos Fab unidos por un enlace disulfuro en la región bisagra; (iii) un fragmento Fd que consiste en los dominios V<sub>H</sub> y CH1; (iv) un fragmento Fv que contiene los dominios V<sub>L</sub> y V<sub>H</sub> de un solo brazo de un anticuerpo, (v) un fragmento de dAb (Ward *et al.*, (1989) *Nature* 341:544 546), que contiene un dominio V<sub>H</sub>; y (vi) una CDR aislada. Además, en esta definición se incluyen los anticuerpos que comprenden una sola cadena pesada y una sola cadena ligera o TCR con una sola cadena alfa o una sola cadena beta.

Los restos "F(ab')<sub>2</sub>" y "Fab" se pueden producir tratando una Ig con una proteasa tal como la pepsina y la papaína, e incluyen fragmentos de anticuerpos generados mediante la digestión de la inmunoglobulina cerca de los enlaces disulfuro que existen entre las regiones bisagra de cada una de las dos cadenas pesadas. Por ejemplo, la papaína escinde la IgG cadena arriba de los enlaces disulfuro existentes entre las regiones bisagra de cada una de las dos cadenas pesadas para generar dos fragmentos de anticuerpo homólogos en los que una cadena ligera compuesta de V<sub>L</sub> y C<sub>L</sub>, y un fragmento de cadena pesada compuesto de V<sub>H</sub> y C<sub>H1</sub> (La región γ1 de la región constante de la cadena pesada) están conectados en sus regiones C terminales a través de un enlace disulfuro. Cada uno de estos dos

fragmentos de anticuerpo homólogos se denomina "Fab". La pepsina también escinde IgG cadena abajo de los enlaces disulfuro existentes entre las regiones bisagra en cada una de las dos cadenas pesadas para generar un fragmento de anticuerpo ligeramente más grande que el fragmento en el que los dos "Fab" mencionados anteriormente están conectados en la región bisagra. Este fragmento de anticuerpo se denomina  $F(ab')_2$ . El fragmento Fab también contiene el dominio constante de la cadena ligera y el primer dominio constante ( $C_{H1}$ ) de la cadena pesada. Los fragmentos "Fab" difieren de los fragmentos Fab en la adición de unos pocos restos en el extremo carboxilo del dominio de cadena pesada  $C_{H1}$  incluyendo una o más cisteínas de la región bisagra del anticuerpo. Fab'-SH es la designación en el presente documento para Fab' en los que el resto (o los restos) de cisteína de los dominios constantes portan un grupo tiol libre. Los fragmentos de anticuerpos  $F(ab')_2$  originalmente se producen como pares de fragmentos Fab' que tienen cisteínas bisagra entre ellos.

"Fv" se refiere a un fragmento de anticuerpo o de TCR que contiene un sitio completo de reconocimiento del antígeno y de unión al antígeno. Esta región consiste en un dímero de un dominio variable de una cadena pesada y una cadena ligera, o una cadena de  $TCR\alpha$  y una cadena de  $TCR\beta$  o una cadena de  $TCR\gamma$  y una cadena de  $TCR\delta$  en una asociación estrecha y no covalente. Es en esta configuración en la que las tres CDR de cada dominio variable interactúan para definir un sitio de unión al antígeno en la superficie del dímero  $V_H-V_L$  o dímero  $V\alpha/V\beta$  o dímero  $V\gamma-V\delta$ . En conjunto, una combinación de una o más de las CDR de cada una de las cadenas  $V_H$  y  $V_L$  o cadenas  $V\alpha-V\beta$  o cadenas  $V\gamma-V\delta$  confiere especificidad de unión al antígeno al anticuerpo o TCR. Por ejemplo, se entendería que, por ejemplo, la CDRH3 y CDRL3 podrían bastar para conferir especificidad de unión al antígeno a un anticuerpo o TCR cuando se transfieren a cadenas  $V_H$  y  $V_L$  o cadenas  $V\alpha$  y  $V\beta$  o cadenas  $V\gamma-V\delta$  de un anticuerpo, TCR o su fragmento de unión al antígeno seleccionado del receptor, y esta combinación de CDR pueden analizarse para determinar su unión, afinidad, etc. Incluso un único dominio variable (o la mitad de un Fv que comprende solo tres CDR específicas para un antígeno) tiene la capacidad de reconocer y unirse al antígeno, aunque es probable que tenga una afinidad inferior que cuando se combina con un segundo dominio variable. Además, aunque los dos dominios de un fragmento Fv ( $V_L$  y  $V_H$  o  $V\alpha$  y  $V\beta$  o  $V\gamma$  y  $V\delta$ ) están codificados por genes separados, se pueden unir usando procedimientos recombinantes mediante un enlazador sintético que les permita formarse como una sola cadena de proteína en la que las regiones de las cadenas  $V_L$  y  $V_H$  o  $V\alpha$  y  $V\beta$  o  $V\gamma$  y  $V\delta$  se emparejan para formar moléculas monovalentes (conocidas como Fv monocatenario (scFv); Bird *et al.* (1988) *Science* 242:423-426; Huston *et al.* (1988) *Proc. Natl. Acad. Sci. EE.UU.* 85:5879-5883; y Osbourn *et al.* (1998) *Nat. Biotechnol.* 16:778). Dichos scFv también pretenden estar abarcados dentro de la expresión "parte de unión al antígeno" de un anticuerpo. Cualquier secuencia de  $V_H$  y  $V_L$  del scFv específico se puede enlazar a un ADNc de la región Fc o secuencias genómicas, para generar vectores de expresión que codifiquen moléculas de Ig (por ejemplo, IgG) completas u otros isotipos.  $V_H$  y  $V_L$  también se pueden usar en la generación de Fab, Fv u otros fragmentos de Ig usando la química de proteínas o la tecnología de ADN recombinante.

Los polipéptidos de unión al antígeno también incluyen dímeros de cadena pesada tales como, por ejemplo, anticuerpos de camélidos y tiburones. Los anticuerpos de camélidos y tiburones comprenden un par homodimérico de dos cadenas de dominios de tipo V y tipo C (ninguno tiene una cadena ligera). Dado que la región  $V_H$  de una IgG de dímero de cadena pesada de un camélido no tiene que realizar interacciones hidrófobas con una cadena ligera, en los camélidos, la región de la cadena pesada que normalmente está en contacto con una cadena ligera se cambia a restos de aminoácidos hidrófilos. Los dominios  $V_H$  de las IgG de dímero de cadena pesada se denominan dominios  $V_{HH}$ . Los Ig-NAR de tiburón comprenden un homodímero de un dominio variable (denominado dominio V-NAR) y cinco dominios constantes similares a C (dominios C-NAR). En los camélidos, la diversidad del repertorio de anticuerpos está determinada por las CDR 1, 2 y 3 de las regiones  $V_H$  o  $V_{HH}$ . La CDR3 de la región VHH del camello se caracteriza por su longitud relativamente larga, con un promedio de 16 aminoácidos (Muyldermans *et al.*, 1994, *Protein Engineering* 7(9): 1129).

Las formas "humanizadas" de anticuerpos o TCR no humanos (por ejemplo, murinos) incluyen anticuerpos o TCR quiméricos que contienen una secuencia mínima derivada de una Ig no humana o un TCR no humano. En su mayoría, los anticuerpos o TCR humanizados son Ig o TCR humanos (anticuerpo o TCR receptor) en los que una o más de las CDR del receptor se reemplazan por CDR de un anticuerpo o TCR de especie no humana (anticuerpo o TCR donante) tal como ratón, rata, conejo o primate no humano que tienen la especificidad, afinidad y función de unión deseadas. En algunos casos, uno o más restos de aminoácidos de FR de la Ig humana o TCR humano se reemplazan por los correspondientes restos de aminoácidos no humanos. Además, los anticuerpos o TCR humanizados pueden contener restos que no se encuentran en el anticuerpo o TCR receptor, o en el anticuerpo o TCR donante. Estas modificaciones se pueden realizar para perfeccionar el rendimiento del anticuerpo o TCR, si es necesario. Un anticuerpo o TCR humanizado puede comprender esencialmente todos de al menos uno y, en algunos casos, dos dominios variables, en los que todas o esencialmente todas las regiones hipervariables corresponden a las de una inmunoglobulina no humana o un TCR no humano, y todos, o esencialmente todos, los FR son los de una secuencia de inmunoglobulina o TCR humana. El anticuerpo o TCR humanizado también puede incluir opcionalmente al menos una parte de una región constante de inmunoglobulina o TCR (Fc), normalmente de una inmunoglobulina humana o de un TCR humano. Véase, por ejemplo, Jones *et al.*, *Nature* 321: 522-525 (1986); Reichmann *et al.*, *Nature* 332: 323-329 (1988); y Presta, *Curr. Op. Struct. Biol.* 2: 593-596 (1992).

Una "secuencia de línea germinal" se refiere a una secuencia genética de la línea germinal (los gametos haploides y las células diploides a partir de las que se forman). El ADN de la línea germinal contiene múltiples segmentos génicos que codifican una sola cadena pesada o ligera de Ig, o una sola cadena de  $TCR\alpha$  o  $TCR\beta$ , o una sola cadena de  $TCR\gamma$

o TCR $\delta$ . Estos segmentos génicos se transportan en las células germinales, pero no se pueden transcribir y traducir hasta que se organizan en genes funcionales. Durante la diferenciación de linfocitos B y linfocitos T en la médula ósea, estos segmentos génicos son redistribuidos aleatoriamente por un sistema genético dinámico capaz de generar más de 10<sup>8</sup> especificidades. La mayoría de estos segmentos genéticos están publicados y recogidos por la base de datos de líneas germinales.

"Afinidad" se refiere a la constante de equilibrio para la unión reversible de dos agentes y se expresa como K<sub>D</sub>. La afinidad de una proteína de unión a un ligando, tal como la afinidad de un anticuerpo por un epítipo, puede ser, por ejemplo, de aproximadamente 100 nanomolar (nM) a aproximadamente 0,1 nM, de aproximadamente 100 nM a aproximadamente 1 picomolar (pM) o de aproximadamente 100 nM a aproximadamente 1 femtomolar (fM). El término "avidez" se refiere a la resistencia de un complejo de dos o más agentes a la disociación después de la dilución.

Un "epítipo" se refiere a la parte de un antígeno u otra macromolécula capaz de formar una interacción de unión con el bolsillo de unión de la región variable de un anticuerpo o TCR. Dichas interacciones de unión pueden manifestarse como un contacto intermolecular con uno o más restos de aminoácidos de una o más CDR. La unión al antígeno puede implicar, por ejemplo, una CDR3, un par de CDR3 o, en algunos casos, interacciones de hasta las seis CDR de las cadenas V<sub>H</sub> y V<sub>L</sub>. Un epítipo puede ser una secuencia peptídica lineal (es decir, "continua") o puede estar compuesto de secuencias de aminoácidos no contiguas (es decir, "conformacional" o "discontinua"). Un anticuerpo o TCR puede reconocer una o más secuencias de aminoácidos; por lo tanto, un epítipo puede definir más de una secuencia de aminoácidos distinta. Los epítipos reconocidos por los anticuerpos y TCR se pueden determinar mediante técnicas de cartografía de péptidos y análisis de secuencias bien conocidas por los expertos en la materia. Las interacciones de unión se manifiestan como contactos intermoleculares con uno o más restos de aminoácidos de una CDR.

"Específico" se refiere a una situación en la que un anticuerpo o TCR no mostrará ninguna unión significativa a moléculas distintas del antígeno que contengan el epítipo reconocido por el anticuerpo o TCR. El término también es aplicable cuando, por ejemplo, un dominio de unión al antígeno es específico de un determinado epítipo que es portado por una serie de antígenos, en cuyo caso el anticuerpo, TCR, o fragmento de unión al antígeno de los mismos, seleccionado que porta el dominio de unión al antígeno será capaz de unirse a los diversos antígenos portadores del epítipo. Las expresiones "se une preferentemente" o "se une específicamente" significan que los anticuerpos, TCR o fragmentos de los mismos se unen a un epítipo con mayor afinidad de la que se une a secuencias de aminoácidos no relacionadas y, si son reactivos de forma cruzada a otros polipéptidos que contienen el epítipo, son no tóxicos a los niveles en los que están formulados para la administración en uso humano. En un aspecto, dicha afinidad es al menos 1 vez superior, al menos 2 veces superior, al menos 3 veces superior, al menos 4 veces superior, al menos 5 veces superior, al menos 6 veces superior, al menos 7 veces superior, al menos 8 veces superior, al menos 9 veces superior, al menos 10 veces superior, al menos 20 veces superior, al menos 30 veces superior, al menos 40 veces superior, al menos 50 veces superior, al menos 60 veces superior, al menos 70 veces superior, al menos 80 veces superior, al menos 90 veces superior, al menos 100 veces superior o al menos 1.000 veces superior a la afinidad del anticuerpo, TCR o fragmento del mismo para secuencias de aminoácidos no relacionadas. El término "unión" se refiere a una asociación directa entre dos moléculas, debida a, por ejemplo, interacciones covalentes, electrostáticas, hidrófobas e iónicas y/o de enlaces de hidrógeno en condiciones fisiológicas, e incluye interacciones tales como puentes salinos y puentes de agua, así como cualquier otro medio de unión convencional.

"Farmacéuticamente aceptable" se refiere a entidades moleculares y composiciones que son fisiológicamente tolerables y normalmente no producen una reacción alérgica o adversa similar, tal como molestias gástricas, mareos y similares, cuando se administran a un ser humano.

Una "dosis unitaria" cuando se usa en referencia a una composición terapéutica se refiere a unidades físicamente individuales como dosificación unitaria para seres humanos, conteniendo cada unidad una cantidad predeterminada de material activo calculada para producir el efecto terapéutico deseado en asociación con el diluyente requerido; es decir, excipiente o vehículo.

Un "material de acondicionamiento" se refiere a una estructura física que alberga los componentes del kit. El material de acondicionamiento puede mantener los componentes de forma estéril y puede estar hecho de material comúnmente usado para dichos fines (por ejemplo, papel, fibra corrugada, vidrio, plástico, folio, ampollas, etc.). La etiqueta o el prospecto del envase pueden incluir instrucciones escritas apropiadas. Los kits, por lo tanto, pueden incluir además etiquetas o instrucciones para usar los componentes del kit en cualquier procedimiento de la invención. Un kit puede incluir un compuesto en un envase o dispensador junto con instrucciones para administrar el compuesto en un procedimiento descrito en el presente documento.

"Prevención" se refiere a la profilaxis, prevención de la aparición de síntomas, prevención de la progresión de una enfermedad o de un trastorno asociado con niveles excesivos de proteína o correlacionado con la actividad de la proteína.

"Inhibición", "tratamiento" y "tratar" se usan indistintamente, y se refieren, por ejemplo, a la estasis de los síntomas, a la prolongación de la supervivencia, a la mejora parcial o completa de los síntomas y a la erradicación parcial o total de una afección, de una enfermedad o de un trastorno asociado con niveles excesivos de proteína o correlacionado con la actividad de la proteína. Por ejemplo, el tratamiento del cáncer incluye, pero sin limitación, la estasis, la

eliminación parcial o total de un tumor o crecimiento canceroso. El tratamiento o la eliminación parcial incluye, por ejemplo, una reducción en el crecimiento o el tamaño y/o el volumen del tumor, tal como de aproximadamente 2 veces, aproximadamente 3 veces, aproximadamente 4 veces, aproximadamente 5 veces, aproximadamente 10 veces, aproximadamente 20 veces, aproximadamente 50 veces, o cualquier reducción en el medio. Asimismo, el tratamiento o la eliminación parcial puede incluir un porcentaje de reducción del crecimiento o tamaño del tumor y/o volumen de aproximadamente el 1 %, 2 %, 3 %, 4 %, 5 %, 10 %, 20 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 %, 95 % o cualquier porcentaje de reducción de entre medias.

Un "anticuerpo neutralizante" o "TCR neutralizante" se refiere a cualquier anticuerpo o TCR que inhibe la replicación de un patógeno, tal como un virus o una bacteria, independientemente del mecanismo mediante el que se logra la neutralización.

Un "repertorio de anticuerpos" o "repertorio de TCR" se refiere a una colección de anticuerpos, TCR o fragmentos de los mismos. Un repertorio de anticuerpos se puede usar, por ejemplo, para seleccionar un determinado anticuerpo o seleccionar una determinada propiedad, tal como la capacidad de unión, la especificidad de unión, la capacidad de transporte gastrointestinal, la estabilidad, la afinidad y similares. El término incluye específicamente bancos de anticuerpos y TCR, incluyendo todas las formas de bancos combinatorias, tales como, por ejemplo, bancos de presentación de anticuerpos en fagos, incluyendo, sin limitación, bancos de presentación de anticuerpos en fagos de Fv monocatenario (scFv) y de Fab de cualquier fuente, incluyendo bancos vírgenes, sintéticos y semisintéticos.

Una "molécula de ácido nucleico diana", "molécula diana", "polinucleótido diana", "molécula de polinucleótido diana" se refiere a cualquier ácido nucleico de interés.

Una reacción en cadena de la polimerasa (PCR) se refiere a una reacción de amplificación *in vitro* de secuencias de polinucleótido mediante la extensión de cebador simultánea de cadenas complementarias de un polinucleótido bicatenario. Las reacciones de PCR producen copias de un polinucleótido de molde flanqueado por sitios de unión a cebadores. El resultado, con dos cebadores, es un aumento exponencial del número de copias de polinucleótidos de molde de ambas cadenas con cada ciclo, porque, con cada ciclo, se replican ambas cadenas. El dúplex de polinucleótido tiene terminaciones correspondientes a los extremos de los cebadores usados. La PCR puede comprender una o más repeticiones de desnaturalización de un polinucleótido de molde, Hibridación de cebadores a sitios de unión a cebadores y extensión de los cebadores mediante una ADN o ARN polimerasa en presencia de nucleótidos. Las temperaturas, las duraciones en cada etapa y las tasas de cambio entre las etapas particulares dependen de muchos factores bien conocidos por los expertos en la materia. (McPherson *et al.*, IRL Press, Oxford (1991 y 1995)). Por ejemplo, en una PCR convencional en la que se usa ADN polimerasa de Taq, se puede desnaturalizar un polinucleótido de molde bicatenario a una temperatura >90 °C, los cebadores se pueden hibridar a una temperatura en el intervalo de 50-75 °C, y los cebadores se pueden extender a una temperatura en el intervalo de 72-78 °C. En algunas realizaciones, la PCR comprende PCR de transcripción inversa (RT-PCR), PCR en tiempo real, PCR anidada, PCR cuantitativa, PCR multiplexada o similares. En algunas realizaciones, la PCR no comprende RT-PCR. (patentes de EE.UU. n.º 5.168.038, 5.210.015, 6.174.670, 6.569.627, y 5.925.517; Mackay *et al.*, *Nucleic Acids Research*, 30: 1292-1305 (2002)). La RT-PCR comprende una reacción de PCR precedida por una reacción de transcripción inversa y se amplifica un ADNc resultante, La PCR anidada comprende una PCR de dos etapas en la que un amplicón de una primera reacción de PCR en la que se usa un primer conjunto de cebadores se convierte en la muestra para una segunda reacción de PCR en la que se usa un segundo conjunto de cebadores, al menos uno de los cuales se une a una ubicación interior de un amplicón de una primera reacción de PCR. La PCR multiplexada comprende una reacción de PCR, en la que una pluralidad de secuencias de polinucleótidos se somete a PCR en la misma mezcla de reacción simultáneamente. Los volúmenes de reacción de PCR pueden estar en cualquier lugar entre 0,2 µl y 1.000 µl. La PCR cuantitativa comprende una reacción de PCR diseñada para medir una cantidad, abundancia o concentración absoluta o relativa de una o más secuencias de una muestra. Las mediciones cuantitativas pueden incluir la comparación de una o más secuencias de referencia o patrones con una secuencia polinucleotídica de interés. (Freeman *et al.*, *Biotechniques*, 26: 112-126 (1999); Becker-Andre *et al.*, *Nucleic Acids Research*, 17: 9437-9447 (1989); Zimmerman *et al.*, *Biotechniques*, 21: 268-279 (1996); Diviacco *et al.*, *Gene*, 122: 3013-3020 (1992); Becker-Andre *et al.*, *Nucleic Acids Research*, 17: 9437-9446 (1989)).

En otras realizaciones, los procedimientos, los kits y las composiciones desvelados en el presente documento pueden comprender un soporte. En algunas realizaciones, los procedimientos, los kits y las composiciones desvelados en el presente documento no comprenden un soporte. Por lo general, un soporte sólido comprende uno o más materiales que comprenden una o más superficies rígidas o semirrígidas. En algunas realizaciones, el soporte es un soporte no sólido. El soporte o sustrato puede comprender una membrana, un papel, un plástico, una superficie recubierta, una superficie plana, un vidrio, un portaobjetos, un chip o cualquier combinación de los mismos. En algunas realizaciones, una o más superficies de un soporte son esencialmente planas, aunque, en algunas realizaciones, puede ser deseable separar físicamente las regiones de síntesis para diferentes compuestos con, por ejemplo, pocillos, regiones elevadas, pasadores, zanjas grabadas o similares. En algunas realizaciones, los soportes sólidos comprenden perlas, resinas, geles, microesferas u otras configuraciones geométricas. Como alternativa, los soportes sólidos pueden comprender chips de sílice, micropartículas, nanopartículas, placas y matrices. El soporte sólido puede comprender el uso de perlas que se autoensamblan en micropocillos. Por ejemplo, el soporte sólido comprende la tecnología BeadArray de Illumina. Como alternativa, el soporte sólido comprende la tecnología Bead Array de Abbott Molecular y el sistema FlexiPlex™

de Applied Microarray. En otros ejemplos, el soporte sólido es una placa. Los ejemplos de placas incluyen, aunque sin limitación, placas de múltiples matrices MSD, placas MSD Multi-Spot®, microplaca, microplaca ProteOn, AlphaPlate, placa DELFIA, IsoPlate y LumaPlate. En algunas realizaciones, un soporte puede comprender una pluralidad de perlas. En algunas realizaciones, un soporte puede comprender una matriz. En algunas realizaciones, un soporte puede comprender un portaobjetos de vidrio. Los procedimientos, sustratos y técnicas aplicables a polímeros (patentes de EE.UU. n.º 5.744.305; 5.143.854; 5.242.974; 5.252.743; 5.324.633; 5.384.261; 5.405.783; 5.424.186; 5.451.683; 5.482.867; 5.491.074; 5.527.681; 5.550.215; 5.571.639; 5.578.832; 5.593.839; 5.599.695; 5.624.711; 5.631.734; 5.795.716; 5.831.070; 5.837.832; 5.856.101; 5.858.659; 5.936.324; 5.968.740; 5.974.164; 5.981.185; 5.981.956; 6.025.601; 6.033.860; 6.040.193; 6.090.555; 6.136.269; 6.269.846 y 6.428.752; publicación de patente de EE.UU. n.º 20090149340, 20080038559, 20050074787; y en las publicaciones PCT n.º WO 00/58516, WO 99/36760 y WO 01/58593). La unión de los polinucleótidos a un soporte puede comprender la reticulación de amina-tiol, la reticulación de maleimida, N-hidroxisuccinimida o N-hidroxisulfosuccinimida, Zenon o SiteClick. La unión de los ácidos nucleicos marcados al soporte puede comprender la unión de biotina a la pluralidad de polinucleótidos y el recubrimiento de una o más perlas con estreptavidina. En algunas realizaciones, el soporte sólido es una perla. Los ejemplos de perlas incluyen, aunque sin limitación, perlas de estreptavidina, perlas de agarosa, perlas magnéticas, Dynabeads®, microperlas MACS®, perlas conjugadas con anticuerpos (p. ej., microesfera anti-inmunoglobulina), perlas conjugadas con proteína A, perlas conjugadas con proteína G, perlas conjugadas con proteína A/G, perlas conjugadas con proteína L, perlas conjugadas con polinucleótido dT, perlas de sílice, perlas de tipo sílice, microperlas anti-biotina, microperla antifluorocromadas y perlas magnéticas terminadas en carboxi BcMag™. El diámetro de las perlas puede ser de aproximadamente 5 µm, 10 µm, 20 µm, 25 µm, 30 µm, 35 µm, 40 µm, 45 µm o 501 µm. El soporte sólido puede ser una matriz o micromatriz. El soporte sólido puede comprender regiones diferenciadas. El soporte sólido puede ser una matriz, por ejemplo, una matriz direccionable.

"Nucleótido", "nucleósido", "resto de nucleótido" y "resto de nucleósido", como se usa en el presente documento, puede significar un resto desoxirribonucleótido o ribonucleótido, u otro análogo de nucleósido similar capaz de servir como un componente de un cebador adecuado para usar en una reacción de amplificación (por ejemplo, reacción de PCR). Dichos nucleósidos y derivados de los mismos pueden usarse como componentes básicos de los cebadores descritos en el presente documento, a menos que se indique lo contrario. Nada en la presente solicitud pretende excluir la utilización de derivados de nucleósidos o bases que se hayan modificado químicamente para potenciar su estabilidad o utilidad en una reacción de amplificación, siempre que la modificación química no interfiera en su reconocimiento por una polimerasa como desoxiguanina, desoxicitosina, desoxitimidina o desoxiadenina, según sea adecuado. En algunas realizaciones, los análogos de nucleótidos pueden estabilizar la formación de híbridos. En algunas realizaciones, los análogos de nucleótidos pueden desestabilizar la formación de híbridos. En algunas realizaciones, los análogos de nucleótidos pueden potenciar la especificidad de hibridación. En algunas realizaciones, los análogos de nucleótidos pueden reducir la especificidad de hibridación.

Un "ácido nucleico", o sus equivalentes gramaticales, se refiere a un solo nucleótido o a al menos dos nucleótidos unidos covalentemente entre sí.

Un "polinucleótido" o sus equivalentes gramaticales se refiere a al menos dos nucleótidos unidos covalentemente entre sí. Un polinucleótido comprende una molécula que contiene dos o más nucleótidos. Un polinucleótido comprende una forma polimérica de nucleótidos de cualquier longitud, ya sea ribonucleótidos, desoxirribonucleótidos o ácidos nucleicos peptídicos (PNA), que comprenden bases de purina y pirimidina u otras bases naturales, modificadas química o bioquímicamente, no naturales o derivados de bases nucleotídicas. La estructura principal del polinucleótido puede comprender azúcares y grupos fosfato, o azúcar o grupos fosfato modificados o sustituidos. Un polinucleótido puede comprender nucleótidos modificados, tales como nucleótidos metilados y análogos de nucleótidos. La secuencia de nucleótidos puede estar interrumpida por componentes no nucleotídicos. Un polinucleótido puede incluir otras moléculas, tales como otro polinucleótido hibridado. Los polinucleótidos incluyen secuencias de ácido desoxirribonucleico (ADN), ácido ribonucleico (ARN) o ambos. Los ejemplos no limitantes de polinucleótidos incluyen un gen, un fragmento de gen, un exón, un intrón, ADN intergénico (incluyendo, sin limitación, ADN heterocromático), ARN mensajero (ARNm), ARN de transferencia, ARN ribosómico, ribozimas, ARN de interferencia pequeño (ARNip), ADNc, polinucleótidos recombinantes, polinucleótidos ramificados, plásmidos, vectores, ADN aislado de una secuencia, ARN aislado de una secuencia, sondas de ácido nucleico y cebadores. Los polinucleótidos se pueden aislar de fuentes naturales, recombinantes o sintetizarse artificialmente.

Un polinucleótido comprende una secuencia específica de cuatro bases de nucleótidos: adenina (A); citosina (C); guanina (G); y timina (T) (uracilo (U) para timina (T) cuando el polinucleótido es ARN). Por lo tanto, una secuencia de polinucleótido es la representación alfabética de una molécula de polinucleótido; como alternativa, el término puede aplicarse a la propia molécula de polinucleótido. Esta representación alfabética puede introducirse en bases de datos en un ordenador que tenga una unidad de procesamiento central, y se puede usar para aplicaciones bioinformáticas tales como genómica funcional, búsqueda de homología, secuencias reunidas, alineación de secuencias y determinación de secuencias de consenso.

Los polinucleótidos pueden incluir nucleótidos no convencionales, tales como análogos de nucleótidos o nucleótidos modificados. En algunas realizaciones, los nucleótidos no convencionales pueden estabilizar la formación de híbridos. En algunas realizaciones, los nucleótidos no convencionales pueden desestabilizar la formación de híbridos. En

algunas realizaciones, los nucleótidos no convencionales pueden mejorar la especificidad de hibridación. En algunas realizaciones, los nucleótidos no convencionales pueden reducir la especificidad de hibridación. Los ejemplos de modificaciones de nucleótidos no convencionales incluyen 2'-O-Me, 2'-O-alilo, 2'-O-propargilo, 2'-O-alquilo, 2'-fluoro, 2'-arabino, 2'-xilo, 2'-fluoro-arabino, fosforotioato, fosforoditioato, fosforoamidatos, 2'-amino, pirimidina sustituida con alquilo en 5, desoxiguanosina 3', pirimidina sustituida con halógeno en 5, purina sustituida con alquilo, purina sustituida con halo, nucleótidos bicíclicos, 2'MOE, moléculas de APN, moléculas de ALN, moléculas como de ALN, diaminopurina, S2T, 5-fluorouracilo, 5-bromouracilo, 5-clorouracilo, 5-yodouracilo, hipoxantina, xantina, 4-acetilcitosina, 5-(carboxihidroxi)metiluracilo, 5-carboximetilaminometil-2-tiouridina, 5-carboximetilaminometiluracilo, dihidrouracilo, beta-D-galactosilqueosina, inosina, N6-isopenteniladenina, 1-metilguanina, 1-metilinosina, 2,2-dimetilguanina, 2-metiladenina, 2-metilguanina, 3-metilcitosina, 5-metilcitosina, N6-adenina, 7-metil-guanina, 5-metilaminometiluracilo, 5-metoxiaminometil-2-tiouracilo, beta-D-mannosilqueosina, 5'-metoxicarboximetiluracilo, 5-metoxiuracilo, 2-metil-2-tiouracilo, 2-tiouracilo, 4-tiouracilo, 5-metiluracilo, metiléster de ácido uracil-5-oxiacético, ácido uracil-5-oxiacético (v), 5-metil-2-tiouracilo, 3-(3-amino-3-N-2-carboxipropil) uracilo, (acp3)w, 2,6-diaminopurina y sus derivados.

Un "sujeto", "individuo", "hospedador" o "paciente" se refiere a organismos vivos tales como mamíferos. Los ejemplos de sujetos y hospedadores incluyen, aunque sin limitación, caballos, vacas, camellos, ovejas, cerdos, cabras, perros, gatos, conejos, cobayas, ratas, ratones (por ejemplo, ratones humanizados), jerbos, primates no humanos (por ejemplo, macacos), seres humanos y similares, no mamíferos, incluyendo, por ejemplo, vertebrados no mamíferos, tales como aves (por ejemplo, pollos o patos), peces (por ejemplo, tiburones) o ranas (por ejemplo, Xenopus) e invertebrados de no mamíferos, así como las especies transgénicas de los mismos. En determinados aspectos, un sujeto se refiere a un solo organismo (por ejemplo, ser humano). En ciertos aspectos, o se proporciona un grupo de individuos que componen una pequeña cohorte que tiene un factor inmunitario común para estudiar y/o una enfermedad, y/o una cohorte de individuos sin la enfermedad (por ejemplo, control negativo/normal). Un sujeto del que se obtienen muestras puede ser afectado con una enfermedad y/o un trastorno (por ejemplo, una o más alergias, infecciones, cánceres o trastornos autoinmunitarios o similares) y puede compararse con un sujeto de control negativo que no está afectado por la enfermedad.

Un "kit" se refiere a un sistema de administración para suministrar materiales o reactivos para llevar a cabo un procedimiento desvelado en el presente documento. En algunas realizaciones, los kits incluyen sistemas que permiten el almacenamiento, transporte o suministro de reactivos de reacción (por ejemplo, sondas, enzimas, etc., en los recipientes apropiados) y/o materiales de soporte (por ejemplo, tampones, instrucciones escritas para realizar el ensayo, etc.) de un lugar a otro. Por ejemplo, los kits incluyen uno o más recintos (por ejemplo, cajas) que contienen los reactivos de reacción y/o materiales de soporte pertinentes. Dichos contenidos se pueden administrar al destinatario deseado juntos o por separado. Por ejemplo, un primer recipiente puede contener una enzima para su uso en un ensayo, mientras que un segundo recipiente contiene una pluralidad de cebadores.

Un "polipéptido" se refiere a una molécula que comprende al menos dos aminoácidos. En algunas realizaciones, el polipéptido consiste en un solo péptido. En algunas realizaciones, un polipéptido comprende dos o más péptidos. Por ejemplo, un polipéptido puede comprender al menos aproximadamente 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900 o 1.000 péptidos o aminoácidos. Los ejemplos de polipéptidos incluyen, aunque sin limitación, cadenas de aminoácidos, proteínas, péptidos, hormonas, sacáridos polipeptídicos, lípidos, glicolípidos, fosfolípidos, anticuerpos, enzimas, quinasas, receptores, factores de transcripción y ligandos.

Una "muestra" se refiere a una muestra biológica, ambiental, médica, de sujeto o paciente, o una muestra que contiene un polinucleótido, tal como un polinucleótido diana.

## 50 MUESTRAS

En los procedimientos descritos en el presente documento, se puede usar cualquier muestra biológica que contenga polinucleótidos. Por ejemplo, una muestra puede ser una muestra biológica de un sujeto que contiene ARN o ADN. Los polinucleótidos pueden extraerse de la muestra biológica, o la muestra puede someterse directamente a los procedimientos sin extracción ni purificación de los polinucleótidos. La muestra puede ser extraída o aislada de ADN o ARN. Una muestra también puede ser ARN o ADN total extraído de una muestra biológica, un banco de ADNc, ADN vírico o genómico. En una realización, los polinucleótidos se aíslan de una muestra biológica que contiene una variedad de otros componentes, tales como proteínas, lípidos y ácidos nucleicos no molde. Las moléculas de molde de ácido nucleico pueden obtenerse de cualquier material celular, obtenido de un animal, una planta, una bacteria, un hongo o cualquier otro organismo celular. En determinadas realizaciones, los polinucleótidos se obtienen de una sola célula. Los polinucleótidos se pueden obtener directamente de un organismo o de una muestra biológica obtenida de un organismo. Se puede usar cualquier muestra de tejido o fluido corporal como fuente de ácido nucleico para su uso en la invención. Los polinucleótidos también se pueden aislar de células cultivadas, tales como un cultivo celular primario o una estirpe celular. Las células o los tejidos a partir de los que se obtienen los ácidos nucleicos de molde pueden infectarse con un virus u otro patógeno intracelular.

En determinadas realizaciones, se pueden aislar células inmunitarias productoras de TCR o anticuerpos de la sangre u otras muestras biológicas de un sujeto u hospedador, tal como de un ser humano u otro animal, tal como un ser humano u otro animal que ha sido inmunizado o que padece una infección, cáncer, una afección autoinmunitaria o cualquier otra enfermedad para identificar un anticuerpo o TCR específico de patógenos, tumores y/o enfermedades de posible importancia clínica. Por ejemplo, el ser humano puede ser diagnosticado con una enfermedad, presentar síntomas de una enfermedad, no ser diagnosticado con una enfermedad o no mostrar síntomas de una enfermedad. Por ejemplo, el ser humano puede ser aquel que estuvo expuesto y/o que puede producir anticuerpos o TCR útiles contra un agente infeccioso (por ejemplo, virus, bacterias, parásitos, priones, etc.), antígeno o enfermedad. Por ejemplo, el animal puede ser aquel que estuvo expuesto y/o que puede producir anticuerpos o TCR útiles contra un agente infeccioso (por ejemplo, virus, bacterias, parásitos, priones, etc.), antígeno o enfermedad. Ciertas células inmunitarias de hospedadores inmunizados producen anticuerpos o TCR hacia uno o más antígenos diana en cuestión y/o uno o más antígenos desconocidos. En la presente invención, el conjunto de linfocitos puede enriquecerse en las células inmunitarias deseadas mediante cualquier procedimiento adecuado, tal como la selección y clasificación de las células mediante la selección de células activadas por fluorescencia (FACS), la selección de células activadas magnéticamente (MACS), la exploración u otro procedimiento de selección para generar una pluralidad de células inmunitarias a partir de una muestra, tal como un banco de células inmunitarias, antes de que se secuencien las cadenas de anticuerpos, se produzcan anticuerpos o se cree un banco de expresión. En contraste con los procedimientos de enriquecimiento de la técnica anterior, que proporcionan solo unos pocos subconjuntos de células inmunitarias que expresan diferentes anticuerpos, y por lo tanto, solo unas pocas combinaciones naturales de dominios variables, el banco de células inmunitarias de la presente invención contiene al menos 2 subconjuntos de o células inmunitarias individuales que expresan diferentes anticuerpos o TCR. Por ejemplo, el banco de células inmunitarias de la presente invención puede contener al menos 5, 10, 100, 250, 500, 750, 1.000, 2.500, 5.000, 10.000, 25.000, 50.000, 75.000, 100.000, 250.000, 500.000, 750.000, 1.000.000, 2.500.000, 5.000.000, 7.500.000 o 10.000.000 subconjuntos de células inmunitarias individuales que expresan diferentes anticuerpos o TCR. Los procedimientos de la presente invención maximizan la recuperación de células inmunitarias y proporcionan una diversidad muy alta.

En algunas realizaciones, se utilizan células inmunitarias de donantes humanos o no humanos no inmunizados. El repertorio virgen de un animal (el repertorio antes de la exposición al antígeno) proporciona al animal anticuerpos o TCR que pueden unirse con afinidad moderada ( $K_A$  de aproximadamente  $1 \times 10^{-6}$  a  $1 \times 10^{-7}$  M) a prácticamente cualquier molécula no propia. La diversidad de secuencia de los sitios de unión del anticuerpo o TCR no se codifica directamente en la línea germinal, sino que se ensambla de manera combinatoria a partir de los segmentos del gen V. Las inmunizaciones hacen que cualquier célula inmunitaria que produce una combinación de  $V_H-V_L$  o  $V\alpha-V\beta$  o  $V\gamma-V\delta$  a unirse al inmunógeno para proliferar (expansión clónica) y para segregar el anticuerpo correspondiente como se ha indicado anteriormente. Sin embargo, el uso de células de bazo y/o células inmunitarias u otros linfocitos de sangre periférica (PBL) de un sujeto no inmunizado puede proporcionar una mejor representación del posible repertorio de anticuerpos o TCR, y también permite la construcción de un banco de anticuerpos o TCR contra los linfocitos B o T posterior usando cualquier especie animal.

En algunos casos, para obtener suficiente ácido nucleico para el análisis, se extrae un volumen de sangre de al menos 0,001, 0,005, 0,01, 0,05, 0,1, 0,5, 1, 2, 3, 4, 5, 10, 20, 25, 30, 35, 40, 45 o 50 ml.

En algunos casos, el material de partida es sangre periférica. Las células de la sangre periférica pueden enriquecerse en un tipo de célula en particular (por ejemplo, células mononucleares; glóbulos rojos; células CD4+; células CD8+; células inmunitarias; linfocitos T, linfocitos NK o similares). Las células de la sangre periférica también se pueden agotar selectivamente de un tipo de célula particular (por ejemplo, células mononucleares; glóbulos rojos; células CD4+; células CD8+; células inmunitarias; linfocitos T, linfocitos NK o similares).

En algunos casos, el material de partida puede ser una muestra de tejido que comprenda un tejido sólido, con ejemplos no limitantes que incluyen cerebro, hígado, pulmón, riñón, próstata, ovario, bazo, ganglio linfático (incluyendo las amígdalas), tiroides, páncreas, corazón, músculo esquelético, intestino, laringe, esófago y estómago. En otros casos, el material de partida puede ser células que contienen ácidos nucleicos, células inmunitarias y, en particular, linfocitos B o linfocitos T. En algunos casos, el material de partida puede ser una muestra que contiene ácidos nucleicos, de cualquier organismo, del que se pueda obtener material genético. En algunos casos, una muestra es un líquido, por ejemplo, sangre, saliva, linfa u orina.

Se puede tomar una muestra de un sujeto con una afección. En algunos casos, el sujeto del que se toma una muestra puede ser un paciente, por ejemplo, un paciente con cáncer o un paciente con sospecha de cáncer. El sujeto puede ser un mamífero, por ejemplo, un ser humano, y puede ser macho o hembra. En algunos casos, la hembra se encuentra en estado de gestación. La muestra puede ser una biopsia tumoral. La biopsia puede ser realizada, por ejemplo, por un proveedor de atención sanitaria, incluido un médico, médico asistente, enfermera, veterinario, dentista, quiropráctico, paramédico, dermatólogo, oncólogo, gastroenterólogo o cirujano.

En algunos casos, los materiales que no son ácidos nucleicos pueden eliminarse del material de partida usando tratamientos enzimáticos (tales como la digestión con proteasas).

En algunos casos, la sangre puede recogerse en un aparato que contiene un quelante de magnesio que incluye, pero

sin limitación, EDTA, y se almacena a 4 °C. Opcionalmente, se puede añadir un quelante de calcio, incluyendo, pero sin limitación, EGTA. En otro caso, se añade un inhibidor de la lisis celular a la sangre, incluido, pero sin limitación, formaldehído, derivados de formaldehído, formol, glutaraldehído, derivados de glutaraldehído, un reticulante proteico, un reticulante de ácido nucleico, un reticulante de proteína y de ácido nucleico, reticulantes reactivos de amina primaria, reticulantes reactivos de sulfhidrilo, adición de sulfhidrilo o reducción de disulfuro, reticulantes reactivos de hidratos de carbono, reticulantes reactivos de carboxilo, reticulantes fotorreactivos o reticulantes escindibles.

En algunos casos, cuando el material extraído comprende ARN monocatenario, ARN bicatenario o ADN-ARN híbrido, estas moléculas pueden convertirse en ADN bicatenario usando técnicas conocidas en el campo. Por ejemplo, la transcriptasa inversa se puede emplear para sintetizar ADN a partir de moléculas de ARN. En algunos casos, la conversión de ARN en ADN puede requerir una etapa de ligadura previa, para ligar un fragmento enlazador al ARN, permitiendo así el uso de cebadores universales para iniciar la transcripción inversa. En otros casos, se puede usar la cola poli-A de una molécula de ARNm, por ejemplo, para iniciar la transcripción inversa. Después de la conversión en el ADN, los procedimientos detallados en el presente documento se pueden usar, en algunos casos, para capturar, seleccionar, marcar o aislar posteriormente una secuencia deseada.

Las moléculas de ácido nucleico incluyen ácido desoxirribonucleico (ADN) y/o ácido ribonucleico (ARN). Las moléculas de ácido nucleico pueden ser sintéticas o derivadas de fuentes naturales. En una realización, las moléculas de ácido nucleico se aíslan de una muestra biológica que contiene una variedad de otros componentes, tales como proteínas, lípidos y ácidos nucleicos no molde. Las moléculas de molde de ácido nucleico pueden obtenerse de cualquier material celular, obtenido de un animal, una planta, una bacteria, un hongo o cualquier otro organismo celular. En determinadas realizaciones, las moléculas de ácido nucleico se obtienen de una sola célula. Las muestras biológicas para su uso en la presente invención incluyen partículas o preparaciones víricas. Las moléculas de ácido nucleico pueden obtenerse directamente de un organismo o de una muestra biológica obtenida de un organismo, por ejemplo, de sangre, orina, líquido cefalorraquídeo, fluido seminal, saliva, esputo, heces y tejido. Se puede usar cualquier muestra de tejido o fluido corporal como fuente de ácido nucleico para su uso en la invención. Las moléculas de ácido nucleico también se pueden aislar de células cultivadas, tales como un cultivo celular primario o una estirpe celular. Las células o los tejidos a partir de los que se obtienen los ácidos nucleicos de molde pueden infectarse con un virus u otro patógeno intracelular.

Una muestra también puede ser ARN total extraído de una muestra biológica, un banco de ADNc, ADN vírico o genómico. En determinadas realizaciones, las moléculas de ácido nucleico están unidas a otras moléculas diana tales como proteínas, enzimas, sustratos, anticuerpos, agentes aglutinantes, perlas, moléculas pequeñas, péptidos o cualquier otra molécula. En general, el ácido nucleico se puede extraer de una muestra biológica mediante una variedad de técnicas tales como las descritas por Sambrook y Russell, "Molecular Cloning: A Laboratory Manual", Tercera Edición, Cold Spring Harbor, N. Y. (2001). Las moléculas de ácido nucleico pueden ser monocatenarias, bicatenarias o bicatenarias con regiones monocatenarias (por ejemplo, estructuras de tronco y de bucle).

Los procedimientos de extracción de ADN son bien conocidos en la técnica. Un protocolo de aislamiento de ADN clásico se basa en la extracción con disolventes orgánicos, tal como una mezcla de fenol y cloroformo, seguido de la precipitación con etanol (J. Sambrook *et al.*, "Molecular Cloning: A Laboratory Manual", 1989, 2ª Ed., Cold Spring Harbour Laboratory Press: Nueva York, N.Y.). Otros procedimientos incluyen: extracción de ADN mediante precipitación de proteínas (P. Sunnucks *et al.*, *Genetics*, 1996, 144: 747-756; S. M. Aljanabi *et al.*, *Nucl. Acids Res.* 1997, 25: 4692-4693), extracción de ADN con sales de bromuro de trimetilamonio (S. Gustincich *et al.*, *BioTechniques*, 1991, 11: 298-302) y extracción de ADN con tiocianato de guanidinio (J. B. W. Hammond *et al.*, *Biochemistry*, 1996, 240: 298-300). Hay una variedad de kits disponibles en el mercado para extraer ADN de muestras biológicas (por ejemplo, BD Biosciences Clontech (Palo Alto, CA); Epicentre Technologies (Madison, WI); Genra Systems, Inc. (Minneapolis, MN); MicroProbe Corp. (Bothell, WA); Organon Teknika (Durham, NC); y Qiagen Inc. (Valencia, CA)).

Los procedimientos de extracción de ARN también son bien conocidos en la técnica (por ejemplo, J. Sambrook *et al.*, "Molecular Cloning: A Laboratory Manual" 1989, 211d Ed., Cold Spring Harbour Laboratory Press: Nueva York) y hay kits para la extracción de ARN de los fluidos corporales disponibles en el mercado (por ejemplo, Ambion, Inc. (Austin, TX); Amersham Biosciences (Piscataway, NJ); BD Biosciences Clontech (Palo Alto, CA); BioRad Laboratories (Hercules, CA); Dynal Biotech Inc. (Lake Success, NY); Epicentre Technologies (Madison, WI); Genra Systems, Inc. (Minneapolis, MN); GIBCO BRL (Gaithersburg, MD); Invitrogen Life Technologies (Carlsbad, CA); MicroProbe Corp. (Bothell, WA); Organon Teknika (Durham, NC); Promega, Inc. (Madison, WI); y Qiagen Inc. (Valencia, CA)).

Una o más muestras pueden ser de una o más fuentes. Una o más de las muestras pueden ser de dos o más fuentes. Una o más de las muestras pueden ser de uno o más sujetos. Una o más de las muestras pueden ser de dos o más sujetos. Una o más de las muestras pueden ser del mismo sujeto. Uno o más sujetos pueden ser de la misma especie. Uno o más sujetos pueden ser de diferentes especies. El uno o más sujetos pueden estar sanos. El uno o más sujetos pueden verse afectados por una enfermedad, un trastorno o una afección.

En algunas realizaciones, una muestra es un líquido, tal como sangre, saliva, linfa, orina, líquido cefalorraquídeo, fluido seminal, esputo, heces u homogeneizados de tejidos.

Se puede tomar una muestra de un sujeto con una afección. En algunas realizaciones, el sujeto del que se toma una muestra puede ser un paciente, por ejemplo, un paciente con cáncer o un paciente con sospecha de cáncer. El sujeto puede ser un mamífero, por ejemplo, un ser humano, y puede ser macho o hembra. En algunas realizaciones, la hembra se encuentra en estado de gestación. La muestra puede ser una biopsia tumoral. La biopsia puede ser realizada, por ejemplo, por un proveedor de atención sanitaria, incluido un médico, médico asistente, enfermera, veterinario, dentista, quiropráctico, paramédico, dermatólogo, oncólogo, gastroenterólogo o cirujano.

En algunas realizaciones, los polinucleótidos están unidos a otras moléculas diana tales como proteínas, enzimas, sustratos, anticuerpos, agentes aglutinantes, perlas, moléculas pequeñas, péptidos o cualquier otra molécula. En algunas realizaciones, los polinucleótidos no están unidos a un soporte sólido. Los ácidos nucleicos pueden extraerse de una muestra biológica mediante una variedad de técnicas (Sambrook *et al.*, "Molecular Cloning: A Laboratory Manual", Tercera Edición, Cold Spring Harbor, N. Y. (2001)).

En algunas realizaciones, la muestra es saliva. En algunas realizaciones, la muestra es sangre entera. En algunas realizaciones, para obtener una cantidad suficiente de polinucleótidos para el análisis, se extrae un volumen de sangre de al menos aproximadamente 0,001, 0,005, 0,01, 0,05, 0,1, 0,5, 1, 2, 3, 4, 5, 10, 20, 25, 30, 35, 40, 45 o 50 ml. En algunas realizaciones, la sangre puede recogerse en un aparato que contiene un quelante de magnesio que incluye, pero sin limitación, EDTA, y se almacena a 4 °C. Opcionalmente, se puede añadir un quelante de calcio, incluyendo, pero sin limitación, EGTA.

En algunas realizaciones, se añade un inhibidor de la lisis celular a la sangre, incluido, pero sin limitación, formaldehído, derivados de formaldehído, formol, glutaraldehído, derivados de glutaraldehído, un reticulante proteico, un reticulante de ácido nucleico, un reticulante de proteína y de ácido nucleico, reticulantes reactivos de amina primaria, reticulantes reactivos de sulfhidrilo, adición de sulfhidrilo o reducción de disulfuro, reticulantes reactivos de hidratos de carbono, reticulantes reactivos de carboxilo, reticulantes fotorreactivos o reticulantes escindibles. En algunas realizaciones, los materiales que no son ácidos nucleicos pueden eliminarse del material de partida usando tratamientos enzimáticos (tales como la digestión con proteasas).

Una pluralidad de muestras puede comprender al menos 2, 3, 4, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90 o 100, o más muestras. La pluralidad de muestras puede comprender al menos aproximadamente 100, 200, 300, 400, 500, 600, 700, 800, 900 o 1.000, o más muestras. La pluralidad de muestras puede comprender al menos aproximadamente 1.000, 2.000, 3.000, 4.000, 5.000, 6.000, 7.000, 8.000 muestras, 9.000 o 10.000 muestras, o 100.000 muestras, o 1.000.000 o más muestras. La pluralidad de muestras puede comprender al menos aproximadamente 10.000 muestras.

El uno o más polinucleótidos de una primera muestra pueden ser diferentes de uno o más polinucleótidos de una segunda muestra. El uno o más polinucleótidos de una primera muestra pueden ser diferentes de uno o más polinucleótidos de una pluralidad de muestras. Uno o más polinucleótidos de una muestra puede comprender al menos aproximadamente el 80 %, 85 %, 90 %, 95 %, 96 %, 97 %, 98 %, 99 % o 100 % de identidad de secuencia. En algunas realizaciones, uno o más polinucleótidos de una muestra pueden diferir en menos de aproximadamente 100, 90, 80, 70, 60, 50, 40, 30, 25, 20, 10, 9, 8, 7, 6, 5, 4, 3, 2 o 1 nucleótido o par de bases. Una pluralidad de polinucleótidos de una o más muestras de la pluralidad de muestras puede comprender dos o más secuencias idénticas. Al menos aproximadamente el 1 %, 2 %, 3 %, 4 %, 5 %, 6 %, 7 %, 8 %, 9 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 97 %, 98 %, 99 % o 100 % del total de polinucleótidos de una o más de la pluralidad de muestras puede comprender la misma secuencia. Una pluralidad de polinucleótidos de una o más muestras de la pluralidad de muestras puede comprender al menos dos secuencias diferentes. Al menos aproximadamente el 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 81 %, 82 %, 83 %, 84 %, 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o 100 % del total de polinucleótidos de una o más de la pluralidad de muestras puede comprender al menos dos secuencias diferentes. En algunas realizaciones, uno o más polinucleótidos son variantes entre sí. Por ejemplo, uno o más polinucleótidos pueden contener polimorfismos de un solo nucleótido u otros tipos de mutaciones. En otro ejemplo, uno o más polinucleótidos son variantes de corte y empalme.

Una primera muestra puede comprender una o más células y la segunda muestra puede comprender una o más células. Los una o más células de la primera muestra pueden ser del mismo tipo de célula que la una o más células de la segunda muestra. La una o más células de la primera muestra pueden ser de un tipo de célula diferente como una o más células diferentes de la pluralidad de muestras.

La pluralidad de muestras puede obtenerse simultáneamente. Se puede obtener una pluralidad de muestras a la vez. La pluralidad de muestras se puede obtener secuencialmente. Se puede obtener una pluralidad de muestras a lo largo de años, por ejemplo, 100 años, 10 años, 5 años, 4 años, 3 años, 2 años o 1 año de obtención de una o más muestras diferentes. Se pueden obtener una o más muestras en el transcurso de aproximadamente un año después de obtener una o más muestras diferentes. Se pueden obtener una o más muestras en el transcurso de 12 meses, 11 meses, 10 meses, 9 meses, 8 meses, 7 meses, 6 meses, 4 meses, 3 meses, 2 meses o 1 mes de obtención de una o más muestras diferentes. Se pueden obtener una o más muestras en el transcurso de 30 días, 28 días, 26 días, 24 días, 21 días, 20 días, 18 días, 17 días, 16 días, 15 días, 14 días, 13 días, 12 días, 11 días, 10 días, 9 días, 8 días, 7 días,

6 días, 5 días, 4 días, 3 días, 2 días o 1 día de obtención de una o más muestras diferentes. Se pueden obtener una o más muestras en el transcurso de aproximadamente 24 horas, 22 horas, 20 horas, 18 horas, 16 horas, 14 horas, 12 horas, 10 horas, 8 horas, 6 horas, 4 horas, 2 horas o 1 hora de obtención de una o más muestras diferentes. Se pueden obtener una o más muestras en el transcurso de aproximadamente 60 segundos, 45 segundos, 30 segundos, 20 segundos, 10 segundos, 5 segundos, 2 segundos o 1 segundo de obtención de una o más muestras diferentes. Se pueden obtener una o más muestras en menos de un segundo después de obtener una o más muestras diferentes.

Los diferentes polinucleótidos de una muestra pueden estar presentes en la muestra a diferentes concentraciones o cantidades (por ejemplo, diferente número de moléculas). Por ejemplo, la concentración o cantidad de un polinucleótido puede ser superior a la concentración o cantidad de otro polinucleótido en la muestra. En algunas realizaciones, la concentración o cantidad de al menos un polinucleótido en la muestra es al menos aproximadamente 1,5, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1.000 o más veces superior a la concentración o cantidad de al menos otro polinucleótido en la muestra. En otro ejemplo, la concentración o cantidad de un polinucleótido es inferior a la concentración o cantidad de otro polinucleótido en la muestra. La concentración o cantidad de al menos un polinucleótido en la muestra puede ser de al menos aproximadamente 1,5, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1.000 o más veces inferior a la concentración o cantidad de al menos otro polinucleótido de la muestra.

En algunas realizaciones, dos o más muestras pueden contener diferentes cantidades o concentraciones de los polinucleótidos. En algunas realizaciones, la concentración o cantidad de un polinucleótido en una muestra puede ser superior a la concentración o cantidad del mismo polinucleótido en una muestra diferente. Por ejemplo, una muestra de sangre puede contener una cantidad más alta de un determinado polinucleótido que una muestra de orina. Como alternativa, una sola muestra puede dividirse en dos o más submuestras. Las submuestras pueden contener diferentes cantidades o concentraciones del mismo polinucleótido. La concentración o cantidad de al menos un polinucleótido en una muestra puede ser de al menos aproximadamente 1,5, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1.000 o más veces superior a la concentración o cantidad del mismo polinucleótido en otra muestra. Como alternativa, la concentración o cantidad de un polinucleótido en una muestra puede ser inferior a la concentración o cantidad del mismo polinucleótido en una muestra diferente. Por ejemplo, La concentración o cantidad de al menos un polinucleótido en una muestra puede ser de al menos aproximadamente 1,5, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1.000 o más veces inferior a la concentración o cantidad del mismo polinucleótido de otra muestra.

### 35 POLINUCLEÓTIDOS DIANAS

En algunos casos, los procedimientos proporcionados en el presente documento están dirigidos a la amplificación y secuenciación de una molécula de polinucleótido diana, tal como una molécula de polinucleótido de una célula. En algunos casos, los procedimientos proporcionados en el presente documento están dirigidos a la amplificación y secuenciación de dos o más regiones de una molécula de polinucleótido diana. En algunos casos, los procedimientos proporcionados en el presente documento están dirigidos a la amplificación y secuenciación de dos o más moléculas de polinucleótido diana. En un aspecto, los polinucleótidos diana son ARN. En un aspecto, los polinucleótidos diana son ácidos nucleicos genómicos. El ADN derivado del material genético en los cromosomas de un determinado organismo puede ser ADN genómico. En realizaciones preferidas, los polinucleótidos diana incluyen secuencias que comprenden regiones variables de un anticuerpo o TCR producido por una célula inmunitaria. En algunas realizaciones, los polinucleótidos diana incluyen secuencias que comprenden una región variable de una cadena pesada de un anticuerpo producido por una célula inmunitaria. En algunas realizaciones, los polinucleótidos diana incluyen secuencias que comprenden una región variable de una cadena ligera de un anticuerpo producido por una célula inmunitaria. En algunas realizaciones, los polinucleótidos diana incluyen secuencias que comprenden una región variable de una cadena alfa de un TCR producido por una célula inmunitaria. En algunas realizaciones, los polinucleótidos diana incluyen secuencias que comprenden una región variable de una cadena beta de un TCR producido por una célula inmunitaria. En algunas realizaciones, los polinucleótidos diana incluyen secuencias que comprenden una región variable de una cadena gamma de un TCR producido por una célula inmunitaria. En algunas realizaciones, los polinucleótidos diana incluyen secuencias que comprenden una región variable de una cadena delta de un TCR producido por una célula inmunitaria.

Los polinucleótidos diana pueden obtenerse de prácticamente cualquier fuente y pueden prepararse usando procedimientos conocidos en la técnica. Por ejemplo, los polinucleótidos diana se pueden aislar directamente sin amplificación usando procedimientos conocidos en la técnica, incluyendo, sin limitación, la extracción de un fragmento de ADN genómico o ARNm de un organismo o una célula (por ejemplo, una célula inmunitaria) para obtener polinucleótidos diana. Un polinucleótido diana también puede abarcar el ADNc generado a partir del ARN (tal como el ARNm) a través de la PCR de transcripción inversa. En algunos casos, un polinucleótido diana es una molécula de ARN. En algunos casos, un polinucleótido diana es una molécula de ARNm, o un ADNc producido a partir de la molécula de ARNm. En algunos casos, un polinucleótido diana es una molécula de ARNm o molécula de ADNc producida a partir de la molécula de ARNm, a partir de una sola célula inmunitaria. En algunos casos, los polinucleótidos diana son moléculas de ARNm o moléculas de ADNc producidas a partir de las moléculas de ARNm,

a partir de células inmunitarias individuales. En algunos casos, los polinucleótidos diana son moléculas de ARNm que codifican una secuencia de anticuerpos de una sola célula inmunitaria. En algunos casos, los polinucleótidos diana son moléculas de ARNm que codifican secuencias de anticuerpos de cadena pesada de células inmunitarias individuales. En algunos casos, los polinucleótidos diana son moléculas de ARNm que codifican una secuencia de anticuerpo de cadena pesada de una sola célula inmunitaria. En algunos casos, los polinucleótidos diana son moléculas de ARNm que codifican secuencias de anticuerpos de cadena ligera de células inmunitarias individuales. En algunos casos, los polinucleótidos diana son moléculas de ARNm que codifican una secuencia de anticuerpo de cadena ligera variable de una sola célula inmunitaria. En algunos casos, los polinucleótidos diana son moléculas de ARNm que codifican secuencias de anticuerpos de cadena ligera variable de células inmunitarias individuales. En algunos casos, los polinucleótidos diana son moléculas de ARNm que codifican una secuencia de anticuerpo de cadena ligera variable de una sola célula inmunitaria. En algunos casos, los polinucleótidos diana son moléculas de ARNm que codifican secuencias de anticuerpos de cadena pesada variable de células inmunitarias individuales. En algunos casos, los polinucleótidos diana son moléculas de ARNm que codifican una secuencia de anticuerpo de cadena pesada variable de una sola célula inmunitaria. En algunos casos, un polinucleótido diana puede ser un ácido nucleico libre de células, por ejemplo, ADN o ARN. En algunos casos, los polinucleótidos diana son moléculas de ARNm que codifican secuencias de TCR de cadena alfa, beta, gamma y/o delta variable de células inmunitarias individuales.

Los procedimientos descritos en el presente documento pueden usarse para generar un banco de polinucleótidos a partir de uno o más polinucleótidos diana para la secuenciación. Los polinucleótidos diana incluyen cualquier polinucleótido de interés que no sea producto de una reacción de amplificación. Por ejemplo, un polinucleótido diana puede incluir un polinucleótido en una muestra biológica. Por ejemplo, los polinucleótidos diana no incluyen productos de una reacción de PCR. Por ejemplo, los polinucleótidos diana pueden incluir un molde de polinucleótido usado para generar productos de una reacción de amplificación, pero no incluyen los propios productos de amplificación. Por ejemplo, los polinucleótidos diana pueden incluir un molde de polinucleótido usado para generar productos de una reacción de transcripción inversa o reacción de extensión del cebador, y también incluyen la reacción de transcripción inversa o los propios productos de reacción de extensión del cebador. Por ejemplo, los polinucleótidos diana incluyen polinucleótidos de interés que pueden someterse a una reacción de transcripción inversa o una reacción de extensión del cebador. Por ejemplo, los polinucleótidos diana incluyen ARN o ADN. Por ejemplo, los polinucleótidos diana incluyen ADNc. En algunas realizaciones, los polinucleótidos de ARN diana son ARNm. En algunas realizaciones, los polinucleótidos de ARN diana están poliadenilados. En algunas realizaciones, los polinucleótidos de ARN no están poliadenilados. En algunas realizaciones, los polinucleótidos diana son polinucleótidos de ADN. Los polinucleótidos de ADN pueden ser ADN genómico. Los polinucleótidos de ADN pueden comprender exones, intrones, regiones no traducidas o cualquier combinación de los mismos.

En algunas realizaciones, pueden generarse bancos a partir de dos o más regiones de un polinucleótido diana. En algunas realizaciones, se pueden generar bancos de procedimientos a partir de dos o más polinucleótidos diana. En algunas realizaciones, los polinucleótidos diana son ácidos nucleicos genómicos o ADN derivado de cromosomas. En algunas realizaciones, los polinucleótidos diana incluyen secuencias que comprenden una variante, tal como un polimorfismo o una mutación. En algunas realizaciones, los polinucleótidos diana incluyen ADN y no ARN. En algunas realizaciones, los polinucleótidos diana incluyen ARN y no ADN. En algunas realizaciones, los polinucleótidos diana incluyen ADN y ARN. En algunas realizaciones, un polinucleótido diana es una molécula de ARNm. En algunas realizaciones, un polinucleótido diana es una molécula de ADN. En algunas realizaciones, un polinucleótido diana es un polinucleótido monocatenario. En algunas realizaciones, un polinucleótido diana es un polinucleótido bicatenario. En algunas realizaciones, un polinucleótido diana es una cadena sencilla de un polinucleótido bicatenario.

Los polinucleótidos diana pueden obtenerse a partir de cualquier muestra biológica y prepararse usando procedimientos conocidos en la técnica. En algunas realizaciones, los polinucleótidos diana se aíslan directamente sin amplificación. Los procedimientos para el aislamiento directo son conocidos en la técnica. Los ejemplos no limitantes incluyen la extracción de ADN genómico o ARNm de una muestra biológica, un organismo o una célula.

En algunas realizaciones, se purifican uno o más polinucleótidos diana a partir de una muestra biológica. En algunas realizaciones, un polinucleótido diana no se purifica a partir de la muestra biológica en la que está contenido. En algunas realizaciones, un polinucleótido diana se aísla de una muestra biológica. En algunas realizaciones, un polinucleótido diana no se aísla de la muestra biológica en la que está contenido. En algunas realizaciones, un polinucleótido diana puede ser un ácido nucleico libre de células. En algunas realizaciones, un polinucleótido diana puede ser un ácido nucleico fragmentado. En algunas realizaciones, un polinucleótido diana puede ser un ácido nucleico transcrito. En algunas realizaciones, un polinucleótido diana es un polinucleótido modificado. En algunas realizaciones, un polinucleótido diana es un polinucleótido no modificado.

En algunas realizaciones, un polinucleótido diana es un polinucleótido de una sola célula. En algunas realizaciones, los polinucleótidos diana son de células individuales. En algunas realizaciones, un polinucleótido diana es un polinucleótido de una muestra que contiene una pluralidad de células.

En algunas realizaciones, un polinucleótido diana codifica una secuencia de biomarcador. En algunas realizaciones, un polinucleótido diana codifica dos o más secuencias de biomarcador. En algunas realizaciones, una pluralidad de polinucleótidos diana codifica una secuencia de biomarcador. En algunas realizaciones, una pluralidad de polinucleótidos diana codifica dos o más secuencias de biomarcador. En algunas realizaciones, una pluralidad de polinucleótidos diana codifica 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90 o 100, o más secuencias de biomarcador.

En algunas realizaciones, una pluralidad de polinucleótidos diana comprende un panel de secuencias de inmunoglobulina. En algunas realizaciones, una pluralidad de polinucleótidos diana comprende un panel de secuencias de TCR. Por ejemplo, un panel de secuencias de inmunoglobulina puede ser secuencias de V<sub>H</sub> y/o V<sub>L</sub>. En algunas realizaciones, un panel de secuencias de inmunoglobulina o TCR contiene 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10 secuencias de inmunoglobulina o TCR. En algunas realizaciones, un panel de secuencias de inmunoglobulina o TCR contiene al menos aproximadamente 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 1.000, 1.500, 2.000, 3.000, 4.000, 5.000, 6.000, 7.000, 8.000, 9.000, 10.000, 11.000, 12.000, 13.000, 14.000, 15.000, 16.000, 17.000, 18.000, 19.000, 20.000, 25.000, 30.000, 35.000, 40.000, 45.000, 50.000, 60.000, 70.000, 80.000, 90.000, 100.000, 200.000, 300.000, 400.000, 500.000, 600.000, 700.000, 800.000, 900.000,  $1 \times 10^6$ ,  $2 \times 10^6$ ,  $3 \times 10^6$ ,  $4 \times 10^6$ ,  $5 \times 10^6$ ,  $6 \times 10^6$ ,  $7 \times 10^6$ ,  $8 \times 10^6$ ,  $9 \times 10^6$ ,  $1 \times 10^7$ ,  $2 \times 10^7$ ,  $3 \times 10^7$ ,  $4 \times 10^7$ ,  $5 \times 10^7$ ,  $6 \times 10^7$ ,  $7 \times 10^7$ ,  $8 \times 10^7$ ,  $9 \times 10^7$ ,  $1 \times 10^8$ ,  $2 \times 10^8$ ,  $3 \times 10^8$ ,  $4 \times 10^8$ ,  $5 \times 10^8$ ,  $6 \times 10^8$ ,  $7 \times 10^8$ ,  $8 \times 10^8$ ,  $9 \times 10^8$ ,  $1 \times 10^9$ ,  $2 \times 10^9$ ,  $3 \times 10^9$ ,  $4 \times 10^9$ ,  $5 \times 10^9$ ,  $6 \times 10^9$ ,  $7 \times 10^9$ ,  $8 \times 10^9$ ,  $9 \times 10^9$ ,  $1 \times 10^{10}$ ,  $2 \times 10^{10}$ ,  $3 \times 10^{10}$ ,  $4 \times 10^{10}$ ,  $5 \times 10^{10}$ ,  $6 \times 10^{10}$ ,  $7 \times 10^{10}$ ,  $8 \times 10^{10}$ ,  $9 \times 10^{10}$ ,  $1 \times 10^{11}$ ,  $2 \times 10^{11}$ ,  $3 \times 10^{11}$ ,  $4 \times 10^{11}$ ,  $5 \times 10^{11}$ ,  $6 \times 10^{11}$ ,  $7 \times 10^{11}$ ,  $8 \times 10^{11}$ ,  $9 \times 10^{11}$ ,  $1 \times 10^{12}$ ,  $2 \times 10^{12}$ ,  $3 \times 10^{12}$ ,  $4 \times 10^{12}$ ,  $5 \times 10^{12}$ ,  $6 \times 10^{12}$ ,  $7 \times 10^{12}$ ,  $8 \times 10^{12}$  o  $9 \times 10^{12}$  secuencias de inmunoglobulina o TCR. En algunas realizaciones, un panel de secuencias de inmunoglobulina o TCR contiene como máximo aproximadamente 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 1.000, 1.500, 2.000, 3.000, 4.000, 5.000, 6.000, 7.000, 8.000, 9.000, 10.000, 11.000, 12.000, 13.000, 14.000, 15.000, 16.000, 17.000, 18.000, 19.000, 20.000, 25.000, 30.000, 35.000, 40.000, 45.000, 50.000, 60.000, 70.000, 80.000, 90.000, 100.000, 200.000, 300.000, 400.000, 500.000, 600.000, 700.000, 800.000, 900.000,  $1 \times 10^6$ ,  $2 \times 10^6$ ,  $3 \times 10^6$ ,  $4 \times 10^6$ ,  $5 \times 10^6$ ,  $6 \times 10^6$ ,  $7 \times 10^6$ ,  $8 \times 10^6$ ,  $9 \times 10^6$ ,  $1 \times 10^7$ ,  $2 \times 10^7$ ,  $3 \times 10^7$ ,  $4 \times 10^7$ ,  $5 \times 10^7$ ,  $6 \times 10^7$ ,  $7 \times 10^7$ ,  $8 \times 10^7$ ,  $9 \times 10^7$ ,  $1 \times 10^8$ ,  $2 \times 10^8$ ,  $3 \times 10^8$ ,  $4 \times 10^8$ ,  $5 \times 10^8$ ,  $6 \times 10^8$ ,  $7 \times 10^8$ ,  $8 \times 10^8$ ,  $9 \times 10^8$ ,  $1 \times 10^9$ ,  $2 \times 10^9$ ,  $3 \times 10^9$ ,  $4 \times 10^9$ ,  $5 \times 10^9$ ,  $6 \times 10^9$ ,  $7 \times 10^9$ ,  $8 \times 10^9$ ,  $9 \times 10^9$ ,  $1 \times 10^{10}$ ,  $2 \times 10^{10}$ ,  $3 \times 10^{10}$ ,  $4 \times 10^{10}$ ,  $5 \times 10^{10}$ ,  $6 \times 10^{10}$ ,  $7 \times 10^{10}$ ,  $8 \times 10^{10}$ ,  $9 \times 10^{10}$ ,  $1 \times 10^{11}$ ,  $2 \times 10^{11}$ ,  $3 \times 10^{11}$ ,  $4 \times 10^{11}$ ,  $5 \times 10^{11}$ ,  $6 \times 10^{11}$ ,  $7 \times 10^{11}$ ,  $8 \times 10^{11}$ ,  $9 \times 10^{11}$ ,  $1 \times 10^{12}$ ,  $2 \times 10^{12}$ ,  $3 \times 10^{12}$ ,  $4 \times 10^{12}$ ,  $5 \times 10^{12}$ ,  $6 \times 10^{12}$ ,  $7 \times 10^{12}$ ,  $8 \times 10^{12}$  o  $9 \times 10^{12}$  secuencias de inmunoglobulina o TCR. En algunas realizaciones, un panel de secuencias de inmunoglobulina o TCR contiene a partir de aproximadamente 10-20, 10-30, 10-40, 10-30, 10-40, 10-50, 10-60, 10-70, 10-80, 10-90, 10-100, 50-60, 50-70, 50-80, 50-90, 50-100, 100-200, 100-300, 100-400, 100-300, 100-400, 100-500, 100-600, 100-700, 100-800, 100-900, 100-1000, 500-600, 500-700, 500-800, 500-900, 500-1.000, 1.000-2.000, 1.000-3.000, 1.000-4.000, 1.000-3.000, 1.000-4.000, 1.000-5.000, 1.000-6.000, 1.000-7.000, 1.000-8.000, 1.000-9.000, 1.000-10.000, 5.000-6.000, 5.000-7.000, 5.000-8.000, 5.000-9.000, 5.000-10.000,  $1-1 \times 10^5$ ,  $1-2 \times 10^5$ ,  $1-3 \times 10^5$ ,  $1-4 \times 10^5$ ,  $1-5 \times 10^5$ ,  $1-6 \times 10^5$ ,  $1-7 \times 10^5$ ,  $1-8 \times 10^5$ ,  $9 \times 10^5$ ,  $1-1 \times 10^6$ ,  $1-2 \times 10^6$ ,  $1-3 \times 10^6$ ,  $1-4 \times 10^6$ ,  $1-5 \times 10^6$ ,  $1-6 \times 10^6$ ,  $1-7 \times 10^6$ ,  $1-8 \times 10^6$ ,  $9 \times 10^6$ ,  $1 \times 10^7$ ,  $1-2 \times 10^7$ ,  $1-3 \times 10^7$ ,  $1-4 \times 10^7$ ,  $1-5 \times 10^7$ ,  $1-6 \times 10^7$ ,  $1-7 \times 10^7$ ,  $1-8 \times 10^7$ ,  $1-9 \times 10^7$ ,  $1-1 \times 10^8$ ,  $1-2 \times 10^8$ ,  $1-3 \times 10^8$ ,  $1-4 \times 10^8$ ,  $1-5 \times 10^8$ ,  $1-6 \times 10^8$ ,  $1-7 \times 10^8$ ,  $1-8 \times 10^8$ ,  $1-9 \times 10^8$ ,  $1-1 \times 10^9$ ,  $1-2 \times 10^9$ ,  $1-3 \times 10^9$ ,  $1-4 \times 10^9$ ,  $1-5 \times 10^9$ ,  $1-6 \times 10^9$ ,  $1-7 \times 10^9$ ,  $1-8 \times 10^9$ ,  $1-9 \times 10^9$ ,  $1-1 \times 10^{10}$ ,  $1-2 \times 10^{10}$ ,  $1-3 \times 10^{10}$ ,  $1-4 \times 10^{10}$ ,  $1-5 \times 10^{10}$ ,  $1-6 \times 10^{10}$ ,  $1-7 \times 10^{10}$ ,  $1-8 \times 10^{10}$ ,  $1-9 \times 10^{10}$ ,  $1-1 \times 10^{11}$ ,  $1-2 \times 10^{11}$ ,  $1-3 \times 10^{11}$ ,  $1-4 \times 10^{11}$ ,  $1-5 \times 10^{11}$ ,  $1-6 \times 10^{11}$ ,  $1-7 \times 10^{11}$ ,  $1-8 \times 10^{11}$ ,  $1-9 \times 10^{11}$ ,  $1-1 \times 10^{12}$ ,  $1-2 \times 10^{12}$ ,  $1-3 \times 10^{12}$ ,  $1-4 \times 10^{12}$ ,  $1-5 \times 10^{12}$ ,  $1-6 \times 10^{12}$ ,  $1-7 \times 10^{12}$ ,  $1-8 \times 10^{12}$  o  $1-9 \times 10^{12}$  secuencias de inmunoglobulina o TCR.

En algunas realizaciones, un polinucleótido diana es de aproximadamente 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 1.000, 1.500, 2.000, 3.000, 4.000, 5.000, 6.000, 7.000, 8.000, 9.000, 10.000, 11.000, 12.000, 13.000, 14.000, 15.000, 16.000, 17.000, 18.000, 19.000 o 20.000 bases o pares de bases de longitud. En algunas realizaciones, un polinucleótido diana es de al menos aproximadamente 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 1.000, 1.500, 2.000, 3.000, 4.000, 5.000, 6.000, 7.000, 8.000, 9.000, 10.000, 11.000, 12.000, 13.000, 14.000, 15.000, 16.000, 17.000, 18.000, 19.000 o 20.000 bases o pares de bases de longitud. En algunas realizaciones, un polinucleótido diana es como máximo de aproximadamente 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 1.000, 1.500, 2.000, 3.000, 4.000, 5.000, 6.000, 7.000, 8.000, 9.000, 10.000, 11.000, 12.000, 13.000, 14.000, 15.000, 16.000, 17.000, 18.000, 19.000 o 20.000 bases o pares de bases de longitud. En algunas realizaciones, un polinucleótido diana es a partir de aproximadamente 10-20, 10-30, 10-40, 10-30, 10-40, 10-50, 10-60, 10-70, 10-80, 10-90, 10-100, 50-60, 50-70, 50-80, 50-90, 50-100, 100-200, 100-300, 100-400, 100-300, 100-400, 100-500, 100-600, 100-700, 100-800, 100-900, 100-1.000, 500-600, 500-700, 500-800, 500-900, 500-1.000, 1.000-2.000, 1.000-3.000, 1.000-4.000, 1.000-3.000, 1.000-4.000, 1.000-5.000, 1.000-6.000, 1.000-7.000, 1.000-8.000, 1.000-9.000, 1.000-10.000, 5.000-6.000, 5.000-7.000, 5.000-8.000, 5.000-9.000 o 5.000-10.000 bases o pares de bases de longitud. En algunas realizaciones, la longitud media de los polinucleótidos diana, o fragmentos de los mismos, puede ser inferior a aproximadamente 100, 200, 300, 400, 500 u 800 pares de bases, o inferior a aproximadamente 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190 o 200 nucleótidos, o inferior a aproximadamente 1, 2, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100 kilobases. En algunas realizaciones, una secuencia diana de un molde relativamente corto, tal como una muestra que contiene un polinucleótido diana, es de aproximadamente 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95 o 100 bases. En determinadas realizaciones, los datos de secuenciación se alinean contra secuencias conocidas o

esperadas usando una base de datos que contiene secuencias de inmunoglobulina o secuencias de TCR asociadas con una enfermedad o afección.

## SECUENCIACIÓN DEL REPERTORIO INMUNITARIO

5 La presente invención usa etapas en las que se manipulan ácidos nucleicos para generar bancos de polinucleótidos para la secuenciación. En algunas realizaciones, la presente invención utiliza etapas en las que se manipulan ácidos nucleicos para producir anticuerpos monoclonales recombinantes. En un sentido general, en algunas realizaciones de la invención, la amplificación del material genético de las células inmunitarias y/o linfocitos T, por ejemplo, la reacción  
10 en cadena de la polimerasa de transcripción inversa (PCR de transcripción inversa) se emplea para generar la amplificación de ADNc del material genético de las células inmunitarias. Para las moléculas de anticuerpos, los genes de inmunoglobulina se pueden obtener a partir de ADN genómico o ARNm de células inmunitarias o linfocitos T. El ARN puede ser una cadena pesada (segmentos V, D, J) o una cadena ligera (segmentos V, J). En algunas realizaciones, el material de partida es ARN de células inmunitarias compuestas por segmentos de genes V, D, J que  
15 codifican un anticuerpo y contienen una región constante.

El material de partida polinucleotídico, tal como ARN, se puede transcribir de manera inversa en ADNc usando uno o un conjunto de polinucleótidos. Los polinucleótidos pueden comprender una parte complementaria a una región del ARN, tal como en una región constante o a una cola poli-A del ARNm. Un código de barras del vaso, que puede ser  
20 un tramo de □20 nucleótidos degenerados con o sin una posición de base intercalada conocida, tal como NNNNWNNNNWNNNNWNNNNW, en el que W significa A o T.

El ADNc resultante de la transcripción inversa se puede marcar con uno o más códigos de barras, por ejemplo, con un código de barras del vaso y un código de barras molecular. Se pueden usar diversos oligonucleótidos de un  
25 determinado diseño para el marcaje. El ADNc marcado resultante de la transcripción inversa puede amplificarse una o más veces, tal como mediante amplificación por PCR. Se pueden usar diversos cebadores de un determinado diseño para la amplificación. Un producto de una primera reacción de amplificación, tal como PCR, puede amplificarse usando una segunda reacción de amplificación, tal como una primera o segunda fase de PCR. Se pueden usar varios cebadores para la etapa de amplificación. Se puede generar un banco de polinucleótidos amplificados usando los  
30 procedimientos descritos en el presente documento. Un banco resultante puede comprender una secuencia completa o parcial de anticuerpo o TCR con códigos de barras moleculares y de vaso apropiados. En otras realizaciones, se puede usar el cambio de molde para generar bancos, tal como para la secuenciación de repertorios inmunitarios. Por ejemplo, se puede emplear el cambio de molde durante la transcripción inversa para generar una región en el producto de la transcripción inversa que sea complementaria a un polinucleótido que albergue  
35 un código de barras, tal como un polinucleótido con código de barras del vaso o un polinucleótido con código de barras molecular. El cambio de molde se puede emplear durante la transcripción inversa para eliminar los problemas de sesgo de la PCR. Estos procedimientos se pueden usar para la secuenciación de anticuerpos, tal como mediante el uso de una plataforma de secuenciación de alto rendimiento.

40 El material de partida puede ser ARN o ADN, tal como a partir de células inmunitarias o linfocitos T que comprenden los segmentos de genes V, D, J que codifican un anticuerpo y contienen la región constante. En algunas realizaciones, el polinucleótido diana comprende segmentos de cadena pesada (segmentos de V, D, J) o segmentos de cadena ligera (segmentos de V, J).

45 Los polinucleótidos diana pueden transcribirse de forma inversa en ADNc usando uno o un conjunto de polinucleótidos. Los ejemplos de cebadores en un conjunto de polinucleótidos para la transcripción inversa de un polinucleótido diana pueden comprender una parte complementaria a una región del polinucleótido diana. En algunas realizaciones, la parte complementaria a una región del polinucleótido diana puede ser complementaria a una región constante o a una cola poli-A del polinucleótido diana, tal como ARNm. Se pueden usar múltiples oligonucleótidos, tales como cebadores,  
50 para hibridar una o más regiones constantes. Se puede emplear una transcriptasa inversa para llevar a cabo la reacción de transcripción inversa. En realizaciones particulares, una transcriptasa inversa puede comprender una actividad de transferasa terminal no molde. Cuando una transcriptasa inversa que comprende una actividad de transferasa terminal no molde alcanza el final de un molde, puede añadir tres o más restos que no sean de molde, tal como tres o más restos de citosina que no son de molde. En algunas realizaciones, se usa la transcriptasa inversa Superscript II™ con este fin. En algunas realizaciones, se usa la transcriptasa inversa Maxima™ con este fin. En algunas realizaciones, se usa la transcriptasa inversa Protoscript II™ con este fin. En algunas realizaciones, se usa la transcriptasa inversa del virus de la leucemia murina de Moloney (MMLV-RT) con este fin. En algunas realizaciones, se usa la transcriptasa inversa HighScriber™ con este fin. En algunas realizaciones, se usa una desoxinucleotidil transferasa terminal con este fin. En algunas realizaciones, se usa la transcriptasa inversa del virus de la mieloblastosis  
55 aviar (AMV) con este fin. Se puede usar cualquier transcriptasa inversa capaz de transcribir ARN que tenga actividad transferasa terminal no molde. Se puede usar cualquier polimerasa inversa capaz de transcribir ARN que tenga actividad transferasa terminal no molde. Se puede usar cualquier polimerasa inversa capaz de transcribir ADN que tenga actividad transferasa terminal no molde.

65 Las reacciones de transcripción inversa, tales como las descritas anteriormente, se pueden realizar en presencia de un polinucleótido de marcaje en 3'. Un polinucleótido de marcaje en 3' puede ser un polinucleótido usado para añadir

ácidos nucleicos a un extremo 3' de un polinucleótido diana, tal como un ADNc. Un polinucleótido de marcaje en 3' puede ser un polinucleótido usado como molde para añadir ácidos nucleicos a un extremo 3' de un polinucleótido diana, tal como un ADNc. Un polinucleótido de marcaje en 3' puede ser un polinucleótido que se hibrida con un extremo 3' de un polinucleótido diana, tal como un ADNc. Un polinucleótido de marcaje en 3' puede ser un polinucleótido que contenga una región 3', tal como una región terminal 3', que se hibrida con un extremo 3' de un polinucleótido diana, tal como un ADNc. Por ejemplo, un polinucleótido de marcaje en 3' puede comprender un segmento, tal como un segmento que se hibride a tres o más restos que no son de molde. En algunas realizaciones, un polinucleótido de marcaje en 3' es un polinucleótido con código de barras molecular. En algunas realizaciones, un polinucleótido de marcaje en 3' puede comprender un código de barras molecular. En algunas realizaciones, un polinucleótido de marcaje en 3' puede comprender 3 restos de ribo-guanina o análogos de los mismos en el extremo 3' (rGrGrG) (bases de ARN) que sean complementarios a y que se hibriden a la cadena producida por la enzima de transcripción inversa. En algunas realizaciones, se pueden usar tres o más restos de guanina en lugar de ribo-guanina (nucleótido de ADN en lugar de nucleótido de ARN). En algunas realizaciones, un polinucleótido de marcaje en 3' puede comprender 1 o 2 restos de ribo-guanina en el extremo 3' y un resto de desoxirribo-guanina o un análogo del mismo en el extremo 3' (rGrGG) que sean complementarios y que se hibriden a la cadena producida por la enzima de transcripción inversa.

Tras la hibridación de un polinucleótido de marcaje en 3' a un CCC de la cadena de ADNc, una transcriptasa inversa puede continuar extendiendo el ADNc al polinucleótido de marcaje, uniendo así un código de barras molecular, o su complemento, a una población diana de polinucleótidos, tales como ADNc, en la reacción. Por ejemplo, el polinucleótido de marcaje en 3' puede ser un polinucleótido que contenga una región 5' a la región 3' que se hibride con un extremo 3' de un polinucleótido diana. La región 5' a la región 3' que se hibrida con un extremo 3' de un polinucleótido diana puede comprender una región que no sea complementaria al polinucleótido diana, tal como un ADNc. La región 5' a la región 3' que se hibrida con un extremo 3' de un polinucleótido diana puede comprender un código de barras molecular. La región 5' a la región 3' que se hibrida con un extremo 3' de un polinucleótido diana puede comprender una región complementaria a un polinucleótido con código de barras del vaso o su complemento. En otros experimentos, el cambio de molde se puede realizar en reacciones separadas. Por ejemplo, se puede añadir un polinucleótido de marcaje en 3' después de la reacción de transcripción inversa, y se pueden usar enzimas tales como una transcriptasa inversa o polimerasa para extenderse en un polinucleótido de marcaje. Debido a que un polinucleótido de marcaje puede albergar un código de barras molecular degenerado único en cada molécula de un vaso, cada ADNc de un vaso se puede marcar de manera única con un código de barras molecular. En algunas realizaciones, el cambio de molde se puede realizar al mismo tiempo que se realiza una reacción de transcripción inversa.

En algunas realizaciones, un polinucleótido de marcaje en 3', tal como un polinucleótido con código de barras molecular, puede comprender además una región 5', tal como una región terminal 5' que sea complementaria a un polinucleótido de marcaje en 3' o su complemento que contenga otro código de barras, tal como un código de barras del vaso. En algunas realizaciones, un polinucleótido diana que contiene un código de barras molecular o su complemento, tal como una molécula de ADNc marcada, puede comprender una región 3', tal como una región terminal 3' que sea complementaria a un polinucleótido de marcaje en 3' o su complemento que contenga otro código de barras, tal como un código de barras del vaso.

En algunas realizaciones, un polinucleótido de marcaje en 3' es un polinucleótido con código de barras del vaso. Tras la generación de un polinucleótido que contiene un código de barras molecular o su complemento a partir de un polinucleótido diana, se puede añadir un código de barras del vaso al polinucleótido diana con código de barras molecular. Un polinucleótido de marcaje en 3' puede ser un polinucleótido usado para añadir ácidos nucleicos a un extremo 3' de un polinucleótido diana, tal como un polinucleótido diana con código de barras molecular. Un polinucleótido de marcaje en 3' puede ser un polinucleótido usado como molde para añadir ácidos nucleicos a un extremo 3' de un polinucleótido diana, tal como un polinucleótido diana con código de barras molecular. Un polinucleótido de marcaje en 3' puede ser un polinucleótido que se hibrida con un extremo 3' de un polinucleótido diana, tal como un polinucleótido diana con código de barras molecular. Un polinucleótido de marcaje en 3' puede ser un polinucleótido que contenga una región 3', tal como una región terminal 3', que se hibrida con un extremo 3' de un polinucleótido diana, tal como un polinucleótido diana con código de barras molecular. Un polinucleótido con código de barras del vaso puede comprender una región 3', tal como una región terminal 3', que se hibrida con un extremo 3' de un polinucleótido diana con código de barras molecular.

Tras la hibridación de un polinucleótido de marcaje 3' a un polinucleótido diana con código de barras molecular, una transcriptasa inversa puede continuar extendiendo el ADNc al polinucleótido de marcaje en 3', tal como un polinucleótido con código de barras del vaso, uniendo así un código de barras del vaso, o su complemento, a una población diana de polinucleótidos, tal como polinucleótidos diana con código de barras molecular, en la reacción. Por ejemplo, el polinucleótido de marcaje en 3' puede ser un polinucleótido que contenga una región 5' a la región 3' que se hibride con un extremo 3' de un polinucleótido diana con código de barras molecular. La región 5' a la región 3' que se hibrida con un extremo 3' de un polinucleótido diana con código de barras molecular puede comprender una región que no sea complementaria al polinucleótido diana o al polinucleótido diana con código de barras molecular. La región 5' a la región 3' que se hibrida con un extremo 3' de un polinucleótido diana con código de barras molecular puede comprender un código de barras del vaso.

En algunas realizaciones, un polinucleótido de marcaje en 3' es un producto amplificado. En algunas realizaciones, un polinucleótido de marcaje en 3' es un producto amplificado que se origina a partir de una sola molécula. En algunas realizaciones, un polinucleótido de marcaje en 3' es un producto amplificado de un polinucleótido con código de barras del vaso. En algunas realizaciones, un polinucleótido de marcaje en 3' es un producto amplificado que se origina a partir de un solo polinucleótido con código de barras del vaso. La región 5' a la región 3' que se hibrida con un extremo 3' de un polinucleótido diana con código de barras molecular puede comprender una región complementaria a un cebador o su complemento. La región 5' a la región 3' que se hibrida con un extremo 3' de un polinucleótido diana con código de barras molecular puede comprender una región complementaria a un cebador o complemento del mismo que se usó para amplificar el polinucleótido con código de barras del vaso.

Un polinucleótido diana con código de barras doble, tal como un ADNc que contiene un código de barras molecular y un código de barras de un vaso, puede entonces amplificarse, tal como por PCR. Luego se puede realizar la PCR, por ejemplo, usando un conjunto de cebadores. Un producto de la reacción de PCR mencionada anteriormente se puede amplificar una o más veces, tal como en una o más series de PCR, o se puede secuenciar directamente.

Un banco producido de acuerdo con los procedimientos descritos en el presente documento puede ser un banco que comprenda una secuencia de TCR o anticuerpo grande o completa con códigos de barras apropiados, tales como códigos de barras del vaso y códigos de barras moleculares, que están secuenciados. En algunas realizaciones, un banco producido de acuerdo con los procedimientos descritos en el presente documento puede contener segmentos de agrupamiento apropiados para la secuenciación. En algunas realizaciones, se pueden generar muchas copias con códigos de barras moleculares idénticos. En algunas realizaciones, se pueden generar muchas copias de polinucleótidos que contienen códigos de barras moleculares idénticos para cada molécula de polinucleótido diana única inicial. En algunas realizaciones, se pueden generar muchas copias de polinucleótidos que contienen códigos de barras moleculares idénticos para cada molécula de polinucleótido diana única inicial marcada con un código de barras del vaso.

Tras la secuenciación, las secuencias con códigos de barras moleculares idénticos pueden aparearse o emparejarse. Tras la secuenciación, las secuencias con códigos de barras del vaso idénticos se pueden aparearse o emparejar. Tras la secuenciación, las secuencias con secuencias diana idénticas pueden aparearse o emparejarse. En algunas realizaciones, las lecturas de secuenciación pueden deshacerse en secuencias de consenso. Al deshacerse las lecturas de secuenciación apareadas o emparejadas en una secuencia de consenso se pueden reducir o eliminar la secuenciación y los errores de la PCR. La secuenciación se puede realizar usando un primer sitio de cebador para una primera lectura. La secuenciación se puede realizar usando el primer sitio de cebador para una segunda lectura. La secuenciación se puede realizar usando un segundo sitio de cebador para una segunda lectura.

Las cadenas pesadas y ligeras de anticuerpos que contienen los mismos códigos de barras del vaso pueden emparejarse y, en algunas realizaciones, clonarse en un sistema de vectores de mamíferos. La construcción de anticuerpo se puede expresar en otras estirpes de células hospedadoras humanas o de mamíferos. La construcción se puede validar luego mediante ensayos de transfección transitoria y análisis de transferencia Western del anticuerpo o TCR expresado de interés.

Los procedimientos de amplificación de ARN o ADN son bien conocidos en la técnica y se pueden usar de acuerdo con la presente invención sin experimentación indebida, basándose en las enseñanzas y orientaciones presentadas en el presente documento. Los procedimientos conocidos de amplificación de ADN o ARN incluyen, aunque sin limitación, la reacción en cadena de la polimerasa (PCR) y procesos de amplificación relacionados (véase, por ejemplo, las patentes de EE.UU. n.º 4.683.195, 4.683.202, 4.800.159, 4.965.188, de Mullis, *et al.*; documentos 4.795.699 y 4.921.794 de Tabor *et al.*; documento 5.142.033 de Innis; documento 5.122.464 de Wilson *et al.*; documento 5.091.310 de Innis; documento 5.066.584 de Gyllensten *et al.*; documento 4.889.818 de Gelfand *et al.*; documento 4.994.370 de Silver *et al.*; documento 4.766.067 de Biswas; 4.656.134 de Ringold) y la amplificación mediada por ARN que usa ARN no codificante para la secuencia diana como molde para la síntesis de ADN bicatenario (patente de EE.UU. N.º 5.130.238 de Malek *et al.*, con el nombre comercial NASBA). (Véase, por ejemplo, Ausubel, citado anteriormente; o Sambrook, citado anteriormente.)

De manera conveniente, las etapas del procedimiento que se describe en el presente documento, tales como la amplificación, la secuenciación y similares, pueden o no pueden llevarse a cabo en un formato de ensayo múltiple que emplee un fase sólida en la que se inmovilice una pluralidad de sustratos, por ejemplo, antígenos y similares, tal como una matriz. En algunas realizaciones, la matriz es un biochip de proteína. Usando biochips de proteína, se pueden analizar cientos e incluso miles de antígenos. Como se usan en el presente documento, "matriz", "micromatriz" o "biochip" se refiere a un sustrato sólido que tiene una superficie generalmente plana a la que se adhiere un adsorbente. Con frecuencia, la superficie del biochip comprende una pluralidad de ubicaciones direccionables, cada una de las cuales tiene el adsorbente unido allí. Los biochips se pueden adaptar para engranarse a una superficie de contacto de sonda y, por lo tanto, funcionan como sondas. Un "biochip de proteína" se refiere a un biochip adaptado para la captura de polipéptidos. Muchos biochips de proteínas se describen en la técnica. Los procedimientos para producir matrices de polipéptidos se describen, por ejemplo, en De Wildt *et al.*, 2000, *Nat. Biotechnol.* 18:989-994; Lueking *et al.*, 1999, *Anal. Biochem.* 270:103-111; Ge, 2000, *Nucleic Acids Res.* 28, e3, 1-VH; MacBeath y Schreiber, 2000, *Science* 289: 1760-1763; documento WO 01/40803 y documento WO 99/51773A1. El uso de matrices permite que

una serie de etapas,

tales como la detección, se realicen de forma robótica y/o de alto rendimiento. Los polipéptidos para la matriz se pueden detectar a alta velocidad, por ejemplo, usando un aparato de robótica disponible en el mercado, por ejemplo, de Genetic MicroSystems o BioRobotics. El sustrato de la matriz puede ser, por ejemplo, nitrocelulosa, plástico, vidrio, por ejemplo, vidrio modificado en la superficie. La matriz también puede incluir una matriz porosa, por ejemplo, acrilamida, agarosa u otro polímero. Tras la captura en un biochip, los analitos pueden detectarse mediante una variedad de procedimientos de detección seleccionados de, por ejemplo, un procedimiento de espectrometría de iones en fase gaseosa, un procedimiento óptico, un procedimiento electroquímico, microscopía de fuerza atómica y un procedimiento de radiofrecuencia. Es de particular interés el uso de la espectrometría de masas, y en particular, de SELDI. Los procedimientos ópticos incluyen, por ejemplo, la detección de fluorescencia, luminiscencia, quimioluminiscencia, absorbancia, reflectancia, transmitancia, birrefringencia o índice de refracción (p. ej., resonancia del plasmón superficial, elipsometría, un procedimiento de espejo resonante, un procedimiento de guía de onda con acoplamiento de rejilla o interferometría). Los procedimientos ópticos incluyen microscopía (confocal y no confocal), procedimientos de imagen y procedimientos de no imagen. Los inmunoensayos en varios formatos (por ejemplo, ELISA) son procedimientos populares para la detección de analitos capturados en una fase sólida. Los procedimientos electroquímicos incluyen procedimientos de voltametría y amperometría. Los procedimientos de radiofrecuencia incluyen la espectroscopia de resonancia multipolar.

En algunas realizaciones de la invención, por ejemplo, el enfoque de diversidad natural para preparar anticuerpos monoclonales, se emplean técnicas que se han establecido para trabajar con células individuales. Una técnica incorpora un accesorio especial que se puede usar en FACS para desviar células individuales a recipientes separados. Dichos accesorios están disponibles en el mercado y son bien conocidos en la técnica. Dichos accesorios son útiles para dispensar células individuales en compartimentos seleccionados de, por ejemplo, placas de cultivo de microtitulación convencionales de 96 pocillos. Como alternativa, las células pueden depositarse en una placa de microtitulación a una dilución limitante para garantizar la deposición de una sola célula.

Una segunda técnica es la PCR realizada en células inmunitarias individuales para amplificar los segmentos de  $V_H$  y  $V_L$ . En el enfoque de diversidad natural, se usa la PCR de una sola célula para retener el emparejamiento nativo de  $V_L$  y  $V_H$  en la célula individual. La especificidad de un anticuerpo está determinada por las regiones determinantes de la complementariedad (CDR) de dentro de la región  $V_L$  y la región  $V_H$ .

Los procedimientos para realizar la PCR de una sola célula son bien conocidos en la técnica (por ejemplo, Larrick, J. W. *et al.*, *Bio/Technology* 7:934 (1989)). Por ejemplo, los linfocitos B productores de anticuerpos del banco de linfocitos B o los linfocitos T productores de TCR del banco de linfocitos T pueden fijarse con una solución fijadora o una solución que contenga un producto químico tal como formaldehído, glutaraldehído o similares. Las células se permeabilizan entonces con una solución de permeabilización que comprende, por ejemplo, un detergente. El proceso de fijación y permeabilización debe proporcionar suficiente porosidad para permitir la entrada de enzimas, nucleótidos y otros reactivos en las células sin la destrucción indebida de los compartimentos celulares o los ácidos nucleicos. La adición de enzimas y nucleótidos puede luego entrar en las células para transcribir de forma inversa el ARNm celular de  $V_H$  y  $V_L$  o  $V_\alpha$  y  $V_\beta$  o  $V_\gamma$  y  $V_\delta$ , por ejemplo, en las secuencias de ADNc correspondientes. La transcripción inversa se puede realizar en una sola etapa, u opcionalmente, junto con un procedimiento de PCR, usando una transcriptasa inversa, cantidades suficientes de los cuatro dNTP y cebadores que se unan al ARNm proporcionando un grupo hidroxilo 3' para la transcriptasa inversa para iniciar la polimerización. Se puede usar cualquier cebador complementario al ARNm, pero se prefiere usar cebadores complementarios a un extremo 3' terminal de las moléculas de  $V_H$  y  $V_L$  o  $V_\alpha$  y  $V_\beta$  o  $V_\gamma$  y  $V_\delta$  para facilitar la selección del ARNm de la región variable. Numerosos estudios han indicado que se pueden preparar polinucleótidos degenerados para servir como los cebadores del extremo 5' para  $V_H$  y  $V_L$  o  $V_\alpha$  y  $V_\beta$  o  $V_\gamma$  y  $V_\delta$ . El procedimiento de banco combinatorio para la generación de moléculas de direccionamiento se basa en dichos cebadores. Además, numerosos experimentos han demostrado que la PCR puede amplificar los segmentos génicos de interés, tales como  $V_H$  y  $V_L$  o  $V_\alpha$  y  $V_\beta$  o  $V_\gamma$  y  $V_\delta$ , de una sola célula. Debido a la capacidad de trabajo incluso con una sola célula, este enfoque de PCR puede generar anticuerpos incluso cuando las células inmunitarias de interés se producen a una baja frecuencia.

En la realización de alta diversidad, tras la clasificación por FACS, se agrupan las células del banco de células inmunitarias y se realiza la PCR de transcripción inversa en todo el conjunto de células. La generación de ARNm para clonar anticuerpos o TCR se realiza fácilmente mediante procedimientos bien conocidos para la preparación y caracterización de anticuerpos o TCR (véase, por ejemplo, "Antibodies: A Laboratory Manual", 1988). Por ejemplo, el ARN total del banco de linfocitos B se extrae mediante procedimientos apropiados que son convencionales en la técnica. El ADNc se sintetiza a continuación a partir del ARN mediante procedimientos apropiados, por ejemplo, usando polinucleótidos hexámeros aleatorios, o cebadores específicos del gen C o la familia del gen C, o cebadores específicos del gen V o de la familia del gen V. De nuevo, estos son procesos conocidos por los expertos en la materia como se ha explicado anteriormente. Los bancos de moléculas de ácido nucleico derivados de bancos de linfocitos B o linfocitos T, por ejemplo, un banco de moléculas de ARN o ADNc derivadas de dichos linfocitos B o T, pueden clonarse en vectores de expresión para formar bancos de expresión. En algunas realizaciones, solo el dominio  $V_H$  o  $V_\alpha$  o  $V_\gamma$  derivado del banco de células inmunitarias se amplifica para generar un banco de dominios  $V_H$  o  $V_\alpha$  o  $V_\gamma$ . El banco de  $V_L$  o  $V_\beta$  o  $V_\delta$  de otra fuente se usa en combinación con el banco de  $V_H$  o  $V_\alpha$  o  $V_\gamma$  para generar anticuerpos o TCR usando los procedimientos descritos en el presente documento. Los bancos de fragmentos de anticuerpo o

TCR pueden construirse combinando los bancos de  $V_H$  y  $V_L$  o  $V\alpha$  y  $V\beta$  o  $V\gamma$  y  $V\delta$  juntos en cualquier número de formas conocidas por los expertos en la materia. Por ejemplo, cada banco se puede crear en diferentes vectores, y los vectores se recombinan *in vitro* o *in vivo*. Como alternativa, los bancos pueden clonarse secuencialmente en el mismo vector, o ensamblarse juntos por PCR y luego clonarse. También se puede usar el ensamblaje de PCR para unir los

5 ADN de  $V_H$  y  $V_L$  o  $V\alpha$  y  $V\beta$  o  $V\gamma$  y  $V\delta$  con el ADN que codifica un espaciador peptídico flexible para formar bancos de Fv monocatenarios (scFv) como se describe en otra parte del presente documento. En otra técnica más, el ensamblaje por PCR en la célula se usa para combinar los genes de  $V_H$  y  $V_L$  o  $V\alpha$  y  $V\beta$  o  $V\gamma$  y  $V\delta$  dentro de los linfocitos mediante PCR y luego clonar los repertorios de los genes enlazados.

## 10 **CÓDIGO DE BARRAS DE UNA SOLA CÉLULA**

Para codificar una sola célula con un código de barras del vaso y un código de barras molecular, se pueden crear vasos, tales como las emulsiones de agua en aceite, de manera que los vasos resultantes contengan 1 célula o menos por vaso. Los vasos se pueden crear de manera que los vasos resultantes también contengan 1 código de barras del

15 vaso por vaso. Los vasos pueden crearse de manera que los vasos resultantes también contengan 1 polinucleótido con código de barras molecular por vaso. Los vasos pueden crearse de manera que los vasos resultantes también contengan dos o más, o una pluralidad de, polinucleótidos con códigos de barras moleculares por vaso. Las células/los vasos pueden someterse a un protocolo de codificación con código de barras único de ARN o ADN como se describe en el presente documento, y el código de barras del vaso y uno o más códigos de barras moleculares de cada vaso

20 pueden fusionarse con una diana de interés, tal como un polinucleótido celular. En algunas realizaciones, los polinucleótidos con código de barras del vaso coincidente pueden fusionarse con componentes celulares presentes en el mismo vaso que el uno o más polinucleótidos con código de barras molecular. Después de la secuenciación, se puede usar la deconvolución del código de barras del vaso y del código de barras molecular para identificar qué ARN (o ADN) se originó a partir de qué célula. En algunas realizaciones, se pueden crear vasos, tales como las emulsiones

25 de agua en aceite, de manera que las emulsiones resultantes contengan 1 célula o más por emulsión. En algunas realizaciones, se pueden crear emulsiones de agua en aceite de manera que las emulsiones resultantes contengan 1 polinucleótido con código de barras del vaso y dos o más polinucleótidos con códigos de barras moleculares por vaso. En algunas realizaciones, se pueden crear vasos de manera que los vasos resultantes contengan más de 1 polinucleótido con código de barras del vaso y dos o más polinucleótidos con código de barras molecular por vaso. En

30 algunas realizaciones, se pueden introducir un código de barras del vaso y un código de barras molecular en los vasos cuando están en solución. En algunas realizaciones, se pueden introducir un código de barras del vaso y un código de barras molecular en los vasos cuando no están unidos a un soporte sólido, tal como una perla.

En algunos aspectos, se pueden aislar células individuales dentro de una emulsión, que puede actuar como un

35 compartimento. Las células pueden ser lisadas, y las transcripciones de la célula pueden recibir un código de barras. Cada una de las transcripciones puede fusionarse con un código de barras molecular o un código de barras del vaso, de manera que, cuando se detecten dos o más transcripciones de ARN con el mismo código de barras del vaso, se puede determinar que se originaron a partir de la misma célula de partida. Esto se puede aplicar a muchos tipos diferentes de secuencias. Una aplicación particular puede ser enlazar las cadenas  $V_H$  y  $V_L$  o  $V\alpha$  y  $V\beta$  o  $V\gamma$  y  $V\delta$  de las

40 secuencias de anticuerpos y TCR.

Se pueden aislar una o más células individuales en una o más emulsiones, en presencia de un código de barras del vaso y códigos de barras moleculares, de modo que una gotita de una o más emulsiones puede contener un máximo de 1 célula o menos. Las células pueden lisarse químicamente mediante un tampón contenido en una emulsión o

45 mediante congelación y descongelación, liberando así el contenido de una célula en una emulsión.

Los ARN de una sola célula se pueden transcribir de forma inversa en ADNc. Se puede realizar una reacción de transcripción inversa con una transcriptasa inversa que posea actividad transferasa terminal no molde que añada aproximadamente 3 restos de citosina como se ha descrito anteriormente. Todos los tampones de transcripción

50 inversa, enzimas y nucleótidos pueden estar presentes cuando se forma una emulsión. En algunas realizaciones, se puede generalizar un cebador (tal como un polinucleótido que comprenda una secuencia poli dT) para dirigirse a todo el ARNm. En algunas realizaciones, se puede usar ADN. En algunas realizaciones, el direccionamiento puede ser hacia más de 2 ARN.

En algunas realizaciones, se puede enlazar un código de barras del vaso a un ARN durante la transcripción inversa. En algunas realizaciones, se puede enlazar un código de barras molecular a un ARN durante la transcripción inversa. En algunas realizaciones, se pueden enlazar un código de barras del vaso y un código de barras molecular a un ARN

55 durante la transcripción inversa.

Se puede llevar a cabo una reacción de transcripción inversa en presencia de un polinucleótido de marcaje en 3'. Un polinucleótido de marcaje en 3' puede comprender un segmento P7 que puede usarse para la hibridación de un

60 cebador de secuenciación. Un polinucleótido de marcaje 3' puede comprender un código de barras del vaso o un código de barras molecular. Un polinucleótido de marcaje en 3' puede comprender 3 restos de ribo-guanina en un extremo 3' (rGrGrG) (bases de ARN) que pueden ser complementarios e hibridarse a una cadena producida por una

65 enzima de transcripción inversa. Por lo tanto, se pueden añadir un código de barras del vaso y un código de barras molecular a un extremo terminal de un ADNc en esta misma emulsión mediante enzimas de transcripción inversa. En

algunas realizaciones, se pueden usar restos de guanina en lugar de ribo-guanina (nucleótido de ADN en lugar de nucleótido de ARN). Tras la hibridación de un polinucleótido de marcaje en 3' a un CCC de una cadena de ADNc, una transcriptasa inversa continúa extendiendo un ADNc a un polinucleótido de marcaje en 3', creando así un marcador con código de barras molecular para todos los ADNc en una reacción. Tras la hibridación de un polinucleótido de marcaje en 3' a una región de un ADNc con código de barras molecular, una transcriptasa inversa o polimerasa continúa extendiendo un ADNc con código de barras molecular en otro polinucleótido de marcaje en 3', creando así un marcador con código de barras del vaso para todos los ADNc en una reacción. En algunas realizaciones, el cambio de molde se puede realizar en una reacción separada en lugar de realizarse al mismo tiempo que se pueda realizar una reacción de transcripción inversa. En algunas realizaciones, se puede añadir un polinucleótido de marcaje en 3' después de una reacción de transcripción inversa, y se pueden usar enzimas tales como una transcriptasa inversa o polimerasa para extenderse en un polinucleótido de marcaje de una manera similar. Dado que un polinucleótido de marcaje en 3' puede albergar un código de barras molecular degenerado único en cada molécula, cada ADNc se puede marcar de manera única con un código de barras molecular. Dado que un polinucleótido de marcaje en 3' puede albergar un mismo código de barras del vaso degenerado en cada molécula de un solo vaso, cada ADNc puede marcarse con un código de barras único para el vaso.

### CLONACIÓN Y EXPRESIÓN DEL MATERIAL GENÉTICO DEL BANCO DE LINFOCITOS B

"Banco de expresión de anticuerpos" o "banco de expresión de TCR" o "banco de expresión", como se usan en el presente documento, pueden referirse a una colección de moléculas (es decir, dos o más moléculas) bien a nivel de ácido nucleico o de proteína. Por lo tanto, este término puede referirse a una colección de vectores de expresión que codifican una pluralidad de moléculas de anticuerpo o TCR (es decir, a nivel de ácido nucleico) o puede referirse a una colección de moléculas de anticuerpo o de TCR tras haberse expresado en un sistema de expresión apropiado (es decir, a nivel de proteína). Como alternativa, los vectores de expresión/el banco de expresión pueden estar contenidos en células hospedadoras adecuadas en las que puedan expresarse. Las moléculas de anticuerpo que son codificadas o expresadas en los bancos de expresión de la invención pueden estar en cualquier formato apropiado, por ejemplo, pueden ser moléculas de anticuerpos o de TCR enteros, o pueden ser fragmentos de anticuerpos o de TCR, por ejemplo, anticuerpos monocatenarios (por ejemplo, anticuerpos scFv), anticuerpos Fv, anticuerpos Fab', fragmentos (Fab')<sub>2</sub>, diacuerpos, etc. La expresión "que codifica/n", como es la secuencia de ácido nucleico "que codifica" o una secuencia codificante de ADN de o una secuencia de nucleótidos "que codifica" una determinada enzima, así como otros términos sinónimos, se refiere a una secuencia de ADN que se transcribe y se traduce a una enzima cuando se coloca bajo el control de secuencias reguladoras apropiadas. Una "secuencia promotora" es una región reguladora del ADN capaz de unirse a la ARN polimerasa en una célula e iniciar la transcripción de una secuencia codificante cadena abajo (sentido 3'). El promotor es parte de la secuencia de ADN. Esta región de secuencia tiene un codón de inicio en su extremo 3'. La secuencia del promotor incluye el número mínimo de bases con elementos necesarios para iniciar la transcripción a niveles detectables por encima del fondo. Sin embargo, después de que la ARN polimerasa se une a la secuencia y se inicia la transcripción en el codón de inicio (extremo 3' con un promotor), la transcripción continúa en el sentido 3'. Dentro de la secuencia promotora se encontrará un sitio de inicio de la transcripción (definido convenientemente mediante cartografía con la nucleasa S1), así como los dominios de unión a proteínas (secuencias de consenso) responsables de la unión de la ARN polimerasa.

Las moléculas de anticuerpo o de TCR identificadas por, derivadas de, seleccionadas de o que se pueden obtener a partir de los bancos de expresión de anticuerpos o TCR de la invención forman un aspecto adicional más de la invención. De nuevo, estas moléculas de anticuerpo o de TCR pueden ser proteínas o ácidos nucleicos que codifican moléculas de anticuerpo o de TCR, cuyos ácidos nucleicos pueden incorporarse, a su vez, a un vector de expresión apropiado y/o estar contenidos en una célula hospedadora adecuada.

El conjunto de ADNc puede someterse a una reacción de PCR con polinucleótidos que se hibridan con una región constante de la cadena pesada de genes de anticuerpos y polinucleótidos que se hibridan con el extremo 5' de la región de la cadena V<sub>H</sub> o V<sub>α</sub> o V<sub>γ</sub> de genes de anticuerpo o TCR. El grupo de ADNc puede someterse a una reacción de PCR con polinucleótidos que se hibridan con una región constante de la cadena pesada, o cadena alfa o gamma de genes de anticuerpos o de TCR y polinucleótidos que se hibridan con la región 5' al extremo 5' de la región de la cadena V<sub>H</sub> o V<sub>α</sub> o V<sub>γ</sub> de un polinucleótido con código de barras que comprende una secuencia de anticuerpo o de TCR. También se puede configurar una reacción de PCR para la amplificación del grupo de cadenas V<sub>L</sub> o V<sub>β</sub> o V<sub>δ</sub> de, por ejemplo, las clases kappa y lambda. El conjunto de ADNc puede someterse a una reacción de PCR con polinucleótidos que se hibridan con una región constante de la cadena ligera de genes de anticuerpos y polinucleótidos que se hibridan con el extremo 5' de la región de la cadena V<sub>L</sub> o V<sub>β</sub> o V<sub>δ</sub> de genes de anticuerpo o TCR. El grupo de ADNc puede someterse a una reacción de PCR con polinucleótidos que se hibridan con una región constante de la cadena ligera de genes de anticuerpos y polinucleótidos que se hibridan con la región 5' al extremo 5' de la región de la cadena V<sub>L</sub> o V<sub>β</sub> o V<sub>δ</sub> de un polinucleótido con código de barras que comprende una secuencia de anticuerpo o de TCR. Dichos oligonucleótidos o cebadores pueden diseñarse basándose en la información de la base de datos de secuencias de genes de inmunoglobulina o TCR conocida y disponible para el público en general.

En algunas realizaciones, las secuencias de V<sub>H</sub> y V<sub>L</sub> o V<sub>α</sub> y V<sub>β</sub> o V<sub>γ</sub> y V<sub>δ</sub> pueden obtenerse convenientemente de un banco de secuencias de V<sub>H</sub> y V<sub>L</sub> o V<sub>α</sub> y V<sub>β</sub> o V<sub>γ</sub> y V<sub>δ</sub> producidas mediante la amplificación por PCR usando uno o más cebadores que no son específicos de los genes de cadena pesada o ligera y, en particular, para una o ambas

regiones terminales de los polinucleótidos  $V_H$  y  $V_L$  o  $V\alpha$  y  $V\beta$  o  $V\gamma$  y  $V\delta$ . En algunas realizaciones, las secuencias de  $V_H$  y  $V_L$  se pueden obtener convenientemente a partir de un banco de secuencias de  $V_H$  y  $V_L$  o  $V\alpha$  y  $V\beta$  o  $V\gamma$  y  $V\delta$  producidas mediante amplificación por PCR usando cebadores específicos de una región del polinucleótido con código de barras del vaso. En algunas realizaciones, las secuencias de  $V_H$  y  $V_L$  pueden obtenerse convenientemente de un

5 banco de secuencias de  $V_H$  y  $V_L$  o  $V\alpha$  y  $V\beta$  o  $V\gamma$  y  $V\delta$  producidas mediante la amplificación por PCR usando cebadores específicos de la familia del gen C o cebadores específicos del gen C. En algunas realizaciones, las secuencias de  $V_H$  y  $V_L$  pueden obtenerse convenientemente de un banco de secuencias de  $V_H$  y  $V_L$  o  $V\alpha$  y  $V\beta$  o  $V\gamma$  y  $V\delta$  producidas mediante la amplificación por PCR usando un conjunto de cebadores con un primer cebador específico de una región del polinucleótido con código de barras del vaso y un segundo cebador o una pluralidad de segundos cebadores que

10 son cebadores específicos de la familia del gen C o cebadores específicos del gen C. En algunas realizaciones, las secuencias de  $V_H$  y  $V_L$  o  $V\alpha$  y  $V\beta$  o  $V\gamma$  y  $V\delta$  pueden obtenerse convenientemente de un banco de secuencias de  $V_H$  y  $V_L$  o  $V\alpha$  y  $V\beta$  o  $V\gamma$  y  $V\delta$  producidas mediante la amplificación por PCR usando un conjunto de cebadores con un primer cebador específico de una región del polinucleótido con código de barras del vaso y un segundo cebador específico de una secuencia universal.

15 En algunas realizaciones, tras la transcripción inversa, las secuencias de ADNc resultantes pueden amplificarse mediante PCR usando uno o más cebadores específicos de los genes de inmunoglobulina y, en particular, para una o ambas regiones terminales de los polinucleótidos  $V_H$  y  $V_L$  o  $V\alpha$  y  $V\beta$  o  $V\gamma$  y  $V\delta$ . En algunas realizaciones, las secuencias de  $V_H$  y  $V_L$  pueden obtenerse de un banco de secuencias de  $V_H$  y  $V_L$  o  $V\alpha$  y  $V\beta$  o  $V\gamma$  y  $V\delta$  producidas mediante la amplificación por PCR usando cebadores específicos de la familia del gen V o cebadores específicos del gen V (Nicholls *et al.*, *J. Immunol. Meth.*, 1993, 165:81; documento WO93/12227) o diseñarse de acuerdo con procedimientos convencionales conocidos en la técnica basándose en la información de secuencias disponible. (Las secuencias de  $V_H$  y  $V_L$  o  $V\alpha$  y  $V\beta$  o  $V\gamma$  y  $V\delta$  se pueden ligar, en general, con una secuencia espaciadora intermedia (por ejemplo, que codifica un espaciador peptídico flexible en fase), formando un casete que codifica un anticuerpo monocatenario). Las secuencias de la región V pueden clonarse convenientemente como ADNc o productos de

20 amplificación por PCR para células de expresión de inmunoglobulina. Las regiones de  $V_H$  y  $V_L$  o  $V\alpha$  y  $V\beta$  o  $V\gamma$  y  $V\delta$  se secuencian, opcionalmente, en los procedimientos descritos en el presente documento y, en particular, tras ciertas etapas como se indica (por ejemplo, después de la PCR de una sola célula; tras la visualización en superficies de células de mamífero o de otro tipo, tras el rastreo FACS y similares). Se puede usar la secuenciación, entre otras

25 razones, para verificar que el nivel de diversidad se encuentre en un nivel aceptable. La secuenciación puede incluir la secuenciación de alto rendimiento, la secuenciación profunda (en la que se secuencia el mismo gen a partir de una pluralidad de muestras individuales para identificar diferencias en las secuencias) o combinaciones de las dos.

30 En algunas realizaciones, no es necesario unir físicamente las combinaciones naturales de  $V_H$  y  $V_L$  o  $V\alpha$  y  $V\beta$  o  $V\gamma$  y  $V\delta$  usando los procedimientos descritos en el presente documento. En algunas realizaciones, los ADNc, los polinucleótidos con código de barras o los ADNc con código de barras amplificados por PCR no están unidos físicamente. En algunas realizaciones, los ADNc, los polinucleótidos con código de barras o los ADNc con código de barras amplificados por PCR no se unen físicamente en la misma reacción o vaso.

35 En algunas realizaciones, las combinaciones naturales de  $V_H$  y  $V_L$  o  $V\alpha$  y  $V\beta$  o  $V\gamma$  y  $V\delta$  se unen físicamente, usando, además de los cebadores de ADNc, un cebador o una pluralidad de cebadores para el extremo 5' del gen  $V_H$  o  $V\alpha$  o  $V\gamma$  y otro cebador o pluralidad de cebadores para el extremo 5' del gen  $V_L$  o  $V\beta$  o  $V\delta$ . Estos cebadores también contienen colas complementarias de secuencia adicional, para permitir el autoensamblaje de los genes  $V_H$  y  $V_L$  o  $V\alpha$  y  $V\beta$  o  $V\gamma$  y  $V\delta$ . Después de la amplificación por PCR y la unión, la posibilidad de obtener productos mixtos, en otras

40 palabras, regiones variables mezcladas, es mínima, porque las reacciones de amplificación y unión se realizaron dentro de cada célula. Además, el riesgo de mezcla se puede reducir por la utilización de reactivos voluminosos tales como nucleótidos marcados con digoxigenina para garantizar además que los pares de ADNc de la región V no abandonan el compartimento celular y se entremezclan, sino que permanecen dentro de la célula para la amplificación por PCR y la unión. Las secuencias amplificadas se unen mediante la hibridación de secuencias terminales complementarias. Después de la unión, se pueden obtener secuencias de las células para usarlas en otras etapas del procedimiento que se describen en el presente documento. Por ejemplo, el ADN recuperado puede amplificarse por PCR usando cebadores terminales, si es necesario, y clonarse en vectores que pueden ser plásmidos, fagos, cósmidos, fagémidos, vectores víricos o combinaciones de los mismos, como se detalla a continuación. Se pueden incorporar sitios de enzimas de restricción convenientes en las secuencias hibridadas para facilitar la clonación. Estos

45 vectores también se pueden guardar como un banco de regiones variables unidas para su posterior uso.

50 En algunas realizaciones en las que se desean proporcionar combinaciones adicionales de  $V_H$  y  $V_L$  o  $V\alpha$  y  $V\beta$  o  $V\gamma$  y  $V\delta$ , se selecciona un sistema de expresión para facilitar esto. Por ejemplo, los sistemas de expresión de bacteriófagos permiten la recombinación aleatoria de secuencias de cadenas pesadas y ligeras. Los expertos en la materia conocen otros sistemas de expresión adecuados.

55 Cabe señalar que, en el caso de las secuencias de  $V_H$  y  $V_L$  o  $V\alpha$  y  $V\beta$  o  $V\gamma$  y  $V\delta$  no derivadas de ser humano, en algunas realizaciones, puede ser preferible quimerizar estas secuencias con un Fc completamente humano. Como se usa en el presente documento, "quimerización" se refiere a una inmunoglobulina o TCR, en la que las regiones variables de cadena pesada y ligera o  $V\alpha$  y  $V\beta$  o  $V\gamma$  y  $V\delta$  no son de origen humano y en la que las regiones constantes de cadena pesada y ligera o  $V\alpha$  y  $V\beta$  o  $V\gamma$  y  $V\delta$  son de origen humano. Esto se ve afectado al amplificar y clonar los

dominios variables en un Fc humano. El Fc humano puede ser parte del vector, o en una molécula separada, y también se podría usar el banco de Fc. En una realización preferida, las moléculas quimerizadas que crecen en células de mamífero tales como células CHO, se seleccionan con FACS dos veces para enriquecer la población celular en células que expresan el anticuerpo de interés. Los anticuerpos quimerizados o TCR se caracterizan, ya sea mediante secuenciación seguida de caracterización funcional, o mediante caracterización funcional directa o cinética. El crecimiento, la detección y la caracterización se describen en detalle a continuación.

Es importante señalar que las reacciones de PCR descritas anteriormente se describen para clonar los anticuerpos en forma de IgG. Se prefieren estos, ya que, en general, se asocian con una respuesta inmunitaria más madura y, en general, muestran una afinidad más alta que los anticuerpos IgM, lo que los hace más deseables para ciertas aplicaciones terapéuticas y de diagnóstico. Como es evidente, sin embargo, se pueden diseñar polinucleótidos que permitan la clonación de una o más de las otras formas de moléculas de inmunoglobulina, por ejemplo, IgM, IgA, IgE e IgD, si se desea o es apropiado.

Una vez que se ha identificado un anticuerpo o TCR, y la población apropiada de dichas células se ha aislado en un momento apropiado y, opcionalmente, enriquecido como se ha descrito anteriormente, no es necesario generar de inmediato los bancos de expresión de anticuerpos o TCR, siempre que el material genético contenido en las células se pueda mantener intacto, lo que permite la creación del banco en una fecha posterior. Por lo tanto, por ejemplo, las células, un lisado celular o un ácido nucleico, por ejemplo, ARN o ADN derivado de los mismos, pueden almacenarse hasta una fecha posterior mediante procedimientos apropiados, por ejemplo, mediante congelación, y generarse bancos de expresión en una fecha posterior cuando se desee.

Una vez que se ha generado el banco de vectores de expresión, las moléculas de anticuerpo codificadas pueden expresarse en un sistema de expresión apropiado y detectarse usando técnicas apropiadas que son bien conocidas y están documentadas en la técnica. Por lo tanto, el procedimiento de la invención definido anteriormente puede comprender las etapas adicionales de expresar el banco de vectores de expresión en un sistema de expresión apropiado y rastrear el banco expresado en busca de anticuerpos con propiedades deseadas, como se explica con más detalle a continuación.

Como se indica en el presente documento, los polinucleótidos preparados mediante los procedimientos de la divulgación que comprenden secuencias de anticuerpos o TCR que codifican polinucleótidos pueden incluir, aunque sin limitación, aquellos que codifican la secuencia de aminoácidos de un fragmento de anticuerpo o de TCR, por sí mismo, la secuencia no codificante para el anticuerpo o TCR completo, o una parte del mismo, la secuencia codificante para un anticuerpo o TCR, fragmento o porción, así como secuencias adicionales, tales como la secuencia codificante de al menos un péptido de fusión o líder señal, con o sin las secuencias codificantes adicionales mencionadas anteriormente, tales como al menos un intrón, junto con secuencias no codificantes adicionales, incluyendo, pero sin limitación, secuencias 5' y 3' no codificantes, tales como las secuencias transcritas, no traducidas, que desempeñan un papel en la transcripción, el procesamiento del ARNm, incluyendo las señales de corte y empalme y poliadenilación (por ejemplo, la unión al ribosoma y la estabilidad del ARNm); una secuencia codificante adicional que codifica aminoácidos adicionales, tales como los que proporcionan funcionalidades adicionales. Por lo tanto, la secuencia que codifica un anticuerpo puede fusionarse con una secuencia marcadora, tal como una secuencia codificante de un péptido que facilite la purificación del anticuerpo o TCR fusionado que comprende un fragmento o una porción de anticuerpo o TCR.

Los productos primarios de la PCR se pueden someter luego opcionalmente a una reacción de PCR secundaria con nuevos conjuntos de polinucleótidos que se hibridan con los extremos 5' y 3' de los dominios variables  $V_H$ ,  $V_L$  kappa y  $V_L$  lambda o  $V_\alpha$  y  $V_\beta$  o  $V_\gamma$  y  $V_\delta$  de los anticuerpos o TCR (según sea apropiado, dependiendo de si la reacción de PCR primaria con la que se usan los nuevos conjuntos de polinucleótidos se diseñó para amplificar porciones de los genes de anticuerpos de cadena pesada o ligera, o genes de TCR de  $V_\gamma$  o  $V_\delta$ ). Estos polinucleótidos incluyen ventajosamente secuencias de ADN específicas para un conjunto definido de enzimas de restricción (es decir, sitios de enzimas de restricción) para la posterior clonación. Las enzimas de restricción seleccionadas deben seleccionarse de modo que no se corten dentro de los segmentos de gen V de anticuerpo o TCR humanos. Dichos polinucleótidos pueden diseñarse en función de la información de la base de datos de secuencias de genes de inmunoglobulina o TCR y de enzimas de restricción conocida y disponible para el público en general. Sin embargo, los sitios de enzimas de restricción preferidos que se deben incluir son NcoI, Hind III, MluI y NotI. Los productos de dichas reacciones de PCR secundarias son repertorios de varios fragmentos/dominios de anticuerpos de V-pesada, V-ligera kappa y V-ligera lambda. Por lo tanto, este tipo de reacción de PCR secundaria, en general, se lleva a cabo cuando el formato de banco de expresión de interés es un formato scFv o Fv, en el que solo están presentes los dominios  $V_H$  y  $V_L$  o  $V_\alpha$  y  $V_\beta$  o  $V_\gamma$  y  $V_\delta$  de un anticuerpo o TCR.

Los productos de PCR también pueden someterse a una reacción de PCR con nuevos conjuntos de cebadores que se hibridan con los extremos 5' y 3' de los polinucleótidos con código de barras. Estos polinucleótidos pueden incluir ventajosamente secuencias de ADN específicas de un conjunto definido de enzimas de restricción (es decir, sitios de enzimas de restricción) para su posterior clonación. Las enzimas de restricción seleccionadas deben seleccionarse de modo que no se corten dentro de los segmentos de gen V de anticuerpo o TCR humanos. Dichos polinucleótidos pueden diseñarse en función de la información de la base de datos de secuencias de genes de inmunoglobulina o

TCR y de enzimas de restricción conocida y disponible para el público en general. Sin embargo, los sitios de enzimas de restricción preferidos que se deben incluir son NcoI, Hind III, MluI y NotI. Los productos de dichas reacciones de PCR secundarias son repertorios de diversos fragmentos/dominios de anticuerpos V<sub>H</sub>, V<sub>L</sub> kappa y V<sub>L</sub> lambda o fragmentos/dominios de TCR V<sub>α</sub> y V<sub>β</sub> o V<sub>γ</sub> y V<sub>δ</sub>.

5 Un experto en la materia reconocerá que, con este sistema, también se pueden usar fragmentos Fv o Fab de cadenas pesadas o ligeras, o cadenas V<sub>α</sub> o V<sub>β</sub> o cadenas V<sub>γ</sub> o V<sub>δ</sub>, o anticuerpos o TCR monocatenarios. Una cadena pesada o ligera, o una cadena V<sub>α</sub> o V<sub>β</sub>, o una cadena V<sub>γ</sub> o V<sub>δ</sub> se pueden mutagenizar, seguido de la adición de la cadena complementaria a la solución. Luego, se permite que las dos cadenas se combinen y formen un fragmento de anticuerpo funcional. La adición de secuencias aleatorias no específicas de cadena ligera o pesada, o V<sub>α</sub> o V<sub>β</sub>, o V<sub>γ</sub> o V<sub>δ</sub> permite la producción de un sistema combinatorio para generar un banco de diversos miembros.

15 Los bancos de dichos repertorios de fragmentos clonados que comprenden las regiones variables de cadena pesada o cadena V<sub>α</sub> o cadena V<sub>γ</sub>, o los fragmentos de las mismas, y/o las regiones variables de cadena ligera o cadena V<sub>β</sub> o cadena V<sub>δ</sub>, o fragmentos de las mismas, de genes de anticuerpos o TCR derivados de los linfocitos B o T de hospedadores con inmunodeficiencia como se definen en el presente documento forman aspectos adicionales de la invención. Estos bancos que comprenden regiones variables clonadas pueden insertarse opcionalmente en vectores de expresión para formar bancos de expresión.

20 En algunas realizaciones, las reacciones de PCR pueden configurarse para retener la totalidad o parte de las regiones constantes de las diversas cadenas de anticuerpos o TCR contenidas en la población de células inmunitarias aisladas. Esto es deseable cuando el formato del banco de expresión es un formato Fab, en el que el componente de cadena pesada o alfa o gamma comprende los dominios V<sub>H</sub> o V<sub>α</sub> o V<sub>γ</sub> y C<sub>H</sub> o C<sub>α</sub> o C<sub>γ</sub>, y el componente de cadena ligera o de cadena V<sub>β</sub> o de cadena V<sub>δ</sub> comprende los dominios de cadena V<sub>L</sub> o V<sub>β</sub> o V<sub>δ</sub> y C<sub>L</sub> o C<sub>β</sub> o C<sub>δ</sub>. De nuevo, los bancos de dichos fragmentos clonados que comprenden la totalidad o parte de las regiones constantes de cadenas de anticuerpo o TCR forman aspectos adicionales de la invención.

30 Estos ácidos nucleicos pueden comprender convenientemente secuencias además de un polinucleótido de la presente invención. Por ejemplo, un sitio de clonación múltiple que comprende uno o más sitios de restricción de endonucleasas puede insertarse en el ácido nucleico para ayudar al aislamiento del polinucleótido. Asimismo, se pueden insertar secuencias traducibles para ayudar en el aislamiento del polinucleótido traducido de la presente invención. Por ejemplo, una secuencia de marcador de hexa-histidina proporciona un medio conveniente para purificar las proteínas de la presente invención. El ácido nucleico de la presente invención, excluyendo la secuencia codificante, es opcionalmente un vector, adaptador o enlazador para la clonación y/o expresión de un polinucleótido de la presente invención.

35 Se pueden añadir secuencias adicionales a dichas secuencias de clonación y/o expresión para optimizar su función en la clonación y/o expresión, para ayudar al aislamiento del polinucleótido o para mejorar la introducción del polinucleótido en una célula. El uso de vectores de clonación, vectores de expresión, adaptadores y enlazadores es bien conocido en la técnica. (Véase, por ejemplo, Ausubel, citado anteriormente; o Sambrook, citado anteriormente).

45 Los bancos descritos en el presente documento pueden usarse en una variedad de aplicaciones. Como se usan en el presente documento, un banco comprende una pluralidad de moléculas. En algunas realizaciones, un banco comprende una pluralidad de polinucleótidos. En algunas realizaciones, un banco comprende una pluralidad de cebadores. En algunas realizaciones, un banco comprende una pluralidad de lecturas de secuencias de uno o más polinucleótidos, amplicones o conjuntos de amplicones. Un banco se puede almacenar y usar varias veces para generar muestras para su análisis. Algunas aplicaciones incluyen, por ejemplo, el genotipado de polimorfismos, el estudio del procesamiento del ARN y la selección de representantes clónicos para realizar la secuenciación de acuerdo con los procedimientos proporcionados en el presente documento. Pueden generarse bancos que comprenden una pluralidad de polinucleótidos, tales como cebadores, o bancos para secuenciación o amplificación, en los que una pluralidad de polinucleótidos comprende al menos aproximadamente 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1.000, 1.500, 2.000, 3.000, 4.000, 5.000, 6.000, 7.000, 8.000, 9.000, 10.000, 11.000, 12.000, 13.000, 14.000, 15.000, 16.000, 17.000, 18.000, 19.000, 20.000, 30.000, 40.000, 50.000, 60.000, 70.000, 80.000, 90.000, 100.000, 200.000, 300.000, 400.000, 500.000, 600.000, 700.000, 800.000, 900.000, 1.000.000, 50.000.000, 100.000.000 o más códigos de barras moleculares o códigos de barras del vaso. En algunas realizaciones, los bancos de polinucleótidos comprenden una pluralidad de al menos aproximadamente 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1.000, 1.500, 2.000, 3.000, 4.000, 5.000, 6.000, 7.000, 8.000, 9.000, 10.000, 11.000, 12.000, 13.000, 14.000, 15.000, 16.000, 17.000, 18.000, 19.000, 20.000, 30.000, 40.000, 50.000, 60.000, 70.000, 80.000, 90.000, 100.000, 200.000, 300.000, 400.000, 500.000, 600.000, 700.000, 800.000, 900.000, 1.000.000, 50.000.000, 100.000.000 o más polinucleótidos únicos, en los que cada polinucleótido único comprende uno o más códigos de barras moleculares y códigos de barras del vaso.

## 65 CÓDIGOS DE BARRAS

Un código de barras puede ser un código de barras molecular o un código de barras del vaso. En algunas realizaciones,

un código de barras, tal como un código de barras molecular o un código de barras del vaso, puede tener cada uno una longitud dentro de un intervalo de 2 a 36 nucleótidos, de 4 a 36 nucleótidos, o de 6 a 30 nucleótidos, o de 8 a 20 nucleótidos, de 2 a 20 nucleótidos, de 4 a 20 nucleótidos, o de 6 a 20 nucleótidos. En determinados aspectos, las temperaturas de fusión de los códigos de barras de un conjunto son de más/menos 10 °C entre sí, más/menos 5 °C entre sí o más/menos 2 °C entre sí. En determinados aspectos, las temperaturas de fusión de los códigos de barras de un conjunto no son de más/menos 10 °C entre sí, más/menos 5 °C entre sí o más/menos 2 °C entre sí. En otros aspectos, los códigos de barras son miembros de un conjunto de hibridación cruzada mínima. Por ejemplo, la secuencia de nucleótidos de cada miembro de dicho conjunto puede ser lo suficientemente diferente de la de todos los demás miembros del conjunto, por lo que ningún miembro puede formar un dúplex estable con el complemento de cualquier otro miembro en condiciones de hibridación rigurosas. En algunas realizaciones, la secuencia de nucleótidos de cada miembro de un conjunto de hibridación cruzada mínima difiere de las de cada otro miembro en al menos dos nucleótidos. Las tecnologías de códigos de barras se describen en Winzeler *et al.* (1999) *Science* 285:901; Brenner (2000) *Genome Biol.* 1:1 Kumar *et al.* (2001) *Nature Rev.* 2:302; Giaever *et al.* (2004) *Proc. Natl. Acad. Sci. EE.UU.* 101:793; Eason *et al.* (2004) *Proc. Natl. Acad. Sci. EE.UU.* 101:11046; y Brenner (2004) *Genome Biol.* 5:240.

Como se usan en el presente documento, un código de barras molecular comprende información que es única para una sola molécula de una sola célula o de un solo vaso, o dos o más moléculas de una pluralidad o de un banco de moléculas de dos o más células individuales, o de dos o más vasos individuales. Como se usan en el presente documento, un código de barras del vaso comprende información que es única para polinucleótidos de una sola célula o de un solo vaso, en comparación con polinucleótidos de una sola célula diferente o de un solo vaso diferente. En algunas realizaciones, la información única comprende una secuencia única de nucleótidos. Por ejemplo, la secuencia del código de barras molecular o del código de barras de un vaso puede determinarse mediante la determinación de la identidad y del orden de la secuencia única o aleatoria de nucleótidos que comprende el código de barras molecular o un código de barras del vaso. En algunas realizaciones, la información única no se puede usar para identificar la secuencia de un polinucleótido diana. Por ejemplo, un código de barras molecular se puede unir a un polinucleótido diana, pero el código de barras molecular no se puede usar para determinar el polinucleótido diana al que está unido. En algunas realizaciones, la información única no es una secuencia conocida enlazada a la identidad de la secuencia de un polinucleótido diana. Por ejemplo, un código de barras del vaso puede estar unido a uno o más polinucleótidos diana, pero el código de barras del vaso no se puede usar para determinar a cuál de los uno o más polinucleótidos diana está unido. En algunas realizaciones, la información única comprende una secuencia aleatoria de nucleótidos. En algunas realizaciones, la información única comprende una o más secuencias únicas de nucleótidos en un polinucleótido. En algunas realizaciones, la información única comprende una secuencia de nucleótidos degenerada o un código de barras degenerado. Un código de barras degenerado puede comprender una secuencia o composición de bases de nucleótidos variable. Por ejemplo, un código de barras degenerado puede ser una secuencia aleatoria. En algunas realizaciones, una secuencia de complemento de un código de barras molecular o un código de barras del vaso también es una secuencia de código de barras molecular o de código de barras del vaso.

Un código de barras molecular o un código de barras del vaso puede comprender cualquier longitud de nucleótidos. Por ejemplo, un código de barras molecular o un código de barras del vaso puede comprender al menos aproximadamente 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 60, 70, 80, 90, 100, 200, 500 o 1.000 nucleótidos. Por ejemplo, un código de barras molecular o un código de barras del vaso puede comprender como máximo aproximadamente 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 60, 70, 80, 90, 100, 200, 500 o 1.000 nucleótidos. En algunas realizaciones, un código de barras molecular o un código de barras del vaso tiene una determinada longitud de nucleótidos. Por ejemplo, un código de barras molecular o un código de barras del vaso pueden ser de aproximadamente 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 60, 70, 80, 90, 100, 200, 500 o 1.000 nucleótidos de longitud.

En algunas realizaciones, cada código de barras molecular o código de barras del vaso de entre una pluralidad de códigos de barras moleculares o códigos de barras del vaso tiene al menos aproximadamente 2 nucleótidos. Por ejemplo, cada código de barras molecular o un código de barras del vaso de entre una pluralidad de códigos de barras moleculares o códigos de barras del vaso pueden ser de al menos aproximadamente 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 60, 70, 80, 90, 100, 200, 500 o 1.000 nucleótidos de longitud. En algunas realizaciones, cada código de barras molecular o código de barras del vaso de entre una pluralidad de códigos de barras moleculares o códigos de barras del vaso tiene como máximo aproximadamente 1.000 nucleótidos. Por ejemplo, cada código de barras molecular o un código de barras del vaso de entre una pluralidad de códigos de barras moleculares o códigos de barras del vaso pueden ser de como máximo aproximadamente 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 60, 70, 80, 90, 100, 200, 500 o 1.000 nucleótidos de longitud. En algunas realizaciones, cada código de barras molecular o código de barras del vaso de entre una pluralidad de códigos de barras moleculares o códigos de barras del vaso tiene la misma longitud de nucleótidos. Por ejemplo, cada código de barras molecular o un código de barras del vaso de entre una pluralidad de códigos de barras moleculares o códigos de barras del vaso pueden ser de 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36,

- 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 60, 70, 80, 90, 100, 200, 500 o 1.000 nucleótidos de longitud. En algunas realizaciones, uno o más códigos de barras moleculares o códigos de barras del vaso de entre una pluralidad de códigos de barras moleculares o códigos de barras del vaso tienen una longitud diferente de nucleótidos. Por ejemplo, uno o más primeros códigos de barras moleculares o códigos de barras del vaso de entre una pluralidad de
- 5 códigos de barras moleculares o códigos de barras del vaso pueden tener aproximadamente o al menos aproximadamente 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 60, 70, 80, 90, 100, 200, 500 o 1.000 nucleótidos, y uno o más segundos códigos de barras moleculares o códigos de barras del vaso de entre una pluralidad de
- 10 códigos de barras moleculares o códigos de barras del vaso pueden tener aproximadamente o al menos aproximadamente 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 60, 70, 80, 90, 100, 200, 500 o 1.000 nucleótidos, en los que el número de nucleótidos de uno o más primeros códigos de barras moleculares o códigos de barras del vaso es diferente de el uno o más segundos códigos de barras moleculares o códigos de barras del vaso.
- 15 El número de códigos de barras moleculares puede ser superior al número total de moléculas que se deben marcar en una pluralidad de vasos. El número de códigos de barras del vaso puede ser superior al número total de moléculas que se deben marcar en una pluralidad de vasos. Por ejemplo, el número con códigos de barras moleculares o códigos de barras del vaso puede ser de al menos aproximadamente 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30, 40, 50, 60, 70, 80, 90 o 100 veces superior al número total de moléculas que se deben marcar en una pluralidad de vasos.
- 20 El número de códigos de barras moleculares diferentes puede exceder el número total de moléculas que se deben marcar en una pluralidad de vasos. En algunas realizaciones, el número de códigos de barras moleculares diferentes es al menos aproximadamente 1, 1,5, 2, 2,5, 3, 3,5, 4, 4,5, 5, 6, 7, 8, 9, 10, 15, 20, 30, 40, 50, 60, 70, 80, 90 o 100 veces superior al número total de moléculas que se deben marcar en una pluralidad de vasos.
- 25 El número de códigos de barras moleculares diferentes de un solo vaso puede ser superior al número de moléculas diferentes que se deben marcar en el vaso individual. En algunas realizaciones, el número de códigos de barras moleculares diferentes de un solo vaso es al menos aproximadamente 1, 1,5, 2, 2,5, 3, 3,5, 4, 4,5, 5, 6, 7, 8, 9, 10, 15, 20, 30, 40, 50, 60, 70, 80, 90 o 100 veces superior al número de moléculas diferentes que se deben marcar en el vaso individual.
- 30 El número de códigos de barras del vaso diferentes puede ser inferior al número total de moléculas que se deben marcar en una pluralidad de vasos. En algunas realizaciones, el número de códigos de barras del vaso diferentes es al menos aproximadamente 1, 1,5, 2, 2,5, 3, 3,5, 4, 4,5, 5, 6, 7, 8, 9, 10, 15, 20, 30, 40, 50, 60, 70, 80, 90 o 100 veces inferior al número total de moléculas que se deben marcar en una pluralidad de vasos.
- 35 El número de moléculas del producto amplificado de una molécula de polinucleótido con código de barras del vaso de un solo vaso puede ser superior al número de moléculas diferentes que se deben marcar en el vaso individual. En algunas realizaciones, el número de moléculas de producto amplificado de una molécula de polinucleótido con código de barras del vaso de un solo vaso es al menos aproximadamente 1, 1,5, 2, 2,5, 3, 3,5, 4, 4,5, 5, 6, 7, 8, 9, 10, 15, 20,
- 40 30, 40, 50, 60, 70, 80, 90 o 100 veces superior al número de moléculas diferentes que se deben marcar en el vaso individual.
- El número de moléculas de polinucleótido con código de barras del vaso de un solo vaso puede ser inferior al número de moléculas diferentes que se deben marcar en el vaso individual. En algunas realizaciones, el número de moléculas de polinucleótido con código de barras del vaso de un solo vaso es al menos aproximadamente 1, 1,5, 2, 2,5, 3, 3,5,
- 45 4, 4,5, 5, 6, 7, 8, 9, 10, 15, 20, 30, 40, 50, 60, 70, 80, 90 o 100 veces inferior al número de moléculas diferentes que se deben marcar en el vaso individual.
- El número de moléculas de polinucleótido con código de barras del vaso de un solo vaso puede ser una molécula. El número de moléculas de polinucleótido con código de barras del vaso no amplificado de un solo vaso puede ser una molécula.
- 50 En algunas realizaciones, al menos aproximadamente un 1 %, 2 %, 3 %, 4 %, 5 %, 6 %, 7 %, 8 %, 9 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 97 % o 100 % de los diferentes los códigos de barras moleculares tienen la misma concentración. En algunas realizaciones, al menos aproximadamente un 1 %, 2 %, 3 %, 4 %, 5 %, 6 %, 7 %, 8 %, 9 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 97 % o 100 % de los diferentes códigos de barras del vaso tienen la misma concentración.
- 55 En algunas realizaciones, al menos aproximadamente un 1 %, 2 %, 3 %, 4 %, 5 %, 6 %, 7 %, 8 %, 9 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 97 % o 100 % de los diferentes códigos de barras moleculares tienen una concentración diferente. En algunas realizaciones, al menos aproximadamente un 1 %, 2 %, 3 %, 4 %, 5 %, 6 %, 7 %, 8 %, 9 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 97 % o 100 % de los diferentes códigos de barras del vaso tienen una concentración diferente.
- 60 de barras del vaso tienen una concentración diferente.
- 65

Los códigos de barras moleculares o códigos de barras del vaso de una población con códigos de barras moleculares o códigos de barras del vaso pueden tener al menos 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1.000 o más secuencias diferentes. Por ejemplo, los códigos de barras moleculares o códigos de barras del vaso de una población pueden tener al menos 2.000, 3.000, 4.000, 5.000, 6.000, 7.000, 8.000, 9.000, 10.000, 15.000, 20.000, 25.000, 30.000, 35.000, 40.000, 45.000, 50.000, 60.000, 70.000, 80.000, 90.000, 100.000, 200.000, 300.000, 400.000, 500.000, 600.000, 700.000, 800.000, 900.000, 1.000.000 o más secuencias diferentes. Por lo tanto, se puede usar una pluralidad de códigos de barras moleculares o códigos de barras del vaso para generar al menos 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1.000 o más secuencias diferentes de uno o más polinucleótidos, tales como polinucleótidos diana. Por ejemplo, se puede usar una pluralidad de códigos de barras moleculares o códigos de barras del vaso para generar al menos 2.000, 3.000, 4.000, 5.000, 6.000, 7.000, 8.000, 9.000, 10.000, 15.000, 20.000, 25.000, 30.000, 35.000, 40.000, 45.000, 50.000, 60.000, 70.000, 80.000, 90.000, 100.000, 200.000, 300.000, 400.000, 500.000, 600.000, 700.000, 800.000, 900.000, 1.000.000, 2.000.000, 3.000.000, 4.000.000, 5.000.000, 6.000.000, 7.000.000, 8.000.000, 9.000.000, 10.000.000, 15.000.000, 20.000.000, 25.000.000, 30.000.000, 35.000.000, 40.000.000, 45.000.000, 50.000.000, 60.000.000, 70.000.000, 80.000.000, 90.000.000, 100.000.000, 200.000.000, 300.000.000, 400.000.000, 500.000.000, 600.000.000, 700.000.000, 800.000.000, 900.000.000, 1 x 10<sup>6</sup>, 2 x 10<sup>6</sup>, 3 x 10<sup>6</sup>, 4 x 10<sup>6</sup>, 5 x 10<sup>6</sup>, 6 x 10<sup>6</sup>, 7 x 10<sup>6</sup>, 8 x 10<sup>6</sup>, 9 x 10<sup>6</sup>, 1 x 10<sup>7</sup>, 2 x 10<sup>7</sup>, 3 x 10<sup>7</sup>, 4 x 10<sup>7</sup>, 5 x 10<sup>7</sup>, 6 x 10<sup>7</sup>, 7 x 10<sup>7</sup>, 8 x 10<sup>7</sup>, 9 x 10<sup>7</sup>, 1 x 10<sup>8</sup>, 2 x 10<sup>8</sup>, 3 x 10<sup>8</sup>, 4 x 10<sup>8</sup>, 5 x 10<sup>8</sup>, 6 x 10<sup>8</sup>, 7 x 10<sup>8</sup>, 8 x 10<sup>8</sup>, 9 x 10<sup>8</sup>, 1 x 10<sup>9</sup>, 2 x 10<sup>9</sup>, 3 x 10<sup>9</sup>, 4 x 10<sup>9</sup>, 5 x 10<sup>9</sup>, 6 x 10<sup>9</sup>, 7 x 10<sup>9</sup>, 8 x 10<sup>9</sup>, 9 x 10<sup>9</sup>, 1 x 10<sup>10</sup>, 2 x 10<sup>10</sup>, 3 x 10<sup>10</sup>, 4 x 10<sup>10</sup>, 5 x 10<sup>10</sup>, 6 x 10<sup>10</sup>, 7 x 10<sup>10</sup>, 8 x 10<sup>10</sup>, 9 x 10<sup>10</sup>, 1 x 10<sup>11</sup>, 2 x 10<sup>11</sup>, 3 x 10<sup>11</sup>, 4 x 10<sup>11</sup>, 5 x 10<sup>11</sup>, 6 x 10<sup>11</sup>, 7 x 10<sup>11</sup>, 8 x 10<sup>11</sup>, 9 x 10<sup>11</sup>, 1 x 10<sup>12</sup>, 2 x 10<sup>12</sup>, 3 x 10<sup>12</sup>, 4 x 10<sup>12</sup>, 5 x 10<sup>12</sup>, 6 x 10<sup>12</sup>, 7 x 10<sup>12</sup>, 8 x 10<sup>12</sup>, 9 x 10<sup>12</sup> o más secuencias diferentes de uno o más polinucleótidos, tales como polinucleótidos diana. Por ejemplo, se puede usar una pluralidad de códigos de barras moleculares o códigos de barras del vaso para generar al menos aproximadamente 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1.000, 2.000, 3.000, 4.000, 5.000, 6.000, 7.000, 8.000, 9.000, 10.000, 15.000, 20.000, 25.000, 30.000, 35.000, 40.000, 45.000, 50.000, 60.000, 70.000, 80.000, 90.000, 100.000, 200.000, 300.000, 400.000, 500.000, 600.000, 700.000, 800.000, 900.000, 1 x 10<sup>6</sup>, 2 x 10<sup>6</sup>, 3 x 10<sup>6</sup>, 4 x 10<sup>6</sup>, 5 x 10<sup>6</sup>, 6 x 10<sup>6</sup>, 7 x 10<sup>6</sup>, 8 x 10<sup>6</sup>, 9 x 10<sup>6</sup>, 1 x 10<sup>7</sup>, 2 x 10<sup>7</sup>, 3 x 10<sup>7</sup>, 4 x 10<sup>7</sup>, 5 x 10<sup>7</sup>, 6 x 10<sup>7</sup>, 7 x 10<sup>7</sup>, 8 x 10<sup>7</sup>, 9 x 10<sup>7</sup>, 1 x 10<sup>8</sup>, 2 x 10<sup>8</sup>, 3 x 10<sup>8</sup>, 4 x 10<sup>8</sup>, 5 x 10<sup>8</sup>, 6 x 10<sup>8</sup>, 7 x 10<sup>8</sup>, 8 x 10<sup>8</sup>, 9 x 10<sup>8</sup>, 1 x 10<sup>9</sup>, 2 x 10<sup>9</sup>, 3 x 10<sup>9</sup>, 4 x 10<sup>9</sup>, 5 x 10<sup>9</sup>, 6 x 10<sup>9</sup>, 7 x 10<sup>9</sup>, 8 x 10<sup>9</sup>, 9 x 10<sup>9</sup>, 1 x 10<sup>10</sup>, 2 x 10<sup>10</sup>, 3 x 10<sup>10</sup>, 4 x 10<sup>10</sup>, 5 x 10<sup>10</sup>, 6 x 10<sup>10</sup>, 7 x 10<sup>10</sup>, 8 x 10<sup>10</sup>, 9 x 10<sup>10</sup>, 1 x 10<sup>11</sup>, 2 x 10<sup>11</sup>, 3 x 10<sup>11</sup>, 4 x 10<sup>11</sup>, 5 x 10<sup>11</sup>, 6 x 10<sup>11</sup>, 7 x 10<sup>11</sup>, 8 x 10<sup>11</sup>, 9 x 10<sup>11</sup>, 1 x 10<sup>12</sup>, 2 x 10<sup>12</sup>, 3 x 10<sup>12</sup>, 4 x 10<sup>12</sup>, 5 x 10<sup>12</sup>, 6 x 10<sup>12</sup>, 7 x 10<sup>12</sup>, 8 x 10<sup>12</sup>, 9 x 10<sup>12</sup> o más secuencias diferentes de al menos aproximadamente 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1.000, 2.000, 3.000, 4.000, 5.000, 6.000, 7.000, 8.000, 9.000, 10.000, 15.000, 20.000, 25.000, 30.000, 35.000, 40.000, 45.000, 50.000, 60.000, 70.000, 80.000, 90.000, 100.000, 200.000, 300.000, 400.000, 500.000, 600.000, 700.000, 800.000, 900.000, 1 x 10<sup>6</sup>, 2 x 10<sup>6</sup>, 3 x 10<sup>6</sup>, 4 x 10<sup>6</sup>, 5 x 10<sup>6</sup>, 6 x 10<sup>6</sup>, 7 x 10<sup>6</sup>, 8 x 10<sup>6</sup>, 9 x 10<sup>6</sup>, 1 x 10<sup>7</sup>, 2 x 10<sup>7</sup>, 3 x 10<sup>7</sup>, 4 x 10<sup>7</sup>, 5 x 10<sup>7</sup>, 6 x 10<sup>7</sup>, 7 x 10<sup>7</sup>, 8 x 10<sup>7</sup>, 9 x 10<sup>7</sup>, 1 x 10<sup>8</sup>, 2 x 10<sup>8</sup>, 3 x 10<sup>8</sup>, 4 x 10<sup>8</sup>, 5 x 10<sup>8</sup>, 6 x 10<sup>8</sup>, 7 x 10<sup>8</sup>, 8 x 10<sup>8</sup>, 9 x 10<sup>8</sup>, 1 x 10<sup>9</sup>, 2 x 10<sup>9</sup>, 3 x 10<sup>9</sup>, 4 x 10<sup>9</sup>, 5 x 10<sup>9</sup>, 6 x 10<sup>9</sup>, 7 x 10<sup>9</sup>, 8 x 10<sup>9</sup>, 9 x 10<sup>9</sup>, 1 x 10<sup>10</sup>, 2 x 10<sup>10</sup>, 3 x 10<sup>10</sup>, 4 x 10<sup>10</sup>, 5 x 10<sup>10</sup>, 6 x 10<sup>10</sup>, 7 x 10<sup>10</sup>, 8 x 10<sup>10</sup>, 9 x 10<sup>10</sup>, 1 x 10<sup>11</sup>, 2 x 10<sup>11</sup>, 3 x 10<sup>11</sup>, 4 x 10<sup>11</sup>, 5 x 10<sup>11</sup>, 6 x 10<sup>11</sup>, 7 x 10<sup>11</sup>, 8 x 10<sup>11</sup>, 9 x 10<sup>11</sup>, 1 x 10<sup>12</sup>, 2 x 10<sup>12</sup>, 3 x 10<sup>12</sup>, 4 x 10<sup>12</sup>, 5 x 10<sup>12</sup>, 6 x 10<sup>12</sup>, 7 x 10<sup>12</sup>, 8 x 10<sup>12</sup>, 9 x 10<sup>12</sup> o más polinucleótidos diana.

En algunas realizaciones, se usan uno o más códigos de barras moleculares para agrupar o reunir secuencias. En algunas realizaciones, se usan uno o más códigos de barras moleculares para agrupar o reunir secuencias, en las que las secuencias de cada compartimento contienen el mismo código de barras molecular. En algunas realizaciones, se usan uno o más códigos de barras moleculares o códigos de barras del vaso para agrupar o reunir secuencias, en las que las secuencias de cada compartimento comprenden un conjunto de amplicones. En algunas realizaciones, se usan uno o más códigos de barras moleculares para agrupar o reunir secuencias, en las que las secuencias de cada compartimento comprenden una pluralidad de secuencias en la que los polinucleótidos a partir de los que se generó la pluralidad de secuencias derivaron de la misma molécula de polinucleótido en una reacción de amplificación.

En algunas realizaciones, se usan uno o más códigos de barras del vaso para agrupar o reunir secuencias. En algunas realizaciones, se usan uno o más códigos de barras del vaso para agrupar o reunir secuencias, en las que las secuencias de cada compartimento contienen el mismo código de barras del vaso. En algunas realizaciones, se usan uno o más códigos de barras del vaso para agrupar o reunir secuencias, en las que las secuencias de cada compartimento comprenden uno o más conjuntos de amplicones. En algunas realizaciones, se usan uno o más códigos de barras del vaso para agrupar o reunir secuencias, en las que las secuencias en cada compartimento comprenden una pluralidad de secuencias en la que los polinucleótidos a partir de los que se generó la pluralidad de secuencias se derivaron de los polinucleótidos de un solo vaso o de una sola célula.

En algunas realizaciones, se usan uno o más códigos de barras moleculares y códigos de barras del vaso para agrupar o reunir secuencias. En algunas realizaciones, se usan uno o más códigos de barras moleculares y códigos de barras del vaso para agrupar o reunir secuencias, en las que las secuencias en cada compartimento contienen el mismo código de barras molecular y el mismo código de barras del vaso. En algunas realizaciones, se usan uno o más códigos de barras moleculares y códigos de barras del vaso para agrupar o reunir secuencias, en las que las secuencias de cada compartimento comprenden uno o más conjuntos de amplicones. En algunas realizaciones, se usan uno o más códigos de barras moleculares y códigos de barras del vaso para agrupar o reunir secuencias, en las que las secuencias de cada compartimento comprenden una pluralidad de secuencias en la que los polinucleótidos a partir de los que se generó la pluralidad de secuencias se derivaron del mismo polinucleótido en una reacción de amplificación y de la misma célula o del mismo vaso individual. En algunas realizaciones, uno o más códigos de barras moleculares y códigos de barras del vaso no se usan para alinear secuencias.

En algunas realizaciones, uno o más códigos de barras moleculares no se usan para alinear secuencias. En algunas realizaciones, se usan uno o más códigos de barras moleculares para alinear secuencias. En algunas realizaciones, se usan uno o más códigos de barras moleculares para agrupar o reunir secuencias, y se usa una región específica de la diana para alinear secuencias. En algunas realizaciones, uno o más códigos de barras del vaso no se usan para alinear secuencias. En algunas realizaciones, se usan uno o más códigos de barras del vaso para agrupar o reunir secuencias, y se usa una región específica de destino para alinear secuencias. En algunas realizaciones, se usan uno o más códigos de barras del vaso para agrupar o reunir secuencias, y se usa una región específica de la diana para alinear secuencias.

En algunas realizaciones, las secuencias alineadas contienen el mismo código de barras molecular. En algunas realizaciones, las secuencias alineadas contienen el mismo código de barras del vaso. En algunas realizaciones, las secuencias alineadas contienen el mismo código de barras molecular y código de barras del vaso. En algunas realizaciones, se usan uno o más códigos de barras moleculares o códigos de barras del vaso para alinear secuencias, en las que las secuencias alineadas comprenden dos o más secuencias de un conjunto de amplicones. En algunas realizaciones, se usan uno o más códigos de barras moleculares o códigos de barras del vaso para alinear secuencias, en las que las secuencias alineadas comprenden una pluralidad de secuencias en la que los polinucleótidos a partir de los que se generó la pluralidad de secuencias derivaron de la misma molécula de polinucleótido en una reacción de amplificación. En algunas realizaciones, se usan uno o más códigos de barras moleculares o códigos de barras del vaso para alinear secuencias, en las que las secuencias alineadas comprenden una pluralidad de secuencias en la que los polinucleótidos a partir de los que se generó la pluralidad de secuencias se derivaron de una sola célula o un solo vaso.

## 25 GENERACIÓN DE GOTITAS

La división de una muestra de una pluralidad de células en pequeños volúmenes de reacción, asociados al código de barras molecular y de vaso de polinucleótidos de, o derivados de, una célula individual de la pluralidad de células puede permitir una secuenciación de alto rendimiento de un repertorio de secuencias, tales como las secuencias de biomarcadores.

La división de una muestra de una pluralidad de células en pequeños volúmenes de reacción, asociados al código de barras molecular y de vaso de polinucleótidos de, o derivados de, una célula individual de la pluralidad de células puede permitir una secuenciación de alto rendimiento de un repertorio de secuencias, tales como secuencias que representan un porcentaje del transcriptoma de un organismo. Por ejemplo, un repertorio de secuencias puede comprender una pluralidad de secuencias que representan al menos aproximadamente el 0,00001 %, 0,00005 %, 0,00010 %, 0,00050 %, 0,001 %, 0,005 %, 0,01 %, 0,05 %, 0,1 %, 0,5 %, 1 %, 2 %, 2,5 %, 3 %, 3,5 %, 4 %, 4,5 %, 5 %, 6 %, 7 %, 8 %, 9 %, 10 %, 15 %, 20 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 98 %, 99 % o 100 % del transcriptoma de un organismo.

La división de una muestra de células inmunitarias en pequeños volúmenes de reacción, asociados al código de barras molecular y de vaso de polinucleótidos de, o derivados de, una célula inmunitaria individual de la pluralidad de células inmunitarias puede permitir la secuenciación de alto rendimiento de un repertorio de secuencias de cadenas pesadas y ligeras. Estos procedimientos también pueden permitir el emparejamiento de las cadenas pesadas y ligeras después de la secuenciación basada en las secuencias con códigos de barras. La división de una muestra en pequeños volúmenes de reacción como se describe en el presente documento también puede permitir el uso de cantidades reducidas de reactivos, lo que reduce el coste de materiales del análisis.

En algunos casos, la reacción de transcripción inversa y/o la reacción de amplificación (por ejemplo, PCR) se llevan a cabo en gotitas, tales como en la PCR digital de gotitas. En determinados aspectos, la invención proporciona compartimentos fluidicos para contener todo o una parte de un material diana. En algunas realizaciones, un compartimento es una gotita. Aunque se hace referencia a "gotitas" a lo largo de la memoria descriptiva, ese término se usa indistintamente con el compartimento de fluido y la partición de fluido a menos que se indique lo contrario. Salvo los casos en los que se indique lo contrario, "gotita" se usa por conveniencia, y se puede usar cualquier partición o compartimento de fluido. Las gotitas usadas en el presente documento pueden incluir composiciones de emulsión (o mezclas de dos o más fluidos inmiscibles), tal como se describe en la patente de EE.UU. n.º 7.622.280. Las gotitas se pueden generar mediante dispositivos descritos en el documento WO/2010/036352. El término emulsión, como se usa en el presente documento, puede referirse a una mezcla de líquidos inmiscibles (tal como aceite y agua). Las emulsiones en fase oleosa y/o de agua en aceite permiten la compartimentación de mezclas de reacción dentro de gotitas acuosas. Las emulsiones pueden comprender gotitas acuosas dentro de una fase oleosa continua. Las emulsiones proporcionadas en el presente documento pueden ser emulsiones de aceite en agua, en las que las gotitas son gotitas de aceite dentro de una fase acuosa continua. Las gotitas proporcionadas en el presente documento están diseñadas para evitar la mezcla entre compartimentos, protegiendo cada compartimento su contenido de la evaporación y uniéndose con el contenido de otros compartimentos.

Las mezclas o emulsiones descritas en el presente documento pueden ser estables o inestables. Las emulsiones

pueden ser relativamente estables y tener una coalescencia mínima. La coalescencia se produce cuando las gotitas pequeñas se combinan para formar progresivamente gotas más grandes. En algunos casos, menos del 0,00001 %, 0,00005 %, 0,00010 %, 0,00050 %, 0,001 %, 0,005 %, 0,01 %, 0,05 %, 0,1 %, 0,5 %, 1 %, 2 %, 2,5 %, 3 %, 3,5 %, 4 %, 4,5 %, 5 %, 6 %, 7 %, 8 %, 9 % o 10% de las gotitas generadas por un generador de gotitas se une con otras gotitas. Las emulsiones también pueden tener una floculación limitada, un proceso mediante el que la fase dispersa sale de la suspensión en escamas.

Se pueden generar gotitas con un diámetro medio de aproximadamente, menos de aproximadamente, o más de aproximadamente, o de al menos aproximadamente 0,001, 0,01, 0,05, 0,1, 1, 5, 10, 20, 30, 40, 50, 60, 70, 80, 100, 120, 130, 140, 150, 160, 180, 200, 300, 400 o 500 micrómetros. Las gotitas pueden tener un diámetro medio de aproximadamente 0,001 a aproximadamente 500, de aproximadamente 0,01 a aproximadamente 500, de aproximadamente 0,1 a aproximadamente 500, de aproximadamente 0,1 a aproximadamente 100, de aproximadamente 0,01 a aproximadamente 100 o de aproximadamente 1 a aproximadamente 100 micrómetros. Se sabe que los procedimientos microfluídicos para producir gotitas de emulsión usando el enfoque de flujo cruzado de microcanales o la agitación física producen emulsiones monodispersas o polidispersas. Las gotitas pueden ser gotitas monodispersas. Las gotitas se pueden generar de modo que el tamaño de las gotitas no varíe en más de más/menos el 5 % del tamaño medio de las gotitas. En algunos casos, las gotitas se generan de manera que el tamaño de las gotitas no varíe en más de más/menos el 2 % del tamaño medio de las gotitas. Un generador de gotitas puede generar una población de gotitas a partir de una sola muestra, en la que ninguna de las gotitas varía en tamaño en más de más/menos aproximadamente el 0,1%, 0,5 %, 1 %, 1,5 %, 2 %, 2,5 %, 3 %, 3,5 %, 4 %, 4,5 %, 5 %, 5,5 %, 6 %, 6,5 %, 7 %, 7,5 %, 8 %, 8,5 %, 9 %, 9,5 % o 10 % del tamaño medio de la población total de gotitas.

Una mayor estabilidad mecánica puede ser útil para manipulaciones microfluídicas y procesamiento fluido de mayor cizalla (por ejemplo, en capilares microfluídicos o mediante giros de 90 grados, tales como válvulas, en la trayectoria fluidica). Las gotitas o cápsulas antes y después de tratarse térmicamente pueden ser estables mecánicamente a las manipulaciones con pipetas y a la centrifugación convencionales.

Una gotita puede formarse haciendo fluir una fase oleosa a través de una muestra acuosa. La fase acuosa puede comprender una solución tamponada y reactivos para realizar una reacción de amplificación, que incluyen células, nucleótidos, análogos de nucleótidos, polinucleótidos con códigos de barras moleculares, cebadores de polinucleótidos con códigos de barras del vaso, ácidos nucleicos de molde y enzimas, tales como una ADN polimerasa, ARN polimerasa y/o transcriptasa inversa.

La fase acuosa puede comprender una solución tamponada y reactivos para realizar una reacción de amplificación con o sin una superficie sólida, tal como una perla. La solución tamponada puede comprender aproximadamente, más que aproximadamente o menos de aproximadamente 1, 5, 10, 15, 20, 30, 50, 100 o 200 mM de Tris. En algunos casos, la concentración de cloruro de potasio puede ser de aproximadamente, más de aproximadamente o menos de aproximadamente 10, 20, 30, 40, 50, 60, 80, 100, 200 mM. La solución tamponada puede comprender Tris aproximadamente 15 mM y KCl aproximadamente 50 mM. Los nucleótidos pueden comprender moléculas de desoxirribonucleótido trifosfato, que incluyen dATP, dCTP, dGTP y dTTP, a concentraciones de aproximadamente, más de aproximadamente, o menos de aproximadamente 50, 100, 200, 300, 400, 500, 600 o 700  $\mu$ M cada una. En algunos casos, el dUTP se añade dentro de la fase acuosa a una concentración de aproximadamente, más de aproximadamente o menos de aproximadamente 50, 100, 200, 300, 400, 500, 600 o 700, 800, 900 o 1.000  $\mu$ M. En algunos casos, se añade cloruro de magnesio o acetato de magnesio ( $MgCl_2$ ) a la fase acuosa a una concentración de aproximadamente, más de aproximadamente o menos de aproximadamente 1,0, 2,0, 3,0, 4,0 o 5,0 mM. La concentración de  $MgCl_2$  puede ser de aproximadamente 3,2 mM. En algunos casos, se usa acetato de magnesio o magnesio. En algunos casos, se usa sulfato de magnesio.

Se puede usar un agente bloqueador no específico, tal como BSA o gelatina de piel bovina, en el que la gelatina o la BSA está presente en un intervalo de concentraciones del aproximadamente 0,1 al 0,9 % p/v. Otros posibles agentes bloqueadores pueden incluir betalactoglobulina, caseína, leche en polvo u otros agentes bloqueadores comunes. En algunos casos, las concentraciones preferidas de BSA y gelatina son del aproximadamente 0,1 % p/v.

Los cebadores para la amplificación dentro de la fase acuosa pueden tener una concentración de aproximadamente, más de aproximadamente o menos de aproximadamente 0,05, 0,1, 0,2, 0,3, 0,4, 0,5, 0,6, 0,7, 0,8, 0,9, 1,0, 1,2, 1,5, 1,7 o 2,0  $\mu$ M. La concentración del cebador dentro de la fase acuosa puede ser de aproximadamente 0,05 a aproximadamente 2, de aproximadamente 0,1 a aproximadamente 1,0, de aproximadamente 0,2 a aproximadamente 1,0, de aproximadamente 0,3 a aproximadamente 1,0, de aproximadamente 0,4 a aproximadamente 1,0 o de aproximadamente 0,5 a aproximadamente 1,0  $\mu$ M. La concentración de los cebadores puede ser de aproximadamente 0,5  $\mu$ M. Los intervalos proporcionables para las concentraciones de ácido nucleico diana en la PCR incluyen, pero sin limitación, entre aproximadamente 1 pg y aproximadamente 500 ng.

En algunos casos, la fase acuosa también puede comprender aditivos que incluyen, pero sin limitación, ácidos nucleicos de bloqueo/de fondo no específicos (por ejemplo, ADN de esperma de salmón), bioconservantes (por ejemplo, azida de sodio), potenciadores de la PCR (por ejemplo, betaína, trehalosa, etc.) e inhibidores (por ejemplo, inhibidores de la ARNsa). Otros aditivos pueden incluir, por ejemplo, dimetilsulfóxido (DMSO), glicerol, (mono)hidrato

de betaína (*N,N,N*-trimetilglicina = [carboxi-metil]trimetilamonio), trehalosa, 7-desaza-2'-desoxiguanosina trifosfato (dC7GTP o 7-desaza-2'-dGTP), BSA (albúmina de suero bovino), formamida (metanamida), cloruro de tetrametilamonio (TMAC), otros derivados de tetralquilamonio (por ejemplo, cloruro de tetraetilamonio (TEA-Cl) y cloruro de tetrapropilamonio (TPrA-Cl), detergente no iónico (por ejemplo, Triton X-100, Tween 20, Nonidet P-40 (NP-40)) o PREXCEL-Q. En algunos casos, la fase acuosa puede comprender 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10 aditivos diferentes. En otros casos, la fase acuosa puede comprender al menos 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10 aditivos diferentes.

En algunos casos, se puede añadir un copolímero de bloques de óxido de etileno/óxido de propileno no iónico a la fase acuosa a una concentración del aproximadamente 0,1 %, 0,2 %, 0,3 %, 0,4 %, 0,5 %, 0,6 %, 0,7 %, 0,8 %, 0,9 % o 1,0 %. Los biotensioactivos comunes incluyen tensioactivos no iónicos tales como Pluronic F-68, Tetronics y Zonyl FSN. Pluronic F-68 puede estar presente a una concentración del aproximadamente 0,5 % p/v.

En algunos casos, el cloruro de magnesio se puede sustituir por el sulfato de magnesio, a concentraciones similares. La solución tamponada se puede sustituir por una amplia selección de tampones de PCR comerciales, comunes, de diversos proveedores.

La emulsión se puede formular para producir gotitas altamente monodispersas que tengan una película interfacial de tipo líquido que se puede convertir calentando en microcápsulas que tienen una película interfacial de tipo sólido; dichas microcápsulas pueden comportarse como biorreactores capaces de retener su contenido a través de un proceso de reacción tal como la amplificación por PCR. La conversión en la forma de microcápsula puede ocurrir tras el calentamiento. Por ejemplo, dicha conversión puede ocurrir a una temperatura superior a aproximadamente 50 °C, 60 °C, 70 °C, 80 °C, 90 °C o 95 °C. En algunos casos, este calentamiento se produce mediante un termociclador. Durante el proceso de calentamiento, se puede usar una capa de fluido o aceite mineral para evitar la evaporación. El exceso de aceite de fase continua puede o no puede eliminarse antes del calentamiento. Las cápsulas biocompatibles pueden ser resistentes a la coalescencia y/o floculación en una amplia selección de procesos térmicos y mecánicos. Tras la conversión, las cápsulas se pueden almacenar a aproximadamente, más de aproximadamente o menos de aproximadamente 3 °C, 4 °C, 5 °C, 6 °C, 7 °C, 8 °C, 9 °C, 10 °C, 15 °C, 20 °C, 25 °C, 30 °C, 35 °C o 40 °C. Estas cápsulas pueden ser útiles en aplicaciones biomédicas, tales como la encapsulación digitalizada y estable de macromoléculas, en particular, fluidos biológicos acuosos que contienen una mezcla de ácidos nucleicos o proteínas, o ambos juntos; Administración de fármacos y vacunas; bancos biomoleculares; aplicaciones clínicas de generación de imágenes y otras.

Las microcápsulas pueden contener uno o más polinucleótidos, y pueden resistir la coalescencia, en particular, a altas temperaturas. Por consiguiente, las reacciones de amplificación por PCR pueden producirse a una densidad muy alta (por ejemplo, número de reacciones por unidad de volumen). En algunos casos, pueden producirse más de 100.000, 500.000, 1.000.000, 1.500.000, 2.000.000, 2.500.000, 5.000.000 o 10.000.000 por ml. En algunos casos, las reacciones se producen en un solo pocillo, por ejemplo, un pocillo de una placa de microtitulación, sin mezclarse entre los volúmenes de reacción. Las microcápsulas también pueden contener otros componentes necesarios para permitir que se produzca una transcripción inversa, una extensión del cebador y/o una reacción de PCR, por ejemplo, cebadores, sondas, dNTP, ADN o ARN polimerasas, etc. Estas cápsulas presentan resistencia a la coalescencia y a la floculación a través de una amplia selección de procesos térmicos y mecánicos.

En algunos casos, la etapa de amplificación se lleva a cabo mediante la realización de una PCR digital, tal como una PCR digital microfluídica o una PCR digital de gotitas.

Las gotitas se pueden generar usando sistemas o dispositivos microfluídicos. Como se usan en el presente documento, el prefijo "micro" (por ejemplo, como "microcanal" o "microfluídico"), en general, se refiere a elementos o artículos que tienen anchos o diámetros de menos de aproximadamente 1 mm y menos de aproximadamente 100 micrómetros en algunos casos. En algunos casos, el elemento o artículo incluye un canal a través del que puede fluir un fluido. Además, "microfluídico", como se usa en el presente documento, se refiere a un dispositivo, aparato o sistema que incluye al menos un canal a microescala.

Los sistemas y dispositivos microfluídicos se han descrito en una variedad de contextos, normalmente, en el contexto del análisis de laboratorio miniaturizado (por ejemplo, clínico). También se han descrito otros usos. Por ejemplo, las publicaciones de solicitud de patente internacional n.º WO 01/89788; WO 2006/040551; WO 2006/040554; WO 2004/002627; WO 2008/063227; WO 2004/091763; WO 2005/021151; WO 2006/096571; WO 2007/089541; WO 2007/081385 y WO 2008/063227.

Una gotita generalmente incluye una cantidad de un primer fluido de muestra en un segundo fluido portador. Con los procedimientos de la invención, se puede usar cualquier técnica conocida en la materia para formar gotitas. Un procedimiento ilustrativo implica hacer fluir una corriente del fluido de muestra que contiene el material diana (por ejemplo, una célula inmunitaria) de manera que interseccione con dos corrientes opuestas de fluido portador que fluye. El fluido portador es inmiscible con el fluido de muestra. La intersección del fluido de muestra con las dos corrientes opuestas del fluido portador que fluye produce la división del fluido de muestra en gotitas de muestra individuales que contienen el material diana.

El fluido portador puede ser cualquier fluido que sea inmiscible con el fluido de muestra. Un fluido portador ilustrativo

es el aceite. En determinadas realizaciones, el fluido portador incluye un tensioactivo.

El mismo procedimiento puede aplicarse para crear gotitas individuales que contienen otros reactivos, tales como los reactivos para una reacción de amplificación, tales como la reacción en cadena de la polimerasa (PCR), o una reacción de amplificación no basada en la PCR, tal como la amplificación por desplazamiento de múltiples cadenas u otros procedimientos conocidos por los expertos en la materia. Los reactivos adecuados para realizar reacciones de amplificación basadas en PCR son conocidos por los expertos en la materia, e incluyen, aunque sin limitación, ADN polimerasas, cebadores directos e inversos, trifosfatos de desoxinucleótidos (dNTP) y uno o más tampones.

En determinadas realizaciones, se forman compartimentos fluidicos proporcionando una primera partición de fluido (por ejemplo, una gotita) que comprende un material diana (por ejemplo, una célula inmunitaria y/o un soporte sólido tal como una perla) y un segundo fluido (por ejemplo, como corriente fluida o dentro de gotitas). El primer y segundo fluido se fusionan para formar una gotita. La fusión se puede lograr mediante la aplicación de un campo eléctrico a los dos fluidos. En determinadas realizaciones, el segundo fluido contiene reactivos para llevar a cabo una reacción de amplificación, tal como una reacción en cadena de la polimerasa o una reacción de amplificación.

En determinados aspectos, la invención proporciona un procedimiento para crear un banco de secuencias de anticuerpos de cadena pesada y ligera y/o secuencias de TCR de cadena alfa y beta y/o secuencias de TCR de cadena gamma y delta con un código de barras único que incluye la obtención de una pluralidad de construcciones de ácido nucleico en la que cada construcción incluye un N-mero único y un N-mero funcional. El N-mero funcional puede ser un N-mero aleatorio, un cebador de PCR, un cebador universal, un anticuerpo, un extremo pegajoso o cualquier otra secuencia. El procedimiento puede incluir la creación de M conjuntos de un número N de compartimentos de fluido, cada uno de los cuales contiene una o más copias de una construcción única. El procedimiento puede crear bancos de códigos de barras de mayor complejidad mediante la adición de una construcción adicional a cada compartimento de un conjunto, y su repetición para que cada conjunto con el fin de producir N x M compartimentos, conteniendo cada uno un par único de construcciones. Los pares se pueden hibridar o ligar para producir nuevas construcciones. En cada construcción en un banco de códigos de barras, cada N-mero único puede adaptarse para su identificación por secuenciación, hibridación de sondas, otros procedimientos o una combinación de procedimientos.

#### BANCOS DE GOTITAS

En general, un banco de gotitas está formado por una serie de elementos del banco que se agrupan en una sola colección. Los bancos pueden variar en cuanto a su complejidad desde un único elemento del banco a  $1 \times 10^{15}$  elementos del banco o más. Cada elemento del banco es uno o más componentes dados a una concentración fija. El elemento puede ser, pero sin limitación, células, perlas, aminoácidos, proteínas, polipéptidos, ácidos nucleicos, polinucleótidos o compuestos químicos de moléculas pequeñas. El elemento puede contener un identificador, tal como un código de barras molecular, un código de barras del vaso o ambos.

Un elemento del banco de células puede incluir, pero sin limitación, hibridomas, linfocitos B, linfocitos T, células primarias, líneas celulares cultivadas, células cancerosas, células madre o cualquier otro tipo de célula. Los elementos del banco celular se preparan encapsulando una serie de células, desde una a decenas de miles, en gotitas individuales. El número de células encapsuladas generalmente viene dado por las estadísticas de Poisson a partir de la densidad numérica de las células y del volumen de la gotita. Sin embargo, en algunos casos, el número se desvía de las estadísticas de Poisson como se describe en Edd *et al.*, "Controlled encapsulation of single-cells into monodisperse picolitre drops". *Lab Chip*, 8(8): 1262-1264, 2008. La naturaleza diferenciada de las células permite la preparación de los bancos en masa con una pluralidad de variantes celulares, tales como las células inmunitarias que producen un anticuerpo o TCR cada una, todas presentes en un solo medio de partida, y luego ese medio se divide en cápsulas de gotitas individuales que contienen una célula como máximo. Las células dentro de las gotitas individuales se lisan, los polinucleótidos de cadena pesada y ligera y/o los polinucleótidos de cadena alfa y beta y/o los polinucleótidos de cadena gamma y delta de las células lisadas se codifican con códigos de barras moleculares y códigos de barras del vaso, y se amplifican y luego se combinan o se agrupan para formar un banco que consiste en elementos de banco de cadena pesada y ligera y/o cadena alfa y beta y/o cadena gamma y delta.

Un elemento de banco basado en perlas contiene una o más perlas, y también puede contener otros reactivos, tales como anticuerpos, enzimas u otras proteínas. En caso de que todos los elementos del banco contengan diferentes tipos de perlas, pero los mismos medios circundantes, los elementos del banco pueden prepararse a partir de un solo fluido de partida o tener una variedad de fluidos de partida. En el caso de los bancos celulares preparados en masa a partir de una colección de variantes, los elementos del banco se prepararán a partir de una variedad de fluidos de partida. Es deseable tener exactamente una célula por gotita, conteniendo solo unas cuantas gotitas más de una célula al comenzar con una pluralidad de células. En algunos casos, se pueden lograr variaciones de las estadísticas de Poisson para proporcionar una carga mejorada de gotitas, de modo que haya más gotitas con exactamente una célula por gotita y pocas excepciones de gotitas vacías o que contengan más de una célula.

En algunas realizaciones, es deseable tener exactamente un polinucleótido con código de barras del vaso por gotita con solo unas cuantas gotitas que contienen más de un polinucleótido con código de barras del vaso cuando se parte de una pluralidad de polinucleótidos con código de barras del vaso. En algunos casos, se pueden lograr variaciones de las estadísticas de Poisson para proporcionar una carga mejorada de gotitas, de modo que haya más gotitas con

exactamente un polinucleótido con código de barras del vaso por gotita y pocas excepciones de gotitas vacías o que contengan más de un polinucleótido con código de barras del vaso.

5 Los ejemplos de bancos de gotitas son colecciones de gotitas que tienen diferentes contenidos, que varían desde perlas, células, moléculas pequeñas, ADN, cebadores, anticuerpos y polinucleótidos con código de barras. Las gotitas varían en tamaño desde aproximadamente 0,5 micrómetros hasta 500 micrómetros de diámetro, lo que corresponde a aproximadamente 1 picolitro a 1 nanolitro. Sin embargo, las gotitas pueden ser tan pequeñas como de 5 micrómetros y tan grandes como de 500 micrómetros. Preferentemente, las gotitas son inferiores a 100 micrómetros, de aproximadamente 1 micrómetro a aproximadamente 100 micrómetros de diámetro. El tamaño más preferido es de 10 aproximadamente 20 a 40 micrómetros de diámetro (10 a 100 picolitros). Las propiedades preferidas examinadas de los bancos de gotitas incluyen el equilibrio de la presión osmótica, el tamaño uniforme y los intervalos de tamaño.

15 Las gotitas comprendidas dentro del banco de gotitas proporcionado por la presente invención son preferentemente de tamaño uniforme. Es decir, el diámetro de cualquier gotita dentro del banco variará menos del 5 %, 4 %, 3 %, 2 %, 1 % o 0,5 % en comparación con el diámetro de otras gotitas dentro del mismo banco. El tamaño uniforme de las gotitas en el banco puede ser crítico para mantener la estabilidad y la integridad de las gotitas, y también puede ser esencial para el uso posterior de las gotitas dentro del banco para los diversos ensayos biológicos y químicos descritos en el presente documento.

20 La presente divulgación proporciona un banco de gotitas que comprende una pluralidad de gotitas acuosas dentro de un fluido inmiscible, en el que cada gotita es preferentemente de tamaño esencialmente uniforme y comprende un elemento del banco diferente. La invención proporciona un procedimiento para formar el banco de gotitas que comprende proporcionar un solo fluido acuoso que comprende diferentes elementos del banco, encapsular cada elemento del banco en una gotita acuosa dentro de un fluido inmiscible.

25 En determinadas realizaciones, diferentes tipos de elementos (por ejemplo, células o perlas), se agrupan en una sola fuente contenida en el mismo medio. Después de la agrupación inicial, los elementos se encapsulan en gotitas para generar un banco de gotitas en el que cada gotita con un tipo diferente de perla o célula es un elemento del banco diferente. La dilución de la solución inicial permite el proceso de encapsulación. En algunas realizaciones, las gotitas formadas contendrán un solo elemento o no contendrán nada, es decir, estarán vacías. En otras realizaciones, las gotitas formadas contendrán múltiples copias de un elemento del banco. Los elementos que se encapsulan son generalmente variantes de un tipo. En un ejemplo, los elementos son células inmunitarias de una muestra de sangre, y cada célula inmunitaria se encapsula para amplificar y codificar con código de barras las secuencias de anticuerpos de los nucleótidos en las células inmunitarias.

35 Por ejemplo, en un tipo de banco de emulsiones, hay elementos del banco que tienen diferentes partículas, es decir, células o polinucleótidos con código de barras en un medio diferente y se encapsulan antes de la agrupación. En un ejemplo, un número específico de elementos del banco, es decir, número n de células diferentes o polinucleótidos con código de barras, está contenido dentro de diferentes medios. Cada uno de los elementos del banco se emulsiona y se agrupa por separado, momento en el que cada uno del número n de elementos del banco diferentes se combinan y se agrupan en una sola agrupación. La agrupación resultante contiene una pluralidad de gotitas de emulsión de agua en aceite que contienen cada una un tipo diferente de partícula.

45 En algunas realizaciones, las gotitas formadas contendrán un solo elemento del banco o no contendrán nada, es decir, estarán vacías. En otras realizaciones, las gotitas formadas contendrán múltiples copias de un elemento del banco. El contenido de las perlas sigue una distribución de Poisson, en la que existe una distribución de probabilidad diferenciada que expresa la probabilidad de que ocurra un número de eventos en un periodo de tiempo fijo si estos eventos ocurren con una velocidad media conocida e independientemente del tiempo transcurrido desde el último evento. Los aceites y tensioactivos usados para crear los bancos impiden el intercambio del contenido del banco entre las gotitas.

## 50 **TRANSCRIPCIÓN REVERSA**

En algunos casos, los polinucleótidos diana se preparan a partir de un ARN mediante transcripción inversa. En algunos casos, los polinucleótidos diana se preparan a partir de un ADN mediante la extensión del cebador, tal como usando una polimerasa.

60 Los procedimientos descritos en el presente documento se pueden usar en PCR de transcripción inversa acoplada (PCR de transcripción inversa). Por ejemplo, la transcripción inversa y la PCR se pueden llevar a cabo en dos etapas distintas. En primer lugar, se puede sintetizar una copia de ADNc del ARNm de muestra usando un cebador polinucleótido dT, un cebador específico de la secuencia, un cebador universal o cualquier cebador descrito en el presente documento.

65 La transcripción inversa y la PCR se pueden llevar a cabo en una sola reacción de vaso cerrado. Por ejemplo, se pueden emplear tres cebadores, uno para la transcripción inversa y dos para la PCR. El cebador para la transcripción inversa puede unirse al ARNm 3' en la posición del amplicón de PCR. Aunque no es esencial, el cebador de transcripción inversa puede incluir restos de ARN o análogos modificados, tales como las bases de ARN 2'-O-metilo,

que no formarán un sustrato para la RNasa H cuando se hibriden con el ARNm.

La temperatura para llevar a cabo la reacción de transcripción inversa depende de la transcriptasa inversa que se use. En algunos casos, se usa una transcriptasa inversa termoestable y la reacción de transcripción inversa se lleva a cabo a de aproximadamente 37 °C a aproximadamente 75 °C, a de aproximadamente 37 °C a aproximadamente 50 °C, a de aproximadamente 37 °C a aproximadamente 55 °C, a de aproximadamente 37 °C a aproximadamente 60 °C, a de aproximadamente 55 °C a aproximadamente 75 °C, a de aproximadamente 55 °C a aproximadamente 60 °C, a de aproximadamente 37 °C o a aproximadamente 60 °C. En algunos casos, se usa una transcriptasa inversa que transfiere 3 o más nucleótidos terminales no molde a un extremo del producto transcrito.

Una reacción de transcripción inversa y la reacción de PCR descritas en el presente documento pueden llevarse a cabo en varios formatos conocidos en la técnica, tal como en tubos, placas de microtitulación, dispositivos microfluídicos o, preferentemente, gotitas.

Una reacción de transcripción inversa se puede llevar a cabo en volúmenes que varían de 5 µl a 100 µl, o en volúmenes de reacción de 10 µl a 20 µl. En gotitas, los volúmenes de reacción pueden variar de 1 pl a 100 nl, o de 10 pl a 1 nl. En algunos casos, la reacción de transcripción inversa se lleva a cabo en una gotita que tiene un volumen que es aproximadamente o inferior a 1 nl. En algunos casos, una reacción de PCR está en una gotita que tiene un intervalo de volumen de reacción de 1 pl a 100 nl, preferentemente de 10 pl a 1 nl. En algunos casos, la reacción de PCR se lleva a cabo en una gotita que tiene un volumen que es aproximadamente o inferior a 1 nl. En algunos casos, una reacción de transcripción inversa y una reacción de PCR se llevan a cabo en la misma gotita que tiene un volumen de reacción de 1 pl a 100 nl o de 10 pl a 1 nl. En algunos casos, la reacción de transcripción inversa y la reacción de PCR se llevan a cabo en una gotita que tiene un volumen que es de aproximadamente o inferior a 1 nl o un volumen que es de aproximadamente o inferior a 1 pl. En algunos casos, una reacción de transcripción inversa y una reacción de PCR se llevan a cabo en una gotita diferente. En algunos casos, una reacción de transcripción inversa y una reacción de PCR se llevan a cabo en una pluralidad de gotitas, cada una con un volumen de reacción que varía de 1 pl a 100 nl o de 10 pl a 1 nl. En algunos casos, la reacción de transcripción inversa y la reacción de PCR se llevan a cabo en una pluralidad de gotitas, cada una con un volumen que es de aproximadamente o inferior a 1 nl.

En algunos casos, una primera reacción de PCR es en una primera gotita que tiene un intervalo de volumen de reacción de 1 pl a 100 nl, preferentemente de 10 pl a 1 nl, y una segunda reacción de PCR es en una segunda gotita que tiene un volumen de reacción en un intervalo de 1 pl a 100 nl, preferentemente de 10 pl a 1 nl. En algunos casos, una primera reacción de PCR es en una primera gotita que tiene un volumen que es de aproximadamente o inferior a 1 nl, y una segunda reacción de PCR es en una segunda gotita que tiene un volumen que es de aproximadamente o inferior a 1 nl.

En algunos casos, una primera reacción de PCR y una segunda reacción de PCR se llevan a cabo en una pluralidad de gotitas, cada una con un volumen de reacción que varía de 1 pl a 100 nl o de 10 pl a 1 nl. En algunos casos, una primera reacción de PCR y una segunda reacción de PCR se llevan a cabo en una pluralidad de gotitas, cada una con un volumen que es de aproximadamente o inferior a 1 nl.

Los polinucleótidos diana, tal como el ARN, pueden transcribirse de forma inversa en ADNc usando uno o más cebadores de transcripción inversa. El uno o más cebadores de transcripción inversa pueden comprender una región complementaria a una región del ARN, tal como una región constante (por ejemplo, una región constante de cadena pesada o ligera o una cola poli-A de ARNm). En algunas realizaciones, los cebadores de transcripción inversa pueden comprender un primer cebador de transcripción inversa con una región complementaria a una región constante de un primer ARN, y un segundo cebador de transcripción inversa con una región complementaria a una región constante de un segundo ARN. En algunas realizaciones, los cebadores de transcripción inversa pueden comprender un primer cebador de transcripción inversa con una región complementaria a una región constante de un primer ARN, y uno o más cebadores de transcripción inversa con una región complementaria a una región constante de uno o más ARN, respectivamente.

En algunas realizaciones, los cebadores de transcripción inversa no comprenden un código de barras.

Los cebadores de transcripción inversa pueden comprender además una región que no sea complementaria de una región del ARN. En algunas realizaciones, la región que no es complementaria a una región del ARN se encuentra en 5' con respecto a una región de los cebadores que es complementaria al ARN. En algunas realizaciones, la región que no es complementaria a una región del ARN se encuentra en 3' con respecto a una región de los cebadores que es complementaria al ARN. En algunas realizaciones, la región que no es complementaria a una región del ARN es una región saliente 5'. En algunas realizaciones, la región que no es complementaria a una región del ARN comprende un sitio de cebador para la amplificación y/o una reacción de secuenciación. Usando uno o más cebadores descritos en el presente documento, las moléculas de ARN se transcriben de manera inversa usando reactivos adecuados conocidos en la técnica.

Después de realizar las reacciones de transcripción inversa de las moléculas de ARN, las moléculas de ADNc resultantes pueden codificarse con un código de barras molecular y un código de barras del vaso y amplificarse

mediante una o más reacciones de PCR, tal como una primera y/o una segunda reacción de PCR. La primera y/o la segunda reacción de PCR pueden utilizar un par de cebadores o una pluralidad de pares de cebadores. La primera y/o la segunda reacción de PCR pueden utilizar una pluralidad de cebadores directos/inversos y un cebador inverso. La primera y/o la segunda reacción de PCR pueden utilizar una pluralidad de cebadores directos/inversos y un cebador directo. Un primer y/o un segundo cebador de una pluralidad de cebadores directos/inversos puede ser un cebador directo/inverso que contenga una región complementaria a las moléculas de ADNc o moléculas de ADNc con código de barras. Un primer y/o un segundo cebador de una pluralidad de cebadores directos/inversos puede ser un cebador directo/inverso que contenga una región complementaria a las moléculas de ADNc con código de barras.

10 En algunas realizaciones, una pluralidad de cebadores directos/inversos comprende uno o más cebadores directos/inversos en los que cada uno de los cebadores directos/inversos de la pluralidad de cebadores directos/inversos comprende una región complementaria a una o más regiones de cadena arriba o de cadena abajo con respecto a un segmento de V de los ADNc o los ADNc con código de barras. Por ejemplo, una pluralidad de cebadores directos/inversos comprende un cebador directo/inverso que comprende una región complementaria a una región cadena arriba o cadena abajo con respecto a un segmento de V de los ADNc o ADNc con código de barras y uno o más cebadores directos/inversos que comprenden una región complementaria a una o más regiones cadena arriba o cadena abajo con respecto a un segmento de V de los ADNc o los ADNc con código de barras. Por ejemplo, una pluralidad de cebadores directos/inversos comprende un primer y/o un segundo cebador directo/inverso que comprenden una región complementaria a una primera y/o una segunda región cadena arriba o cadena abajo con respecto a un segmento de V de los ADNc o los ADNc con código de barras y un segundo cebador directo/inverso que comprende una región complementaria a una segunda región cadena arriba o cadena abajo con respecto a un segmento de V de los ADNc o los ADNc con código de barras. Por ejemplo, una pluralidad de cebadores directos/inversos comprende un primer y/o un segundo cebador directo/inverso que comprenden una región complementaria a una primera y/o una segunda región cadena arriba o cadena abajo con respecto a un segmento de V de los ADNc o los ADNc con código de barras, un segundo cebador directo/inverso que comprende una región complementaria a una segunda región cadena arriba o cadena abajo con respecto a un segmento de V de los ADNc o ADNc con código de barras, y un tercer cebador directo/inverso que comprende una región complementaria a una tercera región cadena arriba o cadena abajo con respecto a un segmento de V de los ADNc o ADNc con código de barras, Etc. Los cebadores de la pluralidad de cebadores directos/inversos se pueden usar para hibridarse a todas las posibles regiones de cadena arriba o cadena abajo de todos los segmentos de V expresados por las células, tales como linfocitos B o linfocitos T inmunitarios, de la muestra.

En algunas realizaciones, una pluralidad de cebadores directos/inversos comprende uno o más cebadores directos/inversos en los que cada uno de los cebadores directos/inversos de la pluralidad de cebadores directos/inversos comprende una región complementaria a una o más regiones de cadena arriba o de cadena abajo a un segmento de C de los ADNc o los ADNc con código de barras. Por ejemplo, una pluralidad de cebadores directos/inversos comprende un cebador directo/inverso que comprende una región complementaria a una región cadena arriba o cadena abajo con respecto a un segmento de C de los ADNc o ADNc con código de barras y uno o más cebadores directos/inversos que comprenden una región complementaria a una o más regiones cadena arriba o cadena abajo con respecto a un segmento de C de los ADNc o los ADNc con código de barras. Por ejemplo, una pluralidad de cebadores directos/inversos comprende un primer y/o un segundo cebador directo/inverso que comprenden una región complementaria a una primera y/o una segunda región cadena arriba o cadena abajo con respecto a un segmento de C de los ADNc o los ADNc con código de barras y un segundo cebador directo/inverso que comprende una región complementaria a una segunda región cadena arriba o cadena abajo con respecto a un segmento de C de los ADNc o los ADNc con código de barras. Por ejemplo, una pluralidad de cebadores directos/inversos comprende un primer y/o un segundo cebador directo/inverso que comprenden una región complementaria a una primera y/o una segunda región cadena arriba o cadena abajo con respecto a un segmento de C de los ADNc o los ADNc con código de barras, un segundo cebador directo/inverso que comprende una región complementaria a una segunda región cadena arriba o cadena abajo con respecto a un segmento de C de los ADNc o ADNc con código de barras, y un tercer cebador directo/inverso que comprende una región complementaria a una tercera región cadena arriba o cadena abajo con respecto a un segmento de C de los ADNc o ADNc con código de barras, Etc. Los cebadores de la pluralidad de cebadores directos/inversos se pueden usar para hibridarse a todas las posibles regiones de cadena arriba o cadena abajo de todos los segmentos de C expresados por las células, tales como linfocitos B o linfocitos T inmunitarios, de la muestra.

En algunas realizaciones, una pluralidad de cebadores directos/inversos comprende uno o más cebadores directos/inversos en los que cada uno de los cebadores directos/inversos de la pluralidad de cebadores directos/inversos comprende una región complementaria a una o más regiones de cadena arriba o de cadena abajo con respecto a un código de barras molecular de los ADNc con código de barras. Por ejemplo, una pluralidad de cebadores directos/inversos comprende un cebador directo/inverso que comprende una región complementaria a una región cadena arriba o cadena abajo con respecto a un código de barras molecular de los ADNc con código de barras y uno o más cebadores directos/inversos que comprenden una región complementaria a una o más regiones cadena arriba o cadena abajo con respecto a un código de barras molecular de los ADNc con código de barras. Por ejemplo, una pluralidad de cebadores directos/inversos comprende un primer y/o un segundo cebador directo/inverso que comprenden una región complementaria a una primera y/o una segunda región cadena arriba o cadena abajo con respecto a un código de barras molecular de los ADNc con código de barras y un segundo cebador directo/inverso

que comprende una región complementaria a una segunda región cadena arriba o cadena abajo con respecto a un código de barras molecular de los ADNc con código de barras. Por ejemplo, una pluralidad de cebadores directos/inversos comprende un primer y/o un segundo cebador directo/inverso que comprenden una región complementaria a una primera y/o una segunda región cadena arriba o cadena abajo con respecto a un código de barras molecular de los ADNc con código de barras, un segundo cebador directo/inverso que comprende una región complementaria a una segunda región cadena arriba o cadena abajo con respecto a un código de barras molecular de los ADNc con código de barras, y un tercer cebador directo/inverso que comprende una región complementaria a una tercera región cadena arriba o cadena abajo con respecto a un código de barras molecular de los ADNc con código de barras, etc. La pluralidad de cebadores directos/inversos se pueden usar para hibridarse a todas las posibles regiones de cadena arriba o cadena abajo de todos los códigos de barras moleculares expresados por las células, tales como linfocitos B o linfocitos T inmunitarios, de la muestra.

En algunas realizaciones, una pluralidad de cebadores directos/inversos comprende uno o más cebadores directos/inversos en los que cada uno de los cebadores directos/inversos de la pluralidad de cebadores directos/inversos comprende una región complementaria a una o más regiones de cadena arriba o de cadena abajo con respecto a un código de barras del vaso de los ADNc con código de barras. Por ejemplo, una pluralidad de cebadores directos/inversos comprende un cebador directo/inverso que comprende una región complementaria a una región cadena arriba o cadena abajo con respecto a un código de barras del vaso de los ADNc con código de barras y uno o más cebadores directos/inversos que comprenden una región complementaria a una o más regiones cadena arriba o cadena abajo con respecto a un código de barras del vaso de los ADNc con código de barras. Por ejemplo, una pluralidad de cebadores directos/inversos comprende un primer y/o un segundo cebador directo/inverso que comprenden una región complementaria a una primera y/o una segunda región cadena arriba o cadena abajo con respecto a un código de barras del vaso de los ADNc con código de barras, un segundo cebador directo/inverso que comprende una región complementaria a una segunda región cadena arriba o cadena abajo con respecto a un código de barras del vaso de los ADNc con código de barras, y un tercer cebador directo/inverso que comprende una región complementaria a una tercera región cadena arriba o cadena abajo con respecto a un código de barras del vaso de los ADNc con código de barras, etc. Los cebadores de la pluralidad de cebadores directos/inversos se pueden usar para hibridarse a todas las posibles regiones de cadena arriba o cadena abajo de todos los códigos de barras del vaso expresados por las células, tales como linfocitos B o linfocitos T inmunitarios, de la muestra.

Los cebadores directos/inversos de la pluralidad de cebadores directos/inversos comprenden además una región que no es complementaria a una región del ARN. En algunas realizaciones, la región que no es complementaria a una región del ARN se encuentra en 5' con respecto a una región de los cebadores directo/inverso que es complementaria al ARN (es decir, una región cadena arriba o cadena abajo de un segmento V). En algunas realizaciones, la región que no es complementaria a una región del ARN se encuentra en 3' con respecto a una región de los cebadores directo/inverso que es complementaria al ARN. En algunas realizaciones, la región que no es complementaria a una región del ARN es una región saliente 5'. En algunas realizaciones, la región que no es complementaria a una región del ARN comprende un sitio de cebador para la amplificación y/o una segunda reacción de secuenciación. En algunas realizaciones, la región que no es complementaria a una región del ARN comprende un sitio de cebador para la amplificación y/o una tercera reacción de secuenciación. En algunas realizaciones, la región que no es complementaria a una región del ARN comprende un sitio de cebador para una segunda y una tercera reacción de secuenciación. En algunas realizaciones, la secuencia del sitio de cebado para la segunda y la tercera reacción de secuenciación es la misma. Usando uno o más cebadores directo/inverso y un cebador inverso como se describe en el presente documento, las moléculas de ADNc se amplifican usando reactivos adecuados conocidos en la técnica. En algunas realizaciones, una región es complementaria a una región del ARN, tal como la región constante o una cola poli-A del ARNm.

## AMPLIFICACIÓN

La muestra que contiene el polinucleótido diana puede comprender ARNm, o fragmentos del mismo, que pueden amplificarse. En algunos casos, la longitud media del ARNm, o fragmentos del mismo, puede ser inferior a aproximadamente 100, 200, 300, 400, 500 u 800 pares de bases, o inferior a aproximadamente 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190 o 200 nucleótidos, o inferior a aproximadamente 1, 2, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100 kilobases. En algunos casos, se amplifica una secuencia diana de un molde relativamente corto, tal como una muestra que contiene un molde que es de aproximadamente 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95 o 100 bases.

Una reacción de amplificación puede comprender uno o más aditivos. En algunos casos, el uno o más aditivos son dimetilsulfóxido (DMSO), glicerol, (mono)hidrato de betaína (*N,N,N*-trimetilglicina = [caroxi-metil]trietilamonio), trehalosa, 7-desaza-2'-desoxiguanosina trifosfato (dC7GTP o 7-desaza-2'-dGTP), BSA (albúmina de suero bovino), formamida (metanamida), cloruro de tetrametilamonio (TMAC), otros derivados de tetralquilamonio (por ejemplo,

cloruro de tetraetilamonio (TEA-Cl) y cloruro de tetrapropilamonio (TPrA-Cl), detergente no iónico (por ejemplo, Triton X-100, Tween 20, Nonidet P-40 (NP-40)) o PREXCEL-Q. En algunos casos, una reacción de amplificación comprende 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10 aditivos diferentes. En otros casos, una reacción de amplificación comprende al menos 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10 aditivos diferentes.

5 Las reacciones de termociclado se pueden realizar en muestras contenidas en volúmenes de reacción (por ejemplo, gotitas). Las gotitas pueden ser polidispersas o, preferentemente, monodispersas, generadas por agitación, ultrasonidos o microfluidicamente a través de una unión al canal T u otros medios por los que están familiarizados con la técnica. Las densidades pueden superar las 20.000 gotitas/40 ul (gotitas de 1 nl), 200.000 gotitas/40 ul (gotitas de 100 pl). Las gotitas pueden permanecer intactas durante el termociclado. Las gotitas pueden permanecer intactas durante el termociclado a densidades superiores a aproximadamente 10.000 gotitas/ $\mu$ l, 100.000 gotitas/ $\mu$ l, 200.000 gotitas/ $\mu$ l, 300.000 gotitas/ $\mu$ l, 400.000 gotitas/ $\mu$ l, 500.000 gotitas/ $\mu$ l, 600.000 gotitas/ $\mu$ l, 700.000 gotitas/ $\mu$ l, 800.000 gotitas/ $\mu$ l, 900.000 gotitas/ $\mu$ l o 1.000.000 gotitas/ $\mu$ l. En otros casos, dos o más gotitas no se unen durante el termociclado. En otros casos, más de 100 o más de 1.000 gotitas no se unen durante el termociclado.

15 Se puede usar cualquier ADN polimerasa que catalice la extensión del cebador, incluyendo, pero sin limitación, la ADN polimerasa de *E.coli*, fragmento Klenow de ADN polimerasa 1 de *E.coli*, ADN polimerasa T7, ADN polimerasa T4, polimerasa Taq, ADN polimerasa Pfu, ADN polimerasa Vent, bacteriófago 29, REDTaq™, ADN polimerasa genómica o secuenciada. En algunos casos, se usa una ADN polimerasa termoestable. También se puede realizar una PCR inicial en caliente en la que la reacción se calienta hasta 95 °C durante dos minutos antes de la adición de la polimerasa o la polimerasa se puede mantener inactiva hasta la primera etapa de calentamiento en el ciclo 1. La PCR inicial en caliente se puede usar para minimizar la amplificación inespecífica. Se puede usar cualquier número de ciclos de PCR para amplificar el ADN, por ejemplo, aproximadamente, más de aproximadamente o menos de aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44 o 45 ciclos. El número de ciclos de amplificación puede ser de aproximadamente 1-45, 10-45, 20-45, 30-45, 35-45, 10-40, 10-30, 10-25, 10-20, 10-15, 20-35, 25-35, 30-35 o 35-40.

30 La amplificación de ácidos nucleicos diana se puede realizar mediante cualquier medio conocido en la técnica. Los ácidos nucleicos diana pueden amplificarse mediante la reacción en cadena de la polimerasa (PCR) o la amplificación isotérmica del ADN. Los ejemplos de técnicas de PCR que se pueden usar incluyen, aunque sin limitación, PCR cuantitativa, PCR fluorescente cuantitativa (QF-PCR), PCR fluorescente multiplex (MF-PCR), PCR en tiempo real (PCR de transcripción inversa), PCR de una sola célula, PCR de polimorfismos de longitud de fragmentos de restricción (PCR-RFLP), PCR-RFLP/PCR-RFLP de transcripción inversa, PCR de inicio en caliente, PCR anidada, PCR de colonias de polimerasa *in situ*, amplificación por círculo rodante (RCA) *in situ*, PCR digital (dPCR), PCR digital de gotitas (ddPCR), PCR de puente, PCR en picolitros y PCR de emulsión. Otros procedimientos de amplificación adecuados incluyen la reacción en cadena de la ligasa (LCR), la amplificación de la transcripción, la PCR de sonda de inversión molecular (MIP), la replicación de secuencias autosostenida, la amplificación selectiva de las secuencias de polinucleótidos diana, la reacción en cadena de la polimerasa cebada con secuencias de consenso (CP-PCR), la reacción en cadena de la polimerasa cebada arbitrariamente (AP-PCR), la PCR cebada con polinucleótidos degenerados (DOP-PCR) y la amplificación de secuencias basada en ácidos nucleicos (NABSA). Otros procedimientos de amplificación que se pueden usar en el presente documento incluyen los descritos en las patentes de EE.UU. n.º 5.242.794; 5.494.810; 4.988.617; y 6.582.938, además de incluir la amplificación de ARN mediada por la replicasa Q beta. La amplificación puede ser una amplificación isotérmica, por ejemplo, una amplificación lineal isotérmica.

45 En algunas realizaciones, la amplificación no se produce en un soporte sólido. En algunas realizaciones, la amplificación no se produce en un soporte sólido en una gotita. En algunas realizaciones, la amplificación se produce en un soporte sólido cuando la amplificación no está en una gotita.

50 Una reacción de amplificación puede comprender uno o más aditivos. En algunas realizaciones, el uno o más aditivos son dimetilsulfóxido (DMSO), glicerol, (mono)hidrato de betaína (*N,N,N*-trimetilglicina = [caroxi-metil]trimetilamonio), trehalosa, 7-desaza-2'-desoxiguanosina trifosfato (dC7GTP o 7-desaza-2'-dGTP), BSA (albúmina de suero bovino), formamida (metanamida), cloruro de tetrametilamonio (TMAC), otros derivados de tetralquilamonio (por ejemplo, cloruro de tetraetilamonio (TEA-Cl) y cloruro de tetrapropilamonio (TPrA-Cl), detergente no iónico (por ejemplo, Triton X-100, Tween 20, Nonidet P-40 (NP-40)) o PREXCEL-Q. En algunas realizaciones, una reacción de amplificación puede comprender 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10 aditivos diferentes. En otros casos, una reacción de amplificación puede comprender al menos 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10 aditivos diferentes.

## CEBADORES

60 En general, se pueden usar uno o más pares de cebadores en una reacción de amplificación; un cebador de un par de cebadores puede ser un cebador directo y un cebador de un par de cebadores puede ser un cebador inverso.

65 En algunos casos, se puede usar un primer par de cebadores en la reacción de amplificación; un cebador del primer par puede ser un cebador directo complementario a una secuencia de una primera molécula de polinucleótido diana y un cebador del primer par puede ser un cebador inverso que puede ser complementario a una segunda secuencia

de la primera molécula de polinucleótido diana, y un primer locus diana puede residir entre la primera secuencia y la segunda secuencia. En algunas realizaciones, el primer locus diana comprende una secuencia  $V_H$  o  $V_\alpha$  o  $V_\gamma$ .

En algunos casos, se puede usar un segundo par de cebadores en la reacción de amplificación; un cebador del segundo par puede ser un cebador directo complementario a una primera secuencia de una segunda molécula de polinucleótido diana y un cebador del segundo par puede ser un cebador inverso complementario a una segunda secuencia de la segunda molécula de polinucleótido diana, y un segundo locus diana puede residir entre la primera secuencia y la segunda secuencia. En algunas realizaciones, el segundo locus diana comprende una secuencia  $V_L$  o  $V_\beta$  o  $V_\delta$ .

En algunos casos, se puede usar un tercer par de cebadores en la reacción de amplificación; un cebador del tercer par puede ser un cebador directo complementario a una primera secuencia de una tercera molécula de polinucleótido diana y un cebador del tercer par puede ser un cebador inverso complementario a una segunda secuencia de la tercera molécula de polinucleótido diana, y un tercer locus diana puede residir entre la primera secuencia y la segunda secuencia. En algunas realizaciones, el tercer locus diana comprende un código de barras, tal como un código de barras molecular o código de barras del vaso.

La longitud del cebador directo y el cebador inverso pueden depender de la secuencia del polinucleótido diana y del locus diana. Por ejemplo, la longitud y/o TM del cebador directo y del cebador inverso se pueden optimizar. En algún caso, un cebador puede ser de aproximadamente, más de aproximadamente o menos de aproximadamente 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59 o 60 nucleótidos de longitud. En algunos casos, un cebador es de aproximadamente 15 a aproximadamente 20, de aproximadamente 15 a aproximadamente 25, de aproximadamente 15 a aproximadamente 30, de aproximadamente 15 a aproximadamente 40, de aproximadamente 15 a aproximadamente 45, de aproximadamente 15 a aproximadamente 50, de aproximadamente 15 a aproximadamente 55, de aproximadamente 15 a aproximadamente 60, de aproximadamente 20 a aproximadamente 25, de aproximadamente 20 a aproximadamente 30, de aproximadamente 20 a aproximadamente 35, de aproximadamente 20 a aproximadamente 40, de aproximadamente 20 a aproximadamente 45, de aproximadamente 20 a aproximadamente 50, de aproximadamente 20 a aproximadamente 55 o de aproximadamente 20 a aproximadamente 60 nucleótidos de longitud.

Un cebador puede ser un ADN monocatenario antes de unirse a un polinucleótido molde. En algunos casos, el cebador inicialmente comprende una secuencia bicatenaria. La longitud apropiada de un cebador puede depender del uso previsto del cebador, pero puede variar de aproximadamente 6 a aproximadamente 50 nucleótidos, o de aproximadamente 15 a aproximadamente 35 nucleótidos. En general, las moléculas de cebador corto pueden requerir temperaturas más frías para formar complejos híbridos suficientemente estables con un molde. En algunas realizaciones, un cebador no necesita reflejar la secuencia exacta del ácido nucleico de molde, pero puede ser suficientemente complementario para hibridarse con un molde. En algunos casos, un cebador puede ser parcialmente bicatenario antes de unirse a un polinucleótido de molde. Un cebador con secuencia bicatenaria puede tener un bucle en horquilla de aproximadamente, más de aproximadamente o menos de aproximadamente 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16., 17, 18, 19 o 20 bases. Una porción bicatenaria de un cebador puede ser de aproximadamente, más de aproximadamente, menos de aproximadamente o al menos de aproximadamente 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49 o 50 pares de bases. El diseño de cebadores adecuados para la amplificación de una secuencia diana dada es bien conocido en la técnica.

Los cebadores pueden incorporar características adicionales que permitan la detección o inmovilización del cebador, pero no modifique una propiedad básica del cebador (por ejemplo, que actúa como un punto de inicio de la síntesis del ADN). Por ejemplo, los cebadores pueden contener una secuencia de ácido nucleico adicional en el extremo 5' que no está hibridada con un ácido nucleico diana, pero que facilita la clonación o amplificación adicional, o la secuenciación de un producto amplificado. Por ejemplo, la secuencia adicional puede comprender un sitio de unión al cebador, tal como un sitio de unión al cebador universal. Una región del cebador que es suficientemente complementaria a un molde para hibridarse puede denominarse en el presente documento una región de hibridación.

En otro caso, un cebador utilizado en los procedimientos y en las composiciones que se describen en el presente documento puede comprender uno o más nucleósidos universales. Son ejemplos no limitantes de nucleósidos universales el 5-nitroindol y la inosina, como se describe en las publicaciones de solicitud de EE.UU n.º 2009/0325169 y 2010/0167353.

Los cebadores pueden diseñarse de acuerdo con parámetros conocidos para evitar estructuras secundarias y la autohibridación. Diferentes pares de cebadores pueden hibridarse y fundirse a aproximadamente las mismas temperaturas, por ejemplo, a más/menos 1 °C, 2 °C, 3 °C, 4 °C, 5 °C, 6 °C, 7 °C, 8 °C, 9 °C o 10 °C de otro par de cebadores. En algunos casos, inicialmente se usan más de 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 100, 200, 500, 1.000, 5.000, 10.000 o más cebadores. Dichos cebadores pueden hibridarse con los polinucleótidos diana descritos en el presente documento.

Los cebadores pueden prepararse mediante una variedad de procedimientos que incluyen, pero sin limitación, la clonación de secuencias apropiadas y la síntesis química directa usando procedimientos bien conocidos en la técnica (Narang *et al.*, *Methods Enzymol.* 68: 90 (1979); Brown *et al.*, *Methods Enzymol.* 68:109 (1979)). Los cebadores también se pueden obtener de fuentes comerciales. Los cebadores pueden tener una temperatura de fusión idéntica. Los cebadores pueden tener temperaturas de fusión no idénticas. Las longitudes de los cebadores pueden extenderse o acortarse en el extremo 5' o en el extremo 3' para producir cebadores con las temperaturas de fusión deseadas. Uno de los cebadores de un par de cebadores puede ser más largo que el otro cebador. Las longitudes de hibridación en 3' de los cebadores, dentro de un par de cebadores, pueden diferir. Asimismo, la posición de hibridación de cada par de cebadores puede diseñarse de modo que la secuencia y la longitud de los pares de cebadores produzcan la temperatura de fusión deseada. Una ecuación para determinar la temperatura de fusión de los cebadores inferiores a 25 pares de bases es la Regla de Wallace ( $TM = 2(A + T) + 4(G + C)$ ). También se pueden usar programas informáticos para diseñar los cebadores. La TM (temperatura de fusión o de hibridación) de cada cebador puede calcularse usando programas informáticos. La temperatura de hibridación de los cebadores se puede recalcular y aumentar después de cualquier ciclo de amplificación, incluyendo, entre otros, los ciclos 1, 2, 3, 4, 5, ciclos 6-10, ciclos 10-15, ciclos 15-20, ciclos 20-25, ciclos 25-30, ciclos 30-35 o ciclos 35-40. Tras los ciclos iniciales de amplificación, la mitad 5' de los cebadores puede incorporarse a los productos de cada locus de interés; por lo tanto, la TM se puede recalcular basándose tanto en las secuencias de la mitad 5' como de la mitad 3' de cada cebador.

La realización de una o más reacciones de los procedimientos desvelados en el presente documento puede comprender el uso de uno o más cebadores. Como se usan en el presente documento, un cebador comprende un polinucleótido bicatenario, monocatenario o parcialmente monocatenario que es suficientemente complementario para hibridarse con un polinucleótido de molde. Un cebador puede ser un ADN monocatenario antes de unirse a un polinucleótido molde. En algunas realizaciones, el cebador inicialmente comprende una secuencia bicatenaria. Un sitio de cebador incluye el área del molde con el que se hibrida un cebador. En algunas realizaciones, los cebadores son capaces de actuar como un punto de inicio para la síntesis de ácido nucleico dirigida por molde. Por ejemplo, los cebadores pueden iniciar la síntesis de ácido nucleico dirigida por molde cuando cuatro nucleótidos diferentes y un agente de polimerización o una enzima, tal como la ADN o ARN polimerasa o la transcriptasa inversa. Un par de cebadores incluye 2 cebadores: un primer cebador con una región cadena arriba 5' que se hibrida con un extremo 5' de una secuencia de molde, y un segundo cebador con una región cadena abajo 3' que se hibrida con el complemento del extremo 3' de la secuencia de molde. Un conjunto de cebadores incluye dos o más cebadores: un primer cebador o una primera pluralidad de cebadores con una región cadena arriba 5' que se hibrida con un extremo 5' de una secuencia de molde o una pluralidad de secuencias de molde, y un segundo cebador o una segunda pluralidad de cebadores con una región cadena abajo 3' que se hibrida con el complemento del extremo 3' de la secuencia de molde o la pluralidad de secuencias de molde. En algunas realizaciones, un cebador comprende una secuencia específica de la diana. En algunas realizaciones, un cebador comprende una secuencia con código de barras de muestra. En algunas realizaciones, un cebador comprende una secuencia de cebado universal. En algunas realizaciones, un cebador comprende una secuencia de cebado de PCR. En algunas realizaciones, un cebador comprende una secuencia de cebado de PCR usada para iniciar la amplificación de un polinucleótido. (Dieffenbach, "PCR Primer: A Laboratory Manual", 2ª edición (Cold Spring Harbor Press, Nueva York (2003)). El sitio o la secuencia de unión al cebador universal permite la unión de un cebador universal a un polinucleótido y/o amplicón. Los cebadores universales son bien conocidos en la técnica, e incluyen, aunque sin limitación, -47F (M13F), alfaMF, AOX3', AOX5', BGHr, CMV-30, CMV-50, CVMf, LACrmt, lamgta gt10F, lambda gt 10R, lambda gt11F, lambda gt11R, M13 rev, M13Forward(-20), M13Reverse, machos, p10SEQPpQE, pA-120, pet4, pGAP directo, pGLRVpr3, pGLpr2R, pKLAC14, pQEFS, pQERS, pucU1, pucU2, reversA, seqIREStam, seqIRESzpet, seqori, seqPCR, seqpIRES-, seqpIRES+, seqpSecTag, seqpSecTag+, seqretro+PSI, SP6, T3-prom, T7-prom y T7-termInv. Como se usa en el presente documento, la unión puede referirse a interacciones covalentes y a interacciones no covalentes. La unión del cebador universal con el sitio de unión al cebador universal se puede usar para la amplificación, detección y/o secuenciación del polinucleótido y/o del amplicón. El sitio de unión del cebador universal puede comprender al menos aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900 o 1.000 nucleótidos o pares de bases. En otro ejemplo, el sitio de unión del cebador universal comprende al menos aproximadamente 1.500, 2.000, 2.500, 3.000, 3.500, 4.000, 4.500, 5.000, 5.500, 6.000, 6.500, 7.000, 7.500, 8.000, 8.500, 9.000, 9.500 o 10.000 nucleótidos o pares de bases. En algunas realizaciones, el sitio de unión del cebador universal comprende 1-10, 10-20, 10-30 o 10-100 nucleótidos o pares de bases. En algunas realizaciones, el sitio de unión del cebador universal comprende a partir de aproximadamente 1-90, 1-80, 1-70, 1-60, 1-50, 1-40, 1-30, 1-20, 1-10, 2-90, 2-80, 2-70, 2-60, 2-50, 2-40, 2-30, 2-20, 2-10, 1-900, 1-800, 1-700, 1-600, 1-500, 1-400, 1-300, 1-200, 1-100, 2-900, 2-800, 2-700, 2-600, 2-500, 2-400, 2-300, 2-200, 2-100, 5-90, 5-80, 5-70, 5-60, 5-50, 5-40, 5-30, 5-20, 5-10, 10-90, 10-80, 10-70, 10-60, 10-50, 10-40, 10-30, 10-20, 10-10, 5-900, 5-800, 5-700, 5-600, 5-500, 5-400, 5-300, 5-200, 5-100, 10-900, 10-800, 10-700, 10-600, 10-500, 10-400, 10-300, 10-200, 10-100, 25-900, 25-800, 25-700, 25-600, 25-500, 25-400, 25-300, 25-200, 25-100, 100-1.000, 100-900, 100-800, 100-700, 100-600, 100-500, 100-400, 100-300, 100-200, 200-1000, 200-900, 200-800, 200-700, 200-600, 200-500, 200-400, 200-300, 300-1000, 300-900, 300-800, 300-700, 300-600, 300-500, 300-400, 400-1.000, 400-900, 400-800, 400-700, 400-600, 400-500, 500-1000, 500-900, 500-800, 500-700, 500-600, 600-1000, 600-900, 600-800, 600-700, 700-1.000, 700-900, 700-800, 800-1000, 800-900 o 900-1.000 nucleótidos o pares de bases.

Los cebadores pueden tener una longitud compatible con su uso en la síntesis de productos de extensión de cebadores. Un cebador puede ser un polinucleótido que tiene una longitud de 8 a 200 nucleótidos. La longitud de un

cebador puede depender de la secuencia del polinucleótido de molde y del locus del molde. Por ejemplo, se puede optimizar la longitud y/o la temperatura de fusión (TM) de un cebador o conjunto de cebadores. En algún caso, un cebador puede ser de aproximadamente, más de aproximadamente o menos de aproximadamente 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59 o 60 nucleótidos de longitud. En algunas realizaciones, los cebadores tienen una longitud de aproximadamente 8-100 nucleótidos, por ejemplo, 10-75, 15-60, 15-40, 18-30, 20-40, 21-50, 22-45, 25-40, 7-9, 12-15, 15-20, 15-25, 15-30, 15-45, 15-50, 15-55, 15-60, 20-25, 20-30, 20-35, 20-45, 20-50, 20-55 o 20-60 nucleótidos de longitud o cualquier longitud de entre medias. En algunas realizaciones, los cebadores son como máximo de aproximadamente 10, 12, 15, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95 o 100 nucleótidos de longitud.

En general, se pueden usar uno o más pares de cebadores en una reacción de amplificación exponencial; un cebador de un par de cebadores puede ser un cebador directo y un cebador de un par de cebadores puede ser un cebador inverso. En algunas realizaciones, se puede usar un primer par de cebadores en la reacción de amplificación exponencial; un cebador del primer par puede ser un cebador directo complementario a una secuencia de un primer polinucleótido de molde y un cebador del primer par puede ser un cebador inverso complementario a una segunda secuencia de la primera molécula de polinucleótido de molde, y un primer locus de molde puede residir entre la primera secuencia y la segunda secuencia. En algunas realizaciones, se puede usar un segundo par de cebadores en la reacción de amplificación; un cebador del segundo par puede ser un cebador directo complementario a una primera secuencia de una segunda molécula de polinucleótido diana y un cebador del segundo par puede ser un cebador inverso complementario a una segunda secuencia de la segunda molécula de polinucleótido diana, y un segundo locus diana puede residir entre la primera secuencia y la segunda secuencia. En algunas realizaciones, el segundo locus diana comprende una secuencia variable de anticuerpo de cadena ligera. En algunas realizaciones, se puede usar un tercer par de cebadores en la reacción de amplificación; un cebador del tercer par puede ser un cebador directo complementario a una primera secuencia de una tercera molécula de polinucleótido de molde y un cebador del tercer par puede ser un cebador inverso complementario a una segunda secuencia de la tercera molécula de polinucleótido de molde, y un tercer locus de molde puede residir entre la primera secuencia y la segunda secuencia.

El uno o más cebadores pueden hibridarse a al menos una porción de una pluralidad de polinucleótidos de molde. El uno o más cebadores pueden hibridarse al extremo 3' y/o al extremo 5' de la pluralidad de polinucleótidos de molde. El uno o más cebadores pueden hibridarse a una región interna de la pluralidad de polinucleótidos de molde. La región interna puede ser de al menos aproximadamente 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 100, 150, 200, 220, 230, 240, 250, 260, 270, 280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 380, 390, 400, 410, 420, 430, 440, 450, 460, 470, 480, 490, 500, 510, 520, 530, 540, 550, 560, 570, 580, 590, 600, 650, 700, 750, 800, 850, 900 o 1.000 nucleótidos de los extremos 3' o los extremos 5' de la pluralidad de polinucleótidos de molde. El uno o más cebadores pueden comprender un panel fijo de cebadores. El uno o más cebadores pueden comprender al menos uno o más cebadores personalizados. El uno o más cebadores pueden comprender al menos uno o más cebadores de control. El uno o más cebadores pueden comprender al menos uno o más cebadores de genes constitutivos. El uno o más cebadores pueden comprender un cebador universal. El cebador universal puede hibridarse a un sitio de unión del cebador universal. En algunas realizaciones, el uno o más cebadores personalizados se hibridan con un SBC, una región específica de la diana, complementos de los mismos o cualquier combinación de los mismos. El uno o más cebadores pueden comprender un cebador universal. El uno o más cebadores pueden diseñarse para amplificar o realizar la extensión del cebador, transcripción inversa, extensión lineal, amplificación no exponencial, amplificación exponencial, PCR o cualquier otro procedimiento de amplificación de uno o más polinucleótidos diana o de molde.

La región específica de la diana puede comprender al menos aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 100, 150, 200, 220, 230, 240, 250, 260, 270, 280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 380, 390, 400, 410, 420, 430, 440, 450, 460, 470, 480, 490, 500, 510, 520, 530, 540, 550, 560, 570, 580, 590, 600, 650, 700, 750, 800, 850, 900 o 1.000 nucleótidos o pares de bases. En otro ejemplo, la región específica de la diana comprende al menos aproximadamente 1.500, 2.000, 2.500, 3.000, 3.500, 4.000, 4.500, 5.000, 5.500, 6.000, 6.500, 7.000, 7.500, 8.000, 8.500, 9.000, 9.500 o 10.000 nucleótidos o pares de bases. En algunas realizaciones, la región específica de la diana comprende a partir de aproximadamente 5-10, 10-15, 10-20, 10-30, 15-30, 10-75, 15-60, 15-40, 18-30, 20-40, 21-50, 22-45, 25-40, 7-9, 12-15, 15-20, 15-25, 15-30, 15-45, 15-50, 15-55, 15-60, 20-25, 20-30, 20-35, 20-45, 20-50, 20-55, 20-60, 2-900, 2-800, 2-700, 2-600, 2-500, 2-400, 2-300, 2-200, 2-100, 25-900, 25-800, 25-700, 25-600, 25-500, 25-400, 25-300, 25-200, 25-100, 100-1000, 100-900, 100-800, 100-700, 100-600, 100-500, 100-400, 100-300, 100-200, 200-1000, 200-900, 200-800, 200-700, 200-600, 200-500, 200-400, 200-300, 300-1000, 300-900, 300-800, 300-700, 300-600, 300-500, 300-400, 400-1000, 400-900, 400-800, 400-700, 400-600, 400-500, 500-1000, 500-900, 500-800, 500-700, 500-600, 600-1000, 600-900, 600-800, 600-700, 700-1000, 700-900, 700-800, 800-1000, 800-900 o 900-1000 nucleótidos o pares de bases.

Los cebadores pueden diseñarse de acuerdo con parámetros conocidos para evitar estructuras secundarias y la autohibridación. En algunas realizaciones, diferentes pares de cebadores pueden hibridarse y fundirse a aproximadamente las mismas temperaturas, por ejemplo, a más/menos 1 °C, 2 °C, 3 °C, 4 °C, 5 °C, 6 °C, 7 °C, 8 °C, 9 °C o 10 °C de otro par de cebadores. En algunas realizaciones, uno o más cebadores de una pluralidad de cebadores

pueden hibridarse y fundirse a aproximadamente las mismas temperaturas, por ejemplo, a más/menos 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10 °C que otro cebador de la pluralidad de cebadores. En algunas realizaciones, uno o más cebadores de una pluralidad pueden hibridarse y fundirse a diferentes temperaturas que otro cebador de la pluralidad de cebadores.

5 Una pluralidad de cebadores para una o más etapas de los procedimientos descritos en el presente documento puede comprender una pluralidad de cebadores que comprende aproximadamente, como máximo aproximadamente o al menos aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1.000, 1.500, 2.000, 3.000, 4.000, 5.000, 6.000, 7.000, 8.000, 9.000, 10.000, 11.000, 12.000, 13.000, 14.000, 15.000, 16.000, 17.000, 18.000, 19.000, 20.000, 30.000, 40.000, 50.000, 60.000, 70.000, 80.000, 90.000, 100.000, 200.000, 300.000, 400.000, 500.000, 600.000, 700.000, 800.000, 900.000, 1.000.000, 50.000.000, 100.000.000 cebadores diferentes. Por ejemplo, cada cebador de una pluralidad de cebadores puede comprender una secuencia o región específica de la diana o del molde diferente.

## 15 SECUENCIACIÓN

Tras realizar uno o más de los procedimientos o de las etapas del procedimiento descritos en el presente documento, se puede secuenciar un banco de polinucleótidos generados.

20 La secuenciación se puede realizar mediante cualquier procedimiento de secuenciación conocido en la técnica. En algunas realizaciones, la secuenciación se puede realizar a un alto rendimiento. Las tecnologías de secuenciación de próxima generación adecuadas incluyen la plataforma 454 Life Sciences (Roche, Branford, CT) (Margulies *et al.*, *Nature*, 437, 376-380 (2005)); el analizador del genoma de Illumina, el ensayo de metilación de GoldenGate o ensayos de metilación de Infinium, es decir, La matriz de perlas 27K de metilación humana Infinium o la matriz de metilación VeraCode GoldenGate (Illumina, San Diego, CA; Bibkova *et al.*, *Genome Res.* 16, 383-393 (2006); y las patentes de EE.UU. n.º 6.306.597, 7.598.035, 7.232.656), o la secuenciación de ADN mediante ligadura, el sistema de ligadura SOLiD (Applied Biosystems/life Technologies; las patentes de EE.UU. n.º 6.797.470, 7.083.917, 7.166.434, 7.320.865, 7.332.285, 7.364.858 y 7.429.453); o la tecnología de secuenciación de ADN de una sola molécula en tiempo real de Helicos (Harris *et al.*, *Science*, 320, 106-109 (2008); las patentes de EE.UU. n.º 7.037.687, 7.645.596 y 7.169.560, y 7.769.400), la tecnología para una sola molécula en tiempo real (SMRT™) de Pacific Biosciences, y la secuenciación (Soni *et al.*, *Clin. Chem.* 53, 1996-2001 (2007)). Estos sistemas permiten la secuenciación multiplexada en paralelo de muchos polinucleótidos aislados de una muestra (Dear, *Brief Funct. Genomic Proteomic*, 1(4), 397-416 (2003) y McCaughan *et al.*, *J. Pathol.*, 220, 297-306 (2010)). En algunas realizaciones, los polinucleótidos se secuencian mediante secuenciación por ligadura de sondas modificadas con colorante, pirosecuenciación o secuenciación de una sola molécula. La determinación de la secuencia de un polinucleótido se puede realizar mediante procedimientos de secuenciación tales como la secuenciación de una sola molécula de Helioscope™, la secuenciación del ADN de Nanopore, la secuenciación masiva de distintivos en paralelo de Lynx Therapeutics (MPSS), pirosecuenciación 454, la secuenciación en tiempo real de una sola molécula (RNAP), secuenciación Illumina (Solexa), secuenciación SOLiD, Ion Torrent™, la secuenciación de semiconductores iónicos, la secuenciación SMRT™ de una sola molécula, la secuenciación de colonias de polimerasa, la secuenciación de ADN Nanoball y el enfoque de VisiGen Biotechnologies. Como alternativa, la determinación de la secuencia de polinucleótidos puede usar plataformas de secuenciación, incluyendo, pero sin limitación, Genome Analyzer IIx, HiSeq, y MiSeq ofrecidas por Illumina, la tecnología en tiempo real para una sola molécula (SMRT™), tal como el sistema PacBio RS ofrecido por Pacific Biosciences (California) y el secuenciador Solexa, La tecnología de secuenciación de una sola molécula en tiempo real (tSMS™) tal como el secuenciador HeliScope™ ofrecido por Helicos Inc. (Cambridge, MA). La secuenciación puede comprender la secuenciación MiSeq. La secuenciación puede comprender la secuenciación HiSeq. En algunas realizaciones, determinar la secuencia de un polinucleótido comprende la secuenciación de extremos pareados, la secuenciación mediante nanoporos, la secuenciación de alto rendimiento, la secuenciación Shotgun, la secuenciación del terminador con colorante, la secuenciación del ADN de múltiples cebadores, el avance de cebadores, la secuenciación didesoxi de Sanger, la secuenciación de Maxim-Gilbert, la pirosecuenciación, la secuenciación de una sola molécula en tiempo real o cualquier combinación de las mismas. Como alternativa, la secuencia de un polinucleótido puede determinarse mediante microscopía electrónica o una matriz de transistores de efecto de campo sensibles a los productos químicos (chemFET).

55 Un procedimiento puede comprender además la secuenciación de uno o más polinucleótidos del banco. Un procedimiento puede comprender además alinear una o más secuencias de polinucleótidos, lecturas de secuencias, secuencias de amplicones o secuencias de conjuntos de amplicones del banco entre sí.

60 Como se usan en el presente documento, la alineación comprende comparar una secuencia de prueba, tal como una secuencia leída, con una o más secuencias de prueba, secuencias de referencia o una combinación de las mismas. En algunas realizaciones, la alineación puede usarse para determinar una secuencia de consenso a partir de una pluralidad de secuencias o secuencias alineadas. En algunas realizaciones, la alineación comprende determinar una secuencia de consenso a partir de una pluralidad de secuencias que tienen cada una un código de barras molecular o un código de barras del vaso idénticos. En algunas realizaciones, la longitud de una secuencia alineada con fines comparativos es al menos un 30 %, al menos el 40 %, al menos el 50 %, al menos el 60 %, al menos el 70 %, al menos el 80 %, al menos el 90 % o al menos el 95 %, de la longitud de una secuencia de referencia. La comparación real de las dos o más secuencias se puede realizar mediante procedimientos bien conocidos, por ejemplo, usando un

algoritmo matemático. Un ejemplo no limitante de dicho algoritmo matemático se describe en Karlin, S. y Altschul, S., *Proc. Natl. Acad. Sci. EE.UU.*, 90-5873-5877 (1993). Dicho algoritmo está incorporado en los programas NBLAST y XBLAST (versión 2.0), como se describe en Altschul, S. *et al.*, *Nucleic Acids Res.*, 25:3389-3402 (1997). Cuando se utilizan los programas BLAST y Gapped BLAST, se puede usar cualquier parámetro relevante de los programas respectivos (por ejemplo, NBLAST). Por ejemplo, los parámetros para la comparación de secuencias se pueden establecer en una puntuación = 100, longitud de palabra = 12, o se pueden variar (por ejemplo, W = 5 o W = 20). Otros ejemplos incluyen el algoritmo de Myers y Miller, CABIOS (1989), ADVANCE, ADAM, BLAT y FASTA. En algunas realizaciones, el porcentaje de identidad entre dos secuencias de aminoácidos se puede lograr usando, por ejemplo, el programa GAP en el paquete informático GCG (Accelrys, Cambridge, RU).

La secuenciación puede comprender la secuenciación de al menos aproximadamente 10, 20, 30, 40, 50, 60, 70, 80, 90, 100 o más nucleótidos o pares de bases de los polinucleótidos. En algunas realizaciones, la secuenciación comprende la secuenciación de al menos aproximadamente 200, 300, 400, 500, 600, 700, 800, 900, 1.000 o más nucleótidos o pares de bases de los polinucleótidos. En otros ejemplos, la secuenciación comprende la secuenciación de al menos aproximadamente 1.500, 2.000, 3.000, 4.000, 5.000, 6.000, 7.000, 8.000, 9.000, 10.000 o más nucleótidos o pares de bases de los polinucleótidos.

La secuenciación puede comprender al menos aproximadamente 200, 300, 400, 500, 600, 700, 800, 900, 1.000 o más lecturas de secuencia por serie. Como se usan en el presente documento, una lectura de secuencia comprende una secuencia de nucleótidos determinada a partir de una secuencia o un flujo de datos generados por una técnica de secuenciación. En algunas realizaciones, la secuenciación comprende la secuenciación de al menos aproximadamente 1.500, 2.000, 3.000, 4.000, 5.000, 6.000, 7.000, 8.000, 9.000, 10.000 o más lecturas de secuenciación por serie. La secuenciación puede comprender más de, menos de o igual a aproximadamente 1.000.000.000 de lecturas de secuencia por serie. La secuenciación puede comprender más de, menos o igual a aproximadamente 200.000.000 lecturas por serie.

En algunas realizaciones, el número de lecturas de secuencia usadas para determinar una secuencia de consenso es de aproximadamente 2-1000 lecturas de secuencia. Por ejemplo, el número de lecturas de secuencia usadas para determinar una secuencia de consenso puede ser de aproximadamente 2-900, 2-800, 2-700, 2-600, 2-500, 2-400, 2-300, 2-200, 2-100, 25-900, 25-800, 25-700, 25-600, 25-500, 25-400, 25-300, 25-200, 25-100, 100-1000, 100-900, 100-800, 100-700, 100-600, 100-500, 100-400, 100-300, 100-200, 200-1000, 200-900, 200-800, 200-700, 200-600, 200-500, 200-400, 200-300, 300-1000, 300-900, 300-800, 300-700, 300-600, 300-500, 300-400, 400-1000, 400-900, 400-800, 400-700, 400-600, 400-500, 500-1000, 500-900, 500-800, 500-700, 500-600, 600-1000, 600-900, 600-800, 600-700, 700-1.000, 700-900, 700-800, 800-1.000, 800-900 o 900-1.000 lecturas de secuencia. En algunas realizaciones, el número de lecturas de secuencia usadas para determinar una secuencia de consenso es de al menos aproximadamente 1.000, 1.500, 2.000, 3.000, 4.000, 5.000, 6.000, 7.000, 8.000, 9.000, 10.000, 11.000, 12.000, 13.000, 14.000, 15.000, 16.000, 17.000, 18.000, 19.000, 20.000, 25.000, 30.000, 35.000, 40.000, 45.000, 50.000, 55.000, 60.000, 65.000, 70.000, 75.000, 80.000, 85.000, 90.000, 95.000, 100.000, 150.000, 200.000, 250.000, 300.000, 350.000, 400.000, 450.000, 500.000, 550.000, 600.000, 650.000, 700.000, 750.000, 800.000, 850.000, 900.000, 950.000, 1.000.000, 50.000.000 o 100.000.000 lecturas. En algunas realizaciones, el número de lecturas de secuencia usadas para determinar una secuencia de consenso es como máximo de aproximadamente 1.000, 1.500, 2.000, 3.000, 4.000, 5.000, 6.000, 7.000, 8.000, 9.000, 10.000, 11.000, 12.000, 13.000, 14.000, 15.000, 16.000, 17.000, 18.000, 19.000, 20.000, 25.000, 30.000, 35.000, 40.000, 45.000, 50.000, 55.000, 60.000, 65.000, 70.000, 75.000, 80.000, 85.000, 90.000, 95.000, 100.000, 150.000, 200.000, 250.000, 300.000, 350.000, 400.000, 450.000, 500.000, 550.000, 600.000, 650.000, 700.000, 750.000, 800.000, 850.000, 900.000, 950.000, 1.000.000, 50.000.000 o 100.000.000 lecturas.

Un procedimiento puede comprender lecturas erróneas de secuenciación. Un procedimiento puede comprender determinar el número de lecturas erróneas, tal como para determinar una condición de reacción o diseñar secuencias de cebadores. La comparación del número de lecturas erróneas generadas en una o más de las primeras condiciones o conjuntos de condiciones se puede usar para determinar una condición o conjunto de condiciones preferidas. Por ejemplo, un primer procedimiento puede llevarse a cabo a una alta concentración de sal durante una reacción de PCR, y un segundo procedimiento puede llevarse a cabo a una baja concentración de sal durante una reacción de PCR, en el que el primer y el segundo procedimiento se llevan a cabo de manera esencialmente similar aparte de la diferencia de concentración de sal. Si el primer procedimiento produce un mayor número de lecturas erróneas, tal como un mayor número de lecturas erróneas para una secuencia o cebador de polinucleótido diana en particular, se puede determinar que se prefiere una condición de reacción con menor concentración de sal para esa secuencia o cebador de polinucleótido diana particular.

## 60 **DIAGNÓSTICO**

La presente divulgación proporciona un procedimiento que puede comprender además diagnosticar, pronosticar, controlar, tratar, mejorar y/o prevenir en un sujeto una enfermedad, un trastorno, un síntoma y/o una afección. La presente divulgación proporciona un procedimiento que puede comprender además diagnosticar, pronosticar, controlar, tratar, mejorar y/o prevenir en un sujeto una enfermedad, un trastorno, un síntoma y/o una afección, basándose en la presencia, ausencia o nivel de un polinucleótido diana. La presente divulgación proporciona un

procedimiento que puede comprender además diagnosticar, pronosticar, controlar, tratar, mejorar y/o prevenir en un sujeto una enfermedad, un trastorno, un síntoma y/o una afección, basándose en la presencia, ausencia o nivel de uno o más polinucleótidos diana.

- 5 La presente divulgación proporciona un procedimiento que puede comprender además diagnosticar, pronosticar, controlar, tratar, mejorar y/o prevenir en un sujeto una enfermedad, un trastorno, un síntoma y/o afección basándose en una presencia, una ausencia, un nivel o una secuencia de una o más de las secuencias obtenidas usando los procedimientos descritos en el presente documento. Por ejemplo, se puede realizar un diagnóstico de una enfermedad basándose en la presencia, la ausencia, el nivel o la secuencia de una secuencia variante obtenida usando los procedimientos descritos en el presente documento. La presente divulgación proporciona un procedimiento que puede comprender además diagnosticar, pronosticar, controlar, tratar, mejorar y/o prevenir en un sujeto una enfermedad, un trastorno, un síntoma y/o afección basándose en una presencia, una ausencia, un nivel o una secuencia, uno o más de las lecturas de secuencia obtenidas usando los procedimientos descritos en el presente documento. La presente divulgación proporciona un procedimiento que puede comprender además diagnosticar, pronosticar, controlar, tratar, mejorar y/o prevenir en un sujeto una enfermedad, un trastorno, un síntoma y/o afección basándose en una presencia, una ausencia, un nivel o una secuencia de una o más de las secuencias de consenso obtenidas usando los procedimientos descritos en el presente documento. La presente divulgación proporciona un procedimiento que puede comprender además diagnosticar, pronosticar, controlar, tratar, mejorar y/o prevenir en un sujeto una enfermedad, un trastorno, un síntoma y/o afección basándose en la determinación de un nivel (por ejemplo, una cantidad o concentración) de un polinucleótido diana en una muestra. Un nivel de un polinucleótido diana en una muestra puede determinarse basándose en una o más lecturas de secuencia, secuencias, secuencias de consenso o cualquier combinación de las mismas. Un nivel de cada uno de una pluralidad de polinucleótidos diana en una muestra puede determinarse usando los procedimientos descritos en el presente documento. Un nivel de cada uno de una pluralidad de polinucleótidos diana en una muestra puede determinarse basándose en una serie de lecturas de secuencias, secuencias, secuencias de consenso o cualquier combinación de los mismos de cada polinucleótido diana de la pluralidad. Por ejemplo, se puede determinar un nivel de un primer polinucleótido diana y un nivel de un segundo polinucleótido diana usando los procedimientos descritos en el presente documento.

- En algunas realizaciones, el primer y el segundo polinucleótido diana de una pluralidad de polinucleótidos diana son los mismo. Por ejemplo, un primer polinucleótido diana puede comprender una primera copia de una molécula de ARNm y un segundo polinucleótido diana puede comprender una segunda copia de una molécula de ARNm. En algunas realizaciones, el primer y el segundo polinucleótido diana son diferentes. Por ejemplo, un primer polinucleótido diana puede comprender una primera molécula de ARNm y un segundo polinucleótido diana puede comprender una segunda molécula de ARNm transcrita de un gen diferente que la primera molécula de ARNm. Por ejemplo, un primer polinucleótido diana puede comprender un primer alelo y un segundo polinucleótido diana puede comprender un segundo alelo. Por ejemplo, un primer polinucleótido diana puede comprender una secuencia de tipo silvestre y un segundo polinucleótido diana puede comprender una secuencia variante.

- En algunas realizaciones, un procedimiento puede comprender además diagnosticar o pronosticar a un sujeto con una enfermedad, un trastorno, un síntoma y/o una afección con al menos un 50 % de confianza. Por ejemplo, un diagnóstico o un pronóstico de un sujeto con una enfermedad, un trastorno, un síntoma y/o una afección se puede determinar con al menos un 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 98 %, 99 % o 100 % de confianza. En algunas realizaciones, un diagnóstico o un pronóstico de un sujeto con una enfermedad, un trastorno, un síntoma y/o una afección se puede determinar con un 50 %-100 % de confianza. Por ejemplo, un diagnóstico o un pronóstico de un sujeto con una enfermedad, un trastorno, un síntoma y/o una afección se puede determinar con al menos un 60 %-100 %, 70 %-100 %, 80 %-100 %, 90 %-100 %, 50 %-90 %, 50 %-80 %, 50 %-70 %, 50 %-60 %, 60 %-90 %, 60 %-80 %, 60 %-70 %, 70 %-90 %, 70 %-80 % u 80 %-90 % de confianza.

- En algunas realizaciones, la presencia, la ausencia, el nivel, la secuencia o cualquier combinación de los mismos, de un polinucleótido diana en el sujeto, tal como un biomarcador, se puede determinar con al menos un 50 % de confianza. Por ejemplo, la presencia, la ausencia, el nivel, la secuencia o cualquier combinación de los mismos, de un polinucleótido diana en el sujeto puede determinarse con al menos un 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 98 %, 99 % o 100 % de confianza. En algunas realizaciones, la presencia, la ausencia, el nivel, la secuencia o cualquier combinación de los mismos, de un polinucleótido diana en el sujeto se puede determinar con un 50 %-100% de confianza. Por ejemplo, la presencia, la ausencia, el nivel, la secuencia o cualquier combinación de los mismos, de un polinucleótido diana en el sujeto se puede determinar con un 60 %-100 %, 70 %-100 %, 80 %-100 %, 90 %-100 %, 50 %-90 %, 50 %-80 %, 50 %-70 %, 50 %-60 %, 60 %-90 %, 60 %-80 %, 60 %-70 %, 70 %-90 %, 70 %-80 % u 80 %-90 % de confianza.

## 60 ENZIMAS

Los procedimientos y kits desvelados en el presente documento pueden comprender una o más enzimas. Los ejemplos de enzimas incluyen, pero sin limitación, ligasas, transcriptasas inversas, polimerasas y nucleasas de restricción.

- 65 En algunas realizaciones, la unión de un adaptador a polinucleótidos comprende el uso de una o más ligasas. Los ejemplos de ligasas incluyen, aunque sin limitación, ADN ligasas tales como ADN ligasa I, ADN ligasa III, ADN ligasa

IV y ADN ligasa T4, y ARN ligasas tales como ARN ligasa I T4 y ARN ligasa II T4.

Los procedimientos y kits desvelados en el presente documento pueden comprender además el uso de una o más transcriptasas inversas. En algunas realizaciones, la transcriptasa inversa es una transcriptasa inversa del VIH-1, la transcriptasa inversa del M-MLV, la transcriptasa inversa del AMV y la transcriptasa inversa de la telomerasa. En algunas realizaciones, la transcriptasa inversa es la transcriptasa inversa del M-MLV.

En algunas realizaciones, los procedimientos y kits desvelados en el presente documento comprenden el uso de una o más proteasas.

En algunas realizaciones, los procedimientos y kits desvelados en el presente documento comprenden el uso de una o más polimerasas. Los ejemplos de polimerasas incluyen, aunque sin limitación, ADN polimerasas y ARN polimerasas. En algunas realizaciones, la ADN polimerasa es una ADN polimerasa I, ADN polimerasa II, holoenzima ADN polimerasa III y ADN polimerasa IV. Las ADN polimerasas disponibles en el mercado incluyen, aunque sin limitación, ADN polimerasa Bst 2.0, ADN polimerasa WarmStart™ Bst 2.0, ADN polimerasa Bst, ADN polimerasa IV Sulfolobus, ADN polimerasa Taq, ADN polimerasa 9 N™m, ADN polimerasa Deep VentR™ (exo-), ADN polimerasa Deep VentR™, Hemo KlenTaq™, ADN polimerasa Taq LongAmp®, ADN polimerasa OneTaq®, ADN polimerasa Phusion®, ADN polimerasa de alta fidelidad Q5™, ADN polimerasa y Therminator™, ADN polimerasa Therminator™, ADN polimerasa Therminator™ II, ADN polimerasa Therminator™ III, ADN polimerasa VentR®, ADN polimerasa VentR® (exo-), ADN polimerasa Bsu, ADN polimerasa phi29, ADN polimerasa T4, ADN polimerasa T7, Transferasa terminal, polimerasa Taq Titanium®, ADN polimerasa Taq KAPA y ADN polimerasa de inicio en caliente Taq KAPA.

En algunas realizaciones, la polimerasa es una ARN polimerasa tal como ARN polimerasa I, ARN polimerasa II, ARN polimerasa III, polimerasa Poli (A) de *E. coli*, ARN polimerasa phi6 (RdRP), polimerasa Poli (U), ARN polimerasa SP6 y ARN polimerasa T7.

## REACTIVOS ADICIONALES

Los procedimientos y kits desvelados en el presente documento pueden comprender el uso de uno o más reactivos. Los ejemplos de reactivos incluyen, aunque sin limitación, reactivos de PCR, reactivos de ligadura, reactivos de transcripción inversa, reactivos enzimáticos, reactivos de hibridación, reactivos de preparación de muestras, reactivos de captura de afinidad, soportes sólidos tales como perlas y reactivos para la purificación y/o el aislamiento de ácido nucleico.

Un soporte sólido puede comprender casi cualquier material sólido o insoluble, y a menudo se selecciona una composición de soporte sólida que sea insoluble en el agua. Por ejemplo, un soporte sólido puede comprender o consistir esencialmente en gel de sílice, vidrio (por ejemplo, vidrio de poro controlado (CPG)), nylon, Sephadex®, Sepharose®, celulosa, una superficie metálica (por ejemplo, acero, oro, plata, aluminio, silicio y cobre), un material magnético, un material plástico (por ejemplo, polietileno, polipropileno, poliamida, poliéster, difluoruro de polivinilideno (PVDF) y similares. Los ejemplos de perlas para su uso de acuerdo con las realizaciones pueden incluir una fracción de afinidad que permita que la perla interactúe con una molécula de ácido nucleico. Una fase sólida (por ejemplo, una perla) puede comprender un miembro de un par de unión (por ejemplo, avidina, estreptavidina o un derivado de las mismas). Por ejemplo, la perla puede ser una perla recubierta con estreptavidina, y una molécula de ácido nucleico para la inmovilización en la perla puede incluir una fracción de biotina. En algunos casos, cada molécula de polinucleótido puede incluir dos fracciones de afinidad, tales como biotina, para estabilizar aún más el polinucleótido. Las perlas pueden incluir características adicionales para su uso en la inmovilización de ácidos nucleicos o que se pueden usar en un proceso de detección o selección cadena abajo. Por ejemplo, la perla puede incluir una fracción de unión, un marcador fluorescente o un inactivador fluorescente. En algunos casos, la perla puede ser magnética. En algunos casos, el soporte sólido es una perla. Los ejemplos de perlas incluyen, aunque sin limitación, perlas de estreptavidina, perlas de agarosa, perlas magnéticas, Dynabeads®, microperlas MACS®, perlas conjugadas con anticuerpos (p. ej., microesfera anti-inmunoglobulina), perlas conjugadas con proteína A, perlas conjugadas con proteína G, perlas conjugadas con proteína A/G, perlas conjugadas con proteína L, perlas conjugadas con polinucleótido dT, perlas de sílice, perlas de tipo sílice, microperlas anti-biotina, microperla antifluorocromadas y perlas magnéticas terminadas en carboxi BcMag™. Las perlas o partículas pueden ser hinchables (por ejemplo, perlas poliméricas tales como la resina Wang) o no hinchables (por ejemplo, CPG). En algunas realizaciones, una fase sólida es esencialmente hidrófila. En algunas realizaciones, una fase sólida (por ejemplo, una perla) es esencialmente hidrófoba. En algunas realizaciones, una fase sólida comprende un miembro de un par de unión (por ejemplo, avidina, estreptavidina o un derivado de las mismas) y es esencialmente hidrófoba o esencialmente hidrófila. En algunas realizaciones, una fase sólida comprende un miembro de un par de unión (por ejemplo, avidina, estreptavidina o un derivado de las mismas) y tiene una capacidad de unión superior a aproximadamente 1.350 picomoles de agente de captura libre (por ejemplo, biotina libre) por mg de soporte sólido. En algunas realizaciones, la capacidad de unión de la fase sólida que comprende un miembro de un par de unión es superior a 800, 900, 1.000, 1.100, 1.200, 1.250, 1.300, 1.350, 1.400, 1.450, 1.500, 1.600, 1.800, 2.000 picomoles de agente de captura libre por mg de soporte sólido. Otros ejemplos de perlas que son adecuados para la invención son coloides de oro o perlas tales como perlas de poliestireno o perlas de sílice. Se puede usar esencialmente cualquier radio de perla. Los ejemplos de perlas pueden incluir perlas que tienen un radio que varía entre 150 nanómetros y 10 micrómetros. También se pueden usar otros

tamaños.

Los procedimientos y kits desvelados en el presente documento pueden comprender el uso de uno o más de tampones. Los ejemplos de tampones incluyen, aunque sin limitación, tampones de lavado, tampones de ligadura, 5 tampones de hibridación, tampones de amplificación y tampones de transcripción inversa. En algunas realizaciones, el tampón de hibridación es un tampón disponible en el mercado, tal como solución Hyb de TMAC, solución de hibridación de SSPE y tampón de hibridación de ECONO™. Los tampones desvelados en el presente documento pueden comprender uno o más detergentes.

10 los procedimientos y kits desvelados en el presente documento pueden comprender el uso de uno o más de vehículos. Los vehículos pueden potenciar o mejorar la eficacia de una o más reacciones desveladas en el presente documento (por ejemplo, la reacción de ligadura, transcripción inversa, amplificación, hibridación). Los vehículos pueden disminuir o prevenir la pérdida inespecífica de las moléculas o cualquiera de sus productos (por ejemplo, un polinucleótido y/o amplicón). Por ejemplo, el vehículo puede reducir la pérdida inespecífica de un polinucleótido a través de la absorción a las superficies. El vehículo puede reducir la afinidad de un polinucleótido a una superficie o sustrato (por ejemplo, 15 recipiente, tubo Eppendorf, punta de pipeta). Como alternativa, el vehículo puede aumentar la afinidad de un polinucleótido a una superficie o sustrato (por ejemplo, perla, matriz, vidrio, portaobjetos, chip). Los vehículos pueden proteger al polinucleótido de la degradación. Por ejemplo, los vehículos pueden proteger una molécula de ARN de las ribonucleasas. Como alternativa, los vehículos pueden proteger una molécula de ADN de una ADNasa. Los ejemplos de vehículos incluyen, aunque sin limitación, polinucleótidos tales como ADN y/o ARN, o polipéptidos. Los ejemplos de vehículos de ADN incluyen plásmidos, vectores, ADN poliadenilado y polinucleótidos de ADN. Los ejemplos de vehículos de ARN incluyen ARN poliadenilado, ARN de fago, ARN de fago MS2, ARN de *E.coli*, ARN de levadura, ARNt de levadura, ARN de mamífero, ARNt de mamífero, ribonucleótidos sintéticos poliadenilados cortos y polinucleótidos de ARN. El vehículo de ARN puede ser un ARN poliadenilado. Como alternativa, el vehículo de ARN 20 puede ser un ARN no poliadenilado. En algunas realizaciones, el vehículo es de una bacteria, levadura o virus. Por ejemplo, el vehículo puede ser un polinucleótido o un polipéptido derivado de una bacteria, de una levadura o de un virus. Por ejemplo, el vehículo es una proteína de *Bacillus subtilis*. En otro ejemplo, el vehículo es un polinucleótido de *Escherichia coli*. Como alternativa, el vehículo es un polinucleótido o péptido de un mamífero (por ejemplo, ser humano, ratón, cabra, rata, vaca, ovejas, cerdo, perro o conejo), ave, anfibio o reptil.

30 los procedimientos y kits desvelados en el presente documento pueden comprender el uso de uno o más de agentes de control. Los agentes de control pueden incluir polinucleótidos de control, enzimas inactivas, competidores inespecíficos. Como alternativa, los agentes de control comprenden hibridación brillante, controles de sonda brillante, moldes de ácido nucleico, controles de adiciones, controles de amplificación por PCR. Los controles de amplificación por PCR pueden ser controles positivos. En otros ejemplos, los controles de amplificación por PCR son controles negativos. Los controles de molde de ácido nucleico pueden ser de concentraciones conocidas. Los agentes de control pueden comprender uno o más marcadores.

40 Los controles de adiciones pueden ser moldes que se añaden a una reacción o muestra. Por ejemplo, se puede añadir un molde de adición a una reacción de amplificación. El molde de adición se puede añadir a la reacción de amplificación en cualquier momento después del primer ciclo de amplificación. En algunas realizaciones, el molde de adición se añade a una reacción de amplificación después del ciclo número 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45 o 50. El molde de adición se puede añadir a la reacción de amplificación en cualquier momento antes del último ciclo de amplificación. El molde de adición puede comprender uno o más nucleótidos o pares de bases de ácido nucleico. El molde de adición puede comprender ADN, ARN o cualquier combinación de los mismos. El molde de adición puede comprender uno o más marcadores.

50 En el presente documento, se desvelan moléculas, materiales, composiciones y componentes que se pueden usar para, se pueden usar junto con, se pueden usar en la preparación para, o son productos de procedimientos y composiciones desvelados en el presente documento. Es entiende que cuando se desvelan combinaciones, subconjuntos, interacciones, grupos, etc. de estos materiales, y aunque la referencia específica a cada una de varias combinaciones colectivas e individuales y la permutación de estos compuestos puede no desvelarse explícitamente, cada uno se contempla específicamente y se describe en el presente documento. Por ejemplo, si se desvela y describe un nucleótido o un ácido nucleico, y se describe una serie de modificaciones que se pueden realizar a varias moléculas, 55 incluyendo el nucleótido o el ácido nucleico, se contemplan específicamente todas y cada una de las combinaciones y permutaciones del nucleótido o del ácido nucleico y las modificaciones que son posibles, a menos que se indique específicamente al contrario. Este concepto se aplica a todos los aspectos de la presente solicitud incluyendo, pero sin limitación, las etapas de los procedimientos para elaborar y usar los procedimientos y las composiciones que se desvelan. Por lo tanto, si hay una variedad de etapas adicionales que se pueden realizar, se entiende que cada una de estas etapas adicionales se puede realizar con cualquier realización específica o combinación de realizaciones específicas de los procedimientos desvelados, y que cada combinación se contempla específicamente y se debe considerar como desvelada.

65 Aunque algunas realizaciones descritas en el presente documento se han mostrado y descrito en el presente documento, dichas realizaciones se proporcionan solo a modo de ejemplo. A los expertos en la materia, se les ocurrirán numerosas variaciones, cambios y sustituciones sin apartarse de la divulgación proporcionada en el presente

documento. Debe entenderse que pueden emplearse diversas alternativas a las realizaciones descritas en el presente documento para poner en práctica los procedimientos descritos en el presente documento.

A menos que se indique lo contrario, todos los términos técnicos y científicos usados en el presente documento tienen el mismo significado comúnmente entendido por un experto habitual en la materia a la que pertenece la presente divulgación. Las siguientes referencias contienen realizaciones de los procedimientos y de las composiciones que se pueden usar en el presente documento: "The Merck Manual of Diagnosis and Therapy", 18ª edición, publicado por Merck Research Laboratories, 2006 (ISBN 0-9119102); Benjamin Lewin, "Genes IX", publicado por Jones & Bartlett Publishing, 2007 (ISBN-13: 9780763740634); Kendrew *et al.* (eds.), "The Encyclopedia of Mol. Biology", publicado por Blackwell Science Ltd., 1994 (ISBN 0-632-02182-9); y Robert A. Meyers (ed.), "Mol. Biology and Biotechnology: a Comprehensive Desk Reference", publicado por VCH Publishers, Inc., 1995 (ISBN 1-56081-569-8).

Los procedimientos convencionales de la presente divulgación se describen, por ejemplo, en Maniatis *et al.*, "Molecular Cloning: A Laboratory Manual", Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., EE.UU. (1982); Sambrook *et al.*, "Molecular Cloning: A Laboratory Manual" (2 ed.), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., EE.UU. (1989); Davis *et al.*, "Basic Methods in Molecular Biology", Elsevier Science Publishing, Inc., Nueva York, EE.UU. (1986); o "Methods in Enzymology: Guide to Molecular Cloning Techniques" Vol. 152, S. L. Berger y A. R. Kimmel (eds.), Academic Press Inc., San Diego, EE.UU. (1987). "Current Protocols in Molecular Biology" (CPMB) (Fred M. Ausubel, *et al.* ed., John Wiley and Sons, Inc.), "Current Protocols in Protein Science (CPPS)" (John E. Coligan, *et al.*, ed., John Wiley and Sons, Inc.), "Current Protocols in Immunology (CPI)" (John E. Coligan, *et al.*, ed. John Wiley and Sons, Inc.), "Current Protocols in Cell Biology (CPCB)" (Juan S. Bonifacino *et al.* ed., John Wiley and Sons, Inc.), "Culture of Animal Cells: A Manual of Basic Technique" por R. Ian Freshney, Publisher: Wiley-Liss; 5ª edición (2005), y "Animal Cell Culture Methods" (Methods in Cell Biology, Vol. 57, editores Jennie P. Mather y David Barnes, Academic Press, 1ª edición, 1998).

25

## Ejemplos

### **Ejemplo 1a - Protocolo para preparar células a fin de realizar una secuenciación masiva de polinucleótidos unicelulares de alto rendimiento a base de una emulsión.**

30

Se obtuvieron poblaciones celulares de interés. Estas incluían PBMC totales, células clasificadas, linfocitos B o T enriquecidos con anticuerpos, u otros tipos de células. Las células tenían una membrana plasmática intacta de modo que no filtraban cantidades excesivas de ARNm al medio circundante. Las células no necesitaban ser viables.

35

Las células se lavaron por centrifugación (200 g durante 10 min para los linfocitos T o linfocitos B) dos veces en tampón celular: 1 x Suero salino tamponado con fosfato de Dulbecco (PBS). Las células se diluyeron luego en tampón celular hasta una concentración celular de  $3,5 \times 10^6$ /ml. La suspensión se pipeteó luego a través de un filtro de células de 20  $\mu$ m.

40

### **Ejemplo 1b: Protocolo para la preparación de tejidos sólidos para la secuenciación masiva de polinucleótidos unicelulares con un alto rendimiento a base de una emulsión**

Se trató un tejido sólido (por ejemplo, una muestra de biopsia tumoral o no tumoral) con varias proteasas, incluyendo la colagenasa III (200 U/ml), DNasa I (200 U/ml) y tripsina (5 mg/ml) y un NEDB (Invitrogen) para producir una mezcla de células individuales y agregados que contienen más de una célula. Brevemente, se añadieron tumores extraídos de los ratones a medios de cultivo fríos y se eliminaron la grasa y el tejido mamario circundante del ratón. Se trituraron los tumores en fragmentos de 2 a 4 mm, que luego se incubaron con las soluciones de disociación o enzimas apropiadas durante 30 minutos a 37 °C. Se mezclaron los fragmentos tumorales con movimientos ascendentes y descendentes cada 10 minutos usando una micropipeta de 1.000 ml con un corte de punta en un diámetro adaptado al tamaño del fragmento tisular. Después de cada período de incubación, los fragmentos se filtraron a través de un filtro de células de malla de nylon de 40 mm. Se centrifugaron las células liberadas a 1.200 rpm durante 2 minutos y se almacenaron en medio frío con FCS al 30 % a 4 °C. Se añadió solución de disociación nueva a los fragmentos de tejido restantes durante 30 min. La disociación se detuvo cuando se dejaron de liberar células. Se presionaron los fragmentos a través de un tamiz y todas las células de todos los períodos de incubación se agruparon y se contaron. Las suspensiones celulares luego se filtran a través de un filtro (por ejemplo, 10, 20, 30, 40  $\mu$ m) para eliminar los agregados de gran tamaño. Las células se lavaron por centrifugación (200 g durante 10 minutos para los linfocitos T o linfocitos B) dos veces en tampón celular: 1 x Suero salino tamponado con fosfato de Dulbecco (PBS). La población celular no se tiñó, se clasificó o se separó de otro modo antes del análisis por emulsión.

60

También se realizó un procedimiento alternativo para preparar los tumores extraídos. Se dispusieron los tumores extraídos en 1 ml de tampón de disociación 1 (100 U/ml de colagenasa de tipo IV y 100  $\mu$ g/ml de ADNasa en RPMI + FBS al 10 %) o tampón de disociación 2 (medio RPMI complementado con FBS al 5 %, colagenasa de tipo I (200 U/ml) y DNasa I (100  $\mu$ g/ml) e incubados durante 30 min a 37 °C. Si las células mieloides debían aislarse posteriormente, se sustituyeron el FBS al 5 % y la colagenasa de tipo I en el tampón de disociación 2 con FBS al 10 % y colagenasa de tipo IV (200 U/ml), respectivamente. Se mezclaron los fragmentos tumorales con movimientos ascendentes y descendentes usando una micropipeta de 1.000 ml. La suspensión se filtró luego a través de un filtro de 70  $\mu$ m y se

65

- 5 lavó 3 veces con tampón de separación de MAC complementado con FBS al 10 % para el aislamiento de las células mieloides. Para tumores muy grandes (> 300 mm<sup>2</sup>), las células inflamatorias pueden enriquecerse previamente mediante centrifugación en gradiente de densidad (Percoll o Ficoll). A continuación, se centrifugó la suspensión celular filtrada a 400 g durante 10 min. Se enjuagó el sedimento con 10 ml de tampón MAC y se centrifugó nuevamente con los mismos ajustes.

**Ejemplo 2: Protocolo para preparar la mezcla de reacción en emulsión para realizar una secuenciación masiva de polinucleótidos unicelulares de alto rendimiento a base de una emulsión.**

- 10 Se mezcló una mezcla de reacción en emulsión que contenía los reactivos y oligonucleótidos de la siguiente Tabla 1 a temperatura ambiente en una campana de limpieza por PCR.

**Tabla 1**

Reactivo	Conc. de reserva (mM)	Conc. final de la gotita (mM)	Conc. final de la fase de reacción (mM)	µl por 200 µl
Tris-Cl, pH 8,0	500,00	50,00	100,00	40,00
MgSO <sub>4</sub>	100,00	3,00	6,00	12,00
DTT	1.000,00	10,00	20,00	4,00
Cada dNTP	10,00	0,50	1,00	20,00
oligo-dT con biotina en 5'	1,40 x 10 <sup>-2</sup>	2,50 x 10 <sup>-4</sup>	5,00 x 10 <sup>-4</sup>	7,14
Oligo de cambio de molde	0,1	1,00 x 10 <sup>-3</sup>	2,00 x 10 <sup>-3</sup>	4,00
Moléculas de molde con DB/µl	1,00 x 10 <sup>6</sup>	1,75 x 10 <sup>4</sup>	3,50 x 10 <sup>4</sup>	7,00
Cebador con DB directo	0,2	5,00 x 10 <sup>-4</sup>	1,00 x 10 <sup>-3</sup>	1,00
Cebador con DB inverso	0,2	7,50 x 10 <sup>-4</sup>	1,50 x 10 <sup>-3</sup>	1,50
Inhibidor de HALT proteasa (X)	200	1,00	2,00	2,00
Inhibidor de RNasas enzimáticas (U/µl)	40	0,40	0,80	4,00
Transcriptasa inversa de RNasa H del MMLV				10,00
ADN polimerasa Phusion HF				10,00
Triton X-100 (% v/v)	2,5	0,25	0,50	40,00
Agua				hasta 200
<b>Secuencias de oligonucleótidos:</b>				
Cebador de transcripción inversa anclado a oligo-dT con biotina en 5'	/5BiosG//iSp18/TTT TTT TTT TTT TTT TTT TTT TTT T V N			
Molde con código de barras de la gotita:	ATCCATCCACGACTGACGGACGTATTAANNNNNWNNNNW NNNNAGATCGGAAGAGCACACGTCTGAACTCCAGTCACC			
oligo de cambio de molde	AATACGTCCGTCAGTCGTGGATGNNTNANNTGrGG			
Código de barras del vaso directo	CATCCACGACTGACGGACGTATT			
Código de barras del vaso inverso	GTGACTGGAGTTCAGACGTGTGCT			
/5Biosg/= modificación con biotina en 5'; /iSp 18/= espaciador de 18 átomos de carbono; V= A, C o G; N = cualquier base; rG = riboguanosina; W = A o T.				

- 15 **Ejemplo 3: Protocolo de generación de emulsiones para realizar una secuenciación masiva de polinucleótidos unicelulares de alto rendimiento a base de emulsiones.**

- 20 Una vez preparadas las células y la mezcla de reacción, se formó la emulsión. Se usó una jeringa Hamilton Microliter de 100 µl para sobrecargar un bucle de muestras de PEEK de 100 µl en dos inyecciones de ~100 µl cada una de la mezcla de reacción. Se usó una jeringa Hamilton Gastight de 100 µl para cargar ~110 µl de la suspensión celular en un bucle de tubo FEP de 0,2 mm de diámetro interno y ~100 µl. Se conectó el bucle a un rotador mecánico que invirtió constantemente el bucle de células aproximadamente una vez cada 1-2 segundos para evitar la acumulación y el agrupamiento de las células. La emulsión se formó por chorro de flujo enfocado a través de un chip de reactivo

Dolomite 2 con recubrimiento fluorófilo interno. Los canales de aceite externos contenían tensioactivo a base de polietilenglicol al 0,5-5,0 % (p/v) en aceite de fluorocarbono HFE7500 (Novec 7500). El chorro de emulsión se aplicó a un caudal constante (igual en los canales de la fase celular y de la fase de reacción). La salida del chip de emulsión se recogió a través de un tubo PEEK de 12 cm y 0,5 mm de diámetro interno, cayendo en tubos de polipropileno para PCR que se mantenían a aproximadamente 0 °C en un bloque refrigerado. Se recogieron cuatro fracciones, cada una de las cuales contenía 50 µl de material acuoso en emulsión (5 minutos de tiempo de ejecución por fracción). Se retiró la mayor parte del aceite sedimentado de la parte inferior de cada tubo con una micropipeta capilar. Cada fracción de emulsión se cubrió suavemente con 40 µl de solución de recubrimiento: Na-EDTA 50 mM, pH 8,0, rojo crisol al 0,002 % (p/v). Las emulsiones se incubaron en un termociclador con el siguiente programa (minutos: segundos):

1. 42,0 °C durante 30:00 (transcripción inversa)
2. 95,0 °C durante 05:00 (moldes de transcriptasa inversa desnaturalizada y de ADN)
3. 95,0 °C durante 00:10
4. 65,0 °C durante 00:30
5. 72,0 °C durante 00:30
6. Pasar a 3, 55 ciclos en total (se amplifica el código de barras del vaso y se fusiona a ADNc)
7. 4,0 °C sin límite de tiempo.

La emulsión se mantuvo a 4,0°C durante la noche.

**Ejemplo 4: Protocolo de ruptura de emulsiones para realizar una secuenciación masiva de polinucleótidos unicelulares de alto rendimiento a base de emulsiones.**

Usando una punta de micropipeta capilar, se retiró la mayor cantidad posible de la solución de recubrimiento sin eliminar el material de emulsión. A cada tubo, se añadieron 12,5 µl de solución de proteasa Qiagen y 2,5 µl de EDTA de Na 0,5 M, pH 8,0. La emulsión se rompió añadiendo 40 µl de FC-40:perfluorooctanol a 1:1 e invirtiendo suavemente aproximadamente 10 veces.

Se centrifugó el contenido del tubo suavemente y se incubó en un termociclador con el siguiente programa (minutos: segundos):

1. 50 °C para 15:00 (digestión con proteasa)
2. 70 °C para 10:00 (inactivación de proteasa)
3. 95 °C para 03:00 (inactivación de proteasa y desnaturalización del ADN)
4. 4,0 °C sin límite de tiempo.

Se centrifugó el tubo, y se trasladaron la fase acuosa superior y la superficie de contacto a un tubo de microcentrífuga nuevo, y se centrifugaron a 15.000 g durante 1 minuto. Se transfirió la fase acuosa superior a un nuevo tubo, sin alterar la superficie de contacto.

**Ejemplo 5: Protocolo de limpieza de los polinucleótidos de emulsiones para realizar una secuenciación masiva de polinucleótidos unicelulares de alto rendimiento a base de emulsiones.**

Se añadieron 0,25 V de perlas de estreptavidina NEB en 2 x BW (Tris-Cl 10 mM, pH 8,0, EDTA 1 mM, NaCl 2 M, Tween-20 al 0,2 % ) y se incubaron a temperatura ambiente durante 15 min. Luego, las perlas se lavaron con 1 x BW, se lavaron tres veces con Tween-20 al 0,001 % y se eluyeron añadiendo 0,25 V de Tween-20 al 0,001 % y calentando a 95 °C durante 3 min. Se añadieron 5 volúmenes de tampón Qiagen PB y se aplicaron a una columna de sílice Zypzy. A continuación, se lavaron las perlas con 0,7 ml de tampón de lavado Zypzy y se eluyeron en 180 µl de: Tris-Cl 5 mM, pH 8,8, EDTA 0,1 mM, Tween-20 al 0,001 %.

**Ejemplo 6 - Protocolo para la primera reacción de PCR (PCR1) de polinucleótidos para la secuenciación de próxima generación a fin de realizar una secuenciación masiva de polinucleótidos unicelulares de alto rendimiento a base de emulsiones.**

Se usaron 163,2 µl de ADNc purificado para la PCR1. En la siguiente Tabla 2, se muestra un ejemplo de configuración para la primera reacción de PCR.

Tabla 2

Banco de la PCR1					
Reactivo	Conc. de reserva	Conc. final	Reacción de 20 µl	Reacción de 60 µl	Reacción de 240 µl
Tampón Q5 x5	5,00 mM	1,00 µM	4,00 µl	12,00 µl	48,00 µl
Cada dNTP	10,00 mM	0,20 µM	0,40 µl	1,20 µl	4,80 µl
Inicio de Q5 caliente	125,00 mM	1,00 µM	0,16 µl	0,48 µl	1,92 µl
Cebador 633	10 µm		0,16 µl	0,48 µl	1,92 µl
Mezcla de cebadores [IgH/TCRα]-[IgL/TCRβ]-[C]	10 µm (cada uno)		0,16 µl	0,48 µl	1,92 µl
ADNc			13,60 µl	40,80 µl	163,20 µl
H <sub>2</sub> O			1,52 µl	4,56 µl	18,24 µl
<b>Secuencias de cebadores IgH/TCRα/TCRγ de mezcla de cebadores [IgH/TCRα /TCRγ]-[IgL/TCRβ/TCRδ]-[C]</b>					
IgM	GGGTTGGGGCGGATGCAC				
IgD	CATCCGGAGCCTTGGTGG				
IgA	CCTTGGGGCTGGTCGGGG				
IgE	CGGATGGGCTCTGTGTGG				
IgG	CCGATGGGCCCTTGGTGG				
TCRα 1	GGATTTAGAGTCTCTCAGCTG				
TCRα2	CACGGCAGGGTCAGGGTTC				
TCRγ	AAAATAGTGGGCTTGGGG				
<b>Secuencias de cebadores IgL/TCRβ/TCRδ de mezcla de cebadores [IgH/TCRα /TCRγ]-[IgL/TCRβ/TCRδ]-[C]</b>					
IgKJ1	TTTGATCTCCACCTTGGTCCCTCCGC				
IgKJ2	TTTGATCTCCAGCTTGGTCCCCTGG				
IgKJ3	TTTGATATCCACTTTGGTCCCAGGGC				
IgKJ4	TTTGATTTCCACCTTGGTCCCTTGGC				
IgKJ5	TTTAATCTCCAGTCGTGTCCCTTGGC				
IgLJ1	GAGGACGGTCACCTTGGTGCCA				
IgLJ2	TAGGACGGTCAGCTTGGTCCCTCC				
IgLJ3	GAGGACGGTCAGCTGGGTGCC				
IgLJ4	TAAAATGATCAGCTGGGTTCTCCAC				
IgLJ5	TAGGACGGTGACCTTGGTCCCAGT				
TCRβ1	GGGAGATCTCTGCTTCTGATG				
TCRβ2	CGACCTCGGGTGGGAACAC				
TCRδ	AGACAAGCGACATTTGTTCCA				
<b>Secuencia de cebador C de mezcla de cebadores [IgH/TCRα /TCRγ]-[IgL/TCRβ/TCRδ]-[C]</b>					
633	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT				

Se dividieron en partes alícuotas cuatro reacciones de 60 µl en los tubos de PCR y se realizó el siguiente programa en un termociclador:

- 5 1. 98 °C durante 01:00
2. 98 °C durante 00:10
- 10 3. 64 °C durante 00:20
4. 72 °C durante 00:20
5. Se pasa a 2 durante un total de 6 ciclos
- 15 6. 4 °C sin límite de tiempo.

El producto de la PCR se purificó con 1,2 volúmenes de AMPure XP, se lavó con etanol al 80 % y se eluyó en 60 µl de tampón de dilución (Tris-Cl 10 mM, pH 8,0, EDTA 0,1 mM)

**Ejemplo 7 - Protocolo para la segunda reacción de PCR (PCR2) de polinucleótidos para la secuenciación de próxima generación a fin de realizar una secuenciación masiva de polinucleótidos unicelulares de alto rendimiento a base de emulsiones.**

## ES 2 727 656 T3

Se usaron 20  $\mu$ l de producto de PCR1 purificado para cada banco secundario (por ejemplo, cadena de IgL o IgH o cadena de TCR $\alpha$  o TCR $\beta$ , o cadena de TCR $\gamma$  o TCR $\delta$ ). En la siguiente Tabla 3, se muestra un ejemplo de configuración para la segunda reacción de PCR.

Tabla 3  
Banco de PCR2

Reactivo	Conc. de reserva	Conc. final	Reacción de 20 µl	Reacción de 50 µl
Tampón Q5 x5	5,00 mM	1,00 µM	4,00 µl	10,00 µl
Cada dNTP	10,00 mM	0,20 µM	0,40 µl	1,00 µl
Inicio de Q5 caliente	125,00 mM	1,00 µM	0,16 µl	0,40 µl
Cebador C7-index-P7	2 µM		1,60 µl	4,00 µl
Mezcla de cebadores [P5-IgH/TCRα/TCRγ]-[P5-IgL/TCRβ/TCRδ]	1 µM (cada uno)		1,60 µl	4,00 µl
ADNc			8,00 µl	20,00 µl
H <sub>2</sub> O			4,24 µl	10,60 µl
<b>Secuencias de cebadores para la mezcla P5-IgH/TCRα/TCRγ (pesada)</b>				
IgM	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGGTTGGGGCGGATGCAC			
IgD	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCATCCGGAGCCTTGGTGG			
IgA	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCCTTGGGGCTGGTCGGGG			
IgE	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCGGATCGGATGGGCTCTGTGG			
IgG	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCCGATCGGATGGGCCCTTGGTGG			
TCRα1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGGATTTAGAGTCTCTCAGCTG			
TCRα2	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCACGGCAGGGTCAGGGTTC			
TCRγ	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGGGAAACATGTCATCAAGT			
<b>Secuencias de cebadores para la mezcla P5-IgL/TCRβ/TCRδ (ligera)</b>				
IgKJ1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTTTGAATCCACCTTGGTCCCTCCGC			
IgKJ2	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTTTGAATCCAGCTTGGTCCCTCCGC			
IgKJ3	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTTTGAATCCACTTGGTCCAGGGC			
IgKJ4	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTTTGAATCCACCTTGGTCCCTTGGC			
IgKJ5	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTTTAAATCTCCAGTGGTCCCTTGGC			
IgLJ1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGAGGACGGTCACCTTGGTGCCA			
IgLJ2	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTAGGACGGTCAGCTTGGTCCCTCC			
<b>Secuencias de cebadores para la mezcla P5-IgL/TCRβ/TCRδ (ligera)</b>				
IgLJ3	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGAGGACGGTCAGCTGGGTGCC			
IgLJ4	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTAAATGATCAGCTGGGTCCCTCCAC			
IgLJ5	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTAGGACGGTGACCTTGGTCCAGT			
IgLJ6	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTAGGACGGTCAGCTCGGTCCCCC			
TCRβ1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGGAGATCTCTGCTCTGATG			
TCRβ2	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCGACCTCGGGTGGGAACAC			
TCRδ	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCGGATGGTTGGTATGAGGC			

Se usó un cebador "P7-index-C7" que comprende la concatenación de secuencias de C7 de Illumina, código de barras de 6 bases y P7:

5' CAAGCAGAAGACGGCATAACGAGAT[NNNNNN]GTGACTGGAGTTCAGACGTGTGCTCT  
TCCGATCT 3'.

5 Se ejecutó el siguiente programa en un termociclador:

1. 98 °C durante 01:00
2. 98 °C durante 00:10
- 10 3. 64 °C durante 00:20
4. 72 °C durante 00:20
5. Se pasa a 2 durante un total de 6 ciclos
6. 4 °C sin límite de tiempo.

15 El producto de la PCR se purificó con 1,2 volúmenes de AMPure y se eluyó en 40 µl de tampón de dilución.

**Ejemplo 8 - Protocolo para la tercera reacción de PCR (PCR3) de polinucleótidos para la secuenciación de próxima generación a fin de realizar una secuenciación masiva de polinucleótidos unicelulares de alto rendimiento a base de emulsiones.**

20 Se usaron 8 µl de producto de PCR2 purificado para qPCR para determinar un número final de ciclos de amplificación. En la siguiente Tabla 4, se muestra una configuración para la tercera reacción de PCR.

**Tabla 4**

<b>Banco de qPCR3a</b>			
<b>Reactivo</b>	<b>Conc. de reserva</b>	<b>Conc. final</b>	<b>Reacción de 20 µl</b>
Tampón Q5 x5	5,00 mM	1,00 µM	4,00 µl
Cada dNTP	10,00 mM	0,20 µM	0,40 µl
Verde SYBR I 1:500	83,00 mM	1,00 µM	0,24 µl
Inicio de Q5 caliente	125,00 mM	1,00 µM	0,16 µl
Cebador C5-P5	10,00 µM	0,40 µM	0,80 µl
Cebador C7	10,00 µM	0,40 µM	0,80 µl
ADNc			8,00 µl
H <sub>2</sub> O			5,60 µl
<b>Secuencias de cebadores</b>			
P5	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT		
C7	CAAGCAGAAGACGGCATAACGAGAT		

25 El siguiente programa se ejecutó en una máquina de qPCR:

1. 98 °C durante 01:00
- 30 2. 98 °C durante 00:10
3. 64 °C durante 00:20
- 35 4. 72 °C durante 00:20
5. Se lee la placa
6. Se pasa a 2 durante un total de 25 ciclos
- 40 7. 4 °C sin límite de tiempo.

Se examinó la representación de la intensidad de la qPCR para determinar el ciclo de amplificación en el que la intensidad de la fluorescencia fue máxima, pero en el que la amplificación exponencial del ADN aún no había finalizado. Este fue el número de ciclos final para el punto final de la PCR3.

45 Se usaron 24,0 µl de producto de PCR2 purificado para la PCR3 del punto final. En la siguiente Tabla 5, se muestra una configuración ilustrativa para la reacción de PCR para determinar el número de ciclos del punto final de la tercera PCR.

Tabla 5

Banco de qPCR3b			
Reactivo	Conc. de reserva	Conc. final	Reacción de 60 $\mu$ l
Tampón Q5 x5	5,00 mM	1,00 $\mu$ M	12,00 $\mu$ l
Cada dNTP	10,00 mM	0,20 $\mu$ M	1,20 $\mu$ l
H <sub>2</sub> O	83,00 mM	1,00 $\mu$ M	0,72 $\mu$ l
Inicio de Q5 caliente	125,00 mM	1,00 $\mu$ M	0,48 $\mu$ l
Cebador C5-P5	10,00 $\mu$ M	0,40 $\mu$ M	2,40 $\mu$ l
Cebador C7	10,00 $\mu$ M	0,40 $\mu$ M	2,40 $\mu$ l
ADNc			24,00 $\mu$ l
H <sub>2</sub> O			16,80 $\mu$ l

Se ejecutó el siguiente programa en un termociclador:

- 5 1. 98 °C durante 01:00
2. 98 °C durante 00:10
3. 64 °C durante 00:20
4. 72 °C durante 00:20
5. Se pasa a 2 para el número de ciclos determinado.
- 10 6. 4 °C sin límite de tiempo.

El producto de la PCR se purificó con 1,2 volúmenes de AMPure y se eluyó en 20  $\mu$ l de tampón de dilución. Los bancos estaban listos para la secuenciación. Se agruparon según lo deseado, con o sin purificación con gel de agarosa para eliminar los amplicones truncados contaminantes y luego se secuenciaron usando una plataforma de tecnología de secuenciación de próxima generación.

#### **Ejemplo 9 - Procesamiento de la lectura y asignación de isotipos**

Las lecturas de MiSeq Illumina se procesaron usando tuberías personalizadas construidas alrededor del paquete pRESTO 1 para generar secuencias de consenso de longitud completa para moléculas de ARNm y gotitas, se anotaron con IgBLAST e IMGT/HighV-QUEST y se procesaron con Scripts personalizados y el paquete Change-O para generar estadísticas y cifras. Las lecturas de MiSeq se desmultiplexaron usando el programa informático Illumina. Las posiciones con una calidad inferior a Phred 5 se enmascararon con N. Se identificaron cebadores específicos del isotipo, códigos de barras de las gotitas (DB) y códigos de barras moleculares (MB) en el amplicón, y se recortaron, usando el corte con los cebadores MaskPrimers de pRESTO con un error máximo de 0,2. Se generaron una secuencia de consenso de lectura 1 y una secuencia de consenso de lectura 2 por separado para cada ARNm a partir de lecturas agrupadas por un identificador molecular único (IMU) que comprende el DB y MB juntos, que son replicados de PCR que surgen de la misma molécula de ARNm original. Los grupos de lectura de IMU se alinearon con MUSCLE y se utilizó pRESTO para crear secuencias de consenso con los siguientes parámetros: maxdiv = 0,1; Cebador bf PRIMER; prfreq = 0,6; error máximo = 0,5; q = 5;  $\geq$  60 % de la coincidencia de secuencias de cebador de PCR identificada por cada grupo de lectura; diversidad máxima de nucleótidos = 0,1; usando la regla de la mayoría en las posiciones de indel; y enmascaramiento de columnas de alineación con baja calidad posterior (consenso). Las secuencias de consenso finales apareadas se cosieron en dos series. En primer lugar, se optimizó la alineación sin huecos de los extremos de la secuencia de consenso de cada par de lectura usando una aproximación de puntuación Z y se puntuó con un valor de  $p$  binomial como se implementó en la alineación de AssemblePairs de pRESTO con los siguientes parámetros: longitud mínima = 8;  $\alpha$   $1 \times 10^{-5}$ ; y error máximo = 0,3. Para los pares de lectura que no se lograron coser de esta manera, se intentó la costura usando los exones humanos de BCR y TCR de la línea germinal V para armar cada lectura antes de coser o unir la unión de las lecturas, usando los parámetros de referencia de AssemblePairs de pRESTO: identidad mínima = 0,5; valor de  $e$   $1 \times 10^{-5}$ .

#### **Ejemplo 10 - Anotación del segmento V(D)J y confirmación del isotipo**

Se usaron IgBLAST, Change-O y scripts personalizados para identificar los genes V(D)J de origen de la línea germinal, recortar secuencias de ARNm en una región de V(D)J, identificar las regiones CDR3 y calcular la mutación de las secuencias de nucleótidos de la línea germinal V. IgBLAST cuenta N como desapareamiento, pero las secuencias de ARNm con más de 6 N de la región V se filtraron para el análisis de mutaciones y el análisis de precisión de los apareamientos de las fracciones cruzadas. Para las cadenas pesadas de IG, la identidad del isotipo se confirmó

haciendo coincidir las regiones C sin cebador (exones de región constante) con las secuencias esperadas usando los parámetros de puntuación MaskPrimers de pRESTO: inicio = 0; error máximo = 0,2. Se descartaron los amplicones con identificaciones de región C con cebador/sin cebador discordantes, a excepción de dos combinaciones de cebador/no cebador en las que se resolvió un evento de interferencia de cebador específico mediante examen visual.

5

**Ejemplo 11 - Agrupación de secuencias de V(D)J en linajes clónicos.**

Las secuencias de V(D)J se agruparon en clones usando agrupamiento de enlace simple con una distancia ponderada entre los clones. El agrupamiento se realizó con el paquete DefineClones de Change-O mediante los parámetros de grupo: modelo = mln; gen = primero; dist = 4,0; norma = ninguna. En primer lugar, se agruparon todas las secuencias de consenso de las gotitas de las cadenas V<sub>H</sub> de Ig funcionales en compartimientos de unión V-J, de manera que las secuencias que posiblemente surgieron del mismo evento de recombinación inicial se combinaron juntas (en función del gen V<sub>H</sub> de Ig de mayor coincidencia, gen J<sub>H</sub> de Ig de mayor coincidencia, y se identificó la longitud de la unión mediante IMGT/HighV-QUEST. El umbral de distancia entre clones se escogió generando un histograma de las distancias del vecino más cercano dentro de cada compartimento de V<sub>H</sub> de Ig usando la función distToNearest del paquete shm de Change-O, y examinando visualmente el histograma para un corte de distancia natural (en el canal de un histograma bimodal). Las agrupaciones de clones de las cadenas ligeras se definieron usando el mismo modelo de distancia y umbral.

10

15

20

**Ejemplo 12 - Filtración de gotitas, cálculo de la fidelidad de emparejamiento**

La confianza del emparejamiento de cadena pesada y ligera se evaluó de dos maneras independientes: usando la coincidencia de secuencias de ARNm intragotita y la coincidencia de pares entre réplicas. La coincidencia del ARNm intragotita se definió como la diferencia media de nucleótidos por pares (pi de Nei <0,02) de las secuencias de V(D)J dentro de un locus. Las secuencias de ARNm se redujeron a secuencias codificantes de nucleótidos de V(D)J usando las anotaciones de IgBLAST. Dentro de cada gotita, se agruparon todas las secuencias de ARNm productivas por locus V. Dentro de cada grupo, se alinearon múltiples secuencias usando MUSCLE como se implementó en AlignSets DE pRESTO usando parámetros predeterminados. Las cadenas de consenso de las gotitas se construyeron a partir de múltiples ARNm por locus usando los parámetros de pRESTO: BuildConsensus.py; div máxima = 0,2; error máximo = 0,5. Se usaron gotitas redistribuidas aleatoriamente para seleccionar el corte de diversidad pi ≤ 0,02. En las gotitas redistribuidas, menos del 0,01 % de los locus de cadena pesada (<0,2 % de los locus de cadena ligera) cumplieron con este criterio. Las gotitas incluidas en los receptores inmunitarios o de múltiples células se separaron para un análisis de precisión adicional.

25

30

35

40

45

50

La precisión del emparejamiento se calculó basándose en la observación del mismo par de clones en múltiples repeticiones (experimentos en emulsiones separados), centrándose en los grupos de VDJ que probablemente solo contienen un solo linaje, es decir, que surgen de una sola reorganización de V(D)J y VJ seguida de la expansión. Los reordenamientos de VDJ similares pueden surgir dentro de múltiples tiempos individuales independientes, lo que conduce al mismo reordenamiento de V(D)J de cadena pesada emparejado de forma nativa con múltiples reordenamientos VJ de cadena ligera diferentes. Debido a que los reordenamientos raros de V(D)J proporcionarían una medida más exacta de la precisión técnica lograda mediante los procedimientos descritos en el presente documento, las CDR3 largas y pesadas (CDR3<sub>H</sub>) son un foco para este análisis (como un proxy para los reordenamientos más raros de V(D)J). También se eliminaron las secuencias con > 6 N para aumentar la confianza de asignación de clones. La precisión del emparejamiento aumentó con la longitud de CDR3<sub>H</sub> a más del 96 % para el cuartil más largo de clones observados en las fracciones (2.604 clones con longitud de unión ≥ 54 nt). Debido a que la probabilidad de coincidencias de pares de los clones es la probabilidad conjunta de pares verdaderos de dos experimentos independientes, la precisión del emparejamiento se estimó como la raíz cuadrada de la coincidencia de emparejamiento en las repeticiones, calculada de la siguiente manera, en la que  $d_{hl}^f$  es el número de códigos de barras de las gotitas d con el clon de cadena pesada h y el clon de cadena ligera l emparejados, y se encuentra en la fracción física f. La precisión media del emparejamiento (al cuadrado) para cada experimento se estima, como media, por encima de los clones de cadena pesada h y todos los pares de fracciones (f, g), la concordancia de clones de cadena ligera emparejados (l, k):

55

$$\begin{aligned}
 (\text{precisión}^2) &= \text{media} (P_f P_g) \\
 &= \frac{\text{pares de cadena pesada y cadena ligera uniformes en las fracciones}}{\text{pares totales en los que se ve el clon de cadena pesada en las fracciones}} \\
 \frac{\text{pares de cadena pesada y ligera uniformes}}{\text{pares uniformes} + \text{pares no uniformes}} &= \frac{\sum_h (\sum_{l=k}^{f \neq g} d_{hl}^f \cdot d_{hk}^g)}{\sum_h (\sum_{l=k}^{f \neq g} d_{hl}^f \cdot d_{hk}^g + \sum_{l=k}^{f=g} d_{hl}^f \cdot d_{hk}^g)} \\
 (\text{precisión}^2) &= \frac{33157}{35922}
 \end{aligned}$$

60

Por lo tanto, la precisión media de cada el experimento, (más/menos la varianza en la precisión entre los experimentos) fue del 96,1 % de acuerdo con este experimento ilustrativo.

5 **Ejemplo 13 - Análisis filogenético del VIH.**

Se descubrieron nuevos anticuerpos de neutralización amplia (bNAb) contra el VIH extrayendo las secuencias procesadas de anticuerpos emparejadas de alto rendimiento de la presente invención en cuanto a la similitud con los bNAb conocidos. Se extrajeron bNAb previamente conocidos de donante de PGT y de otros donantes del material publicado. Se puntuaron todos los ARNm de la IgH del VIH recuperados de las emulsiones en cuanto a la similitud con las secuencias de aminoácidos de CDR3 conocidas mediante tblastx 10. Usando secuencias de ARNm de IgH de un donante sano para generar una distribución de fondo de similitudes de secuencia, se usó un corte de puntuación de 27 bits para separar las CDR3 de tipo bNAb candidatos para su análisis posterior. Las secuencias de V(D)J de las secuencias candidatas se alinearon con los bNAb conocidos usando MUSCLE 11 con los parámetros predeterminados, y en particular, con linajes de donantes de PGT usando los parámetros predeterminados excepto: apertura de hueco = -15. Se generaron árboles con los parámetros predeterminados de PhyML, se manipularon y se visualizaron con Newick Utils y Dendroscope, y se examinaron manualmente para seleccionar las secuencias de la cadena pesada de inmunoglobulina intercaladas con las secuencias de bNAb conocidas. Las secuencias de consenso para cada gotita se construyeron como se ha descrito previamente con el examen manual de las alineaciones de cualquier conflicto de aminoácidos dentro de la gotita usando JALVIEW. Se seleccionaron ocho secuencias de cadena pesada y sus secuencias de anticuerpos de cadena ligera emparejadas de forma nativa para los ensayos de síntesis, clonación, expresión y neutralización.

25 **Ejemplo 14 - Análisis y representación de datos.**

Se generaron gráficos usando los paquetes dplyr y ggplot2 R. Los datos se muestrearon de forma descendente y/o se modificaron aleatoriamente con R para fines de visualización solo en las figuras del diagrama de dispersión. El mínimo de muestreo descendente fue de 20.000 gotitas por isotipo o como se indique. Los puntos se alteraron mediante la adición de ruido vertical y horizontal extraído de la misma distribución de probabilidad uniforme, con máximos  $\leq 0,2$  para las unidades de ARNm y  $\leq 0,6$  % para la mutación.

**REIVINDICACIONES**

1. Un procedimiento que comprende:

- 5       **(a)** formar una pluralidad de vasos comprendiendo cada uno
- (i)** una célula individual de una muestra que comprende una pluralidad de células,  
          **(ii)** una pluralidad de polinucleótidos con código de barras molecular y  
          **(iii)** un polinucleótido con código de barras del vaso;
- 10       **(b)** producir:
- (i)** un primer polinucleótido complementario que es complementario a un primer polinucleótido celular de la  
          célula individual y  
          **(ii)** un segundo polinucleótido complementario que es complementario a un segundo polinucleótido celular de  
          la célula individual;
- 15       **(c)** unir:
- (i)** un primer polinucleótido con código de barras molecular de la pluralidad de polinucleótidos con código de  
          barras molecular con el primer polinucleótido complementario y  
          **(ii)** un segundo polinucleótido con código de barras molecular de la pluralidad de polinucleótidos con código de  
          barras molecular con el segundo polinucleótido complementario,
- 20       formando así un primer polinucleótido con un solo código de barras de una sola célula y un segundo polinucleótido  
          con un solo código de barras de una sola célula; y
- 25       **(d)** unir el polinucleótido con código de barras del vaso, o un producto amplificado del mismo, a
- (i)** el primer polinucleótido con un solo código de barras de una sola célula o un producto amplificado del mismo  
          e  
          **(ii)** el segundo polinucleótido con un solo código de barras de una sola célula o un producto amplificado del  
          mismo, formando así una primera secuencia de código de barras doble de una sola célula y una segunda  
          secuencia de código de barras doble de una sola célula.
- 30

35    2. El procedimiento de la reivindicación 1, que comprende además:

- (a)** amplificar la primera secuencia con código de barras doble de una sola célula y la segunda secuencia con  
          código de barras doble de una sola célula; y/o
- 40       **(b)** secuenciar la primera secuencia con código de barras doble de una sola célula, la segunda secuencia con  
          código de barras doble de una sola célula, sus productos amplificados o cualquier combinación de los mismos; y/o
- (c)** determinar un origen celular del primer polinucleótido celular y del segundo polinucleótido celular para que sea  
          el mismo basándose en un código de barras del vaso del polinucleótido con código de barras del vaso; y/o
- 45       **(d)** determinar una secuencia de línea germinal del primer polinucleótido celular, el segundo polinucleótido celular,  
          o ambos; y/o
- (e)** determinar una varianza de una secuencia del primer o segundo polinucleótido celular a partir de la respectiva  
          secuencia de línea germinal, o ambas; y/o
- 50       **(f)** determinar al menos uno de:
- (i)** un número total de secuencias del primer polinucleótido celular únicas;
- (ii)** un número total de secuencias del segundo polinucleótido celular únicas;
- (iii)** un número total de secuencias del primer y del segundo polinucleótido celular emparejadas únicas;
- (iv)** una frecuencia de una secuencia del primer polinucleótido celular;
- (v)** una frecuencia de una secuencia del segundo polinucleótido celular; y
- 55       **(vi)** una frecuencia de una secuencia del primer polinucleótido celular y una secuencia del segundo  
          polinucleótido celular que están emparejadas; y/o
- (g)** determinar un número de moléculas de partida con una secuencia del primer polinucleótido celular, el segundo  
          polinucleótido celular, o ambos, basándose en un código de barras molecular de la primera secuencia con código  
          de barras doble de una sola célula, un código de barras molecular de la segunda secuencia con código de barras  
          doble de una sola célula, o ambos; y/o
- 60       **(h)** amplificar el polinucleótido con código de barras del vaso para formar una pluralidad de polinucleótidos con  
          código de barra del vaso idénticos antes de (d).

65    3. El procedimiento de cualquiera de las reivindicaciones 1-2, en el que

- (i)** el primer polinucleótido celular es una inmunoglobulina de cadena pesada (IgH), un polinucleótido receptor de

- linfocitos T alfa (TCR $\alpha$ ) o receptor de linfocitos T gamma (TCR $\gamma$ ); y en el que el segundo polinucleótido celular es una inmunoglobulina de cadena ligera (IgL), un polinucleótido receptor de linfocitos T beta (TCR $\beta$ ) o receptor de linfocitos T delta (TCR $\delta$ ) y/o
- 5 (ii) en el que la célula individual es un linfocito B o un linfocito T y/o  
 (iii) en el que la primera secuencia con código de barras doble de una sola célula y la segunda secuencia con código de barras doble de una sola célula comprenden cada una un código de barras molecular diferente y/o  
 (iv) en el que el polinucleótido con código de barras molecular de la pluralidad de polinucleótidos con código de barras molecular no es un producto amplificado y/o  
 10 (v) en el que un código de barras del vaso del polinucleótido con código de barras del vaso o amplicón del mismo de un primer vaso de una pluralidad de vasos es diferente de un código de barras del vaso del polinucleótido con código de barras del vaso o amplicón del mismo de un segundo vaso de la pluralidad de vasos.
4. El procedimiento de la reivindicación 3 (v), en el que la primera secuencia con código de barras doble de una sola célula y la segunda secuencia con código de barras doble de una sola célula comprenden cada una el mismo código de barras del vaso.  
 15
5. El procedimiento de cualquiera de las reivindicaciones 1-4, en el que el polinucleótido con código de barras del vaso de (a) está presente en cada vaso de una pluralidad de vasos como una sola molécula.
- 20 6. El procedimiento de la reivindicación 3 (v), 4 o 5, en el que el polinucleótido con código de barras del vaso de dos o más vasos diferentes de la pluralidad de vasos comprende una primera secuencia de vaso común cadena arriba del código de barras del vaso, una segunda secuencia de vaso común cadena abajo del código de barras del vaso, o ambas.
- 25 7. El procedimiento de las reivindicaciones 3 (iii), 4, 5 o 6, en el que cada uno de los polinucleótidos con código de barras molecular de la pluralidad de polinucleótidos con código de barras molecular de cada vaso de la pluralidad de vasos comprende una primera secuencia molecular común cadena arriba del código de barras molecular, una segunda secuencia molecular común cadena abajo del código de barras molecular, o ambas.
- 30 8. El procedimiento de cualquiera de las reivindicaciones 1-7, en el que el polinucleótido con código de barras del vaso o un complemento del mismo comprende:
- (a) una secuencia complementaria a una región de un complemento de cada polinucleótido con código de barras molecular de la pluralidad de polinucleótidos con código de barras molecular; y/o  
 35 (b) una secuencia complementaria a una región del extremo 3' del primer polinucleótido con código de barras molecular y del segundo polinucleótido con código de barras molecular.
9. El procedimiento de cualquiera de las reivindicaciones 1-8, en el que (b) comprende extender un primer cebador diana hibridado a una secuencia común del primer polinucleótido celular y extender un segundo cebador diana hibridado a una secuencia común del segundo polinucleótido celular, en el que el primer cebador diana y el segundo cebador diana se extienden mediante una transcriptasa inversa que comprende una actividad de transferasa terminal no molde, en el que se añaden 3 o más nucleótidos no molde idénticos a un extremo 3' del primer polinucleótido complementario y del segundo polinucleótido complementario; la unión de (c) comprende hibridar una región del extremo 3' del primer polinucleótido con código de barras molecular y del segundo polinucleótido con código de barras molecular de la pluralidad de polinucleótidos con código de barras molecular a 3 o más nucleótidos no molde presentes en un extremo 3' del primer polinucleótido complementario y del segundo polinucleótido complementario, respectivamente, y opcionalmente, en el que la unión de (c) comprende además extender el extremo 3' del primer y del segundo polinucleótido complementario; y la unión de (d) comprende hibridar una región del polinucleótido con código de barras del vaso o un complemento del mismo a (i) un extremo 3' del primer polinucleótido con código de barras molecular e (ii) un extremo 3' del segundo polinucleótido con código de barras molecular, y opcionalmente, la unión de (d) comprende además extender el extremo 3' del primer polinucleótido con código de barras molecular y extender el extremo 3' del segundo polinucleótido con código de barras molecular.  
 40  
 45  
 50
10. El procedimiento de cualquiera de las reivindicaciones 1-9, en el que  
 55 (i) una pluralidad de vasos comprende una pluralidad de emulsiones y/o  
 (ii) la célula individual se lisa antes de (b) y/o  
 (iii) un primer cebador diana, un segundo cebador diana, el polinucleótido con código de barras del vaso, un polinucleótido con código de barras molecular de la pluralidad de polinucleótidos con código de barras molecular, o cualquier combinación de los mismos, no está unido a un soporte sólido.  
 60
11. El procedimiento de cualquiera de las reivindicaciones 1-10, en el que (i) cada vaso de una pluralidad de vasos no comprende un soporte sólido o (ii) (a)-(d) se realiza en un mismo vaso individual de una pluralidad de vasos; y/o (a)-(d) se realizan en la pluralidad de vasos simultáneamente.  
 65
12. El procedimiento de cualquiera de las reivindicaciones 1-11, en el que el primer polinucleótido celular y el segundo

polinucleótido celular son ARN.

13. El procedimiento de cualquiera de las reivindicaciones 1-12, en el que el primer polinucleótido con un solo código de barra de una sola célula comprende el primer polinucleótido con código de barras molecular de la pluralidad o un producto amplificado del mismo y el primer polinucleótido complementario o un producto amplificado del mismo; el segundo polinucleótido con un solo código de barra de una sola célula comprende el segundo polinucleótido con código de barras molecular de la pluralidad o un producto amplificado del mismo y el segundo polinucleótido complementario o un producto amplificado del mismo; la primera secuencia con código de barras doble de una sola célula comprende el polinucleótido con código de barras del vaso o un producto amplificado del mismo y el primer polinucleótido con un solo código de barras de una sola célula o un producto amplificado del mismo; y la segunda secuencia con código de barras doble de una sola célula comprende el polinucleótido con código de barras del vaso o un producto amplificado del mismo y el segundo polinucleótido con un solo código de barras de una sola célula o un producto amplificado del mismo.

14. Una composición que comprende:

(a) una pluralidad de vasos comprendiendo cada uno

- (i) una pluralidad de polinucleótidos de una sola célula de una muestra que comprende una pluralidad de células,
- (ii) una pluralidad de polinucleótidos con código de barras molecular,
- (iii) un polinucleótido con código de barras del vaso;
- (iv) un primer polinucleótido complementario que es complementario a un primer polinucleótido celular de la pluralidad de polinucleótidos de la célula individual y
- (v) un segundo polinucleótido complementario que es complementario a un segundo polinucleótido celular de la pluralidad de polinucleótidos de la célula individual;

en el que el primer polinucleótido complementario comprende un primer código de barras molecular de la pluralidad de polinucleótidos con código de barras molecular y el código de barras del vaso del polinucleótido con código de barras del vaso o un producto amplificado del polinucleótido con código de barras del vaso, y el segundo polinucleótido complementario comprende un segundo código de barras molecular de la pluralidad de polinucleótidos con código de barras molecular y el código de barras del vaso del polinucleótido con código de barras del vaso o un producto amplificado del polinucleótido con código de barras del vaso.

15. La composición de la reivindicación 14, en la que

- (i) el código de barras molecular del primer polinucleótido con código de barras molecular y el segundo polinucleótido con código de barras molecular son diferentes; y/o el código de barras molecular de cada polinucleótido con código de barras molecular de un primer vaso de la pluralidad de vasos es único y/o
- (ii) el primer polinucleótido complementario y el segundo polinucleótido complementario comprenden cada uno un mismo código de barras del vaso y/o
- (iii) un código de barras del vaso de cada polinucleótido con código de barras del vaso o amplicón del mismo de un segundo vaso de la pluralidad de vasos es un segundo mismo código de barras del vaso y/o
- (iv) un primer mismo código de barras del vaso de un primer vaso de la pluralidad de vasos es diferente de un segundo mismo código de barras del vaso de un segundo vaso de la pluralidad de vasos y/o
- (v) el primer polinucleótido celular es ADN, ARN o ARNm; y/o el segundo polinucleótido celular es ADN, ARN o ARNm y/o
- (vi) el primer polinucleótido complementario es ADNc y/o
- (vii) el primer polinucleótido complementario y/o el segundo polinucleótido complementario comprenden cada uno tres o más nucleótidos no molde añadidos a un extremo 3'.

16. El procedimiento de la reivindicación 14 o 15, en el que un primer polinucleótido con código de barras molecular comprende una región complementaria a los tres o más nucleótidos no molde en el extremo 3' del primer polinucleótido complementario; y/o el segundo polinucleótido con código de barras molecular comprende una región complementaria a los tres o más nucleótidos no molde en el extremo 3' del segundo polinucleótido complementario.

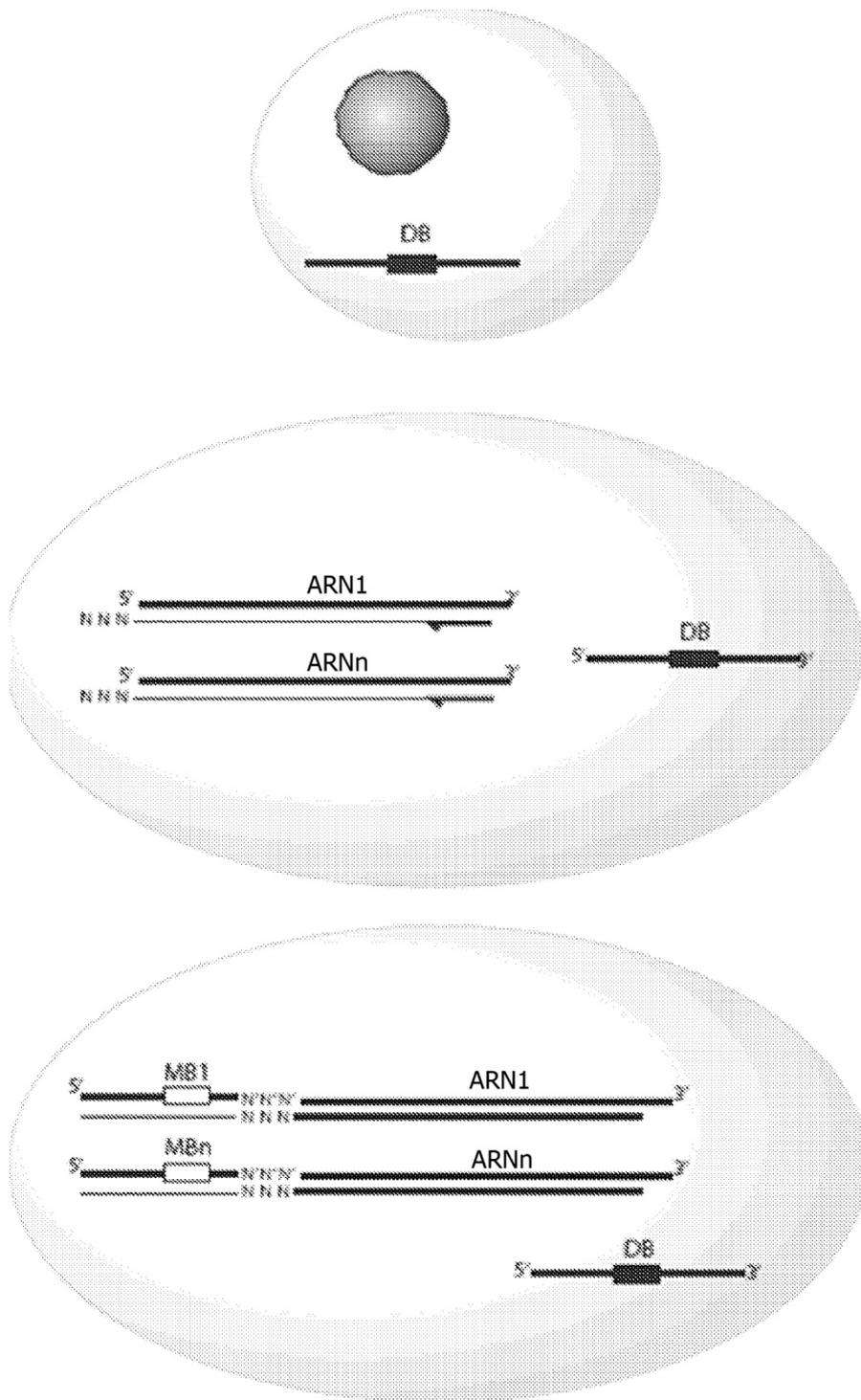


FIG. 1A

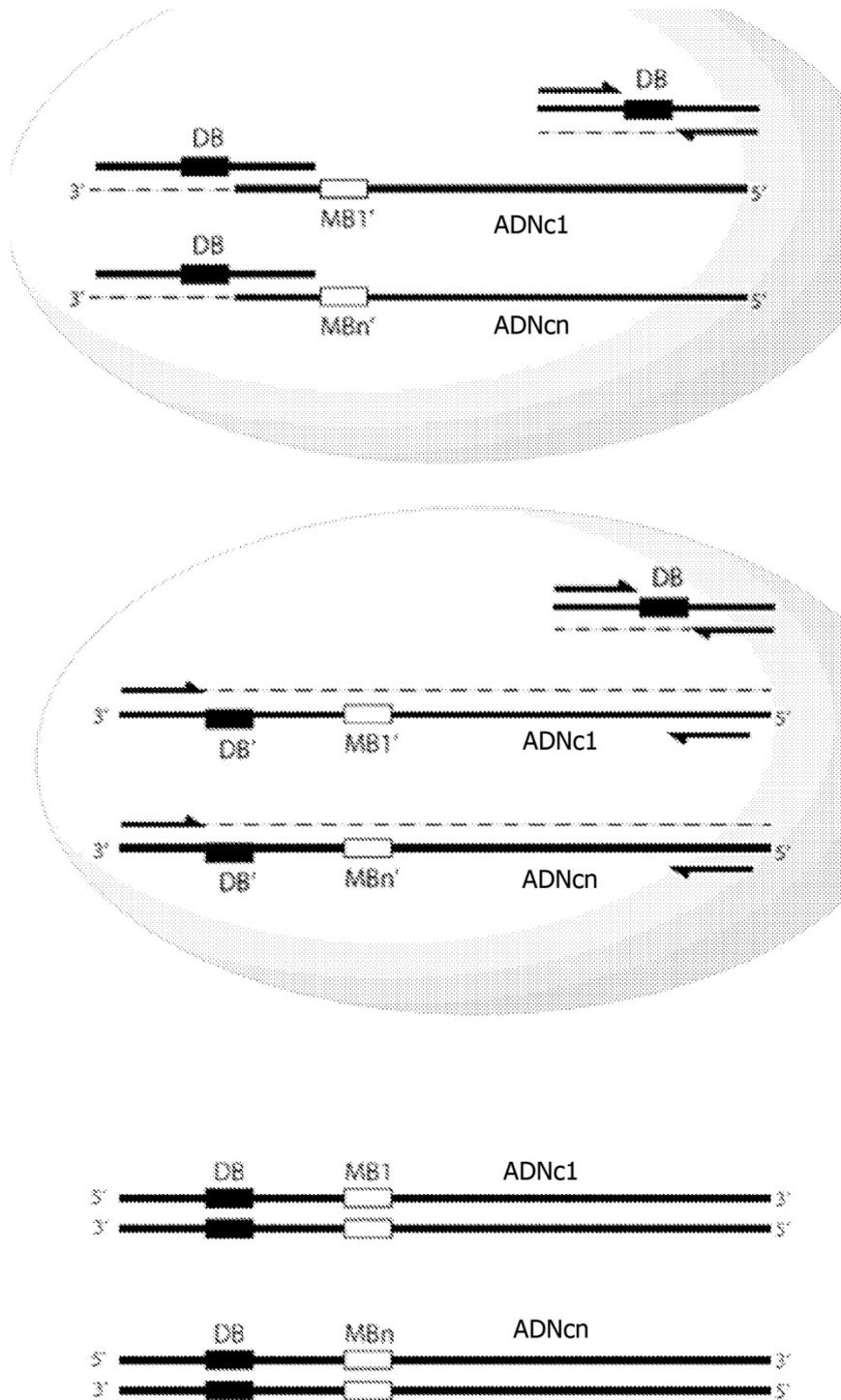


FIG. 1B

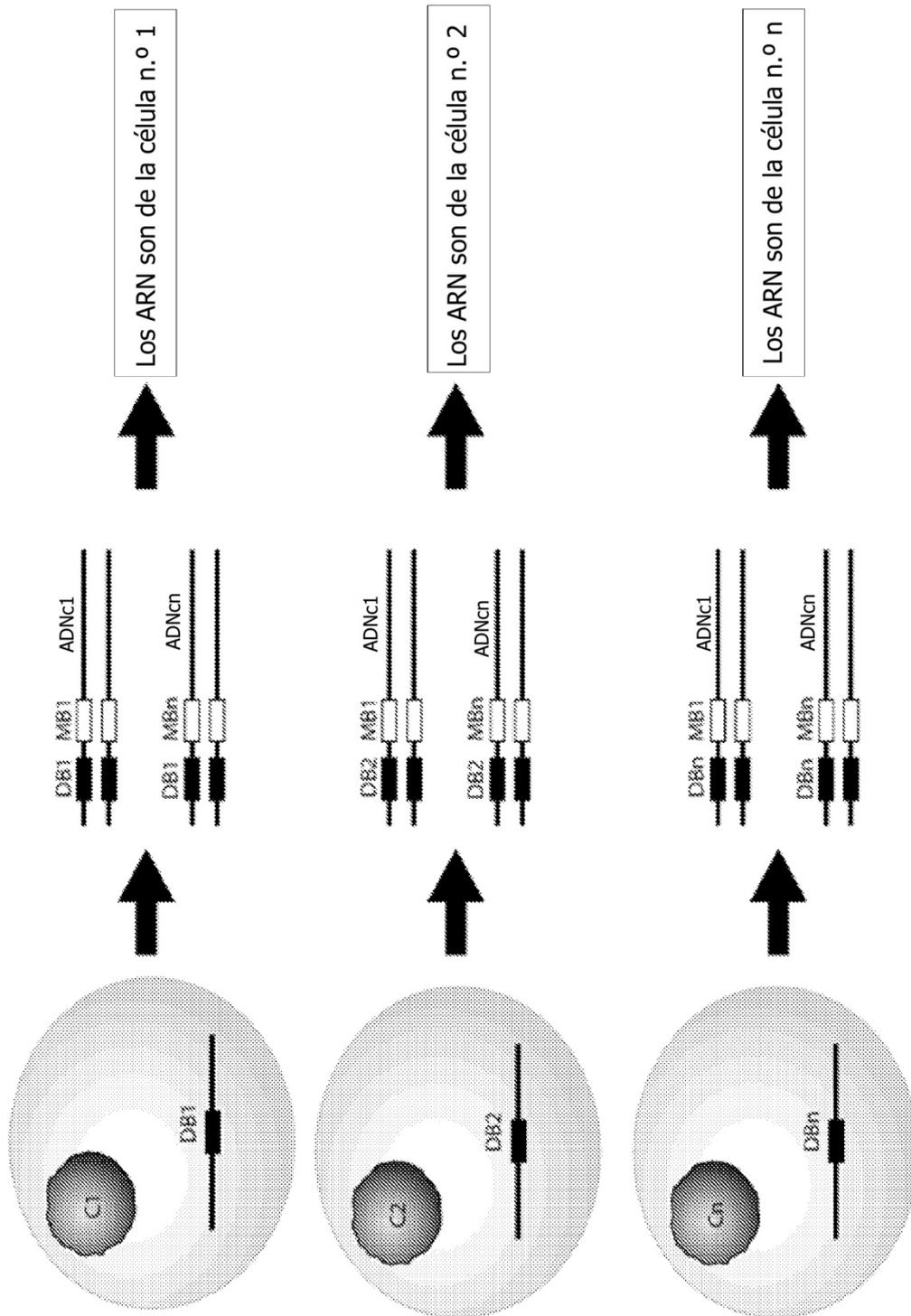


FIG. 2

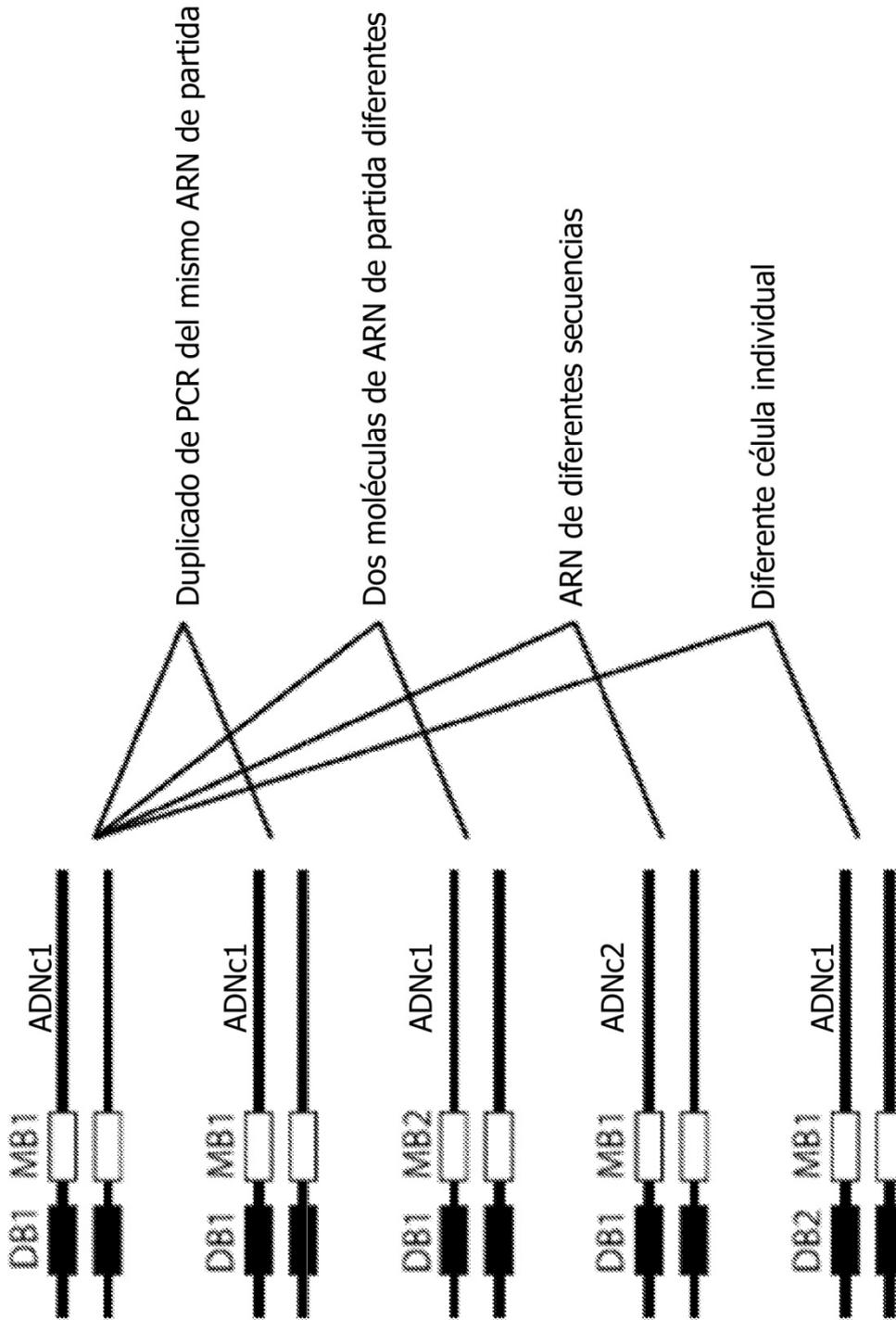


FIG. 3

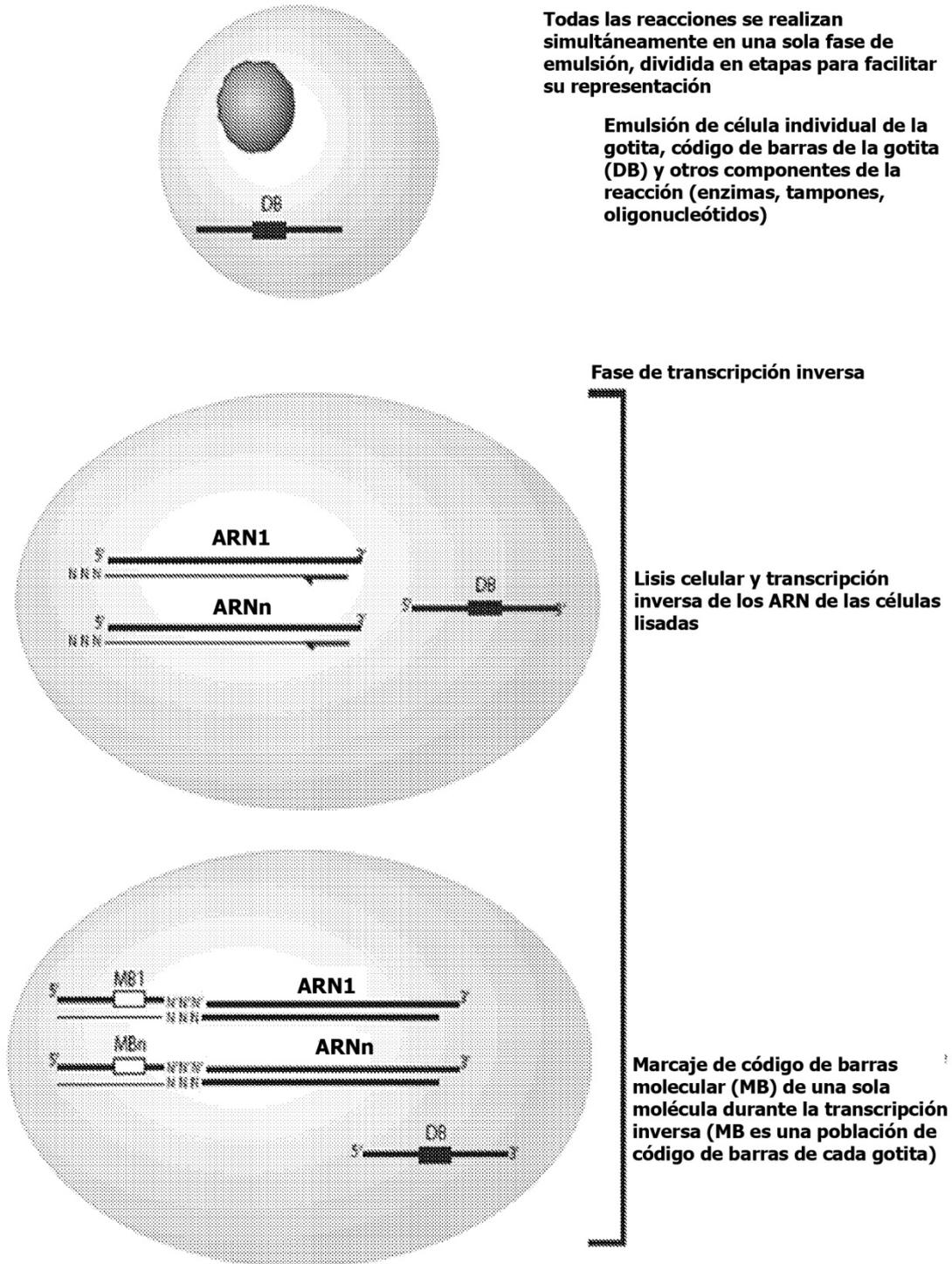


FIG. 4A

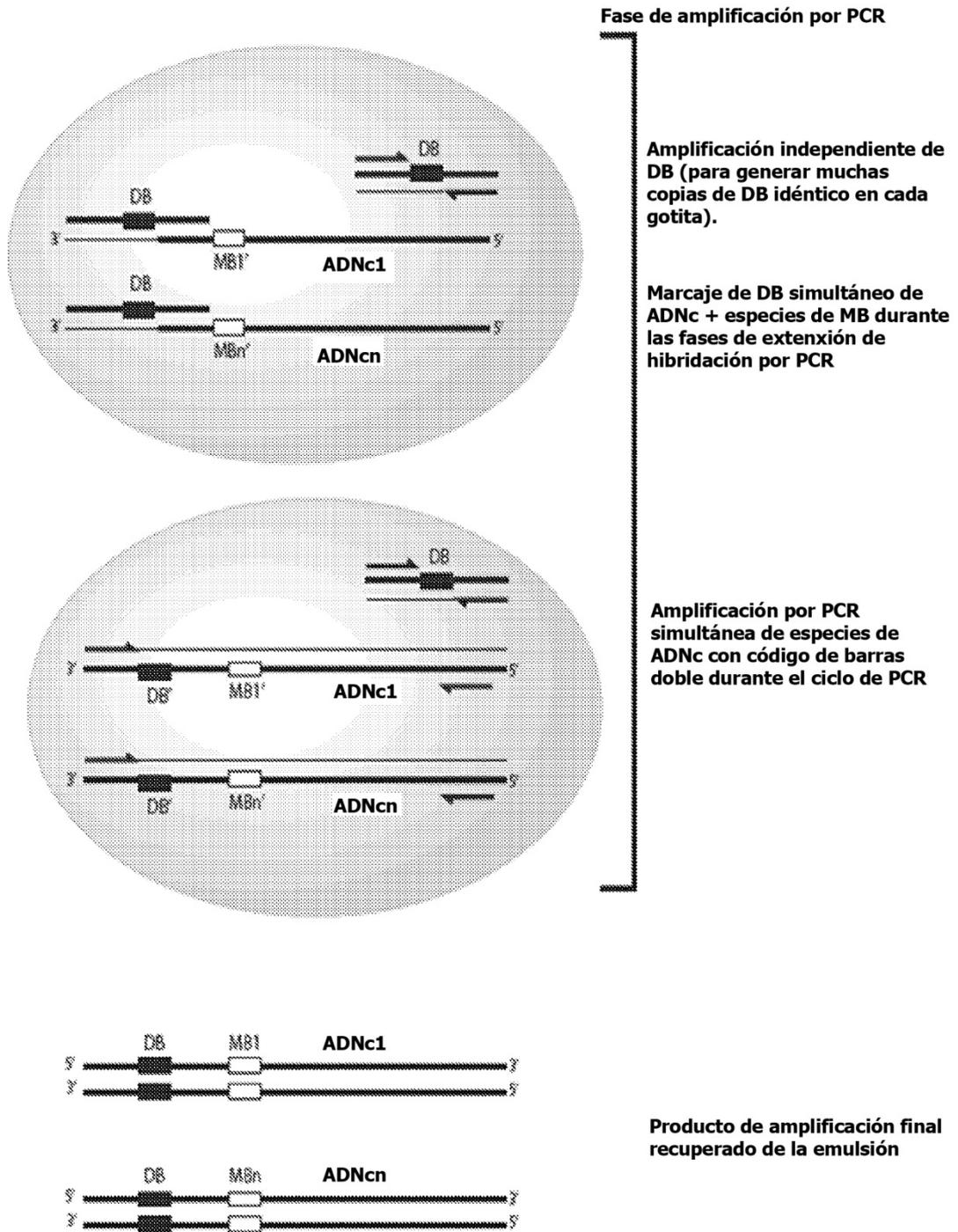


FIG. 4B

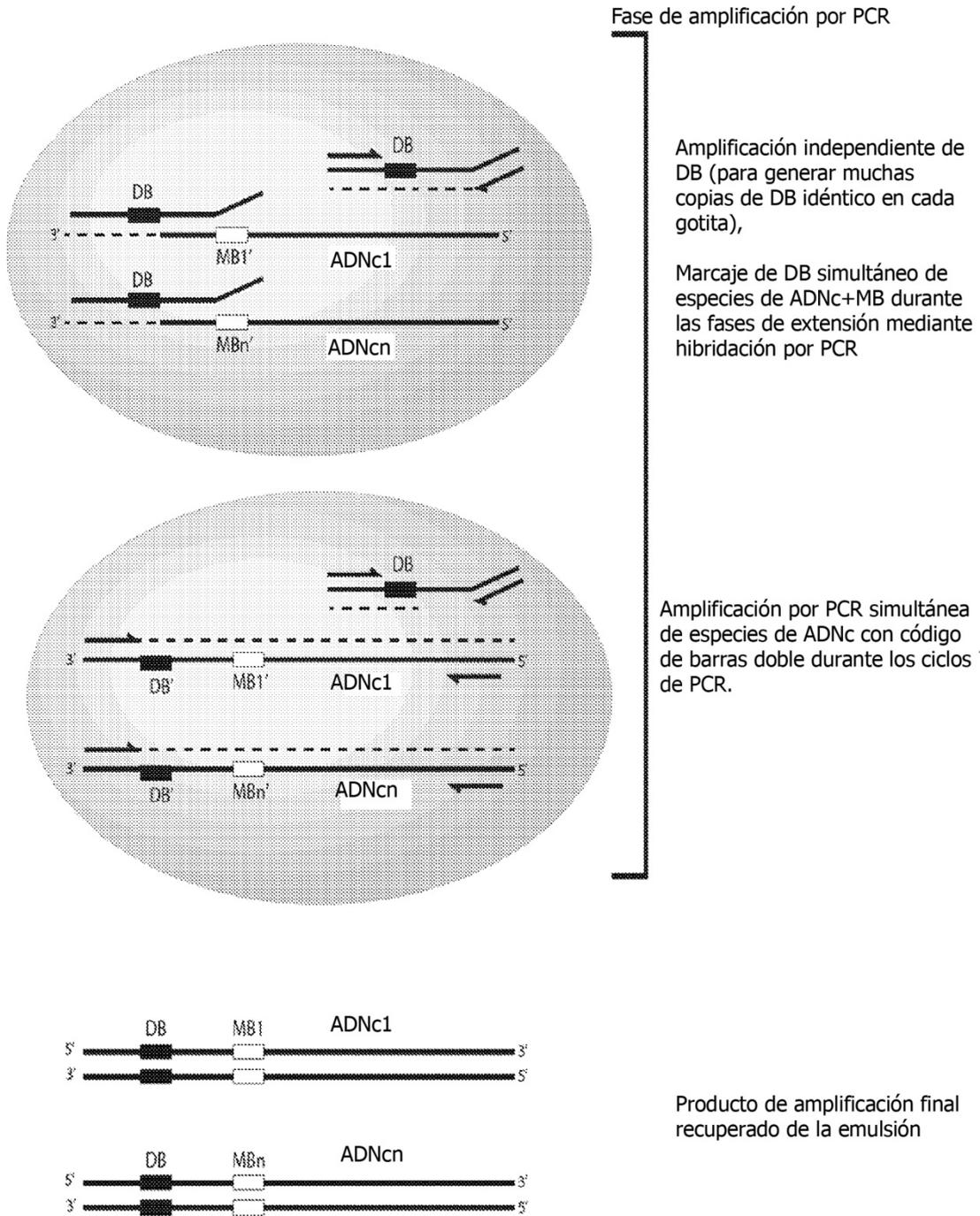
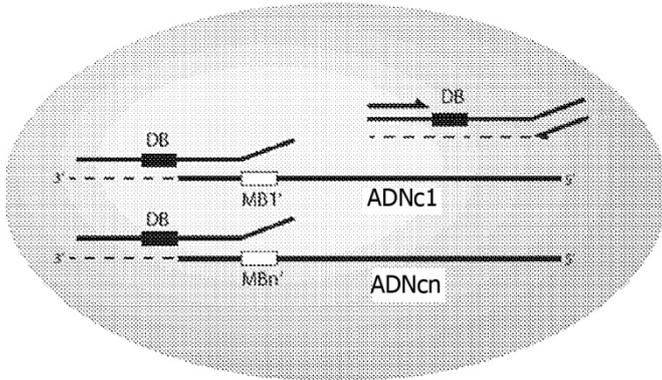


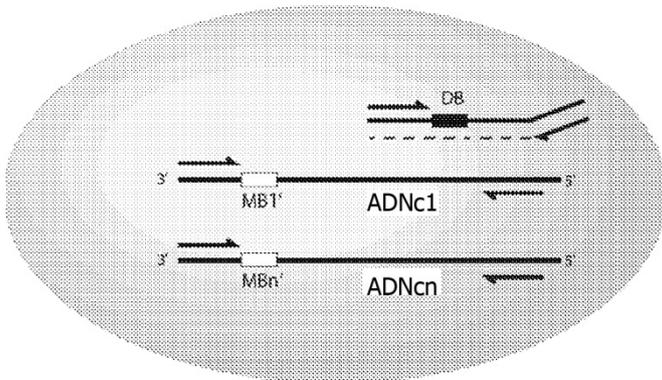
FIG. 4C

Fase de amplificación por PCR

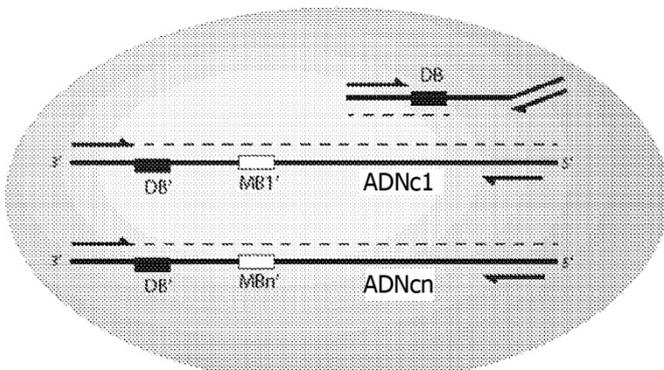


Amplificación independiente de DB (para generar muchas copias de DB idéntico en cada gotita).

Marcaje de DB simultáneo de especies de ADNc+MB durante las fases de extensión mediante hibridación por PCR



Amplificación opcional de ADNc sin marcar antes del marcaje de DB



Amplificación por PCR simultánea de especies de ADNc con código de barras doble durante los ciclos de PCR.



Producto de la amplificación final recuperado de la emulsión.

FIG. 4D

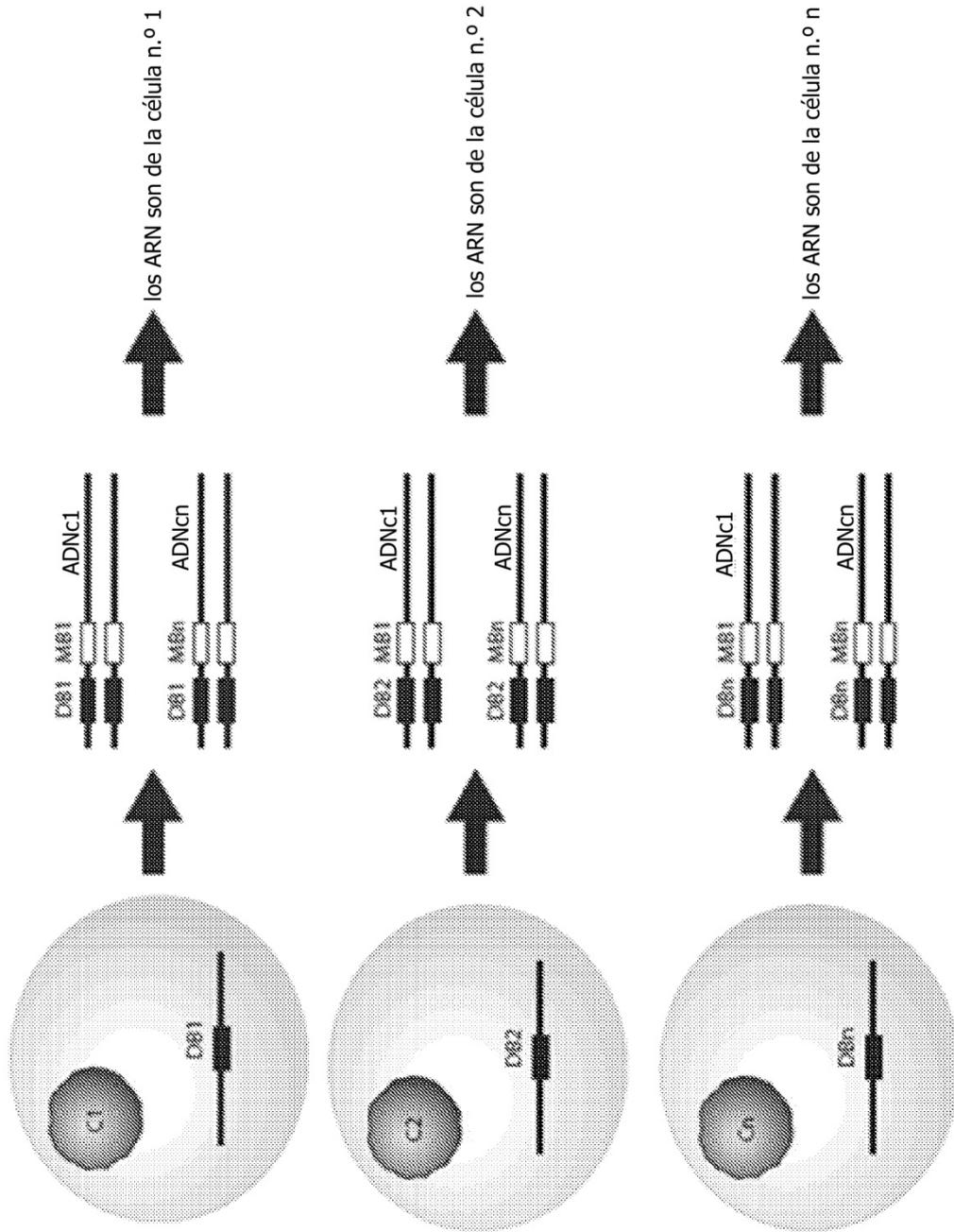


FIG. 5

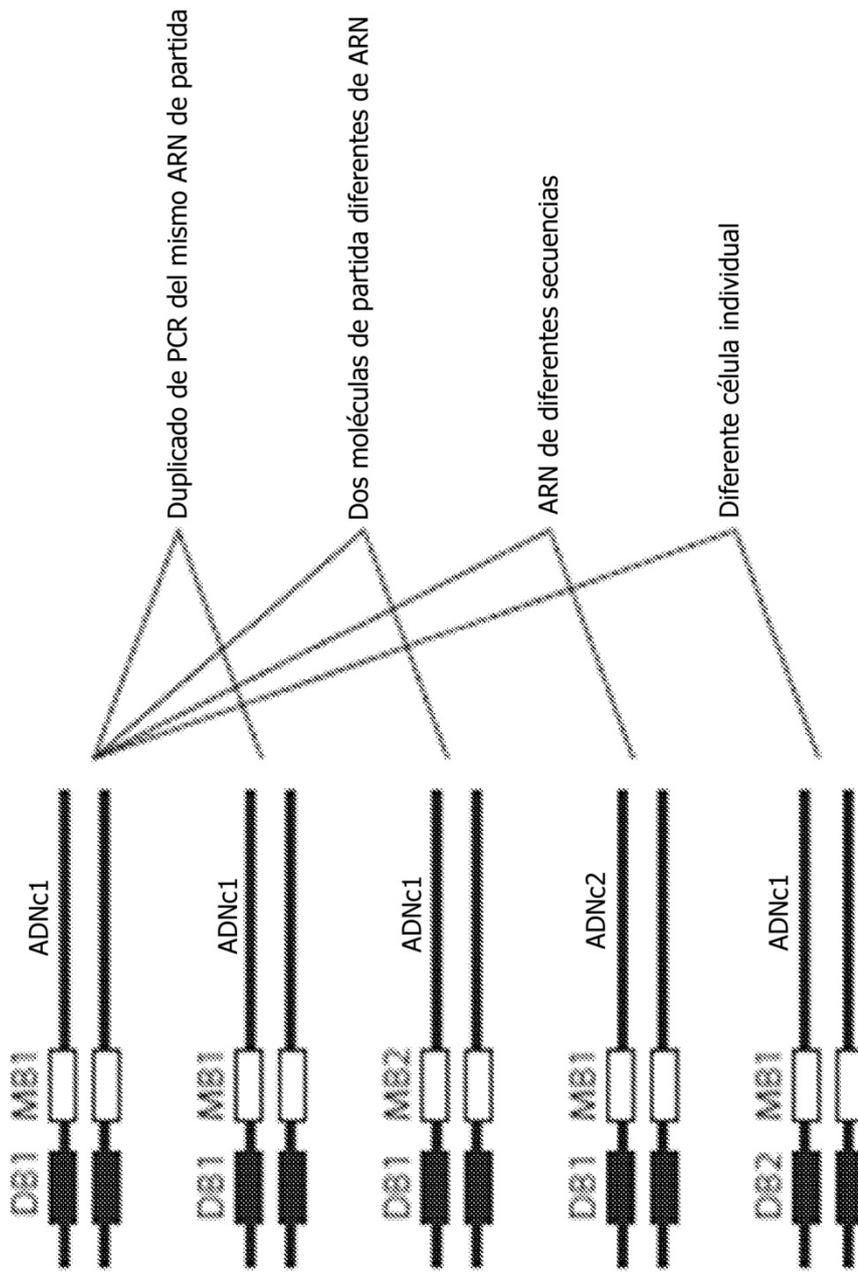


FIG. 6

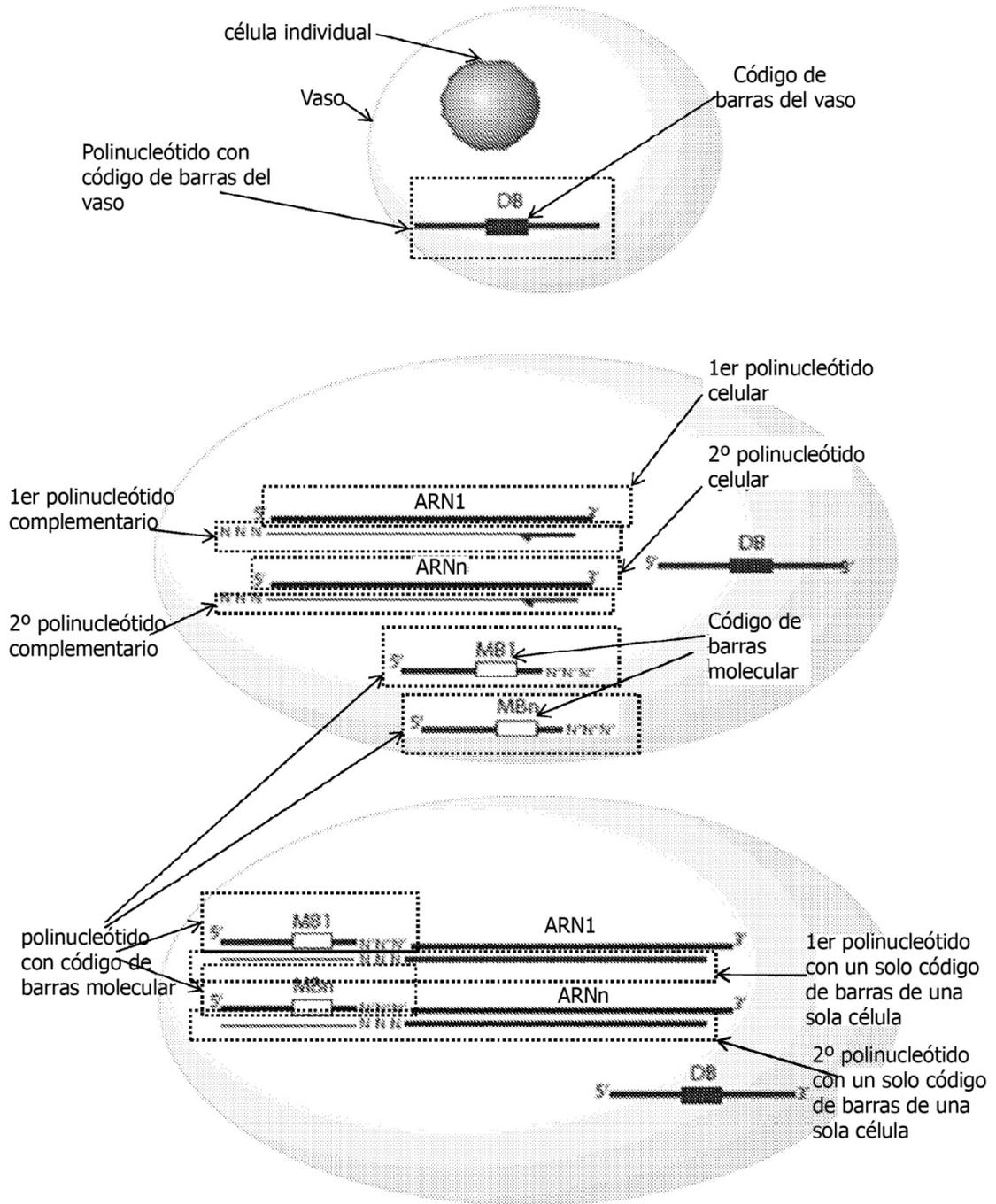


FIG. 7A

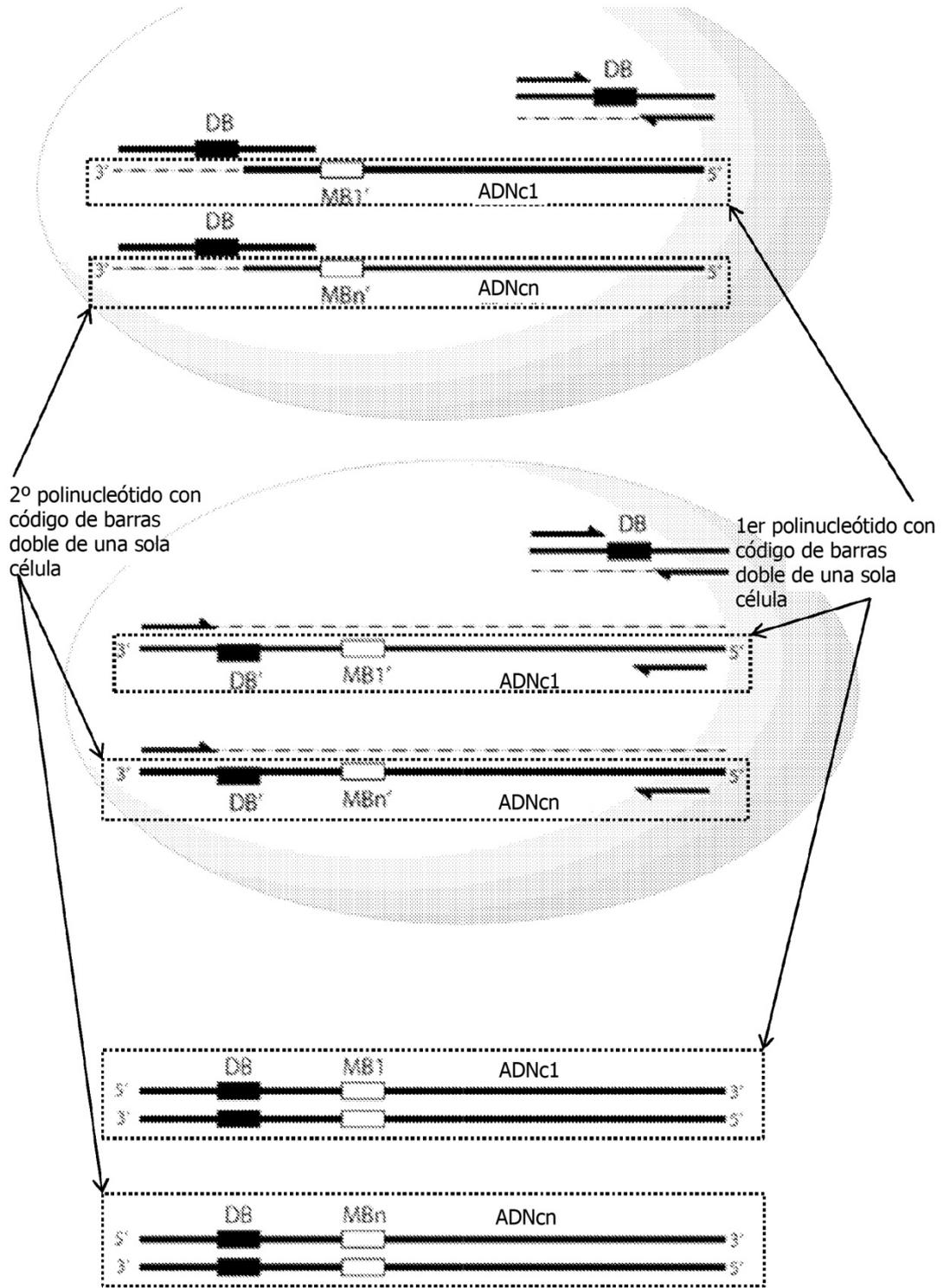


FIG. 7B