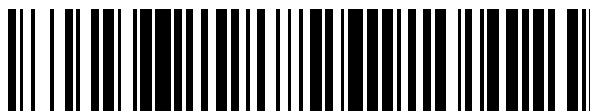


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 711 168**

51 Int. Cl.:

C12Q 1/68 (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **28.08.2014 PCT/US2014/053301**

87 Fecha y número de publicación internacional: **05.03.2015 WO15031691**

96 Fecha de presentación y número de la solicitud europea: **28.08.2014 E 14761937 (3)**

97 Fecha y número de publicación de la concesión europea: **14.11.2018 EP 3039158**

54 Título: **Análisis masivo en paralelo de células individuales**

30 Prioridad:

28.08.2013 US 201361871232 P
12.03.2014 US 201461952036 P
13.06.2014 US 201462012237 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
30.04.2019

73 Titular/es:

BECTON, DICKINSON AND COMPANY (100.0%)
1 Becton Drive
Franklin Lakes, NJ 07417, US

72 Inventor/es:

FAN, CHRISTINA;
FODOR, STEPHEN P.A.;
FU, GLENN K.;
FACER, GEOFFREY RICHARD y
WILHELMY, JULIE

74 Agente/Representante:

IZQUIERDO BLANCO, María Alicia

ES 2 711 168 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

Análisis masivo en paralelo de células individuales

Antecedentes

5 Las masas multicelulares, tales como tejidos y tumores, pueden comprender un medio celular heterogéneo. Estos entornos celulares complejos a menudo pueden presentar múltiples fenotipos, que pueden ser indicativos de múltiples genotipos. Simplificar la complejidad multicelular en variabilidad de células individuales es una faceta importante de la comprensión de la heterogeneidad multicelular. Esta comprensión puede ser importante en el desarrollo de regímenes terapéuticos para combatir enfermedades con múltiples genotipos de resistencia.

Resumen de la invención

10 La invención proporciona un método que comprende: (a) distribuir una muestra que comprende una pluralidad de células en una matriz de micropocillos; (b) poner en contacto una célula individual en un único micropocillo con un soporte sólido individual, en el que el soporte sólido es una microesfera que comprende una pluralidad de oligonucleótidos, comprendiendo cada oligonucleótido un marcador celular, un marcador molecular y una región de unión a diana; (c) hibridar o ligar un ácido nucleico diana de la célula individual con la región de unión a diana de un oligonucleótido de la pluralidad de oligonucleótidos para generar un ácido nucleico diana marcado; (d) opcionalmente amplificar el ácido nucleico diana marcado o complemento del mismo; y (e) determinar una cantidad del ácido nucleico diana o complemento del mismo detectando y/o secuenciando el ácido nucleico diana marcado, el amplicón o complemento del mismo, y contar el número de marcadores moleculares para el ácido nucleico diana, en el que el marcador celular comprende una secuencia de ácido nucleico que proporciona información para la célula con que está en contacto el oligonucleótido, en el que el marcador celular de cada uno de la pluralidad de oligonucleótidos en una microesfera es igual, y en el que los marcadores celulares de los oligonucleótidos en la primera microesfera son diferentes de los marcadores celulares de los oligonucleótidos en una segunda microesfera, en el que el marcador molecular comprende una secuencia de ácido nucleico que proporciona información para el ácido nucleico diana marcado, y en el que cada marcador molecular de cada uno de la pluralidad de oligonucleótidos en una microesfera es diferente, y en el que la región de unión a diana comprende una secuencia de ácido nucleico que puede unirse al ácido nucleico diana.

Resumen de la divulgación

Un aspecto proporcionado es un método, que comprende obtener una muestra que comprende una pluralidad de células; marcar al menos una parte de dos o más moléculas polinucleotídicas, complementos de las mismas o productos de reacción de las mismas, de una primera célula de la pluralidad y una segunda célula de la pluralidad con un primer marcador celular igual específico para la primera célula y un segundo marcador celular igual específico para la segunda célula; y un marcador molecular específico para cada una de las dos o más moléculas polinucleotídicas, complementos de las mismas o productos de reacción de las mismas, en el que cada marcador molecular de las dos o más moléculas polinucleotídicas, complementos de las mismas o productos de reacción de las mismas, de la primera célula son únicos unos con respecto a otros, y en el que cada marcador molecular de las dos o más moléculas polinucleotídicas, complementos de las mismas o producto de reacción de las mismas, de la segunda célula son únicos unos con respecto a otros. En algunas realizaciones, el método comprende además secuenciar la al menos una parte de dos o más moléculas polinucleotídicas, complementos de las mismas o productos de reacción de las mismas. En algunas realizaciones, el método comprende además analizar los datos de secuencia de la secuenciación para identificar un número de moléculas individuales de los polinucleótidos en una específica de las células. En algunas realizaciones, las células son células cancerosas. En algunas realizaciones, las células están infectadas con polinucleótidos víricos. En algunas realizaciones, las células son bacterias u hongos. En algunas realizaciones, la secuenciación comprende secuenciar con longitudes de lectura de al menos 100 bases. En algunas realizaciones, la secuenciación comprende secuenciar con longitudes de lectura de al menos 500 bases. En algunas realizaciones, las moléculas polinucleotídicas son ARNm o microARN y los complementos de las mismas y productos de reacción de las mismas son complementos y productos de reacción de los ARNm o microARN. En algunas realizaciones, los marcadores moleculares están sobre una microesfera. En algunas realizaciones, el marcador específico para una célula individual está en una microesfera. En algunas realizaciones, el marcador específico para una célula individual y los marcadores moleculares están sobre microesferas. En algunas realizaciones, el método se realiza al menos en parte en una emulsión en algunas realizaciones, el método se realiza al menos en parte en un pocillo o micropocillo de una matriz. En algunas realizaciones, se detecta la presencia de un polinucleótido que está asociado con una enfermedad o afección. En algunas realizaciones, la enfermedad o afección es un cáncer. En algunas realizaciones, se detecta al menos una parte de un microARN, complemento del mismo o producto de reacción del mismo. En algunas realizaciones, la enfermedad o afección es una infección vírica. En algunas realizaciones, la infección vírica es de un virus con envuelta. En algunas realizaciones, la infección vírica es de un virus sin envuelta. En algunas realizaciones, el virus contiene ADN vírico que es bicatenario. En algunas realizaciones, el virus contiene ADN vírico que es monocatenario. En algunas realizaciones, el virus se selecciona del grupo que consiste en un poxvirus, un herpesvirus, un virus varicela zoster, un citomegalovirus, un virus de Epstein-Barr, un hepadnavirus, un papovavirus, un polioma virus y cualquier combinación de los mismos. En algunas realizaciones, la primera célula es de una persona que no tiene una enfermedad o afección y la segunda célula es de una persona que tiene la enfermedad o afección. En algunas realizaciones, las personas son diferentes. En algunas realizaciones, las personas son la misma, pero las células se recogen en diferentes puntos temporales. En algunas realizaciones, la primera célula es de una persona que tiene la enfermedad o afección y la segunda célula es de la misma persona. En algunas realizaciones, las células de la

ES 2 711 168 T3

muestra comprenden células de un tejido u órgano. En algunas realizaciones las células de la muestra comprenden células de un timo, glóbulos blancos, glóbulos rojos, células hepáticas, esplenocitos, células pulmonares, células cardíacas, células cerebrales, células de la piel, células pancreáticas, células del estómago, células de la cavidad oral, células de la cavidad nasal, células del colon, células del intestino delgado, células renales, células de una glándula, células cerebrales, células neurales, células de la glía, células del ojo, células de órganos reproductores, células de vejiga, gametos, células humanas, células fetales, células amnióticas o cualquiera combinación de las mismas.

Un aspecto proporcionado es un soporte sólido que comprende una pluralidad de oligonucleótidos que comprenden, cada uno, un marcador celular y un marcador molecular, en el que cada marcador celular de la pluralidad de oligonucleótidos son iguales, y cada marcador molecular de la pluralidad de oligonucleótidos son diferentes; y en el que el soporte es una microsfera, el marcador celular es específico para el soporte sólido, el soporte sólido, cuando se coloca en el centro de un sistema de coordenadas cartesianas tridimensional tiene oligonucleótidos que se extienden en al menos siete de los ocho octantes, o cualquier combinación de los mismos. En algunas realizaciones, la pluralidad de oligonucleótidos comprende además al menos uno de un marcador de muestra; un marcador universal; y una región de unión a ácido nucleico diana. En algunas realizaciones, el soporte sólido comprende la región de unión a ácido nucleico diana, en el que la región de unión a ácido nucleico diana comprende una secuencia seleccionada del grupo que consiste en una secuencia específica de gen, una secuencia de oligodT, un multímero aleatorio y cualquier combinación de los mismos. En algunas realizaciones, el soporte sólido comprende además un ácido nucleico diana o complemento del mismo. En algunas realizaciones, el soporte sólido comprende una pluralidad de ácidos nucleicos diana o complementos de los mismos que comprenden de aproximadamente un 0,01 % a aproximadamente un 100 % de transcritos de un transcriptoma de un organismo o complementos de los mismos, o de aproximadamente de un 0,01 % a aproximadamente un 100 % de genes de un genoma de un organismo o complementos de los mismos. En algunas realizaciones, los marcadores celulares de la pluralidad de oligonucleótidos comprenden una primera secuencia aleatoria conectada a una segunda secuencia aleatoria mediante una primera secuencia de unión a marcador; y los marcadores moleculares de la pluralidad de oligonucleótidos comprenden secuencias aleatorias. En algunas realizaciones, el soporte sólido se selecciona del grupo que consiste en un soporte sólido de polidimetilsiloxano (PDMS), un soporte sólido de poliestireno, un soporte sólido de vidrio, un soporte sólido de polipropileno, un soporte sólido de agarosa, un soporte sólido de gelatina, un soporte sólido magnético, un soporte sólido de pluronic y cualquier combinación de los mismos. En algunas realizaciones, la pluralidad de oligonucleótidos comprende un conector que comprende un grupo funcional de conector, y el soporte sólido comprende un grupo funcional de soporte sólido; en el que el grupo funcional de soporte sólido y el grupo funcional de conector se conectan entre sí. En algunas realizaciones, el grupo funcional de conector y el grupo funcional de soporte sólido se seleccionan individualmente del grupo que consiste en C6, biotina, estreptavidina, una o más aminas primarias, uno o más aldehídos, una o más cetonas y cualquier combinación de los mismos. En algunas realizaciones, los marcadores moleculares de la pluralidad de oligonucleótidos comprenden al menos 15 nucleótidos.

Un aspecto proporcionado es un kit que comprende cualquiera de los soportes sólidos descritos en la presente memoria, e instrucciones para su uso. En algunas realizaciones, el kit comprende además un pocillo. En algunas realizaciones, el pocillo está comprendido en una matriz. En algunas realizaciones, el pocillo es un micropocillo. En algunas realizaciones, el kit comprende además un tampón. En algunas realizaciones, el kit está contenido en un envase. En algunas realizaciones, el envase es una caja. En algunas realizaciones, el envase o caja tiene un volumen de 53,63 l (2 pies cúbicos) o menos. En algunas realizaciones, el envase o caja tiene un volumen de 28,31 l (1 pie cúbico) o menos.

Un aspecto proporcionado es una emulsión que comprende cualquiera de los soportes sólidos descritos en la presente memoria.

Un aspecto proporcionado es una composición que comprende un pocillo y cualquiera de los soportes sólidos descritos en la presente memoria.

Un aspecto proporcionado es una composición que comprende una célula y cualquiera de los soportes sólidos descritos en la presente memoria.

En algunas realizaciones, la emulsión o composición comprende además una célula, en algunas realizaciones, la célula es una célula individual. En algunas realizaciones, el pocillo es un micropocillo. En algunas realizaciones, el micropocillo tiene un volumen que varía de aproximadamente 1000 μm^3 a aproximadamente 120 000 μm^3 .

Un aspecto proporcionado es un método, que comprende poner en contacto una muestra con cualquier soporte sólido divulgado en la presente memoria, hibridar un ácido nucleico diana de la muestra con un oligonucleótido de la pluralidad de oligonucleótidos. En algunas realizaciones, el método comprende además amplificar el ácido nucleico diana o complemento del mismo. En algunas realizaciones el método comprende además secuenciar el ácido nucleico diana o complemento del mismo, en el que la secuenciación comprende secuenciar el marcador molecular del oligonucleótido al que se une el ácido nucleico diana o complemento del mismo. En algunas realizaciones, el método comprende además determinar una cantidad del ácido nucleico diana o complemento del mismo, en el que la determinación comprende cuantificar niveles del ácido nucleico diana o complemento del mismo; contar un número de secuencias que comprenden el mismo marcador molecular; o una combinación de los mismos. En algunas realizaciones, el método no comprende alinear ningún marcador molecular igual o marcador celular igual. En algunas realizaciones, la amplificación comprende la transcripción inversa del ácido nucleico diana. En algunas realizaciones, la amplificación emplea un método seleccionado del grupo que consiste en: PCR, PCR con cebadores

ES 2 711 168 T3

internos, PCR cuantitativa, PCR en tiempo real, PCR digital y cualquier combinación de las mismas. En algunas realizaciones, la amplificación se realiza directamente sobre el soporte sólido; sobre un molde transcrito a partir del soporte sólido; o una combinación de los mismos. En algunas realizaciones, la muestra comprende una célula. En algunas realizaciones, la célula es una célula individual. En algunas realizaciones, el contacto se produce en un pocillo. En algunas realizaciones, el pocillo es un micropocillo y está contenido en una matriz de micropocillos.

Un aspecto proporcionado es un dispositivo, que comprende una pluralidad de micropocillos, en el que cada micropocillo de la pluralidad de micropocillos tiene un volumen que varía de aproximadamente $1000 \mu\text{m}^3$ a aproximadamente $120\,000 \mu\text{m}^3$. En algunas realizaciones, cada micropocillo de la pluralidad de micropocillos tiene un volumen de aproximadamente $20\,000 \mu\text{m}^3$. En algunas realizaciones, la pluralidad de micropocillos comprende aproximadamente 96 a aproximadamente 200 000 micropocillos. En algunas realizaciones, los micropocillos están comprendidos en una capa de un material. En algunas realizaciones, al menos aproximadamente un 10 % de los micropocillos comprenden además una célula. En algunas realizaciones, el dispositivo comprende además cualquiera de los soportes sólidos descritos en la presente memoria.

Un aspecto proporcionado es un aparato que comprende cualquiera de los dispositivos descritos en la presente memoria, y un manipulador de líquidos. En algunas realizaciones, el manipulador de líquidos suministra líquido a la pluralidad de micropocillos en aproximadamente un segundo. En algunas realizaciones, el manipulador de líquidos suministra líquido a la pluralidad de micropocillos desde un único acceso de entrada. En algunas realizaciones, el aparato comprende además un imán. En algunas realizaciones, el aparato comprende además al menos de: un acceso de entrada, un acceso de salida, una bomba, una válvula, una ventilación, un depósito, una cámara de recogida de muestra, un aparato de control de la temperatura o cualquier combinación de los mismos. En algunas realizaciones, el aparato comprende la cámara de recogida de muestras, en el que la cámara de recogida de muestras se puede retirar del aparato. En algunas realizaciones, el aparato comprende además un generador de imágenes ópticas. En algunas realizaciones, el generador de imágenes ópticas produce una señal de salida que se usa para controlar el manipulador de líquidos. En algunas realizaciones, el aparato comprende además un mecanismo de termociclado configurado para realizar una amplificación por reacción en cadena de la polimerasa (PCR) de oligonucleótidos.

Un aspecto proporcionado es un método de producción de un resultado de ensayo de diagnóstico clínico, que comprende producir el resultado de ensayo de diagnóstico clínico con cualquier dispositivo o aparato descrito en la presente memoria; cualquier soporte sólido descrito en la presente memoria; cualquier método descrito en la presente memoria; o cualquier combinación de los mismos. En algunas realizaciones, el resultado de ensayo de diagnóstico clínico se transmite mediante un medio de comunicación.

Un aspecto proporcionado es un método de generación de cualquiera de los soportes sólidos descritos en la presente memoria, que comprende adherir a un soporte sólido: un primer polinucleótido que comprende una primera parte del marcador celular y un primer conector; y poner en contacto un segundo polinucleótido que comprende una segunda parte del marcador celular, una secuencia complementaria al primer conector y el marcador molecular. En algunas realizaciones, el tercer polinucleótido comprende además una región de unión a ácido nucleico diana.

En algunas realizaciones, una emulsión, micropocillo o pocillo contiene únicamente una célula. En algunas realizaciones, de 1 a 2 000 000 de emulsiones, micropocillos o pocillos contienen cada uno únicamente una célula. En algunas realizaciones, el método comprende distribuir como mucho una célula en cada emulsión, micropocillo o pocillo. En algunas realizaciones, se distribuye un soporte sólido individual y una célula individual en una emulsión, micropocillo o pocillo. En algunas realizaciones, de 1 a 2 000 000 de emulsiones, micropocillos o pocillos tienen cada uno distribuidos en los mismos una célula y un soporte sólido. En algunas realizaciones, el método comprende distribuir como mucho un soporte sólido por emulsión, micropocillo o pocillo. En algunas realizaciones, el método comprende distribuir un soporte sólido y una célula a cada uno de 1 a 2 000 000 de micropocillos, emulsiones o pocillos. En algunas realizaciones, la distribución celular es aleatoria o no aleatoria. En algunas realizaciones, la distribución celular es estocástica. En algunas realizaciones, una célula se distribuye por un clasificador celular. En algunas realizaciones, una célula se distribuye poniendo en contacto uno o más pocillos, micropocillos o emulsiones con una disolución diluida de células diluidas de modo que como mucho una célula se distribuya al uno o más pocillos, micropocillos o emulsiones.

En algunas realizaciones, las regiones específicas de diana, regiones específicas de diana de la pluralidad de oligonucleótidos o la región específica de diana de las dos o más moléculas polinucleotídicas, comprenden secuencias complementarias a dos o más dianas de un panel de dianas. En algunas realizaciones, las dos o más dianas del panel de dianas son biomarcadores. En algunas realizaciones, los biomarcadores son biomarcadores para una enfermedad o afección. En algunas realizaciones, la enfermedad o afección es un cáncer, una infección, una infección vírica, una enfermedad inflamatoria, una enfermedad neurodegenerativa, una enfermedad fúngica, una infección bacteriana o cualquier combinación de las mismas. En algunas realizaciones, el panel comprende de: 2-50 000, 2-40 000, 2-30 000, 2-20 000, 2-10 000, 2-9000, 2-8000, 2-7000, 2-6000, 2-5000, 2-1000, 2-800, 2-700, 2-600, 2-500, 2-400, 2-300, 2-200, 2-100, 2-75, 2-50, 2-40, 2-30, 2-20, 2-10 o 2-5 biomarcadores.

Breve descripción de los dibujos

Las características novedosas de la invención se exponen con particularidad en las reivindicaciones anejas. Se obtendrá una mejor comprensión de las características y ventajas de la presente invención por referencia a la siguiente descripción detallada que expone realizaciones ilustrativas, en que se utilizan los principios de la invención, y los dibujos adjuntos de los que:

ES 2 711 168 T3

La **FIG. 1** representa un soporte sólido ejemplar conjugado con un oligonucleótido ejemplar.

La **FIG. 2A-C** representa un flujo de trabajo ejemplar para sintetizar microesferas acopladas a oligonucleótido usando síntesis de división-combinación.

La **FIG. 3** representa una microesfera acoplada a oligonucleótido ejemplar.

5 La **FIG. 4** ilustra una realización ejemplar de una matriz de micropocillos.

La **FIG. 5** representa una distribución ejemplar de soportes sólidos en una matriz de micropocillos.

10 La **FIG. 6A-C** muestra la distribución ejemplar de células en matrices de micropocillos. La FIG. 6A muestra la distribución de células K562 (tamaño grande de célula). La FIG. 6B muestra la distribución de células Ramos (tamaño pequeño de célula). La FIG. 6C muestra la distribución de células Ramos y microesferas acopladas a oligonucleótido en matrices de micropocillos, con flechas rellenas que apuntan a las células Ramos y flechas discontinuas que apuntan a las microesferas acopladas a oligonucleótido.

La **FIG. 7** muestra la estadística ejemplar del volumen de micropocillo, el volumen de soporte sólido y la cantidad de material biológico obtenido de la lisis.

15 La **FIG. 8A-C** ilustra una realización ejemplar del sellado con tapa de microesfera. La FIG. 8A-B muestra imágenes de un pocillo de micromatriz con células y microesferas de oligonucleótido distribuidas en pocillos de un pocillo de micromatriz y con microesferas de sephadex más grandes usadas para sellar los pocillos. Las flechas de puntos apuntan a las células, las flechas discontinuas apuntan a las microesferas acopladas a oligonucleótido y las flechas rellenas apuntan a las microesferas sephadex. La FIG. 8C representa un esquema de la célula y microesfera de oligonucleótido (por ejemplo, oligomicroesfera) depositadas dentro de un pocillo con una microesfera de sephadex usada para sellar el pocillo.

20 La **FIG. 9** representa un gráfico de barras que compara la eficacia de amplificación de GAPHD y RPL 19 amplificados a partir de micropocillos y tubos. Las barras grises representan los datos del micropocillo. Las barras blancas representan los datos del tubo.

25 La **FIG. 10** representa un gel de agarosa que compara la especificidad de amplificación de tres genes diferentes directamente en un soporte sólido.

La **FIG. 11A-I** muestra representaciones gráficas de los resultados de secuenciación.

La **FIG. 12A-C** muestra un histograma de los resultados de secuenciación para la muestra de únicamente K562, la muestra de únicamente Ramos y la muestra de mezcla de K562 + Ramos, respectivamente.

30 La **FIG. 12D-E** muestran un gráfico del número de copias para genes enumerados en la tabla 3 para la muestra de únicamente células Ramos y la muestra de únicamente células K562, respectivamente.

La **FIG. 12F-I** muestra el número de copias para genes individuales.

La **FIG. 12J-M** muestra gráficos del número de moléculas únicas por gen (eje de ordenadas) para las microesferas con 100 combinaciones de códigos de barras únicas.

35 La **FIG. 12N-O** muestra gráficos ampliados de dos microesferas que representan el patrón general de perfiles de expresión génica para los dos tipos celulares

La **FIG. 12P** muestra un diagrama de dispersión de los resultados basado en el análisis de componentes principales del perfil de expresión génica de microesferas 768 con > 30 moléculas por microesfera de la muestra de mezcla de K562 + Ramos.

40 La **FIG. 12Q-R** muestra histogramas del número de copias por amplicón por microesfera para las células de tipo K562 (las microesferas a la izquierda del primer componente principal basado en la FIG. 12P) y las células de tipo Ramos (las microesferas a la derecha del primer componente principal basado en la FIG. 12P), respectivamente.

45 La **FIG. 12S-T** muestra el número de copias por microesfera o célula individual de los genes individuales para las células de tipo K562 (las microesferas a la izquierda del primer componente principal basado en la FIG. 12P) y células de tipo Ramos (las microesferas a la derecha del primer componente principal basado en la FIG. 12P), respectivamente.

La **FIG. 13A** representa los patrones generales de expresión génica para las células de ratón y Ramos.

La **FIG. 13B-C** muestra diagramas de dispersión de los resultados basados en el análisis de componentes principales del perfil de expresión génica de la muestra de alta densidad y la muestra de baja densidad, respectivamente.

50 La **FIG. 13D-E** representa gráficos de la lectura por combinación de código de barras (bc) (eje de ordenadas) frente a la combinación de código de barras única, clasificadas por el número total de moléculas por combinación de bc (eje de abscisas) para células de tipo Ramos y células de tipo murino de la muestra de alta densidad, respectivamente.

ES 2 711 168 T3

La **FIG. 13F-G** representa gráficos del número de moléculas por combinación de código de barras (bc) (eje de ordenadas) frente a la combinación de código de barras única, clasificadas por el número total de moléculas por combinación de bc (eje de abscisas) para células de tipo Ramos y células de tipo murino de la muestra de alta densidad, respectivamente.

5 La **FIG. 13H-I** representa gráficos de la lectura por combinación de código de barras (bc) (eje de ordenadas) frente a la combinación de código de barras única, clasificadas por el número total de moléculas por combinación de código de barras (eje de abscisas) para células de tipo Ramos y células de tipo murino de la muestra de baja densidad, respectivamente.

10 La **FIG. 13J-K** representa gráficos del número de moléculas por combinación de código de barras (eje de ordenadas) frente a la combinación de código de barras única, clasificadas por el número total de moléculas por combinación de código de barras (eje de abscisas) para células de tipo Ramos y células de tipo murino de la muestra de baja densidad, respectivamente.

La **FIG. 14** muestra un gráfico que representa los genes en el eje de abscisas y el log₁₀ de los números de lecturas.

15 La **FIG. 15A** muestra un gráfico de la distribución de genes detectados por marcador celular de tres partes (por ejemplo, código de barras celular).

La **FIG. 15B** muestra un gráfico de la distribución de moléculas únicas detectadas por microesfera (que expresa el panel de genes).

La **FIG. 16** representa los grupos celulares basados en los genes asociados con un código de barras celular.

20 La **FIG. 17A-D** muestra el análisis de marcadores específicos de monocitos. La **FIG. 17E** muestra el grupo celular representado en la FIG. 16.

La **FIG. 18A-B** muestra el análisis de marcadores específicos de linfocitos T. La **FIG. 18C** muestra el grupo celular representado en la FIG. 16.

La **FIG. 19A-B** muestra el análisis de los marcadores específicos de linfocitos T CD8+. La **FIG. 19C** muestra el grupo celular representado en la FIG. 16.

25 La **FIG. 20A** muestra el análisis de marcadores específicos de linfocitos CD4+. La **FIG. 20B** muestra el grupo celular representado en la FIG. 16.

La **FIG. 21A-D** muestra el análisis de marcadores específicos de linfocitos citolíticos naturales (NK). La **FIG. 20E** muestra el grupo celular representado en la FIG. 16.

30 La **FIG. 22A-E** muestra el análisis de marcadores específicos de linfocitos B. La **FIG. 22F** muestra el grupo celular representado en la FIG. 16.

La **FIG. 23A-F** muestra el análisis de receptores de tipo Toll. Los receptores de tipo Toll se expresan principalmente por monocitos y algunos linfocitos B. La FIG. 23G muestra el grupo celular representado en la FIG. 16.

La **FIG. 24** representa un gráfico de los genes frente al log₁₀ del número de lecturas.

35 La **FIG. 25A-D** muestra gráficos del código de barras molecular frente al número de lecturas o log₁₀ del número de lecturas para dos genes.

40 La **FIG. 26A** muestra un gráfico del número de genes en el panel expresado por código de barras celular frente al número de códigos de barras celulares únicos/célula individual. La **FIG. 26B** muestra un histograma del número de moléculas únicas detectadas por microesfera frente a la frecuencia del número de células por código de barras celular único que porta un número dado de moléculas. La **FIG. 26C** muestra un histograma del número de moléculas GAPDH únicas detectadas por microesfera frente a la frecuencia del número de células/código de barras celular único que porta un número dado de moléculas.

La **FIG. 27** muestra un diagrama de dispersión de las 856 células.

La **FIG. 28** muestra un mapa térmico de expresión de las 100 principales (en términos del número total de moléculas detectadas).

45 La **FIG. 29** muestra un flujo de trabajo para el ejemplo 12.

La **FIG. 30** muestra un flujo de trabajo para el ejemplo 13.

50 **FIG. 31A-C.** Agrupamiento de células individuales en mezclas controladas que contienen dos tipos celulares distintos. A. Agrupamiento de una mezcla 1:1 de células K562 y Ramos por análisis de componentes principales de la expresión de 12 genes. El gráfico doble muestra dos grupos distintos, expresando un grupo genes específicos de Ramos y expresando el otro, genes específicos de K562. B. Análisis de componentes principales de una mezcla que contiene un pequeño porcentaje de células Ramos en un fondo de linfocitos B primarios de un individuo sano usando un panel de 111 genes. El color de cada dato puntual indica el número total de moléculas de transcrito únicas detectadas entre el panel de genes completo. Un conjunto de 18 células (en círculo) de las 1198 células

presenta un perfil de expresión génica distinto y con niveles de transcripción mucho mayores. C. Mapa térmico que muestra el nivel de expresión de cada gen en las 100 células principales en la muestra de la FIG. 31B, clasificadas por el número total de moléculas de transcrito detectadas en el panel de genes. Los genes están ordenados mediante agrupación jerárquica en términos de correlación. Las 18 células principales, indicadas por la barra roja horizontal, expresaban preferentemente un conjunto de genes que se sabe que está asociado con linfoma folicular, como se indica por la barra roja vertical.

FIG. 31D. Análisis de PCA de linfocitos B primarios con células Ramos añadidas. El color de cada dato puntual (célula individual) indica el logaritmo del número de moléculas de transcrito que porta cada célula para el gen particular. Siete filas superiores: Genes que se expresan preferentemente por el subconjunto de 18 células que son probablemente células Ramos. Los genes de la primera fila (de izquierda a derecha) incluyen GAPDH, TCL1A, MKI67 y BCL6. Los genes de la segunda fila (de izquierda a derecha) incluyen MYC, CCND3, CD81 y GNAI2. Los genes de la tercera fila (de izquierda a derecha) incluyen IGBP1, CD20, BLNK y DOCK8. Los genes de la cuarta fila (de izquierda a derecha) incluyen IRF4, CD22, IGHM y AURKB. Los genes de la quinta fila (de izquierda a derecha) incluyen CD38, CD10, LEF1 y AICDA. Los genes de la sexta fila (de izquierda a derecha) incluyen CD40, CD27, IL4R y PRKCD. Los genes de la séptima de la fila (de izquierda a derecha) incluyen RGS1, MCL1, CD79a y HLA-DRA. Última fila: Genes que se expresan preferentemente por un subconjunto de linfocitos B primarios, pero no especialmente enriquecidos en esas 18 células. Los genes en la última fila (de izquierda a derecha) incluyen IL6, CD23a, CCR7 y CXCR5.

FIG. 32. Expresión de GAPDH. El color indica el logaritmo natural del número de moléculas de transcrito únicas observadas por célula.

La **FIG. 33A-F** muestra el análisis de componentes principales (PCA) para genes asociados a monocitos. La FIG. 33A muestra el PCA para CD16. La FIG. 33B muestra el PCA para CCRvarA. La FIG. 33C muestra el PCA para CD14. La FIG. 33D muestra el PCA para S100A12. La FIG. 33E muestra el PCA para CD209. La FIG. 33F muestra el PCA para IFNGR1.

La **FIG. 34A-B** muestra el análisis de componentes principales (PCA) para marcadores de linfocitos T globales (CD3). La FIG. 34A muestra el PCA para CD3D y la FIG. 34B muestra el PCA para CD3E.

La **FIG. 35A-E** muestra el análisis de componentes principales (PCA) para genes asociados a linfocitos T CD8. La FIG. 35A muestra el PCA para CD8A. La FIG. 35B muestra el PCA para EOMES. La FIG. 35C muestra el PCA para CD8B. La FIG. 35D muestra el PCA para PRF1. La FIG. 35E muestra el PCA para RUNX3.

La **FIG. 36A-C** muestra el análisis de componentes principales (PCA) para genes asociados a linfocitos T CD4. La FIG. 36A muestra el PCA para CD4. La FIG. 36B muestra el PCA para CCR7. La FIG. 36C muestra el PCA para CD62L.

La **FIG. 37A-F** muestra el análisis de componentes principales (PCA) para genes asociados a linfocitos B. La FIG. 37A muestra el PCA para CD20. La FIG. 37B muestra el PCA para IGHD. La FIG. 37C muestra el PCA para PAX5. La FIG. 37D muestra el PCA para TCL1A. FIG. 37E muestra el PCA para IGHM. La FIG. 37F muestra el PCA para CD24.

La **FIG. 38A-C** muestra el análisis de componentes principales (PCA) para genes asociados a linfocitos citolíticos naturales. La FIG. 38A muestra el PCA para KIR2DS5. La FIG. 38B muestra el PCA para CD16. La FIG. 38C muestra el PCA para CD62L.

FIG. 39. Identificación simultánea de tipos celulares principales en una muestra de PBMC humana (632 células) por análisis de PCA de 81 genes ensayados por CytoSeq. Las células con perfil de expresión altamente correlacionada se codifican con un color similar.

FIG. 40A-B. Análisis de correlación de perfil de expresión génica de células individuales de la muestra de PBMC. 40A. Una matriz que muestra el coeficiente de correlación por pares entre 632 células en la muestra. Las células se ordenan de modo que aquellas con un perfil de expresión génica altamente correlacionado se agrupan juntas. 40B. Mapa térmico que muestra la expresión de cada gen por cada célula. Las células (columnas) están ordenadas de la misma manera que la matriz de correlación anterior. Los genes (filas) están ordenados de modo que los genes que comparten un patrón de expresión altamente similar entre las células se agrupan juntos. El tipo celular de cada grupo de células puede identificarse por el grupo de genes que coexpresan las células. Dentro de cada grupo celular principal, hay un grado sustancial de heterogeneidad en términos de expresión génica.

La **FIG. 41** representa datos de 731 células de un experimento replicado de la muestra de PBMC del mismo donador. Las células con perfil de expresión génica similar (basado en la agrupación jerárquica usando el coeficiente de correlación) se dibujan con color similar.

La **FIG. 42** muestra un mapa térmico que demuestra la correlación en el perfil de expresión génica entre los genes.

FIG. 43. Descripción de CytoSeq. A. Procedimiento experimental de CytoSeq. B. Estructura de oligonucleótidos adheridos a microesferas.

FIG. 44. Disección de subpoblaciones de linfocitos T CD3+. A. El PCA de la muestra no estimulada del donador 1 revela dos ramificaciones principales de células. El nivel de expresión (logaritmo de molécula de transcrito única) de un gen particular dentro de cada célula se indica con color. Los genes de citocinas asociadas a linfocito T auxiliares

y genes efectores están enriquecidos en células en la rama inferior, mientras que los genes asociados a linfocitos T citotóxicos están enriquecidos en la rama superior. En esta ocasión se muestran genes representativos. La primera fila muestra genes relacionados con linfocitos T auxiliares e incluyen (de izquierda a derecha) CD4, SELL y CCR7. La segunda fila muestra genes relacionados con linfocitos T citotóxicos e incluyen (de izquierda a derecha) CD8A, NKG2D y EOMES. B. El PCA de la muestra estimulada con anti-CD3/anti-CD8 del donador 1 que muestra enriquecimiento de expresión de los genes indicado para una de las dos ramas principales que representan linfocitos T auxiliares y citotóxicos. Estos genes se representan a bajas cantidades en la muestra no estimulada. Las dos primeras filas muestran genes que se sabe que están asociados con linfocitos T activados e incluyen (de izquierda a derecha) en la primera fila IRF4, CD69 y MYC y en la segunda fila GAPDH, TNF e IFNG. La tercera fila muestra genes que se sabe que están asociados con linfocitos T auxiliares activados e incluyen (de izquierda a derecha) IL2, LTA y CD40LG. La cuarta fila muestra genes que se sabe que están asociados con linfocitos T citotóxicos activados e incluyen (de izquierda a derecha) CCL4, CCL3 y GZMB. C. Número de células que contribuyen al nivel de expresión global de genes que muestran cambios de gran factor cuando se comparan muestras estimuladas sobre no estimuladas en datos agregados. Para varias citocinas (flechas rojas), la contribución de únicamente un pequeño número de células es responsable de un gran cambio en la expresión génica global en la población completa.

FIG. 45. Diagramas de PCA de muestras de linfocitos que han experimentado estimulación con microesferas anti-CD28/anti-CD3 en los dos donadores, y las muestras no estimuladas correspondientes, con énfasis sobre la expresión de genes que muestran claramente expresión preferente en subconjuntos auxiliares o citotóxicos en las muestras no estimuladas. El color de cada dato puntual (célula individual) indica $\log(\text{número de molécula de transcrito única})$ por célula para el gen indicado. Para cada par de gráficos estimulados y no estimulados en cada donador, la gama de color se ajusta para que sea igual. A. Genes que se sabe que están asociados tanto con linfocitos T auxiliares como citotóxicos. B. Genes que se sabe que están asociados con linfocito T citotóxicos. C. Genes que se sabe que están asociados con linfocitos T auxiliares.

FIG. 46A-D. Diagramas de PCA de muestras de linfocito T que han experimentado estimulación con microesferas anti-CD28/anti-CD3 en los dos donadores, y las muestras no estimuladas correspondientes con énfasis sobre la expresión de genes que se expresan en las muestras estimuladas, pero a un nivel bajo o indetectable en las muestras no estimuladas. El color de cada dato puntual (célula individual) indica $\log(\text{número de molécula de transcrito única})$ por célula para el gen indicado. Para cada par de gráficos estimulados y no estimulados en cada donador, la gama de color se ajusta para que sea igual. 46A y 46D. Genes que se expresan por ambas ramas de células tras la activación. 46B. Genes que se expresan preferentemente por células en la rama superior tras la activación. Estos genes, se sabe que están asociados con linfocitos T citotóxicos activados. 46C. Genes que se expresan preferentemente por células en la rama inferior tras la activación. Estos genes, se sabe que están asociados con linfocitos auxiliares activados.

FIG. 47. La agrupación de datos de linfocitos T CD3+ no estimulados del donador 1 muestra separaciones de células CD4 y CD8, así como un grupo de células que expresa granzima K y granzima A, pero poco CD8. Parte superior: Mapa térmico que muestra la correlación entre cada par de células. Las células que están altamente correlacionadas se agrupan juntas. Parte inferior: Mapa térmico que muestra el nivel de expresión de cada gen de cada célula. Las células y los genes están ordenados mediante agrupación jerárquica bidireccional.

FIG. 48. Similar a la FIG. 47, pero muestra datos de muestra de linfocitos T CD3+ estimulados con anti-CD3/anti-CD28 del donador 1. Parte superior: Mapa térmico que muestra la correlación entre cada par de células. Las células que están altamente correlacionadas se agrupan juntas. Parte inferior: Mapa térmico que muestra el nivel de expresión de cada gen de cada célula. Las células y los genes están ordenados mediante agrupación jerárquica bidireccional.

FIG. 49A-C. En el donador 1, se observó un cambio factorial global grande para varias citocinas en la muestra estimulada con anti-CD28/anti-CD3, en comparación con la no estimulada. A-B: Los cambios factoriales grandes de estas citocinas estaban compuestos principalmente por únicamente unas pocas células individuales (los puntos que están encerrados con cuadrados o círculos). A varias de estas citocinas contribuía el mismo número pequeño de células. C: Los patrones de coexpresión de estas citocinas coinciden con la combinación de citocinas característica para los subconjuntos Th2 y Th17 de linfocitos T auxiliares.

FIG. 50A-B. Disección de subpoblaciones de linfocitos T CD8+. A. La agrupación de datos CytoSeq define dos grupos principales de células CD8+, un grupo expresa genes compartidos por células de memoria centrales/virgenes, y el otro grupo expresa genes compartidos por células de memoria efectoras/efectoras. En esta ocasión se muestran los datos de la muestra no estimulada del donador 2. Parte superior: Mapa térmico que muestra la correlación entre cada par de células. Parte inferior: Mapa térmico que muestra el nivel de expresión de cada gen en cada célula. Las células y los genes están ordenados mediante agrupación jerárquica bidireccional. B. Identificación de linfocitos T específicos de antígeno infrecuentes mediante la expresión de interferón gamma (IFNG) en linfocitos T CD8+ de dos donadores después de estimulación con combinación de péptidos CMV. Cada célula se representa en el espacio de componentes principales 2D. Las células que expresan IFNG (en círculo) están habitualmente entre aquellas con los transcritos más detectados totales en el panel (indicados por el color). En el donador 2, la célula de mayor expresión (cuadrado) no produce IFNG, pero expresa citocinas IL6 e IL1B. El siguiente número a cada círculo indica la posición en orden descendente del número de moléculas de transcrito únicas, totales detectadas para esa célula.

FIG. 51. Similar a la figura 50A excepto que los datos en esta ocasión representan los de una muestra estimulada por CMV del donador 2. A. La agrupación de datos CytoSeq define dos grupos principales de células CD8+, un

grupo expresa genes compartidos por células de memoria centrales/virgenes y el otro grupo expresa genes compartidos por células de memoria efectoras/efectoras. En esta ocasión se muestran los datos de una muestra no estimulada del donador 2. Parte superior: Mapa térmico que muestra la correlación entre cada par de células. Parte inferior: Mapa térmico que muestra el nivel de expresión de cada gen en cada célula. Las células y los genes están ordenados mediante agrupación jerárquica bidireccional.

5 **FIG. 52.** Los datos se representan en el espacio de componentes principales. El color indica el log(número de molécula de transcrito única detectada) para el gen particular. A. Genes que parecen estar expresados por la proporción mayor de células tras estimulación por combinación de péptidos de CMV. B. Genes que están enriquecidos en una rama de células. Estos genes, también se sabe que están asociados con linfocitos T CD8+ virgenes y de memoria centrales. C. Genes que están enriquecidos en la otra rama de células. Estos genes, se sabe que están asociados con linfocitos T CD8+ efectoras y de memoria efectoras. D. Las células que expresan granzima K ocupan una región entre las células virgenes/de memoria centrales y efectoras/de memoria efectoras en el espacio PC. E. Las células que expresan HLA-DRA constituyen un subconjunto especial. F. Genes que se expresan en ambas ramas de células.

15 **FIG. 53.** Igual que la FIG. 50B, excepto que los datos representan los de los controles no estimulados. En ninguna de las células en la muestra del donador 1 expresó IFNG, mientras que una célula en la muestra del donador 2 expresó IFNG, aunque con expresión baja global entre el panel de genes completo (posición 1069). La escala de color se ajusta para que coincida con la del gráfico respectivo para la muestra estimulada.

20 **FIG. 54.** Mapas térmicos que muestran la expresión heterogénea del panel de genes en células que expresan interferón gamma (IFNG) en linfocitos T CD8+ estimulados por CMV de los donadores 1 y 2. También se muestra la célula que porta la mayoría de transcritos totales detectados en el donador 2. Esta célula particular no expresa IFNG, pero expresa fuertemente IL6, IL1B y CCL4. Las células y los genes están ordenados por agrupación jerárquica bidireccional basado en la correlación. La ID de la célula se refiere la posición en el número total de transcritos detectados del panel de genes, y se indican en los diagramas de PCA de la FIG. 50.

25 **FIG. 55.** Esquema de amplificación. La primera PCR amplifica moléculas adheridas a la microesfera usando un cebador específico de gen y un cebador contra la secuencia del cebador 1 de secuenciación Illumina universal. La segunda PCR amplifica los productos de la primera PCR usando un cebador específico de gen interno flanqueado por la secuencia del cebador 2 de secuenciación Illumina y un cebador contra la secuencia del cebador 1 de secuenciación Illumina universal. La tercera PCR añade P5 y P7 y el índice de muestra para convertir productos de PCR en la colección de secuencia Illumina. La secuencia de 150 pb x 2 revela el marcador celular y el marcador molecular en la lectura 1, el gen en la lectura 2 y el índice de muestra en la lectura de índice 1.

La **FIG. 56** representa un esquema de un flujo de trabajo para analizar moléculas de una muestra.

La **FIG. 57** representa un esquema de un flujo de trabajo para analizar moléculas de una muestra.

La **FIG. 58A-B** representa geles de agarosa de productos de PCR.

35 La **FIG. 59** representa un diagrama de lecturas de secuenciación para una pluralidad de genes.

La **FIG. 60A-D** representa diagramas de las lecturas observadas por marcador detectado (RPLD) para Lys, Phe, Thr y los controles de adición Dap, respectivamente. La FIG. 60E representa un diagrama de lecturas frente a entrada.

La **FIG. 61** representa un diagrama de las lecturas observadas por marcador detectado (RPLD) para diversos genes.

La **FIG. 62** representa un diagrama de las lecturas observadas por marcador detectado (RPLD) para diversos genes.

40 La **FIG. 63** representa un diagrama de lecturas totales (marcadores) frente a rpld para diversos genes.

La **FIG. 64** representa un diagrama de RPKM para genes no detectados.

La **FIG. 65** representa un esquema para la síntesis de códigos de barras moleculares.

La **FIG. 66A-C** representa esquemas para la síntesis de códigos de barras moleculares.

La **FIG. 67** muestra un esquema de un flujo de trabajo para el marcaje estocástico de ácidos nucleicos.

45 La **FIG. 68** es un esquema de un flujo de trabajo para el marcaje estocástico de ácidos nucleicos.

La **FIG. 69** ilustra una fijación mecánica dentro de la que pueden anclarse sustratos de matriz de micropocillos, formando de ese modo una cámara o pocillo de reacción en que pueden pipetearse muestras y reactivos para realizar experimentos combinados de marcaje estocástico de células individuales/indización molecular. Parte superior: Vista ampliada que muestra las partes superior e inferior de la fijación y una junta elastomérica para formar un sellamiento a prueba de filtraciones con el sustrato de matriz de micropocillos. Parte inferior: Vista lateral ampliada de la fijación.

La **FIG. 70** ilustra una fijación mecánica que crea dos cámaras o pocillos de reacción cuando se fija un sustrato de matriz de micropocillos dentro de la fijación.

La **FIG. 71** ilustra dos ejemplos de juntas elastoméricas (por ejemplo, polidimetilsiloxano) para su uso con las

fijaciones mecánicas ilustradas en las FIG. 69 y 70. Las juntas elastoméricas proporcionan un sellamiento a prueba de filtraciones con el sustrato de matriz de micropocillos para crear un pocillo de reactivo alrededor de la matriz de micropocillos. Las juntas pueden contener una (parte superior), dos (parte inferior) o más aberturas para crear pocillos de reactivo.

- 5 La **FIG. 72** representa una realización de un cartucho dentro del que se compacta una matriz de micropocillos. Parte izquierda: Una vista ampliada del cartucho que ilustra (desde la parte inferior hasta la parte superior) el sustrato de matriz de micropocillos, una junta que define la celda de flujo o cámara de matriz, un componente de depósito de reactivo y/o residuos para definir compartimentos para que contengan reactivos de ensayo precargados o reactivos usados de almacenamiento, y una cubierta para sellar los depósitos de reactivo y residuos y definir los accesos de entrada y salida de muestra. Parte derecha: Una vista ensamblada de una realización del diseño de cartucho que ilustra el relieve para poner un imán externo en cercana proximidad con la matriz de micropocillos.

La **FIG. 73** representa una realización de un cartucho diseñado para incluir reactivos de ensayo internos con la matriz de micropocillos compactada.

- 15 La **FIG. 74** proporciona una ilustración esquemática de un sistema instrumental para realizar ensayo combinado de marcaje estocástico de células individuales/indización molecular. El sistema instrumental puede proporcionar una diversidad de capacidades de control y análisis, y puede compactarse como módulos individuales o como un sistema completamente integrado. Las matrices de micropocillos pueden integrarse con celdas de flujo que son un componente fijo del sistema o se pueden retirar, o pueden compactarse dentro de cartuchos que se pueden retirar que comprenden además depósitos de reactivo de ensayo precargados y otra funcionalidad.

- 20 La **FIG. 75** ilustra una realización de las etapas del proceso a realizar por un sistema automatizado para realizar ensayos combinados de marcaje estocástico de células individuales/indización molecular.

La **FIG. 76** ilustra una realización de un sistema informático o procesador para proporcionar capacidades de control del instrumento y análisis de datos para el sistema de ensayo divulgado actualmente.

- 25 La **FIG. 77** muestra un diagrama de bloques que ilustra un ejemplo de una arquitectura de sistema informático que puede usarse en relación con realizaciones ejemplares de los sistemas de ensayo de la presente divulgación.

La **FIG. 78** representa un diagrama que muestra una red con una pluralidad de sistemas informáticos, teléfonos móviles, asistente de datos personales y almacenamiento conectado a red (NAS), que pueden usarse con realizaciones ejemplares de los sistemas de ensayo de la presente divulgación.

- 30 La **FIG. 79** representa un diagrama de bloques de un sistema informático multiprocesador que puede usarse con realizaciones ejemplares de los sistemas de ensayo de la presente divulgación.

La **FIG. 80** representa un diagrama de análisis de una muestra de ensayo y comunicación del resultado de ensayo obtenido de la muestra de ensayo mediante un medio de comunicación.

Descripción detallada

- 35 En la presente memoria se divulgan métodos, kits y composiciones para analizar moléculas en una pluralidad de muestras. Generalmente, los métodos, kits y composiciones comprenden (a) moléculas de marcaje estocástico en dos o más muestras con códigos de barras moleculares para producir moléculas marcadas; y (b) detección de las moléculas marcadas. Los códigos de barras moleculares pueden comprender una o más regiones específicas de diana, regiones de marcador, regiones de índice de muestra, regiones de PCR universales, adaptadores, conectores o una combinación de los mismos. Las moléculas marcadas pueden comprender a) una región molecular; b) una región de índice de muestra; y c) una región de marcador. La región molecular puede comprender al menos una parte de la molécula del código de barras molecular que se adhirió originalmente. La región molecular puede comprender un fragmento de la molécula del código de barras molecular que se adhirió originalmente. La región de índice de muestra puede usarse para determinar la fuente de la región molecular. La región de índice de muestra puede usarse para determinar la muestra de la que es originaria la región molecular. La región de índice de muestra puede usarse para diferenciar regiones moleculares de dos o más muestras diferentes. La región de marcador puede usarse para conferir una identidad única a regiones molecular idénticas originarias de la misma fuente. La región de marcador puede usarse para conferir una identidad única a regiones moleculares idénticas originarias de la misma muestra.

- 50 El método para analizar moléculas en una pluralidad de muestras puede comprender: a) producir una pluralidad de ácidos nucleicos de muestra con marca mediante: i) poner en contacto una primera muestra que comprende una pluralidad de ácidos nucleicos con una pluralidad de primeras marcas de muestra para producir una pluralidad de primeros ácidos nucleicos de muestra con marca; y ii) poner en contacto una segunda muestra que comprende una pluralidad de ácidos nucleicos con una pluralidad de segundas marcas de muestra para producir una pluralidad de segundos ácidos nucleicos de muestra con marca, en el que la pluralidad de segundas marcas de muestra son diferentes de las primeras marcas de muestra; b) poner en contacto la pluralidad de ácidos nucleicos con marca de muestra con una pluralidad de marcadores identificadores moleculares para producir una pluralidad de ácidos nucleicos marcados; y c) detectar al menos una parte de los ácidos nucleicos marcados, determinando de ese modo un recuento de una pluralidad de ácidos nucleicos en una pluralidad de muestras. La pluralidad de muestras puede comprender una célula individual.

Como alternativa, el método para analizar moléculas en una pluralidad de muestras puede comprender: a) producir una pluralidad de ácidos nucleicos marcados que comprende: i) poner en contacto una primera muestra con una primera pluralidad de marcas de muestra, en el que la primera pluralidad de marcas de muestra comprende secuencias de ácido nucleico idénticas; ii) poner en contacto la primera muestra con una primera pluralidad de marcadores identificadores moleculares que puede comprender diferentes secuencias de ácido nucleico, en el que el contacto de la primera muestra con la primera pluralidad de marcas de muestra o la primera pluralidad de marcadores identificadores moleculares se produce simultánea o secuencialmente para producir una pluralidad de primeros ácidos nucleicos marcados; iii) poner en contacto una segunda muestra con una segunda pluralidad de marcas de muestra, en el que la segunda pluralidad de marcas de muestra puede comprender secuencias de ácido nucleico idénticas; iv) poner en contacto la segunda muestra con una segunda pluralidad de marcadores identificadores moleculares que puede comprender diferentes secuencias de ácido nucleico, en el que el contacto de la segunda muestra con la segunda pluralidad de marcas de muestra o segunda pluralidad de marcadores identificadores moleculares se produce simultánea o secuencialmente para producir una pluralidad de segundos ácidos nucleicos marcados, en el que la pluralidad de ácidos nucleicos marcados puede comprender la pluralidad de primeros ácidos nucleicos marcados y los segundos ácidos nucleicos marcados; y b) determinar un número de diferentes ácidos nucleicos marcados, determinando de ese modo un recuento de una pluralidad de ácidos nucleicos en una pluralidad de muestras.

El método para analizar moléculas en una pluralidad de muestras puede comprender: a) poner en contacto una pluralidad de muestras que puede comprender dos o más ácidos nucleicos diferentes con una pluralidad de marcas de muestra y una pluralidad de marcadores identificadores moleculares para producir una pluralidad de ácidos nucleicos marcados, en el que: i) la pluralidad de ácidos nucleicos marcados puede comprender dos o más ácidos nucleicos adheridos a dos o más marcas de muestra y dos o más marcadores identificadores moleculares; ii) las marcas de muestra adheridas a los ácidos nucleicos de una primera muestra de la pluralidad de muestras son diferentes de las marcas de muestra adheridas a moléculas de ácidos nucleico de una segunda muestra de la pluralidad de muestras; y iii) dos o más ácidos nucleicos idénticos en la misma muestra se adhieren a dos o más marcadores identificadores moleculares diferentes; y b) detectar al menos una parte de los ácidos nucleicos marcados, determinando de ese modo un recuento de dos o más ácidos nucleicos diferentes en la pluralidad de muestras.

La FIG. 56 representa un flujo de trabajo ejemplar para la cuantificación de moléculas de ARN en una muestra. Como se muestra en la etapa 1 de la FIG. 56, las moléculas de ARN (110) pueden transcribirse de forma inversa para producir moléculas de ADNc (105) mediante la hibridación estocástica de un conjunto de marcadores identificadores moleculares (115) a la región de cola de poliA de las moléculas de ARN. Los marcadores identificadores moleculares (115) pueden comprender una región de oligodT (120), región de marcador (125) y región de PCR universal (130). El conjunto de marcadores identificadores moleculares puede contener 960 tipos diferentes de regiones de marcador. Como se muestra en la etapa 2 de la FIG. 56, las moléculas de ADNc marcadas (170) pueden purificarse para eliminar el exceso de marcadores identificadores moleculares (115). La purificación puede comprender purificación con microesferas Ampure. Como se muestra en la etapa 3 de la FIG. 56, las moléculas de ADNc marcadas (170) pueden amplificarse para producir un amplicón marcado (180). La amplificación puede comprender amplificación por PCR combinada. La amplificación puede comprender una amplificación por PCR combinada con 96 cebadores combinados en un único volumen de reacción. La amplificación puede comprender un cebador personalizado (135) y un cebador universal (140). El cebador personalizado (135) puede hibridar con una región dentro de la porción de ADNc (105) de la molécula de ADNc marcada (170). El cebador universal (140) puede hibridar con la región de PCR universal (130) de la molécula de ADNc marcada (170). Como se muestra en la etapa 4, los amplicones marcados (180) pueden amplificarse adicionalmente por PCR con cebadores internos. Las PCR con cebadores internos puede comprender PCR combinada con 96 cebadores combinados en un único volumen de reacción. La PCR con cebadores internos puede comprender un cebador personalizado (145) y un cebador universal (140). El cebador personalizado (135) puede hibridar con una región dentro de la porción de ADNc (105) del amplicón marcado (180). El cebador universal (140) puede hibridar con la región de PCR universal (130) del amplicón marcado (180). Como se muestra en la etapa 5, puede adherirse uno o más adaptadores (150, 155) al amplicón marcado (180) para producir un amplicón marcado con adaptador (190). El uno o más adaptadores pueden adherirse al amplicón marcado (180) mediante ligamiento. Como se muestra en la etapa 6, el uno o más adaptadores (150, 155) pueden usarse para realizar uno o más ensayos adicionales sobre el amplicón marcador con adaptador (190). El uno o más adaptadores (150, 155) pueden hibridarse con el uno o más cebadores (160, 165). El uno o más cebadores (160, 165) pueden ser cebadores de amplificación por PCR. El uno o más cebadores (160, 165) pueden ser cebadores de secuenciación. El uno o más adaptadores (150, 155) pueden usarse para amplificación adicional de los amplicones marcados con adaptador. El uno o más adaptadores (150, 155) pueden usarse para secuenciar el amplicón marcado con adaptador.

La FIG. 57 representa un esquema ejemplar de un flujo de trabajo para analizar ácidos nucleicos de dos o más muestras. Como se muestra en la FIG. 57, un método para analizar ácidos nucleicos de dos o más muestras puede comprender seleccionar dos o más genes para su análisis y diseñar cebadores personalizados basados en los genes seleccionados (210). El método puede comprender además complementar una o más muestras que comprenden ácidos nucleicos (por ejemplo, ARN) con uno o más controles de adición (220). Los ácidos nucleicos de la muestra pueden amplificarse por RT-PCR combinada (230) con códigos de barras moleculares (o marcas de muestra o marcadores identificadores moleculares) y los cebadores personalizados para producir amplicones marcados. Los amplicones marcados pueden tratarse adicionalmente con uno o más adaptadores de secuenciación para producir amplicones marcados con adaptador (240). Los amplicones marcados con adaptador puede analizarse (250). Como se muestra en la FIG. 57, el análisis de los amplicones marcados (250) puede comprender uno o más

ES 2 711 168 T3

de (1) detección de una secuencia de cebador de PCR universal, poliA y/o código de barras molecular (o marca de muestra, marcador identificador molecular); (2) lectura de mapa en el extremo de los amplicones marcados con adaptador (por ejemplo, 96 genes y controles de adición) que no está adherido al adaptador y/o código de barras (por ejemplo, código de barras molecular, marca de muestra, marcador identificador molecular); y (3) recuento y/o resumen del número de amplicones marcados con adaptador diferentes.

La FIG. 67 muestra un esquema de un flujo de trabajo para el marcaje estocástico de ácidos nucleicos con códigos de barras moleculares (1220). Como se muestra en la etapa 1 de la FIG. 67, las moléculas de ARN pueden marcarse de forma estocástica con un conjunto de códigos de barras moleculares (1220). Los códigos de barras moleculares (1220) pueden comprender una región de unión a diana (1221), región de marcador (1222), región de índice de muestra (1223) y región de PCR universal (1224). En algunos casos, la región de unión a diana comprende una secuencia de oligodT que hibrida con una secuencia de poliA en las moléculas de ARN. La región de marcador (1222) puede contener una secuencia única que puede usarse para distinguir dos o más códigos de barras moleculares diferentes. Cuando el código de barras molecular hibrida con una molécula de ARN, la región de marcador puede usarse para conferir una identidad única a moléculas de ARN idénticas. La región de índice de muestra (1223) puede ser idéntica para un conjunto de códigos de barras moleculares. La región de índice de muestra (1223) puede usarse para distinguir ácidos nucleicos marcados de diferentes muestras. La región de PCR universal (1224) puede servir como sitio de unión al cebador para la amplificación de las moléculas marcadas. Una vez que las moléculas de ARN están marcadas con los códigos de barras moleculares, las moléculas de ARN pueden transcribirse de forma inversa para producir moléculas de ADNc marcadas (1230) que contienen una copia de ADNc de la molécula de ARN (1210) y el código de barras molecular (1220).

Como se muestra en la etapa 2 de la FIG. 67, el exceso de oligos (por ejemplo, códigos de barras moleculares) puede eliminarse por purificación con microesferas Ampure. Como se muestra en la etapa 3 de la FIG. 67, las moléculas de ADNc marcadas pueden amplificarse por PCR combinada. La PCR combinada de las moléculas de ADNc marcadas puede realizarse usando un primer conjunto de cebadores directos (F1, 1235 en la FIG. 67) y cebadores universales (1240) en un volumen de reacción único para producir amplicones marcados (1245). Como se muestra en la etapa 4 de la FIG. 67, los amplicones marcados pueden amplificarse adicionalmente por PCR combinada usando cebadores internos. La amplificación con cebadores internos de los amplicones marcados puede realizarse usando un segundo conjunto de cebadores directos (F2, 1250 en la FIG. 67) y cebadores universales (1240) en un único volumen de reacción para producir amplicones de PCR con cebadores internos marcados. En algunos casos, los cebadores F2 (1250) contienen un adaptador (1251) y una región de unión a diana (1252). La región de unión a diana (1252) de los cebadores F2 puede hibridar con los amplicones marcados y puede cebar la amplificación de los amplicones marcados. El adaptador (1251) y la región de PCR universal (1224) de los amplicones de PCR con cebadores internos puede usarse en la secuenciación de los amplicones de PCR con cebadores internos marcados. Los amplicones pueden secuenciarse por MiSeq. Como alternativa, los amplicones pueden secuenciarse por HiSeq.

La FIG. 68 muestra un esquema de un flujo de trabajo para el marcaje estocástico de ácidos nucleicos. Como se muestra en la etapa 1 de la FIG. 68. Las moléculas de ARN (1305) pueden marcarse de forma estocástica con un conjunto de códigos de barras moleculares (1320). Los códigos de barras moleculares pueden comprender una región de unión a diana (1321), región de marcador (1322) y región de PCR universal (1323). Una vez que los códigos de barras moleculares se han adherido a las moléculas de ARN, las moléculas de ARN (1305) pueden transcribirse de forma inversa para producir moléculas de ADNc marcadas (1325) que comprenden una copia de ADNc de la molécula de ARN (1310) y el código de barras molecular (1320). Como se muestra en la etapa 2 de la FIG. 68, las moléculas de ADNc marcadas pueden purificarse por purificación con microesferas Ampure para eliminar el exceso de oligos (por ejemplo, códigos de barras moleculares). Como se muestra en la etapa 3 de la FIG. 68, los amplicones marcados pueden amplificarse por PCR combinada. La PCR combinada de las moléculas de ADNc marcadas puede realizarse usando un primer conjunto de cebadores directos (F1, 1330 en la FIG. 68) y cebadores universales (1335) en un único volumen de reacción para producir amplicones marcados (1360). Como se muestra en la etapa 4 de la FIG. 67, los amplicones marcados pueden amplificarse adicionalmente por PCR combinada usando cebadores internos. La amplificación con cebadores internos de los amplicones marcados puede realizarse usando un segundo conjunto de cebadores directos (F2, 1340 en la FIG. 68) y cebadores de índice de muestra (1350) en un único volumen de reacción para producir amplicones de PCR con cebadores internos marcados. En algunos casos, los cebadores F2 (1340) contienen un adaptador (1341) y una región de unión a diana (1342). La región de unión a diana (1342) de los cebadores F2 puede hibridar con los amplicones marcados y puede cebar la amplificación de los amplicones marcados. Los cebadores de índice de muestra (1350) pueden comprender una región de cebador universal (1351), región de índice de muestra (1352) y región adaptadora (1353). Como se muestra en la etapa 4 de la FIG. 68, la región de cebador universal (1351) del cebador de índice de muestra puede hibridar con la región de PCR universal de los amplicones marcados. La región de índice de muestra (1352) del cebador de índice de muestra puede usarse para distinguir dos o más muestras. Las regiones adaptadoras (1341, 1353) pueden usarse para secuenciar los amplicones de PCR con cebadores internos marcados. Los amplicones pueden secuenciarse por MiSeq. Como alternativa, los amplicones pueden secuenciarse por HiSeq.

En la presente memoria se divulgan además métodos de producción de una o más colecciones. La una o más colecciones pueden comprender una pluralidad de moléculas marcadas. La una o más colecciones pueden comprender una pluralidad de amplicones marcados. La una o más colecciones pueden comprender una pluralidad de moléculas enriquecidas o un derivado de las mismas (por ejemplo, moléculas marcadas, amplicones marcados). En general, el método de producción de una o más colecciones comprende (a) marcar de forma estocástica una pluralidad de moléculas de dos o más muestras para producir una pluralidad de moléculas marcadas, en el que las moléculas marcadas comprenden una región molecular, una región de índice de muestra y región de marcador; y (b)

producir una o más colecciones a partir de la pluralidad de moléculas marcadas, en el que (i) la una o más colecciones comprenden dos o más moléculas marcadas diferentes, (ii) las dos o más moléculas marcadas diferentes difieren en la región molecular, región de índice de muestra, región de marcador o una combinación de las mismas.

5 El método para producir una o más colecciones puede comprender: a) producir una pluralidad de ácidos nucleicos con marca de muestra mediante: i) poner en contacto una primera muestra que comprende una pluralidad de ácidos nucleicos con una pluralidad de primeras marcas de muestra para producir una pluralidad de primeros ácidos nucleicos con marca de muestra; y ii) poner en contacto una segunda muestra que comprende una pluralidad de ácidos nucleicos con una pluralidad de segundas marcas de muestra para producir una pluralidad de segundos ácidos nucleicos con marca de muestra, en el que la pluralidad de primeras marcas de muestra son diferentes de las segundas marcas de muestra; y b) poner en contacto la pluralidad de ácidos nucleicos con marca de muestra con una pluralidad de marcadores identificadores moleculares para producir una pluralidad de ácidos nucleicos marcados, produciendo de ese modo una colección de ácidos nucleicos marcados.

15 El contacto con una muestra puede ser aleatorio o no aleatorio. Por ejemplo, el contacto de una muestra con marcas de muestra puede ser un contacto aleatorio o no aleatorio. En algunas realizaciones, la muestra se pone en contacto con marcas de muestra aleatoriamente. En algunas realizaciones, la muestra se pone en contacto con marcas de muestra de forma no aleatoria. El contacto con una pluralidad de ácidos nucleicos puede ser aleatorio o no aleatorio. Por ejemplo, el contacto de una pluralidad de ácidos nucleicos con marcas de muestra puede ser un contacto aleatorio o no aleatorio. En algunas realizaciones, la pluralidad de ácidos nucleicos se pone en contacto con marcas de muestra de forma aleatoria. En algunas realizaciones, la pluralidad de ácidos nucleicos se pone en contacto con marcas de muestra de forma no aleatoria.

25 En la presente memoria se divulgan además métodos de producción de uno o más conjuntos de microesferas marcadas. El método de producción del uno o más conjuntos de microesferas marcadas puede comprender la adhesión de uno o más ácidos nucleicos a una o más microesferas, produciendo de ese modo uno o más conjuntos de microesferas marcadas. El uno o más ácidos nucleicos pueden comprender uno o más códigos de barras moleculares. El uno o más ácidos nucleicos pueden comprender una o más marcas de muestra. El uno o más ácidos nucleicos pueden comprender uno o más marcadores identificadores moleculares. El uno o más ácidos nucleicos pueden comprender a) una región de cebador; b) una región de índice de muestra; y c) un conector o región adaptadora. El uno o más ácidos nucleicos pueden comprender a) una región de cebador; b) una región de marcador; y c) un conector o región adaptadora. El uno o más ácidos nucleicos pueden comprender a) una región de índice de muestra; y b) una región de marcador. El uno o más ácidos nucleicos pueden comprender además una región específica de diana. El uno o más ácidos nucleicos pueden comprender además una región de conector. El uno o más ácidos nucleicos pueden comprender además una región adaptadora. El uno o más ácidos nucleicos pueden comprender además una región de índice de muestra. El uno o más ácidos nucleicos pueden comprender además una región de marcador.

40 En la presente memoria se divulgan además métodos para seleccionar uno o más cebadores personalizados. El método de selección de un cebador personalizado para analizar moléculas en una pluralidad de muestras puede comprender: a) un primer pase, en el que los cebadores elegidos pueden comprender: i) no más de tres guaninas secuenciales, no más de tres citosinas secuenciales, no más de cuatro adeninas secuenciales y no más de cuatro timinas secuenciales; ii) al menos 3, 4, 5, o 6 nucleótidos que son guaninas o citosinas; y iii) una secuencia que no forma fácilmente una estructura de horquilla; b) un segundo pase, que comprende: i) una primera ronda de elección de una pluralidad de secuencias que tienen alta cobertura de todos los transcritos; y ii) una o más rondas posteriores, que seleccionan una secuencia que tiene la mayor cobertura de transcritos restantes y una puntuación complementaria con otras secuencias elegidas de no más de 4; y c) añadir secuencias a un conjunto escogido hasta que se sature la cobertura o el número total de cebadores personalizados sea menor de o igual a aproximadamente 96.

50 En la presente memoria se divulgan además kits para su uso en el análisis de dos o más moléculas de dos o más muestras. El kit puede comprender (a) un primer recipiente que comprende un primer conjunto de códigos de barras moleculares, en el que (i) un código de barras molecular del primer conjunto de códigos de barras moleculares comprende una región de índice de muestra y una región de marcador; (ii) la región de índice de muestra de dos o más códigos de barras del primer conjunto de códigos de barras moleculares es igual; y (iii) la región de marcador de dos o más códigos de barras del primer conjunto de códigos de barras moleculares son diferentes; y (b) un segundo recipiente que comprende un segundo conjunto de códigos de barras moleculares, en el que (i) un código de barras molecular del segundo conjunto de códigos de barras moleculares comprende una región de índice de muestra y una región de marcador; (ii) la región de índice de muestra de dos o más códigos de barras del segundo conjunto de códigos de barras moleculares son iguales; (iii) las regiones de marcador de dos o más códigos de barras del segundo conjunto de códigos de barras moleculares son diferentes; (iv) la región de índice de muestra de los códigos de barras del segundo conjunto de códigos de barras moleculares son diferentes de las regiones de índice de muestra de los códigos de barras del primer conjunto de códigos de barras moleculares; y (v) la región de marcador de dos o más códigos de barras del segundo conjunto de códigos de barras moleculares son idénticas a la región de marcador de dos o más códigos de barras del primer conjunto de códigos de barras moleculares.

65 Como alternativa, el kit comprende: a) una pluralidad de microesferas, en las que una o más microesferas de la pluralidad de microesferas pueden comprender al menos uno de una pluralidad de ácidos nucleicos, en la que al menos uno de una pluralidad de ácidos nucleicos puede comprender: i) al menos una secuencia de cebador, en la

que la secuencia de cebador de al menos uno de la pluralidad de ácidos nucleicos es igual para la pluralidad de microesferas; ii) una secuencia específica de microesfera, en la que la secuencia específica de microesfera de uno cualquiera de la pluralidad de ácidos nucleicos es igual, y en la que la secuencia específica de microesfera es diferente para una cualquiera de la pluralidad de microesferas; y iii) una secuencia estocástica, en la que la secuencia estocástica es diferente para uno cualquiera de la pluralidad de ácidos nucleicos; b) un cebador puede comprender una secuencia complementaria a la secuencia de cebador; y c) uno o más agentes de amplificación adecuados para la amplificación de ácidos nucleicos.

Como alternativa, el kit comprende: a) un primer recipiente que comprende un primer conjunto de marcas de muestra, en el que (i) una marca de muestra del primer conjunto de marcas de muestra comprende una región de índice de muestra; y (ii) las regiones de índice de muestra de las marcas de muestra del primer conjunto de marcas de muestra son al menos aproximadamente un 80 % idénticas; y b) un segundo recipiente que comprende un primer conjunto de marcadores identificadores moleculares, en el que (i) un marcador identificador molecular del primer conjunto de marcadores identificadores moleculares comprende una región de marcador; y (ii) al menos aproximadamente un 30 % de las regiones de los marcadores identificadores moleculares totales del primer conjunto de marcadores identificadores moleculares son diferentes.

Antes de describir en mayor detalle los presentes métodos, kits y composiciones, debe entenderse que esta invención no se limita al método, kit o composición particular descrito, ya que puede variar, por supuesto. También debe entenderse que la terminología usada en la presente memoria es con el fin de describir realizaciones particulares únicamente, y no se pretende que sea limitante, ya que el alcance de la presente invención se limitará únicamente por las reivindicaciones anejas. Los ejemplos se exponen para proporcionar a los expertos en la materia una divulgación completa y descripción de la manera en que hacer y usar la presente invención, y no se pretende que limiten el alcance de lo que los autores de la invención consideran su invención ni se pretende que represente que los experimentos siguientes son todos o los únicos experimentos realizados. Se han hecho esfuerzos por asegurar la precisión con respecto a los números usados (por ejemplo, cantidades, temperatura, etc.) pero deben tenerse en cuenta algunos errores experimentales y desviaciones. Salvo que se indique de otro modo, las partes son partes en peso, la masa molecular es masa molecular promedio en masa, la temperatura es en grados centígrados y la presión es en o casi atmosférica.

Se proporcionan métodos, kits y composiciones para el marcaje estocástico de ácidos nucleicos en una pluralidad de muestras o en una preparación de ácidos nucleicos compleja. Estos métodos, kits y composiciones encuentran uso en desentrañar los mecanismos de respuesta celular, diferenciación o transducción de señales en la realización de una amplia diversidad de mediciones clínicas. Estos y otros objetivos, ventajas y características de la invención llegarán a ser evidentes para los expertos en la materia tras la lectura de los detalles de los métodos, kits y composiciones que se describe más completamente a continuación.

Los métodos divulgados en la presente memoria comprenden adherir uno o más códigos de barras moleculares, marcas de muestra y/o marcadores identificadores moleculares a dos o más moléculas de dos o más muestras. Los códigos de barras moleculares, marcas de muestra y/o marcadores identificadores moleculares pueden comprender uno o más oligonucleótidos. En algunos casos, la adhesión de códigos de barras moleculares, marcas de muestra y/o marcadores identificadores moleculares a las moléculas comprende el marcaje estocástico de las moléculas. Los métodos para el marcaje estocástico de moléculas pueden encontrarse, por ejemplo, en los documentos de Estados Unidos con número de serie 12/969581 y 13/327526. En general, el método de marcaje estocástico comprende la adhesión aleatoria de una pluralidad de oligonucleótidos de marca y de marcador a una o más moléculas. Los códigos de barras moleculares, marca de muestra y/o marcadores identificadores moleculares se proporcionan en exceso de la una o más moléculas a marcar. En el marcaje estocástico, cada molécula individual a marcar tiene una probabilidad individual de adhesión a la pluralidad de los códigos de barras moleculares, marcas de muestra y/o marcadores identificadores moleculares. La probabilidad de que cada molécula individual a marcar se adhiera a un código de barras molecular, marca de muestra y/o marcador identificador molecular particular puede ser aproximadamente igual que cualquier otra molécula individual a marcar. Por consiguiente, en algunos casos, la probabilidad de que cualquiera de las moléculas en una muestra encuentre cualquiera de las marcas y marcadores se asume que es igual, una suposición que puede usarse en cálculos matemáticos para estimar el número de moléculas en la muestra. En algunas circunstancias, la probabilidad de adhesión puede manipularse, por ejemplo, eligiendo marcas y marcadores con diferentes propiedades que aumentarían o disminuirían la eficacia de unión de ese código de barras molecular, marca de muestra y/o marcador identificador molecular con una molécula individual. Las marcas y marcadores también pueden variarse en números para alterar la probabilidad de que un código de barras molecular, marca de muestra y/o marcador identificador molecular particular encuentre un compañero de unión durante el marcaje estocástico. Por ejemplo, un marcador está sobrerrepresentado en una combinación de marcadores, aumentando de ese modo las posibilidades de que el marcador sobrerrepresentado encuentre al menos un compañero de unión.

Los métodos divulgados en la presente memoria pueden comprender además la combinación de dos o más muestras. Los métodos divulgados en la presente memoria pueden comprender además combinar una o más moléculas de dos o más muestras. Por ejemplo, los métodos divulgados en la presente memoria comprenden combinar una primera muestra y una segunda muestra. Las dos o más muestras pueden combinarse después de realizar uno o más procedimientos de marcaje estocástico. Las dos o más muestras pueden combinarse después de la adhesión de uno o más conjuntos de códigos de barras moleculares a dos o más moléculas de las dos o más muestras. Las dos o más muestras pueden combinarse después de la adhesión de uno o más conjuntos de marcadores identificadores moleculares a dos o más moléculas

de las dos o más muestras. Por ejemplo, la primera y segunda muestras se combinan antes de entrar en contacto con la pluralidad de marcadores identificadores moleculares.

5 Como alternativa, las dos o más muestras pueden combinarse antes de realizar uno o más procedimientos de marcaje estocástico. Las dos o más muestras pueden combinarse antes de la adhesión de uno o más conjuntos de códigos de barras moleculares a dos o más moléculas de las dos o más muestras. Las dos o más muestras pueden combinarse antes de la adhesión de uno o más conjuntos de marcas de muestra a dos o más moléculas de las dos o más muestras. Las dos o más muestras pueden combinarse antes de la adhesión de uno o más conjuntos de marcadores identificadores moleculares a dos o más moléculas de las dos o más muestras.

10 Las dos o más muestras pueden combinarse después de realizar uno o más ensayos en dos o más moléculas o derivados de las mismas (por ejemplo, moléculas marcadas, amplicones) de las dos o más muestras. El uno o más ensayos pueden comprender una o más reacciones de amplificación. El uno o más ensayos pueden comprender uno o más ensayos de enriquecimiento. El uno o más ensayos pueden comprender uno o más ensayos de detección. Por ejemplo, la primera y segunda muestras se combinan después de detectar los ácidos nucleicos marcados.

15 Las dos o más muestras pueden combinarse antes de realizar uno o más ensayos en dos o más moléculas o derivados de las mismas (por ejemplo, moléculas marcadas, amplicones) de las dos o más muestras. El uno o más ensayos pueden comprender una o más reacciones de amplificación. El uno o más ensayos pueden comprender uno o más ensayos de enriquecimiento. El uno o más ensayos pueden comprender uno o más ensayos de detección. Por ejemplo, la primera y segunda muestras se combinan antes de detectar los ácidos nucleicos marcados.

Soportes

La presente divulgación comprende composiciones y métodos para el análisis combinado de secuencias de células individuales. Los métodos y composiciones de la presente divulgación proporcionan el uso de soportes sólidos. En algunos casos, los métodos, kits y composiciones divulgados en la presente memoria comprenden un soporte.

25 Las expresiones "soporte", "soporte sólido", "soporte semisólido" y "sustrato" pueden usarse indistintamente y se refieren a un material o grupo de materiales que tiene una superficie o superficies rígidas o semirrígidas. Un soporte puede referirse a cualquier superficie que sea transferible de una disolución a otra o forme una estructura para realizar ensayos basados en oligonucleótidos. El soporte o sustrato puede ser un soporte sólido. Como alternativa, el soporte es un soporte no sólido. Un soporte puede referirse a un material insoluble, semisoluble o soluble. Un soporte puede mencionarse como "funcionalizado" cuando incluye un conector, una estructura, un componente principal u otro resto reactivo adherido al mismo, mientras que un soporte sólido puede estar "no funcionalizado" cuando carece de dicho resto reactivo adherido al mismo. El soporte puede emplearse libre en disolución, tal como en un formato de pocillos de microvaloración; en un formato de flujo continuo, tal como en una columna; o en una tira reactiva.

35 El soporte o sustrato puede comprender una membrana, papel, plástico, superficie recubierta, superficie plana, vidrio, portaobjetos, chip o cualquier combinación de los mismos. En muchas realizaciones, al menos una superficie del soporte puede ser sustancialmente plana, aunque en algunas realizaciones puede ser deseable separar físicamente las regiones de síntesis para diferentes compuestos con, por ejemplo, pocillos, regiones elevadas, clavija, surcos grabados, o similares. De acuerdo con otras realizaciones, el soporte o los soportes sólidos pueden adoptar la forma de resinas, geles, microesferas u otras configuraciones geométricas. Como alternativa, el soporte o los soportes sólidos comprenden chips de sílice, micropartículas, nanopartículas, placas y matrices. Los soportes sólidos pueden incluir microesferas (por ejemplo, gel de sílice, vidrio de poro controlado, microesferas magnéticas, Dynabeads, resina Wang; resina Merrifield, microesferas de Sephadex/Sepharose, microesferas de celulosa, microesferas de poliestireno, etc.), capilares, soportes planos tales como filtros de fibra de vidrio, superficies de vidrio, superficies metálicas (acero, plata dorada, aluminio, silicio y cobre), soportes de vidrio, soportes de plásticos, soporte de silicio, chips, filtros, membranas, placas de micropocillos, portaobjetos o similares, materiales de plásticos incluyendo placas de micropocillos o membranas (por ejemplo, formadas de polietileno, polipropileno, poliamida, fluoruro de polivinilideno), obleas, peines, clavijas o agujas (por ejemplo, matrices de clavijas adecuadas para la síntesis o análisis combinatorio) o microesferas en una matriz de pozos o pocillos de nanolitros de superficies planos tales como obleas (por ejemplo, obleas de silicio), obleas con pozos con o sin fondos de filtro.

55 Se han descrito métodos y técnicas aplicables a la síntesis de matrices poliméricas (incluyendo proteínas) en la publicación de patente de Estados Unidos n.º 20050074787, la publicación internacional WO 00/58516, las patentes de Estados Unidos n.º 5143854, 5242974, 5252743, 5324633, 5384261, 5405783, 5424186, 5451683, 5482867, 5491074, 5527681, 5550215, 5571639, 5578832, 5593839, 5599695, 5624711, 5631734, 5795716, 5831070, 5837832, 5856101, 5858659, 5936324, 5968740, 5974164, 5981185, 5981956, 6025601, 6033860, 6040193, 6090555, 6136269, 6269846 y 6428752, en la publicación PCT n.º WO 99/36760 y WO 01/58593. Las patentes que describen técnicas de síntesis en realizaciones específicas incluyen las patentes de Estados Unidos n.º 5412087, 6147205, 6262216, 6310189, 5889165 y 5959098. Se describen matrices de ácidos nucleicos en muchas de las patentes anteriores, pero muchas de las mismas técnicas pueden aplicarse a matrices de polipéptidos. Se divulgan sustratos ejemplares adicionales en la patente de Estados Unidos n.º 5744305 y en las publicaciones de patente de Estados Unidos n.º 20090149340 y 20080038559.

60 La adhesión de los ácidos nucleicos marcados al soporte puede comprender reticulación de amina-tiol, reticulación de maleimida, N-hidroxisuccinimida o N-hidroxisulfosuccinimida, Zenon o SiteClick. La adhesión de los ácidos

ES 2 711 168 T3

nucleicos marcados al soporte puede comprender la adhesión de biotina a la pluralidad de ácidos nucleicos marcados y recubrimiento de la una o más microesferas con estreptavidina.

5 En algunos casos, un soporte sólido puede comprender una estructura molecular. Las estructuras moleculares
ejemplares pueden incluir anticuerpos, antígenos, reactivos de afinidad, polipéptidos, ácidos nucleicos, orgánulos
celulares y similares. Las estructuras moleculares pueden unirse juntas (por ejemplo, un soporte sólido puede
comprender una pluralidad de estructuras moleculares conectadas). Las estructuras moleculares pueden unirse
juntas mediante un conector de aminoácidos, un conector de ácido nucleico, una unión de molécula pequeña (por
ejemplo, biotina y avidina) y/o una unión de matriz (por ejemplo, PEG o glicerol). Las uniones pueden ser no
covalentes. Las uniones pueden ser covalentes. En algunos casos, las estructuras moleculares pueden no unirse.
10 Puede usarse una pluralidad de estructuras moleculares individuales en los métodos de la divulgación.

15 En algunos casos, un soporte puede comprender una nanopartícula. La nanopartícula puede ser una nanopartícula
de níquel, oro, plata, carbono, cobre, silicato, cobalto y platino, óxido de cinc, dióxido de sílice cristalino y/o plata.
Como alternativa o adicionalmente, la nanopartícula puede ser una nanopartícula de oro incluida en un óxido de
manganeso poroso. La nanopartícula puede ser una nanopartícula de hierro. La nanopartícula puede ser un
nanotetrapodo tachonado con nanopartículas de carbono.

20 Un soporte puede comprender un polímero. Un polímero puede comprender una matriz. Una matriz puede
comprender además una o más microesferas. Un polímero puede comprender PEG, glicerol, polisacárido o una
combinación de los mismos. Un polímero puede ser un plástico, caucho, nailon, silicona, neopreno y/o poliestireno.
Un polímero puede ser un polímero natural. Ejemplos de polímeros naturales incluyen, aunque sin limitación,
gomalaca, ámbar, lana, seda, celulosa y caucho natural. Un polímero puede ser un polímero sintético. Ejemplos de
polímeros sintéticos incluyen, aunque sin limitación, caucho sintético, resina de fenol formaldehído (o baquelita),
neopreno, nailon, poli(cloruro de vinilo) (PVC o vinilo), poliestireno, polietileno, polipropileno, poliacrilonitrilo, PVB y
silicona.

25 Un soporte puede ser un soporte semisólido. Un soporte puede comprender un gel (por ejemplo, un hidrogel). Los
términos "hidrogel", "gel" y similares se usan indistintamente en la presente memoria y pueden referirse a un material
que no es un líquido de flujo fácil y no es un sólido, sino un gel que es un gel comprendido de un 0,5 % o más y
preferiblemente menos de un 40 % en peso de material de soluto que forma gel y de un 95 % o menos y
preferiblemente más de un 55 % de agua. Los geles pueden formarse mediante el uso de un soluto que es
preferiblemente un soluto sintético (pero podría ser un soluto natural, por ejemplo, para formar gelatina) que forma
30 celdas interconectadas que se unen a, quedan atrapadas, se absorben y/o mantienen de otro modo agua y crean de
ese modo un gel en combinación con agua, donde el agua incluye agua unida y no unida. El gel puede ser la
estructura básica del parche de hidrogel que incluirá componentes adicionales más allá del material de soluto que
forma el gel y agua, tales como una enzima y una sal, que son componentes adicionales que se describen en la
presente memoria. El gel puede ser un gel polimérico.

35 Un soporte solio puede comprender una nanoestructura estructurada. Por ejemplo, la nanoestructura estructurada
puede comprender recipientes de captura (por ejemplo, un panal en miniatura) que puede comprender los
oligonucleótidos para capturar la célula y/o los contenidos de la célula. En algunos casos las nanoestructuras
estructuradas pueden no necesitar la adición de reactivos exógenos.

40 El soporte comprende una microesfera. Una microesfera puede abarcar cualquier tipo de esfera sólida o hueca,
bola, rodamiento, cilindro u otra configuración similar compuesta de plástico, cerámica, metal o material polimérico
sobre el que puede inmovilizarse un ácido nucleico (por ejemplo, covalente o no covalentemente). Una microesfera
puede comprender una hebra o hebras de nailon. Una microesfera puede ser de forma esférica. Una microesfera
puede ser de forma no esférica. Las microesferas pueden estar sin pulir o, si se están pulidas, la microesfera pulida
puede ser áspera antes del tratamiento (por ejemplo, con un agente alquilante). Una microesfera puede comprender
45 una partícula concreta que puede ser esférica (por ejemplo, microesferas) o tener forma irregular. Las microesferas
pueden comprender una diversidad de materiales incluyendo, aunque sin limitación, materiales paramagnéticos,
cerámica, plástico, vidrio, poliestireno, metilistireno, polímeros acrílicos, titanio, látex, sepharose, celulosa, nailon y
similares. Una microesfera puede adherirse a o incluirse en uno o más soportes. Una microesfera puede adherirse a
un gel o hidrogel. Una microesfera puede incluirse en un gel o hidrogel. Una microesfera puede adherirse a una
matriz. Una microesfera puede incluirse en una matriz. Una microesfera puede adherirse a un polímero. Una
microesfera puede incluirse en un polímero. La posición espacial de una microesfera dentro del soporte (por
ejemplo, gel, matriz, estructura o polímero) puede identificarse usando el oligonucleótido presente en la microesfera
que sirve como dirección de ubicación. Los ejemplos de microesferas incluyen, aunque sin limitación, microesferas
de estreptavidina, microesferas de agarosa, microesferas magnéticas, Dynabeads®, microesferas MACS®,
55 microesferas conjugadas con anticuerpo (por ejemplo, microesfera antiinmunoglobulina), microesferas conjugadas
con proteína A, microesferas conjugadas con proteína G, microesferas conjugadas con proteína A/G, microesferas
conjugadas con proteína L, microesferas conjugadas con oligodT, microesferas de sílice, microesferas de tipo sílice,
microesferas antibiotina, microesferas antifuorocromo y microesferas magnéticas terminadas en carboxi BcMag™.
El diámetro de las microesferas puede ser aproximadamente 5 µm, 10 µm, 20 µm, 25 µm, 30 µm, 35 µm, 40 µm,
60 45 µm o 50 µm. Una microesfera puede referirse a cualquier estructura tridimensional que pueda proporcionar un
área superficial aumentada para la inmovilización de partículas biológicas y macromoléculas, tales como ADN y
ARN.

Un soporte puede ser poroso. Un soporte puede ser permeable o semipermeable. Un soporte puede ser sólido. Un
soporte puede ser semisólido. Un soporte puede ser maleable. Un soporte puede ser flexible. En algunos casos, un

soporte puede moldearse en una forma. Por ejemplo, un soporte puede colocarse sobre un objeto y el soporte puede adoptar la forma del objeto. En algunos casos, el soporte se coloca sobre un órgano y adopta la forma del órgano. En algunos casos, el soporte se produce por impresión 3D.

5 El soporte (por ejemplo, microesferas, nanopartículas) puede ser de al menos aproximadamente 0,1, 0,5, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 100, 500, 1000 o 2000 o más micrómetros de diámetro. Los soportes sólidos (por ejemplo, microesferas) pueden ser como mucho de aproximadamente 0,1, 0,5, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 100, 500, 1000 o 2000 o más micrómetros de diámetro. El diámetro de la microesfera puede ser de aproximadamente 20 micrómetros.

10 En algunos casos un soporte sólido comprende un dendrímero. Un dendrímero puede ser más pequeño que una microesfera. Un dendrímero puede ser subcelular. Un dendrímero puede ser de menos de 1, 0,9, 0,8, 0,7, 0,6, 0,5, 0,4, 0,3, 0,2 o 0,1 micrómetros de diámetro. Un dendrímero puede ser de menos de 0,09, 0,08, 0,07, 0,06, 0,05, 0,04, 0,03 o 0,01 micrómetros de diámetro. Un dendrímero puede comprender tres partes principales, un núcleo, una cubierta interior y una cubierta exterior. Un dendrímero puede sintetizarse para que tenga diferente funcionalidad en cada una de estas partes. La diferente funcionalidad de las partes del dendrímero puede controlar propiedades tales como solubilidad, estabilidad térmica y adhesión de compuestos para aplicaciones particulares. Un dendrímero puede procesarse sintéticamente. Un dendrímero puede sintetizarse por síntesis divergente. La síntesis divergente puede comprender el ensamblaje de un dendrímero a partir de un núcleo multifuncional, que se amplía hacia afuera mediante una serie de reacciones. La síntesis divergente puede comprender una serie de reacciones de Michael. Como alternativa, un dendrímero puede sintetizarse por síntesis convergente. La síntesis convergente puede comprender construir dendrímeros a partir de moléculas pequeñas que finalizan en la superficie de la esfera, y las reacciones pueden proseguir hacia adentro y finalmente se adhieren a un núcleo. Los dendrímeros también pueden prepararse por química clic. La química clic puede comprender reacciones de Diels-Alder, reacciones de tiorino, reacciones de azida-alquino, o una combinación de las mismas. Los ejemplos de dendrímeros incluyen, aunque sin limitación, dendrímero de poli(amidoamina) (PAMAM), dendrímero de PEG-núcleo, dendrímero de fósforo, dendrímero de polipropiliminina y dendrímero de polilisina. Un dendrímero puede ser un dendrímero quirral. Como alternativa, un dendrímero puede ser un dendrímero aquiral.

Un soporte sólido puede comprender una parte de un dendrímero. La parte del dendrímero puede comprender un dendrón. Un dendrón puede comprender secciones de dendrímero con forma de cuña monodispersas con múltiples grupos terminales y una única función de reacción en el punto focal. Un soporte sólido puede comprender un dendrón de poliéster. Los ejemplos de dendrones incluyen, aunque sin limitación, dendrón de poliéster-8-hidroxil-1-acetileno bis-MPA, dendrón de poliéster-16-hidroxil-1-acetileno bis-MPA, dendrón de poliéster-32-hidroxil-1-acetileno bis-MPA, dendrón de poliéster-8-hidroxil-1-carboxilo bis-MPA, dendrón de poliéster-16-hidroxil-1-carboxilo bis-MPA y dendrón de poliéster-32-hidroxil-1-carboxilo bis-MPA.

Un soporte sólido puede comprender un polímero hiperramificado. Un polímero hiperramificado puede comprender macromoléculas dendríticas polidispersas que poseen propiedades de tipo dendrímero. A menudo, los polímeros hiperramificados se preparan en una única etapa de polimerización sintética. El polímero hiperramificado puede estar basado en monómero de ácido 2,2-bis(hidroximetil)propanoico (bis-MPA). Los ejemplos de polímeros hiperramificados incluyen, aunque sin limitación, bis-MPA poliéster-16-hidroxilo hiperramificado, bis-MPA poliéster-32-hidroxilo hiperramificado y bis-MPA poliéster-64-hidroxilo hiperramificado.

El soporte sólido puede ser una matriz o micromatriz. El soporte sólido puede comprender regiones concretas. El soporte sólido puede ser una matriz abordable. En algunos casos, la matriz comprende una pluralidad de sondas fijadas en una superficie sólida. La pluralidad de sondas posibilita la hibridación de la molécula marcada y/o el amplicón marcado a la superficie sólida. La pluralidad de sondas comprende una secuencia que es complementaria al menos una parte de la molécula marcada y/o el amplicón marcado. En algunos casos, la pluralidad de sondas comprende una secuencia que es complementaria a al menos una parte de la marca de muestra, el marcador identificador molecular, el ácido nucleico o una combinación de los mismos. En otros casos, la pluralidad de sondas comprende una secuencia que es complementaria a la unión formada por la adhesión de la marca de muestra o el marcador identificador molecular al ácido nucleico.

La matriz puede comprender una o más sondas. Las sondas pueden estar en una diversidad de formatos. La matriz puede comprender una sonda que comprende una secuencia que es complementaria a al menos una parte del ácido nucleico diana y una secuencia que es complementaria a la región identificadora única de una marca de muestra o marcador identificador molecular, en el que la marca de muestra o marcador identificador molecular comprende un oligonucleótido. La secuencia que es complementaria a al menos una parte del ácido nucleico diana puede adherirse a la matriz. La secuencia que es complementaria a la región de identificador único puede adherirse a la matriz. La matriz puede comprender una primera sonda que comprende una secuencia que es complementaria a al menos una parte del ácido nucleico diana y una segunda sonda que es complementaria a la región de identificador único. Hay varias maneras en que el ácido nucleico marcado de forma estocástica puede hibridar con las matrices. Por ejemplo, la unión de la región de identificador único y el ácido nucleico diana del ácido nucleico marcado de forma estocástica puede hibridar con la sonda en la matriz. Puede haber un hueco en las regiones del ácido nucleico marcado de forma estocástica que puede hibridar con la sonda en la matriz. Diferentes regiones del ácido nucleico marcado de forma estocástica pueden hibridar con dos o más sondas en la matriz. Por tanto, las sondas de la matriz pueden estar en muchos formatos diferentes. Las sondas de la matriz pueden comprender una secuencia que es complementaria a una región de identificar único, una secuencia que es complementaria al ácido nucleico diana o una combinación de los mismos. La hibridación del ácido nucleico marcado de forma estocástica con la matriz puede producirse mediante una diversidad de maneras. Por ejemplo, dos o más nucleótidos del ácido nucleico marcado de

ES 2 711 168 T3

forma estocástica pueden hibridar con una o más sondas en la matriz. Los dos o más nucleótidos del ácido nucleico marcado de forma estocástica que hibridan con las sondas pueden ser nucleótidos consecutivos, nucleótidos no consecutivos o una combinación de los mismos. El ácido nucleico marcado de forma estocástica que hibrida con la sonda puede detectarse mediante cualquier método conocido en la técnica. Por ejemplo, los ácidos nucleicos marcados de forma estocástica pueden detectarse directamente. La detección directa del ácido nucleico marcado de forma estocástica puede comprender la detección de un fluoróforo, hapteno o marcador detectable. Las moléculas marcadas de forma estocástica pueden detectarse indirectamente. La detección indirecta del ácido nucleico marcado de forma estocástica puede comprender el ligamiento u otros métodos enzimáticos o no enzimáticos.

La matriz puede estar en una diversidad de formatos. Por ejemplo, la matriz puede estar en un formato de 16, 32, 48, 64, 80, 96, 112, 128, 144, 160, 176, 192, 208, 224, 240, 256, 272, 288, 304, 320, 336, 352, 368, 384 o 400. Como alternativa, la matriz está en un formato de 8x60K, 4x180K, 2x400K, 1x1M. En otros casos, la matriz está en un formato de 8x15K, 4x44K, 2x105K, 1x244K.

La matriz puede comprender una única matriz. La única matriz puede estar en un único sustrato. Como alternativa, la matriz está en múltiples sustratos. La matriz puede comprender múltiples formatos. La matriz puede comprender una pluralidad de matrices. La pluralidad de matrices puede comprender dos o más matrices. Por ejemplo, la pluralidad de matrices puede comprender al menos aproximadamente 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95 o 100 matrices. En algunos casos al menos dos matrices de la pluralidad de matrices son idénticas. Como alternativa, al menos dos matrices de la pluralidad de matrices son diferentes.

En algunos casos, la matriz comprende zonas con cámara simétricas. Por ejemplo, la matriz comprende zonas con cámara de 0,5 x 0,5 milímetros (mm), 1 x 1 mm, 1,5 x 1,5 mm, 2 x 2 mm, 2,5 x 2,5 mm, 3 x 3 mm, 3,5 x 3,5 mm, 4 x 4 mm, 4,5 x 4,5 mm, 5 x 5 mm, 5,5 x 5,5 mm, 6 x 6 mm, 6,5 x 6,5 mm, 7 x 7 mm, 7,5 x 7,5 mm, 8 x 8 mm, 8,5 x 8,5 mm, 9 x 9 mm, 9,5 x 9,5 mm, 10 x 10 mm, 10,5 x 10,5 mm, 11 x 11 mm, 11,5 x 11,5 mm, 12 x 12 mm, 12,5 x 12,5 mm, 13 x 13 mm, 13,5 x 13,5 mm, 14 x 14 mm, 14,5 x 14,5 mm, 15 x 15 mm, 15,5 x 15,5 mm, 16 x 16 mm, 16,5 x 16,5 mm, 17 x 17 mm, 17,5 x 17,5 mm, 18 x 18 mm, 18,5 x 18,5 mm, 19 x 19 mm, 19,5 x 19,5 mm o 20 x 20 mm. En algunos casos la matriz comprende zonas con cámara de 6,5 x 6,5 mm. Como alternativa, la matriz comprende zonas con cámara asimétricas. Por ejemplo, la matriz comprende zonas con cámara de 6,5 x 0,5 mm, 6,5 x 1 mm, 6,5 x 1,5 mm, 6,5 x 2 mm, 6,5 x 2,5 mm, 6,5 x 3 mm, 6,5 x 3,5 mm, 6,5 x 4 mm, 6,5 x 4,5 mm, 6,5 x 5 mm, 6,5 x 5,5 mm, 6,5 x 6 mm, 6,5 x 6,5 mm, 6,5 x 7 mm, 6,5 x 7,5 mm, 6,5 x 8 mm, 6,5 x 8,5 mm, 6,5 x 9 mm, 6,5 x 9,5 mm, 6,5 x 10 mm, 6,5 x 10,5 mm, 6,5 x 11 mm, 6,5 x 11,5 mm, 6,5 x 12 mm, 6,5 x 12,5 mm, 6,5 x 13 mm, 6,5 x 13,5 mm, 6,5 x 14 mm, 6,5 x 14,5 mm, 6,5 x 15 mm, 6,5 x 15,5 mm, 6,5 x 16 mm, 6,5 x 16,5 mm, 6,5 x 17 mm, 6,5 x 17,5 mm, 6,5 x 18 mm, 6,5 x 18,5 mm, 6,5 x 19 mm, 6,5 x 19,5 mm o 6,5 x 20 mm.

La matriz puede comprender manchas de al menos aproximadamente 1 micrómetro (μm), 2 μm , 3 μm , 4 μm , 5 μm , 6 μm , 7 μm , 8 μm , 9 μm , 10 μm , 15 μm , 20 μm , 25 μm , 30 μm , 35 μm , 40 μm , 45 μm , 50 μm , 55 μm , 60 μm , 65 μm , 70 μm , 75 μm , 80 μm , 85 μm , 90 μm , 95 μm , 100 μm , 125 μm , 150 μm , 175 μm , 200 μm , 225 μm , 250 μm , 275 μm , 300 μm , 325 μm , 350 μm , 375 μm , 400 μm , 425 μm , 450 μm , 475 μm o 500 μm . En algunos casos, la matriz comprende manchas de 70 μm .

La matriz puede comprender inclinación de característica de al menos aproximadamente 1 μm , 2 μm , 3 μm , 4 μm , 5 μm , 6 μm , 7 μm , 8 μm , 9 μm , 10 μm , 15 μm , 20 μm , 25 μm , 30 μm , 35 μm , 40 μm , 45 μm , 50 μm , 55 μm , 60 μm , 65 μm , 70 μm , 75 μm , 80 μm , 85 μm , 90 μm , 95 μm , 100 μm , 125 μm , 150 μm , 175 μm , 200 μm , 225 μm , 250 μm , 275 μm , 300 μm , 325 μm , 350 μm , 375 μm , 400 μm , 425 μm , 450 μm , 475 μm , 500 μm , 525 μm , 550 μm , 575 μm , 600 μm , 625 μm , 650 μm , 675 μm , 700 μm , 725 μm , 750 μm , 775 μm , 800 μm , 825 μm , 850 μm , 875 μm , 900 μm , 925 μm , 950 μm , 975 μm , 1000 μm . En algunos casos la matriz comprende una inclinación de característica de 161 μm .

La matriz puede comprender una o más sondas. En algunos casos, la matriz comprende al menos aproximadamente 5, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90 o 100 sondas. Como alternativa, la matriz comprende al menos aproximadamente 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2300, 2400, 2500, 2600, 2700, 2800, 2900 o 3000 sondas. La matriz puede comprender al menos aproximadamente 3500, 4000, 4500, 5000, 5500, 6000, 6500, 7000, 7500, 8000, 8500, 9000, 9500 o 10 000 sondas. En algunos casos la matriz comprende al menos aproximadamente 960 sondas. Como alternativa, la matriz comprende al menos aproximadamente 2780 sondas. Las sondas pueden ser específica para la pluralidad de marcas oligonucleotídicas. Las sondas pueden ser específicas para al menos una parte de la pluralidad de marcas oligonucleotídicas. Las sondas pueden ser específicas para al menos aproximadamente un 5 %, 10 %, 20 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 %, 95 %, 97 % o 100 % del número total de la pluralidad de marcas oligonucleotídicas. Como alternativa, las sondas son específicas para al menos aproximadamente un 5 %, 10 %, 20 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 %, 95 %, 97 % o 100 % del número total de marcas oligonucleotídicas diferentes de la pluralidad de marcas oligonucleotídicas. Las sondas pueden ser oligonucleótidos. Los oligonucleótidos pueden ser de al menos aproximadamente 1, 2, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100 nucleótidos de longitud. En otros casos las sondas son sondas no específicas. Por ejemplo, las sondas pueden ser específicas para un marcador detectable que se adhiere a la molécula marcada. La sonda puede ser estreptavidina.

La matriz puede ser una matriz impresa. En algunos casos, la matriz impresa comprende uno o más oligonucleótidos adheridos a un sustrato. Por ejemplo, la matriz impresa comprende oligonucleótidos modificados con 5' amina adheridos a un sustrato de epoxi silano.

Como alternativa, la matriz comprende un portaobjetos con uno o más pocillos. El portaobjetos puede comprender al menos aproximadamente 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95 o 100 pocillos. Como alternativa, el portaobjetos comprende al menos aproximadamente 125, 150, 175, 200, 250, 300, 350, 400, 450, 500, 650, 700, 750, 800, 850, 900, 950 o 1000 pocillos. En algunos casos, el portaobjetos comprende 16 pocillos. Como alternativa, el portaobjetos comprende 96 pocillos. En otros casos, el portaobjetos comprende al menos aproximadamente 80, 160, 240, 320, 400, 480, 560, 640, 720, 800, 880 o 960 pocillos.

En algunos casos, el soporte sólido es una matriz de marca Affymetrix 3K, una micromatriz impresa sin contacto Arrayjet o matriz Applied Microarrays Inc (AMI). Como alternativa, el soporte comprende una impresora de contacto, impresora de impacto, impresora de puntos o impresora de clavijas.

El soporte sólido puede comprender el uso de microesferas que se autoensamblan en micropocillos. Por ejemplo, el soporte sólido comprende la tecnología BeadArray de Illumina. Como alternativa, el soporte sólido comprende la tecnología BeadArray de Abbott Molecular y el sistema FlexiPlex TM de Applied Microarray.

En otros casos, el soporte sólido es una placa. Ejemplos de placas incluyen, aunque sin limitación, placas de multimatriz de MSD, placas Multi-Spot® de MSD, microplacas, microplaca ProteOn, AlphaPlate, placa DELFIA, IsoPlate y Luma-Plate.

El método puede comprender además adherir al menos uno de una pluralidad de ácidos nucleicos marcados a un soporte. El soporte puede comprender una pluralidad de microesferas. El soporte puede comprender una matriz. El soporte puede comprender un portaobjetos de vidrio.

El portaobjetos de vidrio puede comprender uno o más pocillos. El uno o más pocillos pueden estar grabados en el portaobjetos de vidrio. El uno o más pocillos puede comprender al menos 960 pocillos. El portaobjetos de vidrio puede comprender una o más sondas. La una o más sondas pueden estar impresas en el portaobjetos de vidrio. El uno o más pocillos pueden comprender además una o más sondas. La una o más sondas pueden estar impresas dentro del uno o más pocillos. La una o más sondas pueden comprender 960 ácidos nucleicos.

Los métodos y kits divulgados en la presente memoria pueden comprender además la distribución de la pluralidad de primeras marcas de muestra, la pluralidad de segundas marcas de muestra, la pluralidad de marcadores identificadores moleculares o cualquier combinación de los mismos en una placa de micropocillos. Los métodos y kits divulgados en la presente memoria pueden comprender además la distribución de una o más microesferas en la placa de micropocillos. Los métodos y kits divulgados en la presente memoria pueden comprender además la distribución de la pluralidad de muestras en una pluralidad de pocillos de una placa de micropocillos. La una o más de la pluralidad de muestras pueden comprender una pluralidad de células. Una o más de la pluralidad de muestras pueden comprender una pluralidad de ácidos nucleicos. El método puede comprender además la distribución de una o menos células a la pluralidad de pocillos. La pluralidad de células puede lisarse en la placa de micropocillos. El método puede comprender además sintetizar ADNc en la placa de micropocillos. La síntesis del ADNc puede comprender transcripción inversa de ARNm. La placa de micropocillos puede comprender una placa de micropocillos fabricada en PDMS por litografía blanda, grabarse en una oblea de silicio, grabarse en un portaobjetos de vidrio, estampado fotoprotector en un portaobjetos de vidrio o una combinación de los mismos. El micropocillos puede comprender un orificio en una placa microcapilar. La placa de micropocillos puede comprender una emulsión de agua en aceite. La placa de micropocillos puede comprender al menos uno o más pocillos. La placa de micropocillos puede comprender al menos aproximadamente 6 pocillos, 12 pocillos, 48 pocillos, 96 pocillos, 384 pocillos, 960 pocillos o 1000 pocillos.

Los métodos y kits pueden comprender además un chip. La placa de micropocillos puede adherirse al chip. El chip puede comprender al menos aproximadamente 6 pocillos, 12 pocillos, 48 pocillos, 96 pocillos, 384 pocillos, 960 pocillos, 1000 pocillos, 2000 pocillos, 3000 pocillos, 4000 pocillos, 5000 pocillos, 6000 pocillos, 7000 pocillos, 8000 pocillos, 9000 pocillos, 10 000 pocillos, 20 000 pocillos, 30 000 pocillos, 40 000 pocillos, 50 000 pocillos, 60 000 pocillos, 70 000 pocillos, 80 000 pocillos, 90 000 pocillos, 100 000 pocillos, 200 000 pocillos, 500 000 pocillos o un millón de pocillos. Los pocillos pueden comprender un área de al menos aproximadamente 300 μm^2 , 400 μm^2 , 500 μm^2 , 600 μm^2 , 700 μm^2 , 800 μm^2 , 900 μm^2 , 1000 μm^2 , 1100 μm^2 , 1200 μm^2 , 1300 μm^2 , 1400 μm^2 , 1500 μm^2 . El método puede comprender además la distribución entre aproximadamente 10 000 y 30 000 muestras en el chip.

Superficies funcionalizadas y oligonucleótidos

La microesfera puede comprender una superficie funcionalizada. Una superficie funcionalizada puede referirse a una superficie del soporte sólido que comprende un grupo funcional. Un grupo funcional puede ser un grupo que puede formar una adhesión con otro grupo funcional. Por ejemplo, un grupo funcional puede ser biotina, que puede formar una adhesión con estreptavidina, otro grupo funcional. Los grupos funcionales ejemplares pueden incluir, aunque sin limitación, aldehídos, cetonas, grupos carboxi, grupos amino, biotina, estreptavidina, ácidos nucleicos, moléculas pequeñas (por ejemplo, para química clic), reactivos homo- y heterobifuncionales (por ejemplo, aminobenzoato de N-succinimidil(4-yodoacetilo) (SIAB), dimaleimida, ácido ditio-bis-nitrobenzoico (DTNB), tioacetato de N-succinimidil-S-acetilo (SATA), N-succinimidil-3-(2-pirilditio)propionato (SPDP), 4-(N-maleimidometil)-ciclohexano-1-carboxilato de succinimidilo (SMCC) y 6-hidracinonicotimida (HYNIC) y anticuerpos. En algunos casos, el grupo funcional es un grupo carboxi (por ejemplo, COOH).

Pueden adherirse oligonucleótidos (por ejemplo, ácidos nucleicos) a soportes sólidos funcionalizados. Los oligonucleótidos inmovilizados en soportes sólidos o estructuras similares pueden servir como sondas de ácidos

nucleico, y pueden realizarse ensayos de hibridación en los que pueden detectarse ácidos nucleicos diana específicos en muestras biológicas complejas.

El soporte sólido (por ejemplo, microesferas) puede funcionalizarse para la inmovilización de oligonucleótidos. Un oligonucleótido puede conjugarse a un soporte sólido mediante enlace amida covalente formado entre el soporte sólido y el oligonucleótido.

Un soporte puede conjugarse a al menos aproximadamente 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900 o 1000 o más oligonucleótidos. Un soporte puede conjugarse a al menos aproximadamente 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 o 1 000 0000, 100 000 000, 500 000 000, 1 000 000 000 o más oligonucleótidos. Un soporte puede conjugarse a al menos aproximadamente 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 o 10 000 000, 100 000 000, 500 000 000, 1 000 000 000 o más oligonucleótidos. Un soporte puede conjugarse a al menos 1 millón de oligonucleótidos. Un soporte puede conjugarse a al menos 10 millones de oligonucleótidos. Un soporte puede conjugarse a al menos 25 millones de oligonucleótidos. Un soporte puede conjugarse al menos 50 millones de oligonucleótidos. Un soporte puede conjugarse al menos 100 millones de oligonucleótidos. Un soporte puede conjugarse al menos 250 millones de oligonucleótidos. Un soporte puede conjugarse al menos 500 millones de oligonucleótidos. Un soporte puede conjugarse al menos 750 millones de oligonucleótidos. Un soporte puede conjugarse al menos aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 o 15 mil millones de oligonucleótidos. Un soporte puede conjugarse a al menos 1000 millones de oligonucleótidos. Un soporte puede conjugarse a al menos 5000 millones de oligonucleótidos.

Los oligonucleótidos pueden adherirse al soporte (por ejemplo, microesferas, polímeros, geles) mediante un conector. La conjugación puede comprender adhesión covalente o no covalente. La conjugación puede introducir un espaciador variable entre las microesferas y los ácidos nucleicos. El conector entre el soporte y el oligonucleótido puede ser escindible (por ejemplo, conector fotoescindible, conector inestable en ácido, conector sensible al calor y conector escindible enzimáticamente).

Los agentes de reticulación para su uso para conjugar moléculas a soportes pueden incluir agentes que pueden reaccionar con un grupo funcional presente en una superficie del soporte sólido y con un grupo funcional presente en la molécula. Los reactivos con capacidad de dicha reactividad pueden incluir aldehídos, acetonas, grupos carboxi, grupos amino, biotina, estreptavidina, ácidos nucleicos, moléculas pequeñas (por ejemplos para química clic), reactivos homo- y heterobifuncionales (por ejemplo, aminobenzoato de N-succinimidil(4-yodoacetilo) (SIAB), dimaleimida, ácido ditio-bis-nitrobenzoico (DTNB), tioacetato de N-succinimidil-S-acetilo (SATA), N-succinimidil-3-(2-pirilditio)propionato (SPDP), 4-(N-maleimidometil)-ciclohexano-1-carboxilato de succinimidilo (SMCC) y 6-hidracinonicotimida (HYNIC).

Una microesfera puede funcionalizarse con un grupo funcional carboxi y un oligonucleótido puede funcionalizarse con un grupo funcional amino.

Un soporte puede ser liso. Como alternativa, o adicionalmente, un soporte puede comprender terrones, crestas o pocillos. Un soporte puede comprender una matriz de micropocillos. Una matriz de micropocillos puede funcionalizarse con grupos funcionales que facilitan la adhesión de oligonucleótidos. Los grupos funcionales en la matriz de micropocillos pueden ser diferentes para diferentes posiciones en la matriz de micropocillos. Los grupos funcionales en la matriz de micropocillos pueden ser iguales para todas las regiones de la matriz de micropocillos.

Componentes del sistema de ensayo

Matrices de micropocillos

Como se describe anteriormente, se usan matrices de micropocillos para atrapar células individuales y microesferas (una microesfera por célula) dentro de una pequeña cámara de reacción de volumen definido. Cada microesfera comprende una colección de sondas oligonucleotídicas para su uso en marcaje estocástico y recuento digital del complemento completo de moléculas de ARNm celulares, que se liberan tras la lisis de la célula. En una realización de la presente divulgación, las matrices de micropocillos son un componente consumible del sistema de ensayo. En otras realizaciones, las matrices de micropocillos pueden ser reutilizables. En cualquier caso, pueden configurarse para usarse como un dispositivo independiente para su uso en la realización de ensayos manualmente, o pueden configurarse para que comprendan un componente desmontable o fijo de un instrumento que proporciona automatización completa o parcial del procedimiento de ensayo.

Los micropocillos de la matriz pueden fabricarse en una diversidad de formas y tamaños, que se eligen para optimizar la eficacia de atrapamiento de una célula individual y la microesfera en cada pocillo. Las geometrías de pocillo apropiadas incluyen, aunque sin limitación, cilíndrica, cónica, hemisférica, rectangular o poliédrica (por ejemplo, geometrías tridimensionales compuestas de varias caras planas, por ejemplo, columnas hexagonales, columnas octogonales, pirámides triangulares invertidas, pirámides cuadradas invertidas, pirámides pentagonales invertidas, pirámides hexagonales invertidas o pirámides trucadas invertidas). Los micropocillos pueden comprender una forma que combina dos o más de estas geometrías. Por ejemplo, en una realización, puede ser parcialmente cilíndrica, teniendo el resto la forma de un cono invertido. En otra realización, puede incluir dos cilindros paralelos, uno de diámetro más grande que el otro, que están conectados mediante un canal vertical (es decir, paralelo a los ejes de los cilindros) que se extiende por toda la longitud (profundidad) de los cilindros. En general, el extremo

abierto (o boca) de cada micropocillo estará ubicado en una superficie superior de la matriz de micropocillos, pero en algunas realizaciones las aberturas pueden estar ubicadas en una superficie inferior de la matriz. En general, el extremo cerrado (o fondo) del micropocillo será plano, pero también son posibles superficies curvadas (por ejemplo, convexas o cóncavas). En general, la forma (y tamaño) de los micropocillos se determinará basándose en los tipos de células y/o microesferas a atrapar en los micropocillos.

Las dimensiones de los micropocillos pueden caracterizarse en términos del diámetro y la profundidad del pocillo. Tal como se usa en la presente memoria, el diámetro del micropocillo se refiere al círculo más grande que puede inscribirse dentro de la sección transversal plana de la geometría del micropocillo. En una realización de la presente divulgación, el diámetro de los micropocillos puede variar de aproximadamente 0,1 a aproximadamente 5 veces el diámetro de las células y/o microesferas a atrapar dentro de los micropocillos. En otras realizaciones, el diámetro de los micropocillos es al menos 0,1 veces, al menos 0,5 veces, al menos 1 vez, al menos 2 veces, al menos 3 veces, al menos 4 veces o al menos 5 veces el diámetro de las células y/o microesferas a atrapar dentro de los micropocillos. En otras realizaciones más, el diámetro de los micropocillos es como mucho de 5 veces, como mucho de 4 veces, como mucho de 3 veces, como mucho de 2 veces, como mucho de 1 vez, como mucho de 0,5 veces, o como mucho de 0,1 veces el diámetro de las células y/o microesferas a atrapar dentro de los micropocillos. En una realización, el diámetro de los micropocillos es aproximadamente 2-5 veces el diámetro de las células y/o microesferas a atrapar dentro de los micropocillos. Los expertos en la materia apreciarán que el diámetro de los micropocillos puede estar dentro de cualquier intervalo limitado por cualquiera de estos valores (por ejemplo, de aproximadamente 0,2 veces a aproximadamente 3,5 veces el diámetro de las células y/o microesferas a atrapar dentro de los micropocillos). Como alternativa, el diámetro de los micropocillos puede especificarse en términos de dimensiones absolutas. En una realización de la presente divulgación, el diámetro de los micropocillos puede variar de aproximadamente 5 a aproximadamente 50 micrómetros. En otras realizaciones, el diámetro de los micropocillos es de al menos 5 micrómetros, al menos 10 micrómetros, al menos 15 micrómetros, al menos 20 micrómetros, al menos 25 micrómetros, al menos 30 micrómetros, al menos 35 micrómetros, al menos 40 micrómetros, al menos 45 micrómetros o al menos 50 micrómetros. En otras realizaciones más el diámetro de los micropocillos es como mucho de 50 micrómetros, como mucho de 45 micrómetros, como mucho de 40 micrómetros, como mucho de 35 micrómetros, como mucho de 30 micrómetros, como mucho de 25 micrómetros, como mucho de 20 micrómetros, como mucho de 15 micrómetros, como mucho de 10 micrómetros o como mucho de 5 micrómetros. En una realización, el diámetro de los micropocillos es de aproximadamente 30 micrómetros. Los expertos en la materia apreciarán que el diámetro de los micropocillos puede estar dentro de cualquier intervalo limitado por cualquiera de estos valores (por ejemplo, de aproximadamente 28 micrómetros a aproximadamente 34 micrómetros).

La profundidad del micropocillo se elige para optimizar la eficacia de atrapamiento de la célula y la microesfera mientras se proporciona al mismo tiempo un intercambio eficaz de tampones de ensayo y otros reactivos contenidos dentro de los pocillos. En una realización de la presente divulgación, la profundidad de los micropocillos puede variar de aproximadamente 0,1 a aproximadamente 5 veces el diámetro de las células y/o microesferas a atrapar dentro de los micropocillos. En otras realizaciones, la profundidad de los micropocillos es al menos 0,1 veces, al menos 0,5 veces, al menos 1 vez, al menos 2 veces, al menos 3 veces, al menos 4 veces o al menos 5 veces el diámetro de las células y/o microesferas a atrapar dentro de los micropocillos. En otras realizaciones más, la profundidad de los micropocillos es como mucho de 5 veces, como mucho de 4 veces, como mucho de 3 veces, como mucho de 2 veces, como mucho de 1 vez, como mucho de 0,5 veces o como mucho de 0,1 veces el diámetro de las células y/o microesferas a atrapar dentro de los micropocillos. En una realización, la profundidad de los micropocillos es aproximadamente 2-5 veces el diámetro de las células y/o microesferas a atrapar dentro de los micropocillos. Los expertos en la materia apreciarán que la profundidad de los micropocillos puede estar dentro de cualquier intervalo limitado por cualquiera de estos valores (por ejemplo, de aproximadamente 0,2 veces a aproximadamente 3,5 veces el diámetro de las células y/o microesferas a atrapar dentro de los micropocillos). Como alternativa, el diámetro de los micropocillos puede especificarse en términos de dimensiones absolutas. En una realización de la presente divulgación, la profundidad de los micropocillos puede variar de aproximadamente 10 a aproximadamente 60 micrómetros. En otras realizaciones, la profundidad de los micropocillos es de al menos 10 micrómetros, al menos 20 micrómetros, al menos 25 micrómetros, al menos 30 micrómetros, al menos 35 micrómetros, al menos 40 micrómetros, al menos 50 micrómetros o al menos 60 micrómetros. En otras realizaciones más la profundidad de los micropocillos es como mucho de 60 micrómetros, como mucho de 50 micrómetros, como mucho de 40 micrómetros, como mucho de 35 micrómetros, como mucho de 30 micrómetros, como mucho de 25 micrómetros, como mucho de 20 micrómetros o como mucho de 10 micrómetros. En una realización, la profundidad de los micropocillos es de aproximadamente 30 micrómetros. Los expertos en la materia apreciarán que la profundidad de los micropocillos puede estar dentro de cualquier intervalo limitado por cualquiera de estos valores (por ejemplo, de aproximadamente 24 micrómetros a aproximadamente 36 micrómetros).

Los pocillos de la matriz de micropocillos están dispuestos en una matriz unidimensional, bidimensional o tridimensional, donde pueden conseguirse matrices tridimensionales, por ejemplo, apilando una serie de dos o más matrices bidimensionales (es decir, apilando dos o más sustratos que comprenden matrices de micropocillos). El patrón y espaciado entre los pocillos se elige para optimizar la eficacia de atrapamiento de una célula individual y microesfera en cada pocillo, así como para maximizar el número de pocillos por unidad de área de la matriz. Los pocillos pueden distribuirse de acuerdo con una diversidad de patrones aleatorios o no aleatorios, por ejemplo, pueden distribuirse de forma completamente aleatoria a través de la superficie del sustrato de matriz, o pueden disponerse en una rejilla cuadrada, una rejilla rectangular o una rejilla hexagonal. En una realización de la presente divulgación, la distancia de un centro a otro (o espaciado) entre los pocillos puede variar de aproximadamente 15 micrómetros a aproximadamente 75 micrómetros. En otras realizaciones, el espaciado entre los pocillos es de al menos 15 micrómetros, al menos 20 micrómetros, al menos 25 micrómetros, al menos 30 micrómetros, al menos

ES 2 711 168 T3

35 micrómetros, al menos 40 micrómetros, al menos 45 micrómetros, al menos 50 micrómetros, al menos 55 micrómetros, al menos 60 micrómetros, al menos 65 micrómetros, al menos 70 micrómetros o al menos 75 micrómetros. En otras realizaciones más, el espaciado de los micropocillos es como mucho de 75 micrómetros, como mucho de 70 micrómetros, como mucho de 65 micrómetros, como mucho de 60 micrómetros, como mucho de 55 micrómetros, como mucho de 50 micrómetros, como mucho de 45 micrómetros, como mucho de 40 micrómetros, como mucho de 35 micrómetros, como mucho de 30 micrómetros, como mucho de 25 micrómetros, como mucho de 20 micrómetros o como mucho de 15 micrómetros. En una realización, el espaciado de los micropocillos es de aproximadamente 55 micrómetros. Los expertos en la materia apreciarán que la profundidad de los micropocillos puede estar dentro de cualquier intervalo limitado por cualquiera de estos valores (por ejemplo, de aproximadamente 18 micrómetros a aproximadamente 72 micrómetros).

La matriz de micropocillos puede comprender características de superficie entre los micropocillos que están diseñadas para ayudar a guiar las células y las microesferas al interior de los pocillos y/o para evitar que sedimenten sobre las superficies entre los pocillos. Los ejemplos de características de superficie adecuadas incluyen, aunque sin limitación, características de superficie abovedadas, estriadas o puntiagudas que rodean los pocillos y/o traspasan la superficie entre los pocillos.

El número total de pocillos en la matriz de micropocillos se determina mediante el patrón y espaciado de los pocillos y las dimensiones globales de la matriz. En una realización de la presente divulgación, el número de micropocillos en la matriz puede variar de aproximadamente 96 a aproximadamente 5 000 000 o más. En otras realizaciones, el número de micropocillos en la matriz es de al menos 96, al menos 384, al menos 1536, al menos 5000, al menos 10 000, al menos 25 000, al menos 50 000, al menos 75 000, al menos 100 000, al menos 500 000, al menos 1 000 000 o al menos 5 000 000. En otras realizaciones más, el número de micropocillos en la matriz es como mucho de 5 000 000, como mucho de 1 000 000, como mucho de 75 000, como mucho de 50 000, como mucho de 25 000, como mucho de 10 000, como mucho de 5000, como mucho de 1536, como mucho de 384 o como mucho de 96 pocillos. En una realización, el número de micropocillos en la matriz es de aproximadamente 96. En otra realización, el número de micropocillos es de aproximadamente 150 000. Los expertos en la materia apreciarán que el número de micropocillos en la matriz puede estar dentro de cualquier intervalo limitado por cualquiera de estos valores (por ejemplo, de aproximadamente 100 a 325 000).

Las matrices de micropocillos pueden fabricarse usando cualquiera de varias técnicas de fabricación conocidas para los expertos en la materia. Ejemplos de métodos de fabricación que pueden usarse incluyen, aunque sin limitación, técnicas de micromecanizado masivo tal como fotolitografía y grabado químico en húmedo, grabado con plasma o grabado en profundidad con iones reactivos; micromoldeo y microrrealce; micromecanizado por láser; impresión 3D u otros procesos de fabricación de escritura directa usando materiales curables; y técnicas similares.

Las matrices de micropocillos pueden fabricarse a partir de cualquiera de varios materiales de sustrato conocidos para los expertos en la materia, donde la elección del material depende típicamente de la elección de la técnica de fabricación, y viceversa. Ejemplos de materiales adecuados incluyen, aunque sin limitación, silicio, sílice fundida, vidrio, polímeros (por ejemplo, agarosa, gelatina, hidrogeles, polidimetilsiloxano (PDMS; elastómero), polimetilmetacrilato (PMMA), policarbonato (PC), polipropileno (PP), polietileno (PE), polietileno de alta densidad (HDPE), poliimida, polímeros de olefina cíclicos (COP), copolímeros de olefina cíclicos (COC), tereftalato de polietileno (PET) y resinas epoxi), metales o películas metálicas (por ejemplo, aluminio acero inoxidable, cobre, níquel, cromo y titanio) y similares. Típicamente, se desea un material hidrófilo para la fabricación de las matrices de micropocillos (para potenciar la capacidad de humedecimiento y minimizar la unión no específica de células y otro material biológico), pero también pueden usarse materiales hidrófobos que pueden tratarse o recubrirse (por ejemplo, por tratamiento con plasma de oxígeno, o injerto de una capa superficial de óxido de polietileno). El uso de materiales hidrófilos porosos para la fabricación de la matriz de micropocillos puede ser deseable para facilitar la evacuación/ventilación capilar de las burbujas de aire atrapadas en el dispositivo. En algunas realizaciones, la matriz de micropocillos se fabrican con un adhesivo óptico. En algunas realizaciones, la matriz de micropocillos se fabrica con material tratado con plasma o corona. El uso de materiales tratados con plasma o corona puede hacer que el material sea hidrófilo. En algunas realizaciones, los materiales tratados con plasma o corona, tales como material hidrófilo, pueden ser más estables que los materiales no tratados. En algunas realizaciones, la matriz de micropocillos se fabrica a partir de un único material. En otras realizaciones, la matriz de micropocillos puede comprender dos o más materiales diferentes que se han unido juntos o juntados mecánicamente.

Puede usarse una diversidad de tratamientos superficiales y técnicas de modificación superficial para alterar las propiedades de las superficies de la matriz de micropocillos. Los ejemplos incluyen, aunque sin limitación, tratamientos con plasma de oxígeno para hacer que las superficies de material hidrófobo sean más hidrófilas, el uso de técnicas de grabado en húmedo o seco para alisar (o arrugar) superficies de vidrio y silicio, adsorción y/o injerto de capas de óxido de polietileno u otros polímeros en superficies de sustrato para hacer que sean más hidrófilas y menos propensas a adsorción no específica de biomoléculas y células, el uso de reacciones de silano para injertar grupos funcionales químicamente reactivos para insertar en superficies de silicio y vidrio por lo demás inerte, etc. Pueden usarse técnicas de fotodesprotección para activar selectivamente grupos funcionales químicamente reactivos en ubicaciones específicas en la estructura de la matriz, por ejemplo, puede usarse la adición o activación selectiva de grupos funcionales químicamente reactivos tales como aminas primarias o grupos carboxilo en las paredes interiores de los micropocillos para acoplar covalentemente sondas oligonucleotídicas, péptidos, proteínas u otras biomoléculas a las paredes de los micropocillos. En general, la elección del tratamiento superficial o modificación superficial utilizada dependerá tanto del tipo de propiedad superficial que se desea como del tipo de material a partir del que se hace la matriz de micropocillos.

En algunas realizaciones, puede ser ventajoso sellar las aberturas de los micropocillos durante, por ejemplo, las etapas de lisis celular, para evitar la hibridación cruzada del ácido nucleico diana entre micropocillos adyacentes. El micropocillo puede sellarse usando una tapa tal como un soporte sólido o una microesfera, donde el diámetro de la microesfera es más grande que el diámetro del micropocillo. Por ejemplo, una microesfera usada como tapa puede ser al menos aproximadamente un 10, 20, 30, 40, 50, 60, 70, 80 o 90 % más grande que el diámetro del micropocillo. Como alternativa, una tapa puede ser como mucho aproximadamente un 10, 20, 30, 40, 50, 60, 70, 80 o 90 % más grande que el diámetro del micropocillo.

Una microesfera usada como tapa puede comprender microesferas de dextrano reticuladas (por ejemplo, Sephadex). El dextrano reticulado puede variar de aproximadamente 10 micrómetros a aproximadamente 80 micrómetros. El dextrano reticulado de la microesfera de tapa puede ser de 20 micrómetros a aproximadamente 50 micrómetros. Una tapa puede comprender, por ejemplo, membranas inorgánicas de nanoporos (por ejemplo, óxidos de aluminio), membranas de diálisis, portaobjetos de vidrio, cubreobjetos y/o película de plástico hidrófila (por ejemplo, película recubierta con una película delgada de agarosa hidratada con tampón de lisis).

En algunas realizaciones, la tapa puede permitir que el tampón pase hacia adentro y hacia afuera del micropocillo, evitando al mismo tiempo que las macromoléculas (por ejemplo, ácidos nucleicos) migren al exterior del pocillo. Una macromolécula de al menos aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 13, 14, 15, 16, 17, 18, 19 o 20 o más nucleótidos puede bloquearse para que no migre hacia adentro o hacia afuera del micropocillo por la tapa. Una macromolécula de como mucho aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 13, 14, 15, 16, 17, 18, 19 o 20 o más nucleótidos puede bloquearse para que no migre hacia adentro o hacia afuera del micropocillo por la tapa.

En algunas realizaciones, una matriz de micropocillos sellados puede comprender una única capa de microesferas en la parte superior de los micropocillos. En algunas realizaciones, una matriz de micropocillos sellados puede comprender múltiples capas de microesferas en la parte superior de los micropocillos. Una matriz de micropocillos sellados puede comprender aproximadamente 1, 2, 3, 4, 5 o 6 o más capas de microesferas.

Fijaciones mecánicas

Cuando se realizan ensayos combinados de marcaje estocástico de células individuales/indización molecular de forma manual, es conveniente montar la matriz de micropocillos en una fijación mecánica para crear una cámara de reacción y facilitar el pipeteo o distribución de suspensiones celulares y reactivos de ensayo en la matriz (**FIG. 69 y 70**). En el ejemplo ilustrado en la **FIG. 69**, la fijación acepta una matriz de micropocillos fabricada sobre un sustrato de 1 mm de grosor, y proporciona soporte mecánico en forma de junta de silicona para confinar los reactivos de ensayo en una cámara de reacción que es de 16 mm de anchura x 35 mm de longitud x aproximadamente 4 mm de profundidad, posibilitando de ese modo el uso de 800 microlitros a 1 mililitro de suspensión celular y suspensión de microesferas (que comprende marcadores oligonucleotídicos basados en microesfera) para realizar el ensayo.

La fijación consiste en placas superiores e inferiores rígidas, mecanizadas (por ejemplo, aluminio) y una junta comprimible (por ejemplo, silicona, polidimetilsiloxano) para crear las paredes de la cámara o pocillo. Las características de diseño incluyen: (i) Bordes de abertura biselados y transparencia para objetivos giratorios de microscopio en y fuera de posición según lo necesario (para visualizar la matriz de micropocillos a diferentes aumentos). (ii) Compresión controlada de la junta de silicona para asegurar una formación uniforme y repetible de un sellamiento a prueba filtraciones con el sustrato de matriz de micropocillos. (iii) Presillas de retención para el funcionamiento conveniente. (iv) Un mecanismo de sujeción de ubicación para fijar y ubicar de forma repetible la matriz. (v) Desensamblaje conveniente para la eliminación de la matriz para la retirada de la matriz durante las etapas de aclarado.

Las placas superiores e inferiores pueden fabricarse usando cualquiera de una diversidad de técnicas (por ejemplo, mecanizado convencional, mecanizado CNC, moldeo por inyección, impresión 3D, etc.) usando una diversidad de materiales (por ejemplo, aluminio, aluminio anodizado, acero inoxidable, teflón, polimetilmetacrilato (PMMA), policarbonato (PC) o materiales poliméricos rígidos similares).

La junta de silicona (polidimetilsiloxano; PDMS) puede configurarse para crear múltiples cámaras (véase la **FIG. 71**) para ejecutar controles y experimentos (o experimentos replicados, o múltiples experimentos independientes) en paralelo. La junta se moldea a partir de PDMS o material elastomérico similar usando un molde de teflón que incluye ángulos proyección para las paredes de la junta verticales para proporcionar buenas características de liberación. Como alternativa, pueden mecanizarse moldes a partir de aluminio u otros materiales (por ejemplo, delrin negro, polieterimida (ultem), etc.), y recubrirse con teflón si fuera necesario para proporcionar buenas características de liberación. Los diseños del molde de la junta se invierten, es decir, de modo que la superficie superior de la parte moldeada (es decir, la superficie en la parte de contacto con el portaobjetos de vidrio u oblea de silicio usada para cubrir el molde durante el moldeo) se convierta en la superficie para crear un sellamiento con el sustrato de matriz de micropocillos durante su uso, evitando de ese modo problemas potenciales con la rugosidad de la superficie del molde y la contaminación superficial a la hora de crear una superficie de junta lisa (para asegurar un sellamiento a prueba de filtraciones con el sustrato de matriz) y también proporcionando una elección flexible de materiales de sustrato y la opción de preensamblaje usando el sustrato de matriz de micropocillos como una base durante el moldeo. Los diseños de molde de junta también pueden incluir rebordes de concentración de la fuerza en los límites de las zonas de pocillo, es decir, la mesa o mesas centrales en el molde (que forman el pocillo o los pocillos) tienen rebordes elevados en las ubicaciones que se convertirán en el perímetro del pocillo o pocillos, de modo que una cubierta colocada sobre la parte superior del molde después del rellenado descansa sobre una zona de contacto pequeña en la ubicación precisa donde un buen perfil de borde es crítico para formar un sellamiento a prueba de

filtraciones entre la junta y el sustrato durante su uso.

Sistemas instrumentales

La presente divulgación también incluye sistemas instrumentales y consumibles para dar soporte a la automatización de ensayos combinados de marcaje estocástico de células individuales/indización molecular. Dichos sistemas pueden incluir cartuchos consumibles que incorporan matrices de micropocillos integradas con celdas de flujo, así como la instrumentación necesaria para proporcionar funcionalidad de control y análisis tal como (i) control de fluidos, (ii) control de la temperatura, (iii) mecanismos de distribución y recogida de células y/o microesferas, (iv) mecanismos de lisis celular, (v) capacidad de imágenes y (vi) procesamiento de imágenes. En algunas realizaciones, la entrada para el sistema comprende una muestra celular y la salida comprende una suspensión de microesferas que comprenden microesferas que tienen adheridos oligonucleótidos que incorporan marcas de muestra, marcas celulares y marcas de indización molecular. En otras realizaciones, el sistema puede incluir funcionalidad adicional, tal como capacidad de termociclado para realizar amplificación por PCR, en cuyo caso la entrada para el sistema comprende una muestra celular y la salida comprende una colección de oligonucleótidos resultante de la amplificación de los oligonucleótidos que incorporan marcas de muestra, marcas celulares y marcas de indización molecular que se adherieron originalmente a las microesferas. En otras realizaciones más, el sistema también puede incluir capacidad de secuenciación, con o sin la necesidad de amplificación de oligonucleótidos, en cuyo caso la entrada para el sistema es una muestra celular y la salida comprende un conjunto de datos que comprende además las secuencias de todas las marcas de muestra, marcas celulares y marcas de indización molecular asociadas con las secuencias diana de interés.

20 *Celdas de flujo de la matriz de micropocillos*

En muchas realizaciones del sistema de ensayo automatizado, el sustrato de la matriz de micropocillos se compactará dentro de una celda de flujo que proporciona una superficie de contacto conveniente con el resto del sistema de manipulación de fluidos y facilita el intercambio de fluidos, por ejemplo, suspensiones celulares y de microesferas, tampones de lisis, tampones de aclarado, etc., que se suministran a la matriz de micropocillos. Las características de diseño pueden incluir: (i) uno o más accesos de entrada para introducir muestras celulares, suspensiones de microesferas y/u otros reactivos de ensayo; (ii) una o más cámaras de matriz de micropocillos diseñadas para proporcionar relleno uniforme e intercambio eficaz de fluidos minimizando al mismo tiempo los remolinos posteriores o zonas muertas, y (iii) uno o más accesos de salida para suministrar fluidos a un punto de recogida de muestras y/o un depósito de residuos. En algunas realizaciones, el diseño de la celda de flujo puede incluir una pluralidad de cámaras de micromatriz que comunica con una pluralidad de matrices de micropocillos de modo que puede procesarse una o más muestras celulares en paralelo. En algunas realizaciones, el diseño de la celda de flujo puede incluir además características para crear perfiles uniformes de velocidad del flujo, es decir, "flujo de taponado", a través de la anchura de la cámara de matriz para proporcionar un suministro más uniforme de células y microesferas a los micropocillos, por ejemplo, usando una barrera porosa ubicada cerca de la entrada de la cámara y anterior a la matriz de micropocillos como un "difusor de flujo" o dividiendo cada cámara de matriz en varias subsecciones que cubren colectivamente la misma zona de matriz total, pero a través de la que la corriente de fluido de entrada dividida fluye en paralelo. En algunas realizaciones, la celda de flujo puede encerrar o incorporar más de un sustrato de matriz de micropocillos. En algunas realizaciones, el ensamblaje integrado de matriz de micropocillos/celda de flujo puede constituir un componente fijo del sistema. En algunas realizaciones, el ensamblaje de matriz de micropocillos/celda de flujo puede retirarse del instrumento.

En general, las dimensiones de los canales de fluidos y la cámara o cámaras de la matriz en los diseños de celda de flujo se optimizarán para (i) proporcionar suministro uniforme de células y microesferas a la matriz de micropocillos y (ii) minimizar el consumo de muestra y reactivo. En algunas realizaciones, la anchura de los canales de fluidos será entre 50 micrómetros y 20 mm. En otras realizaciones, la anchura de los canales de fluidos puede ser de al menos 50 micrómetros, al menos 100 micrómetros, al menos 200 micrómetros, al menos 300 micrómetros, al menos 400 micrómetros, al menos 500 micrómetros, al menos 750 micrómetros, al menos 1 mm, al menos 2,5 mm, al menos 5 mm, al menos 10 mm o al menos 20 mm. En otras realizaciones más, la anchura de los canales de fluidos puede ser como mucho de 20 mm, como mucho de 10 mm, como mucho de 5 mm, como mucho de 2,5 mm, como mucho de 1 mm, como mucho de 750 micrómetros, como mucho de 500 micrómetros, como mucho de 400 micrómetros, como mucho de 300 micrómetros, como mucho de 200 micrómetros, como mucho de 100 micrómetros o como mucho de 50 micrómetros. En una realización, la anchura de los canales de fluidos es de aproximadamente 2 mm. Los expertos en la materia apreciarán que la anchura de los canales de fluidos puede estar dentro de cualquier intervalo limitado por cualquiera de estos valores (por ejemplo, de aproximadamente 250 micrómetros a aproximadamente 3 mm).

En algunas realizaciones, la profundidad de los canales de fluidos será entre 50 micrómetros y 10 mm. En otras realizaciones, la profundidad de los canales de fluidos puede ser de al menos 50 micrómetros, al menos 100 micrómetros, al menos 200 micrómetros, al menos 300 micrómetros, al menos 400 micrómetros, al menos 500 micrómetros, al menos 750 micrómetros, al menos 1 mm, al menos 1,25 mm, al menos 1,5 mm, al menos 1,75 mm, al menos 2 mm, al menos 2,5 mm, al menos 3 mm, al menos 3,5 mm, al menos 4 mm, al menos 4,5 mm, al menos 5 mm, al menos 5,5 mm, al menos 6 mm, al menos 6,5 mm, al menos 7 mm, al menos 7,5 mm, al menos 8 mm, al menos 8,5 mm, al menos 9 mm o al menos 9,5 mm. En otras realizaciones, la profundidad de los canales de fluidos puede ser como mucho de 10 mm, como mucho de 9,5 mm, como mucho de 9 mm, como mucho de 8,5 mm, como mucho de 8 mm, como mucho de 7,5 mm, como mucho de 7 mm, como mucho de 6,5 mm, como mucho de 6 mm, como mucho de 5,5 mm, como mucho de 5 mm, como mucho de 4,5 mm, como mucho de 4 mm, como mucho de 3,5 mm, como mucho de 3 mm, como mucho de 2 mm, como mucho de 1,75 mm, como mucho de

1,5 mm, como mucho de 1,25 mm, como mucho de 1 mm, como mucho de 750 micrómetros, como mucho de 500 micrómetros, como mucho de 400 micrómetros, como mucho de 300 micrómetros, como mucho de 200 micrómetros, como mucho de 100 micrómetros o como mucho de 50 micrómetros. En una realización, la profundidad de los canales de fluidos es de aproximadamente 1 mm. Los expertos en la materia apreciarán que la profundidad de los canales de fluidos puede estar dentro cualquier intervalo limitado por cualquiera de estos valores (por ejemplo, de aproximadamente 800 micrómetros a aproximadamente 1 mm).

Las celdas de flujo pueden fabricarse usando una diversidad de técnicas y materiales conocidos para los expertos en la materia. En general, la celda de flujo se fabricará como una parte separada y posteriormente se fijará mecánicamente o se unirá permanentemente al sustrato de matriz de micropocillos. Los ejemplos de técnicas de fabricación adecuadas incluyen mecanizado convencional, mecanizado CNC, moldeo por inyección, impresión 3D, alineación y laminación de una o más capas de películas poliméricas cortadas con láser o troquel, o cualquiera de varias técnicas de microfabricación tales como fotolitografía y grabado químico en húmedo, grabado en seco, grabado en profundidad por iones reactivos o micromecanizado por láser. Una vez que se ha fabricado la parte de celda de flujo puede adherirse al sustrato de matriz de micropocillos mecánicamente, por ejemplo, fijándola contra el sustrato de matriz de micropocillos (con o sin el uso de una junta), o puede unirse directamente al sustrato de matriz de micropocillos usando cualquiera de una diversidad de técnicas (dependiendo de la elección de materiales usados) conocidas para los expertos en la materia, por ejemplo, mediante el uso de unión anódica, unión térmica, soldadura ultrasónica o cualquiera de una diversidad de adhesivos o películas adhesivas, incluyendo adhesivos basados en epoxi, basados en acrílico, basados en silicona, curables por UV, basados en poliuretano o basados en cianoacrilato.

Las celdas de flujo pueden fabricarse usando una diversidad de materiales conocidos para los expertos en la materia. Los ejemplos de materiales adecuados incluyen, aunque sin limitación, silicio, sílice fundida, vidrio, cualquiera de una diversidad de polímeros, por ejemplo, polidimetilsiloxano (PDMS; elastómero), polimetilmetacrilato (PMMA), policarbonato (PC) polipropileno (PP), polietileno (PE) polietileno de alta densidad (HDPE) poliimida, polímeros de olefina cíclicos (COP), copolímeros de olefina cíclicos (COC), tereftalato de polietileno (PET), resinas epoxi, metales (por ejemplo, aluminio, acero inoxidable, cobre, níquel, cromo y titanio) o una combinación de estos materiales.

Cartuchos

En muchas realizaciones del sistema de ensayo automatizado, la matriz de micropocillos, con o sin una celda de flujo adherida, se compactará dentro de un cartucho consumible que comunica con el sistema instrumental y que puede incorporar funcionalidad adicional. Las características de diseño de los cartuchos pueden incluir (i) uno o más accesos de entrada para crear conexiones fluidas con el instrumento y/o introducir manualmente muestras celulares, suspensiones de microesferas y/u otros reactivos de ensayo en el cartucho, (ii) uno o más canales de desviación, es decir, para automedir las muestras celulares y las suspensiones de microesferas para evitar el sobrellenado y/o retroflujo, (iii) uno o más ensamblajes de matriz de micropocillos/celda de flujo integrados, o una o más cámaras dentro de la que se ubica el sustrato o sustratos de micromatriz, (iv) bombas en miniatura integradas u otros mecanismos de accionamiento de fluidos para controlar el flujo de fluidos a través del dispositivo, (v) válvulas en miniatura integradas para la compartimentación de reactivos precargados y/o el control del flujo de fluidos a través del dispositivo, (vi) ventilaciones para proporcionar una trayectoria de escape para el aire atrapado, (vii) uno o más depósitos de muestra y/o residuos de reactivo, (viii) uno o más accesos de salida para crear conexiones fluidas con el instrumento y/o proporcionar un punto de recogida de muestra procesada, (ix) características de conexión mecánica para ubicar de forma reproducible el cartucho desmontable, consumible con respecto al sistema instrumental y para proporcionar acceso de modo que imanes externos puedan ponerse en cercana proximidad con la matriz de micropocillos, (x) componentes integrados de control de la temperatura y/o una conexión térmica para proporcionar un buen contacto térmico con el sistema instrumental y (xi) características de conexión óptica, por ejemplo, una ventana transparente, para su uso en consulta óptica de la matriz de micropocillos. En algunas realizaciones, el cartucho se diseña para procesar más de una muestra en paralelo. En algunas realizaciones del dispositivo, el cartucho puede comprender además una o más cámaras de recogida de muestra desmontable que son adecuadas para conectar con termocicladores de PCR y/o instrumentos de secuenciación independientes. En algunas realizaciones del dispositivo, el propio cartucho es adecuado para conectar con termocicladores de PCR y/o instrumentos de secuenciación independientes.

En algunas realizaciones del dispositivo, el cartucho puede comprender además componentes que están diseñados para crear barreras físicas y/o química para evitar la difusión de (o aumento de las longitudes de trayectoria y tiempos de difusión para) moléculas grandes para minimizar la contaminación cruzada entre micropocillos. Los ejemplos de dichas barreras incluyen, aunque sin limitación, un patrón de canales en serpentin usados para el suministro de células y microesferas a la matriz de micropocillos, una platina retráctil o membrana deformable que se presiona en contacto con la superficie del sustrato de matriz de micropocillos durante las etapas de lisis o incubación, el uso de microesferas más grandes, por ejemplo, microesferas de Sephadex como se describe previamente, para bloquear las aberturas de los micropocillos o la liberación de un fluido hidrófobo inmiscible desde un depósito dentro del cartucho durante las etapas de lisis o incubación, para separar y compartimentar de forma eficaz cada micropocillo en la matriz. Todas y cada una de estas barreras, o una realización sin dichas barreras, pueden combinarse con la elevación de la viscosidad de la disolución en y adyacente a los micropocillos, por ejemplo, mediante la adición de componentes de disolución tales como glicerol o polietilenglicol.

En general, las dimensiones de los canales de fluidos y la cámara o cámaras de la matriz en los diseños de cartucho se optimizarán para (i) proporcionar suministro uniforme de células y microesferas a la matriz de micropocillos, y (ii)

ES 2 711 168 T3

minimizar el consumo de muestra y reactivo. En algunas realizaciones, la anchura de los canales de fluidos será entre 50 micrómetros y 20 mm. En otras realizaciones, la anchura de los canales de fluidos puede ser de al menos 50 micrómetros, al menos 100 micrómetros, al menos 200 micrómetros, al menos 300 micrómetros, al menos 400 micrómetros, al menos 500 micrómetros, al menos 750 micrómetros, al menos 1 mm, al menos 2,5 mm, al menos 5 mm, al menos 10 mm o al menos 20 mm. En otras realizaciones más, la anchura de los canales de fluidos puede ser como mucho de 20 mm, como mucho de 10 mm, como mucho de 5 mm, como mucho de 2,5 mm, como mucho de 1 mm, como mucho de 750 micrómetros, como mucho de 500 micrómetros, como mucho de 400 micrómetros, como mucho de 300 micrómetros, como mucho de 200 micrómetros, como mucho de 100 micrómetros o como mucho de 50 micrómetros. En una realización, la anchura de los canales de fluidos es de aproximadamente 2 mm. Los expertos en la materia apreciarán que la anchura de los canales de fluidos puede estar dentro de cualquier intervalo limitado por cualquiera de estos valores (por ejemplo, de aproximadamente 250 micrómetros a aproximadamente 3 mm).

En algunas realizaciones, la profundidad de los canales de fluidos en los diseños de cartucho será entre 50 micrómetros y 10 mm. En otras realizaciones, la profundidad de los canales de fluidos puede ser de al menos 50 micrómetros, al menos 100 micrómetros, al menos 200 micrómetros, al menos 300 micrómetros, al menos 400 micrómetros, al menos 500 micrómetros, al menos 750 micrómetros, al menos 1 mm, al menos 1,25 mm, al menos 1,5 mm, al menos 1,75 mm, al menos 2 mm, al menos 2,5 mm, al menos 3 mm, al menos 3,5 mm, al menos 4 mm, al menos 4,5 mm, al menos 5 mm, al menos 5,5 mm, al menos 6 mm, al menos 6,5 mm, al menos 7 mm, al menos 7,5 mm, al menos 8 mm, al menos 8,5 mm, al menos 9 mm o al menos 9,5 mm. En otras realizaciones más, la profundidad de los canales de fluidos puede ser como mucho de 10 mm, como mucho de 9,5 mm, como mucho de 9 mm, como mucho de 8,5 mm, como mucho de 8 mm, como mucho de 7,5 mm, como mucho de 7 mm, como mucho de 6,5 mm, como mucho de 6 mm, como mucho de 5,5 mm, como mucho de 5 mm, como mucho de 4,5 mm, como mucho de 4 mm, como mucho de 3,5 mm, como mucho de 3 mm, como mucho de 2 mm, como mucho de 1,75 mm, como mucho de 1,5 mm, como mucho de 1,25 mm, como mucho de 1 mm, como mucho de 750 micrómetros, como mucho de 500 micrómetros, como mucho de 400 micrómetros, como mucho de 300 micrómetros, como mucho de 200 micrómetros, como mucho de 100 micrómetros o como mucho de 50 micrómetros. En una realización, la profundidad de los canales de fluidos es de aproximadamente 1 mm. Los expertos en la materia apreciarán que la profundidad de los canales de fluidos puede estar dentro de cualquier intervalo limitado por cualquiera de estos valores (por ejemplo, de aproximadamente 800 micrómetros a aproximadamente 1 mm).

Los cartuchos pueden fabricarse usando una diversidad de técnicas y materiales conocidos para los expertos en la materia. En general, los cartuchos se fabricarán como una serie de partes de componentes separados (**FIG. 72**) y posteriormente se ensamblarán (**FIG. 72 y 73**) usando cualquiera de varias técnicas de ensamblaje o unión mecánica. Los ejemplos de técnicas de fabricación adecuadas incluyen, aunque sin limitación, mecanizado convencional, mecanizado CNC, moldeo por inyección, termoformación e impresión 3D. Una vez que se han fabricado los componentes del cartucho, pueden ensamblarse mecánicamente usando tornillos, pinzas y similares, o unirse permanentemente usando cualquiera de una diversidad de técnicas (dependiendo de la elección de los materiales usados), por ejemplo, mediante el uso de unión/soldadura térmica o ultrasónica o cualquiera de una diversidad de adhesivos o películas adhesivas, incluyendo adhesivos basados en epoxi, basados en acrílico, basados en silicona, curables por UV, basados en poliuretano o basados en cianoacrilato.

Los componentes del cartucho pueden fabricarse usando cualquiera de varios materiales adecuados, incluyendo aunque sin limitación, silicio, sílice fundida, vidrio, cualquiera de una diversidad de polímeros, por ejemplo, polidimetilsiloxano (PDMS; elastómero), polimetilmetacrilato (PMMA), policarbonato (PC) polipropileno (PP), polietileno (PE) polietileno de alta densidad (HDPE) poliimida, polímeros de olefina cíclicos (COP), copolímeros de olefina cíclicos (COC), tereftalato de polietileno (PET), resinas epoxi o metales (por ejemplo, aluminio, acero inoxidable, cobre, níquel, cromo y titanio).

Como se describe anteriormente, las características de entrada y salida del cartucho pueden diseñarse para proporcionar conexiones de fluidos convenientes y a prueba de filtraciones con el instrumento, o pueden servir como depósitos abiertos para el pipeteo manual de muestras y reactivos al interior o al exterior del cartucho. Los ejemplos de diseños mecánicos convenientes para los conectores de acceso de entrada y salida incluyen, aunque sin limitación, conectores roscados, conectores estampados, conectores de conexión de Luer, conectores deslizantes de Luer o de "punta deslizante", conectores de ajuste a presión y similares. En algunas realizaciones, los accesos de entrada y salida del cartucho pueden comprender además tapas, cubiertas o cierres con resorte, materiales de cambio de fase o membranas poliméricas que pueden abrirse o perforarse cuando el cartucho se coloca en el instrumento, y que sirven para evitar la contaminación de las superficies internas del cartucho durante el almacenamiento y/o que evitan que los fluidos se derramen cuando el cartucho se retira del instrumento. Como se indica anteriormente, en algunas realizaciones el uno o más accesos de salida del cartucho pueden comprender además una cámara de recogida de muestra desmontable que es adecuada para conectar con termocicladores de PCR y/o instrumentos de secuenciación independientes.

Como se indica anteriormente, en algunas realizaciones el cartucho puede incluir bombas en miniatura integradas u otros mecanismos de accionamiento de fluidos para controlar el flujo de fluidos a través del dispositivo. Los ejemplos de bombas en miniatura adecuadas o mecanismos de accionamiento de fluidos incluyen, aunque sin limitación, mecanismos de jeringa o émbolo en miniatura accionados de forma electromecánica o neumática, propulsores químicos, bombas de diafragma de membrana accionados neumáticamente o mediante un pistón externo, bolsas o vejigas de reactivo accionadas neumáticamente o bombas electroosmóticas.

Como se describe anteriormente, en algunas realizaciones el cartucho puede incluir válvulas en miniatura para compartimentar reactivos precargados y/o controlar el flujo de fluidos a través del dispositivo. Los ejemplos de válvulas en miniatura adecuadas incluyen, aunque sin limitación, "válvulas" de una sola acción fabricadas usando tapones de cera o poliméricos que pueden fundirse o disolverse, o membranas poliméricas que pueden perforarse; 5 válvulas de pellizco construidas usando una membrana deformable y accionamiento neumático, hidráulico, magnético, electromagnético o electromecánico (solenoides), válvulas de una vía construidas usando solapas de membrana deformables y válvulas de compuerta en miniatura.

Como se indica anteriormente, en algunas realizaciones el cartucho puede incluir ventilaciones para proporcionar una trayectoria de escape para el aire atrapado. Las ventilaciones pueden construirse de acuerdo con una 10 diversidad de técnicas conocidas para los expertos en la materia, por ejemplo, usando un tapón poroso de polidimetilsiloxano (PDMS) u otro material hidrófobo que permita la absorción capilar de aire, pero que bloquea la penetración de agua. Las ventilaciones también pueden construirse como aberturas a través de materiales de barrera hidrófoba, de modo que no se produzca humedecimiento de las paredes de la abertura a las presiones usadas durante el funcionamiento.

En general, las características de conexión mecánica del cartucho proporcionan una colocación fácilmente 15 desmontable, pero altamente precisa y repetible del cartucho respecto al sistema instrumental. Las características de superficie de contacto mecánicas adecuadas incluyen, aunque sin limitación, pernos de alineación, guías de alineación, topes mecánicos y similares. En algunas realizaciones, las características de diseño mecánicas incluirán características de relieve para poner un aparato externo, por ejemplo, imanes o componentes ópticos, en cercana 20 proximidad con la cámara de matriz de micropocillos (FIG. 72).

En algunas realizaciones, el cartucho también incluirá componentes de control de la temperatura o característica de 25 conexión térmica para acoplarse a módulos de control de temperatura externos. Ejemplos de elementos de control de la temperatura adecuados incluyen, aunque sin limitación, elementos de calentamiento resistivo, fuentes de luz de emisión de infrarrojos en miniatura, dispositivos de calentamiento o enfriamiento de Peltier, pozos de calor, termistores, termopares y similares. Las características de conexión térmica típicamente se fabricarán a partir de materiales que son buenos conductores térmicos (por ejemplo, cobre, oro, plata, aluminio, etc.) y típicamente comprenderán una o más superficies planas que pueden hacer un buen contacto térmico con los bloques de 30 calentamiento o bloques de enfriamiento externos.

En muchas realizaciones, el cartucho incluirá características de conexión óptica para su uso en imágenes ópticas o 35 consulta espectroscópica de la matriz de micropocillos. Típicamente, el cartucho incluirá una ventana ópticamente transparente, por ejemplo, el propio sustrato de micropocillo o el lateral de la celda de flujo o cámara de micromatriz que es opuesta a la matriz de micropocillos, se fabrica de un material que cumple los requisitos espectrales para la técnica de imágenes o espectroscópica usada para sondear la matriz de micropocillos. Los ejemplos de materiales de ventana óptica adecuados incluyen, aunque sin limitación, vidrio, sílice fusionada, polimetilmetacrilato (PMMA), policarbonato (PC), polímeros de olefina cíclicos (COP) o copolímeros de olefina cíclicos (COC). Típicamente, el 40 cartucho incluirá una segunda ventana o región ópticamente transparente o traslúcida que puede usarse para iluminar la matriz de micropocillos en orientación de iluminación transversal, reflejada u oblicua.

Instrumentos

La presente divulgación también incluye instrumentos para su uso en la automatización de ensayos combinados de 45 marcaje estocástico de células individuales/indización molecular. Como se indica anteriormente, estos instrumentos pueden proporcionar control y análisis de la funcionalidad tal como (i) control de fluidos, (ii) control de la temperatura, (iii) distribución de las células y/o las microesferas y mecanismos de recogida, (iv) mecanismos de lisis celular, (v) control de campo magnético, (vi) capacidad de imágenes y (vii) procesamiento de imágenes. En algunas realizaciones, el sistema instrumental puede comprender uno o más módulos (una posible realización de ello se 50 ilustra esquemáticamente en la FIG. 74), donde cada módulo proporciona uno o más conjuntos de características funcionales específicas para el sistema. En otras realizaciones, el sistema instrumental puede empaquetarse de modo que toda la funcionalidad del sistema resida dentro del mismo paquete. La FIG. 75 proporciona una ilustración esquemática de las etapas del proceso incluidas en una realización del sistema automatizado. Como se indica anteriormente, en algunas realizaciones el sistema puede comprender unidades funcionales adicionales, como 55 componentes integrados o como componentes modulares del sistema, que expanden las capacidades funcionales del sistema para incluir amplificación por PCR (u otros tipos de técnicas de amplificación de oligonucleótidos) y secuenciación de oligonucleótidos.

En general, el sistema instrumental proporcionará capacidad de fluidos para suministrar muestras y/o reactivos a la 60 una o más cámaras de micromatriz o una o más celda de flujo dentro de uno o más cartuchos de ensayo conectados al sistema. Los reactivos de ensayo y tapones pueden almacenarse en frascos, cartuchos de reactivo y tampón u otros recipientes adecuados que se conectan a las entradas del cartucho. El sistema también puede incluir depósitos de residuos en forma de frascos, cartuchos de residuos u otros recipientes de residuos adecuados para recoger los fluidos posteriores del cartucho o cartuchos de ensayo. El control del flujo de fluidos a través del sistema típicamente se realizará mediante el uso de bombas (u otros mecanismos de accionamiento de fluidos) y válvulas. Los ejemplos de bombas adecuadas incluyen, aunque sin limitación, bombas de jeringa, bombas de jeringa programables, bombas peristálticas, bombas de diafragma y similares. En algunas realizaciones, el flujo de fluidos a través del sistema puede controlarse mediante la aplicación de presión neumática positiva en una o más entradas de los recipientes de reactivo y tampón, o en las entradas del cartucho o cartuchos de ensayo. En algunas realizaciones, el flujo de fluidos a través del sistema puede controlarse mediante la extracción de un vacío en la una o más salidas de

los depósitos de residuos, o en las salidas del cartucho o cartuchos de ensayo. Los ejemplos de válvulas adecuadas incluyen, aunque sin limitación, válvulas de retención, válvulas de dos vías o tres vías electromecánicas, válvulas de dos vías y tres vías neumáticas y similares. En algunas realizaciones, el flujo pulsátil puede aplicarse durante las etapas de lavado/aclarado del ensayo para facilitar el intercambio completo y eficaz de fluidos dentro de la una o más celdas de flujo o cámaras de la matriz de micropocillos.

Como se indica anteriormente, en algunas realizaciones el sistema instrumental puede incluir mecanismos para facilitar adicionalmente la distribución uniforme de células y microesferas sobre la matriz de micropocillos. Los ejemplos de dichos mecanismos incluyen, aunque sin limitación, balanceo, agitación, giro, flujo de recirculación, agitación de baja frecuencia (por ejemplo, usando una placa de balanceo o mediante pulsación de una membrana flexible (por ejemplo, silicona) que forma una pared de la cámara o canal de fluidos cercano) o agitación a alta frecuencia (por ejemplo, mediante el uso de transductores piezoeléctricos). En algunas realizaciones, uno o más de estos mecanismos se utilizan en combinación con estructuras o características físicas en las paredes interiores de la celda de flujo o cámara de matriz, por ejemplo, estructuras de entresuelo/sombrero de copa, bandas en forma de V o series de crestas para facilitar la mezcla y/o para ayudar a evitar la combinación de células o microesferas dentro de la cámara de matriz. Pueden usarse varillas de potenciación del flujo en las superficies superior o inferior de la celda de flujo o cámara de matriz para controlar los perfiles de caudal y reducir el corte entre las aberturas de los micropocillos (es decir, para evitar que las células o microesferas se empujen desde los micropocillos durante las etapas de intercambio de reactivo y aclarado).

En algunas realizaciones, el sistema instrumental puede incluir la capacidad de lisis celular mecánica como una alternativa al uso de detergentes u otros reactivos. La sonicación usando un transductor piezoeléctrico de alta frecuencia es un ejemplo de una técnica adecuada.

En algunas realizaciones, el sistema instrumental incluirá funcionalidad de control de la temperatura con el fin de facilitar la precisión y reproducibilidad de los resultados del ensayo, por ejemplo, el enfriamiento de la celda de flujo o cámara de la matriz de micropocillos puede ser ventajosa para minimizar la difusión molecular entre los micropocillos. Los ejemplos de componentes de control de la temperatura que pueden incorporarse en el sistema instrumental incluyen, aunque sin limitación, elementos de calentamiento resistivo, fuentes de luz infrarroja, dispositivos de calentamiento o enfriamiento de Peltier, pozos de calor, termistores, termopares y similares. En algunas realizaciones del sistema, el controlador de la temperatura puede proporcionar cambios programables en la temperatura durante intervalos de tiempo especificados.

Como se indica en otra parte en esta divulgación, muchas realizaciones de los métodos divulgados utilizan campos magnéticos para retirar las microesferas de los micropocillos tras completarse el ensayo. En algunas realizaciones, el sistema instrumental puede comprender además el uso de campos magnéticos para transportar microesferas al interior o al exterior de la celda de flujo o cámara de la matriz de pocillos. Los ejemplos de medios adecuados para proporcionar control de los campos magnéticos incluyen, aunque sin limitación, el uso de electroimanes en una o más posiciones fijas respecto al cartucho, o el uso de imanes permanentes que se reubican mecánicamente según lo necesario. En algunas realizaciones del sistema instrumental, la fuerza del campo o campos magnéticos aplicados se variará variando la cantidad de corriente aplicada al uno o más electroimanes. En algunas realizaciones del sistema instrumental, la fuerza de los campos magnéticos aplicados variará cambiando la posición de uno o más imanes permanentes respecto a la posición de la cámara o cámaras de la micromatriz usando, por ejemplo, accionadores lineales activados por motor de pasos, accionadores lineales activados por servomotor o mecanismos de árbol de levas. En algunas realizaciones del sistema instrumental, el uso de campos magnéticos por impulsos puede ser ventajoso, por ejemplo, para evitar la agrupación de microesferas magnéticas. En algunas realizaciones, un imán en cercana proximidad a la matriz o cámara puede moverse, una vez o múltiples veces entre al menos dos posiciones respecto a la matriz de micropocillos. El movimiento de los imanes puede servir para agitar las microesferas dentro de los micropocillos, para facilitar la retirada de microesferas de los micropocillos o para recoger las microesferas magnéticas en una ubicación deseada.

Como se indica anteriormente, en muchas realizaciones el sistema instrumental incluirá capacidades de imágenes ópticas y/u otras capacidades espectroscópicas. Dicha funcionalidad puede ser útil, por ejemplo, para la inspección de matrices de micropocillos para determinar si la matriz se ha poblado de forma uniforme y óptima o no con células y/o microesferas. Puede utilizarse cualquiera de una diversidad de modos de imágenes, incluyendo, aunque sin limitación, imágenes de campo brillante, de campo oscuro y de fluorescencia/luminiscencia. La elección del modo de imágenes afectará al diseño de las matrices de micropocillos, las celdas de flujo y las cámaras de cartucho por que el sustrato de la matriz y/o la pared opuesta de la celda de flujo o cámara de la matriz necesariamente tiene que ser transparente o traslúcida en el intervalo espectral de interés. En algunas realizaciones, pueden tomarse imágenes de cada matriz de micropocillos en su totalidad dentro de una única imagen. En algunas realizaciones, puede "colocar en mosaico" una serie de imágenes para crear una imagen de alta resolución de la matriz completa. En algunas realizaciones, una única imagen que representa una subsección de la matriz puede usarse para evaluar las propiedades, por ejemplo, distribuciones de células o microesferas, para la matriz como conjunto. En algunas realizaciones, pueden realizarse imágenes de doble longitud de onda de excitación y emisión (o de múltiples longitudes de onda de excitación y/o emisión). Puede usarse cualquiera de una diversidad de fuentes de luz para proporcionar la luz para las imágenes y/o excitación incluyendo, aunque sin limitación, lámparas de tungsteno, lámparas de tungsteno-halógeno, lámparas de arco, láseres, diodos emisores de luz (LED) o diodos de láser. Puede usarse cualquiera de una diversidad de detectores de imágenes con fines de imágenes, incluyendo, aunque sin limitación, series de fotodiodos, cámaras de dispositivo acoplado a carga (CCD) o detectores de imágenes CMOS. El sistema óptico típicamente incluirá una diversidad de componentes ópticos para dirigir, conformar, filtrar y/o enfocar los rayos de luz a través del sistema. Los ejemplos de componentes ópticos adecuados incluyen, aunque sin

limitación, lentes, espejos, prismas, rejillas de difracción, filtros de vidrio coloreados, filtros de interferencia de banda estrecha, filtros de interferencia de banda ancha, reflectores microscópicos, fibras ópticas, guías de onda ópticas y similares. En algunas realizaciones, el sistema instrumental puede usar un sustrato de micromatriz ópticamente transparente como una guía de onda para suministrar luz de excitación a la matriz de micropocillos. La elección del modo de imágenes también puede posibilitar la ejecución de otros tipos de ensayo en paralelo con ensayos de marcaje estocástico/indización molecular, por ejemplo, el uso de ensayos de células vivas/células muertas con azul tripano con imágenes de campo brillante, el uso de ensayos de células vivas/células muertas basados en fluorescencia con imágenes de fluorescencia, etc. La correlación de los datos de viabilidad para células individuales con la marca celular asociada con cada microesfera en el micropocillo asociado puede proporcionar un nivel adicional de discriminación en el análisis de los datos de los ensayos combinados de células individuales. Como alternativa, los datos de viabilidad en forma de estadística para múltiples células pueden emplearse para potenciar las capacidades analíticas y asegurar la calidad del ensayo.

En algunas realizaciones, el sistema puede comprender capacidades que no son de imágenes y/o no son ópticas para sondear la matriz de micropocillos. Los ejemplos de técnicas que no son de imágenes y/o no son ópticas para detectar burbujas de aire atrapadas, determinar la distribución de células y/o microesferas sobre la matriz, etc., incluyen, aunque sin limitación, mediciones de dispersión de luz, mediciones de absorción de ultravioleta/visible/infrarrojos (por ejemplo, usando células teñidas y/o microesferas que incorporan tintes), dispersión de raman coherente y mediciones de conductancia (por ejemplo, usando matrices microfabricadas de electrodos en registro con las matrices de micropocillos).

20 *Procesador del sistema y programa informático*

En general, los sistemas instrumentales diseñados para dar soporte a la automatización de ensayos combinados de marcaje estocástico de células individuales/indización molecular incluirán un procesador u ordenador, junto con un programa informático para proporcionar (i) funcionalidad de control del instrumento, (ii) capacidad de procesamiento y análisis de imágenes y (iii) funcionalidad de almacenamiento de datos, análisis y presentación.

En muchas realizaciones, el sistema instrumental comprenderá un ordenador (o procesador) y medio legible por ordenador que incluye un código para proporcionar una interfaz de usuario, así como control manual, semiautomatizado o completamente automatizado de todas las funciones del sistema, es decir, control del sistema de fluidos, sistema de control de la temperatura, funciones de distribución de células y/o microesferas, funciones de manipulación de las microesferas magnéticas y el sistema de imágenes. Los ejemplos de funciones de control de fluidos proporcionadas por el programa informático de control del instrumento incluyen, aunque sin limitación, caudales volumétricos de fluido, velocidades de flujo de fluidos, la cronología y duración para las etapas de adición de muestra y microesferas, adición de reactivo y aclarado. Los ejemplos de funciones de control de la temperatura proporcionadas por el programa informático de control del instrumento incluyen, aunque sin limitación, especificar uno o más puntos establecidos de la temperatura y control de la cronología, duración y velocidad de rampa para cambios de temperatura. Los ejemplos de funciones de distribución de células y/o microesferas proporcionadas por el programa informático de control del instrumento incluyen, aunque sin limitación, control de los parámetros de agitación tales como amplitud, frecuencia y duración. Los ejemplos de funciones del campo magnético proporcionadas por el programa informático de control del instrumento incluyen, aunque sin limitación, la cronología y duración del campo o campos magnéticos aplicados, y en el caso de electroimanes, la fuerza del campo magnético también. Los ejemplos de funciones de control del sistema de imágenes proporcionadas por el programa informático de control del instrumento incluyen, aunque sin limitación, capacidad de autoenfoco, control de la iluminación y/o tiempos e intensidades de exposición a luz de excitación, control de tasa de adquisición de imágenes, tiempo de exposición y opciones de almacenamiento de datos.

En algunas realizaciones del sistema instrumental, el sistema comprenderá además un medio legible por ordenador que incluye un código para proporcionar capacidad de procesamiento y análisis de imágenes. Los ejemplos de capacidad de procesamiento y análisis de imágenes proporcionada por el programa informático incluyen, aunque sin limitación, ajuste manual, semiautomatizado o completamente automatizado de la exposición de la imagen (por ejemplo, equilibrio de blancos, ajuste del contraste, promediación de la señal y otra capacidad de reducción del ruido, etc.), identificación automatizada de objetos (es decir, para identificar células y microesferas en la imagen), análisis estadístico automatizado (es decir, para determinar el número de células y/o microesferas identificadas por unidad de área de la matriz de micropocillos, o para identificar pocillos que contienen más de una célula o más de una microesfera), y capacidades manuales de medición (por ejemplo, para medir distancias entre objetos, etc.). En algunas realizaciones, el programa informático de control del instrumento y procesamiento/análisis de imágenes estará escrito como módulos de programa informático separados. En algunas realizaciones, el programa informático de control del instrumento y procesamiento/análisis de imágenes estará incorporado en un paquete integrado. En algunas realizaciones, el programa informático del sistema puede proporcionar un análisis de imágenes en tiempo real y control del instrumento integrados, de modo que las etapas de carga de muestra de células y microesferas pueda prolongarse o repetirse hasta que se consigan las distribuciones óptimas de células/microesferas.

En algunas realizaciones del sistema instrumental, el sistema comprenderá un medio legible por ordenador que incluye un código para proporcionar análisis de datos de secuencias. Los ejemplos de funcionalidad de análisis de datos de secuencias que puede proporcionarse por el programa informático de análisis de datos incluyen aunque sin limitación, (i) algoritmos para determinar el número de lecturas por gen por célula y el número de moléculas de transcrito únicas por gen por célula, basado en los datos proporcionados por la secuenciación de la colección de oligonucleótidos creada ejecutando el ensayo, (ii) análisis estadístico de los datos de secuenciación, por ejemplo, análisis de componentes principales, para predecir los intervalos de confianza para determinaciones del número de

moléculas de transcrito por gen por célula, etc., (iii) capacidades de alineación de secuencias para la alineación de datos de secuencias génicas con secuencias de referencia conocidas, (iv) descodificación/separación de códigos de barras de muestra, códigos de barras celulares y códigos de barras moleculares y (v) agrupación automatizada de marcadores moleculares para compensar los errores de amplificación o secuenciación.

5 En general, el ordenador o procesador incluido en los sistemas instrumentales en este momento divulgados, como se ilustra en la **FIG. 76**, puede entenderse además como un aparato lógico que puede leer instrucciones del medio 511 y/o un puerto de red 505, que puede conectarse opcionalmente al servidor 509 que tiene el medio fijo 512. El sistema 500, tal como se muestra en la **FIG. 76** puede incluir una CPU 501, unidades de disco 503, dispositivos de entrada opcionales tales como teclado 515 y/o ratón 516 y un monitor opcional 507. La comunicación de datos
10 puede conseguirse a través del medio de comunicación indicado hasta un servidor en una ubicación local o remota. El medio de comunicación puede incluir cualquier medio de transmisión y/o recepción de datos. Por ejemplo, el medio de comunicación puede ser una conexión de red, una conexión inalámbrica o una conexión a internet. Dicha conexión puede proporcionar comunicación en la red mundial. Se prevé que los datos referidos a la presente divulgación pueden transmitirse mediante dichas redes o conexiones para la recepción y/o revisión por un interesado
15 522 como se ilustra en la **FIG. 76**.

La **FIG. 77** es un diagrama de bloques que ilustra una primera arquitectura ejemplar de un sistema de ordenador 100 que puede usarse en conexión con las realizaciones ejemplares de la presente divulgación. Como se representa en la **FIG. 77**, el sistema de ordenador ejemplar puede incluir un procesador 102 para procesar instrucciones. Ejemplos no limitantes de procesadores incluyen: procesador Intel, procesador Xeon™, procesador AMD Opteron™, procesador Samsung 32-bit RISC ARM 1176JZ(F)-S v1.0TM, procesador ARM Cortex-A8 Samsung S5PC100TM, procesador ARM Cortex-A8 Apple A4™, procesador Marvell PXA 930™ o un procesador funcionalmente equivalente. Pueden usarse múltiples sentencias de ejecución para el procesamiento en paralelo. En algunas realizaciones, también pueden usarse múltiples procesadores o procesadores con múltiples núcleos, ya sea en un único sistema de ordenador, en un grupo o distribuido entre sistemas sobre una red que comprende una pluralidad de ordenadores, teléfonos móviles y/o dispositivos de asistentes de datos personales.
20
25

Como se ilustra en la **FIG. 77**, puede conectarse una antememoria de alta velocidad 104, o incorporarse, en el procesador 102 para proporcionar una memoria de alta velocidad para las instrucciones o datos que se han usado recientemente o se usan frecuentemente por el procesador 102. El procesador 102 se conecta a un puente norte 106 mediante un bus del procesador 108. El puente norte 106 se conecta a memoria de acceso aleatorio (RAM) 110 mediante un bus de memoria 112 y maneja el acceso a la RAM 110 mediante el procesador 102. El puente norte 106 también se conecta a un puente sur 114 mediante un bus del grupo de chips 116. El puente sur 114 se conecta, a su vez a un bus periférico 118. El bus periférico puede ser, por ejemplo, PCI, PCI-X, PCI Express u otro bus periférico. El puente norte y el puente sur a menudo se mencionan como el grupo de chips del procesador y tratan la transferencia de datos entre el procesador, RAM y los componentes periféricos en el bus periférico 118. En algunas arquitecturas alternativas, la funcionalidad del puente norte puede incorporarse en el procesador en lugar de usar un chip de puente norte diferente.
30
35

En algunas realizaciones, el sistema 100 puede incluir una tarjeta aceleradora 122 unida al bus periférico 118. El acelerador puede incluir series de puertos programables de campo (FPGA) u otro soporte físico para acelerar determinados procesamientos. Por ejemplo, un acelerador puede usarse para la reestructuración adaptativa de datos o para evaluar expresiones algebraicas usadas en el procesamiento de conjuntos extendidos.
40

El programa informático y los datos se almacenan en el almacenamiento externo 124 y pueden cargarse la RAM 110 y/o la antememoria 104 para su uso por el procesador. El sistema 100 incluye un sistema operativo para tratar los recursos del sistema; ejemplos no limitantes de sistemas operativos incluyen: Linux, Windows®, MACOS®, BlackBerry OS®, iOS® y otros sistemas operativos funcionalmente equivalentes, así como el programa informático de aplicación que se ejecuta sobre el sistema operativo para administrar el almacenamiento y la optimización de datos de acuerdo con realización ejemplares de la presente invención.
45

En este ejemplo, el sistema 100 también incluye tarjetas de interfaz de red (NIC) 120 y 121 conectadas al bus periférico para proporcionar interfaces de red al almacenamiento externo, tal como el almacenamiento unido a red (NAS) y otros sistemas de ordenador que pueden usarse para el procesamiento en paralelo distribuido.

50 La **FIG. 78** es un diagrama que muestra una red 200 con una pluralidad de sistema de ordenador 202a y 202b, una pluralidad de teléfonos móviles y asistentes de datos personales 202c y almacenamiento unido a red (NAS) 204a y 204b. En realizaciones ejemplares, los sistemas 212a, 212b y 212c pueden tratar el almacenamiento de los datos y optimizar el acceso a los datos para los datos almacenados en el almacenamiento unido a red (NAS) 214a y 214b. Puede usarse un modelo matemático para los datos y evaluarse usando el procesamiento en paralelo distribuido entre los sistemas de ordenador 212a y 212b, y los sistemas de teléfono móvil y asistente de datos personal 212c.
55 Los sistemas de ordenador 212a y 212b, y los sistemas de teléfono móvil y asistente de datos personal 212c también pueden proporcionar procesamiento en paralelo para la reestructuración adaptativa de datos de los datos almacenados en el almacenamiento unido a red (NAS) 214a y 214b. La **FIG. 78** ilustra un ejemplo únicamente, y puede usarse una amplia diversidad de otras arquitecturas de ordenador y sistemas junto con las diversas realizaciones de la presente invención. Por ejemplo, puede usarse un servidor blade para proporcionar procesamiento paralelo. Los blades de procesador pueden conectarse mediante un plano de respaldo para proporcionar procesamiento paralelo. El almacenamiento también puede conectarse al plano de respaldo como un almacenamiento unido a red (NAS) mediante una interfaz de red separada.
60

En algunas realizaciones ejemplares, los procesadores pueden mantener espacios de memoria separados y transmitir datos a través de interfaces de red, plano de respaldo u otros conectores para el procesamiento paralelo por otros procesadores. En otras realizaciones, algunos o todos los procesadores pueden usar un espacio de memoria de dirección virtual compartida.

5 La **FIG. 79** es un diagrama de bloques de un sistema de ordenador de multiprocesador 300 que usa un espacio de memoria de dirección virtual compartida de acuerdo con una realización ejemplar. El sistema incluye una pluralidad de procesadores 302a-f que puede acceder a un subsistema de memoria compartido 304. El sistema incorpora una pluralidad de procesadores de algoritmo de memoria (MAP) 306a-f de soporte físico programable en el subsistema de memoria 304. Cada MAP 306a-f puede comprender una memoria 308a-f y una o más series de puerto programable de campo (FPGA) 310a-f. El MAP proporciona una unidad funcional configurable y algoritmos particulares o pueden proporcionarse partes de algoritmos a los FPGA 310a-f para el procesamiento en cercana coordinación con un procesador respectivo. Por ejemplo, los MAP pueden usarse para evaluar expresiones algebraicas respecto al modelo de datos y para realizar reestructuración adaptativa de los datos en realizaciones ejemplares. En este ejemplo, cada MAP es globalmente accesible por todos los procesadores para estos fines. En una configuración, cada MAP puede usar un acceso de memoria directo (DMA) para acceder a una memoria asociada 308a-f, que le permite ejecutar tareas independientemente de, y de forma asíncrona de, el microprocesador 302a-f respectivo. En esta configuración, un MAP puede suministrar resultados directamente a otro MAP para la canalización y la ejecución paralela de algoritmos.

20 Las arquitecturas de ordenador anteriores y sistemas son ejemplos únicamente, y puede usarse una amplia diversidad de otras arquitecturas de ordenador, teléfono móvil y asistente de datos personales y sistemas en relación con realizaciones ejemplares, incluyendo sistemas que usan cualquier combinación de procesadores generales, coprocesadores, FPGA y otros dispositivos lógicos programables, sistema en chips (SOC), circuitos integrados específicos de aplicación (ASIC) y otros elementos de procesamiento y lógicos. En algunas realizaciones, todo o parte del sistema de ordenador puede implementarse en programa informático o soporte físico. Cualquier diversidad de medio de almacenamiento de datos puede usarse en relación con realizaciones ejemplares, incluyendo memoria de acceso aleatorio, unidades de disco duro, memoria flash, unidades de cinta, matriz de discos, almacenamiento unido a red (NAS) y otros dispositivos y sistemas de almacenamiento de datos locales o distribuidos.

30 En realizaciones ejemplares, el subsistema de ordenador de la presente divulgación puede implementarse usando módulos de programa informático que se ejecutan en cualquiera de las arquitecturas de ordenadores y sistemas anteriores u otros. En otras realizaciones, las funciones del sistema pueden implementarse parcial o completamente en soporte lógico, dispositivos lógicos programables tales como series de puerto programable de campo (FPGA) como se menciona en la **FIG. 79**, sistema en chips (SOC), circuitos integrados específicos de aplicación (ASIC) u otros elementos de procesamiento y lógicos. Por ejemplo, puede implementarse el procesador de conjunto y el optimizador con aceleración del soporte físico mediante el uso de una tarjeta aceleradora de soporte físico, tal como la tarjeta aceleradora 122 ilustrada en la **FIG. 77**.

Oligonucleótidos (por ejemplo, códigos de barras moleculares)

40 Los métodos y kits divulgados en la presente memoria pueden comprender uno o más oligonucleótidos o usos de los mismos. Los oligonucleótidos pueden adherirse a un soporte sólido divulgado en la presente memoria. La adhesión del oligonucleótido al soporte sólido puede producirse mediante pares de grupos funcionales en el soporte sólido y el oligonucleótido. El oligonucleótido puede mencionarse como código de barras molecular. El oligonucleótido puede mencionarse como marcador (por ejemplo, marcador molecular, marcador celular) o marca (por ejemplo, marca de muestra).

45 Los oligonucleótidos pueden comprender un marcador universal. Un marcador universal puede ser igual para todos los oligonucleótidos en una muestra. Un marcador universal puede ser igual para los oligonucleótidos en un conjunto de oligonucleótidos. Un marcador universal puede ser igual para dos o más conjuntos de oligonucleótidos. Un marcador universal puede comprender una secuencia de ácidos nucleicos que pueden hibridar con un cebador de secuenciación. Los cebadores de secuenciación pueden usarse para secuenciar oligonucleótidos que comprenden un marcador universal. Los cebadores de secuenciación (por ejemplo, cebadores de secuenciación universales) pueden comprender cebadores de secuenciación asociados con plataformas de secuenciación de alto rendimiento. Un marcador universal puede comprender una secuencia de ácidos nucleicos que puede hibridar con un cebador de PCR. Un marcador universal puede comprender una secuencia de ácidos nucleicos que puede hibridar con un cebador de secuenciación y un cebador de PCR. La secuencia de ácidos nucleicos del marcador universal que puede hibridar con un cebador de secuenciación y/o de PCR puede mencionarse como sitio de unión a cebador. Un marcador universal puede comprender una secuencia que puede usarse para iniciar la transcripción del oligonucleótido. Un marcador universal puede comprender una secuencia que puede usarse para la prolongación del oligonucleótido o una región dentro del oligonucleótido. Un marcador universal puede ser de al menos aproximadamente 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50 o más nucleótidos de longitud. Un marcador universal puede comprender al menos aproximadamente 10 nucleótidos. Un marcador universal puede ser como mucho de aproximadamente 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50 o más nucleótidos de longitud.

Los oligonucleótidos comprenden un marcador celular. Un marcador celular comprende una secuencia de ácido nucleico que puede proporcionar información para la célula que se pone en contacto con el oligonucleótido (por ejemplo, determinación del ácido nucleico que se origina de determinada célula). Al menos un 60 %, 70 %, 80 %, 85 %, 90 %, 95 %, 97 %, 99 % o 100 % de los oligonucleótidos en el mismo soporte sólido pueden comprender el

mismo marcador celular. Al menos un 60 % de los oligonucleótidos en el mismo soporte sólido pueden comprender el mismo marcador celular. Al menos un 95 % de los oligonucleótidos en el mismo soporte sólido pueden comprender el mismo marcador celular. Todos los oligonucleótidos en un mismo soporte sólido comprenden el mismo marcador celular. El marcador celular de los oligonucleótidos en un primer soporte sólido, son diferentes a los marcadores celulares de los oligonucleótidos en el segundo soporte sólido.

Un marcador celular puede ser de al menos aproximadamente 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50 o más nucleótidos de longitud. Un marcador puede ser como mucho de aproximadamente 300, 200, 100, 90, 80, 70, 60, 50, 40, 30, 20, 15, 12, 10, 9, 8, 7, 6, 5, 4 o menos o más nucleótidos de longitud. Un marcador celular puede comprender entre aproximadamente 5 y aproximadamente 200 nucleótidos. Un marcador celular puede comprender entre aproximadamente 10 y aproximadamente 150 nucleótidos. Un marcador celular puede comprender entre aproximadamente 20 y aproximadamente 125 nucleótidos de longitud.

Los oligonucleótidos comprenden un marcador molecular. Un marcador molecular comprende una secuencia de ácido nucleico que puede proporcionar información de identificación para la especie de ácido nucleico específica hibridada con el oligonucleótido. Los oligonucleótidos conjugados a un mismo soporte sólido comprenden diferentes marcadores moleculares. De esta manera, el marcador molecular puede distinguir los tipos de ácidos nucleicos diana (por ejemplo, genes), que hibridan con los diferentes oligonucleótidos. Un marcador molecular puede ser de al menos aproximadamente 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50 o más nucleótidos de longitud. Un marcador molecular puede ser como mucho de aproximadamente 300, 200, 100, 90, 80, 70, 60, 50, 40, 30, 20, 15, 12, 10, 9, 8, 7, 6, 5, 4 o menos nucleótidos de longitud.

Los oligonucleótidos pueden comprender un marcador de muestra (por ejemplo, índice de muestra). Un marcador de muestra puede comprender una secuencia de ácido nucleico que puede proporcionar información acerca de donde se originó un ácido nucleico diana. Por ejemplo, un marcador de muestra puede ser diferente en diferentes soportes sólidos usados en diferentes experimentos. Un marcador de muestra puede ser de al menos aproximadamente 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50 o más nucleótidos de longitud. Un marcador de muestra puede ser como mucho de aproximadamente 300, 200, 100, 90, 80, 70, 60, 50, 40, 30, 20, 15, 12, 10, 9, 8, 7, 6, 5, 4 o menos nucleótidos de longitud.

Un oligonucleótido puede comprender un marcador universal, un marcador celular, un marcador molecular y un marcador de muestra o cualquier combinación de los mismos. En combinación, el marcador de muestra puede usarse para distinguir ácidos nucleicos diana entre muestras, el marcador celular puede usarse para distinguir ácidos nucleicos diana de diferentes células en la muestra, el marcador molecular puede usarse para distinguir los diferentes ácidos nucleicos diana en la célula (por ejemplo, diferentes copias del mismo ácido nucleico diana) y el marcador universal puede usarse para amplificar y secuenciar los ácidos nucleicos diana.

Un marcador universal, un marcador molecular, un marcador celular, un marcador de conector y/o un marcador de muestra puede comprender una secuencia aleatoria de nucleótidos. Una secuencia aleatoria de nucleótidos puede generarse por ordenar. Una secuencia aleatoria de nucleótidos puede no tener un patrón asociado con la misma. Un marcador universal, un marcador molecular, un marcador celular, marcador de conector y/o un marcador de muestra puede comprender una secuencia no aleatoria (por ejemplo, los nucleótidos comprenden un patrón) de nucleótidos. Las secuencias del marcador universal, un marcador molecular, un marcador celular, marcador de conector y/o un marcador de muestra pueden ser secuencias disponibles en el mercado. Las secuencias del marcador universal, un marcador molecular, un marcador celular, marcador de conector y/o un marcador de muestra pueden comprender secuencias aleatorias. Las secuencias aleatorias pueden referirse a secuencias oligonucleotídicas compuestas de todas las secuencias posibles para una longitud dada del oligómero aleatorio. Como alternativa, o adicionalmente, un marcador universal, un marcador molecular, un marcador celular, marcador de conector y/o un marcador de muestra pueden comprender una secuencia predeterminada de nucleótidos.

La **FIG. 1** muestra un oligonucleótido ejemplar de la divulgación que comprende un marcador universal, un marcador celular y un marcador molecular.

La **FIG. 3** muestra un soporte sólido acoplado a oligonucleótido ejemplar que comprende un soporte sólido (301) acoplado a un oligonucleótido (312). El oligonucleótido (312) comprende un grupo químico (5' amina, 302), un marcador universal (303), un marcador celular (311), un marcador molecular (BC molecular, 311) y una región de unión a diana (oligodT, 310). En este esquema, el marcador celular (311) comprende un primer marcador celular (CL parte 1, 304), un primer conector (conector 1, 305), un segundo marcador celular (CL parte 2, 306), un segundo conector (conector 2, 307), un tercer marcador celular (CL parte 3, 308). El marcador celular (311) es común para cada oligonucleótido en el soporte sólido. Los marcadores celulares (311) para dos o más microesferas pueden ser diferentes. Los marcadores celulares (311) para dos o más microesferas pueden diferir en los marcadores celulares (por ejemplo, CL parte 1 (304), CL parte 2 (306), CL parte 3 (308)). Los marcadores celulares (311) para dos o más microesferas pueden diferir en el primer marcador celular (304), segundo marcador celular (306), tercer marcador celular (308) o una combinación de los mismos. El primer y segundo conectores (303, 305) de los marcadores celulares (311) pueden ser idénticos para dos o más soportes sólidos acoplados a oligonucleótido. El marcador universal (303) puede ser idéntico para dos o más soportes sólidos acoplados a oligonucleótido. El marcador universal (303) puede ser idéntico para dos o más oligonucleótidos en el mismo soporte sólido. El marcador molecular (311) puede ser diferente para al menos dos o más oligonucleótidos en el soporte sólido. El soporte sólido puede comprender 100 o más oligonucleótidos. El soporte sólido puede comprender 1000 o más oligonucleótidos. El soporte sólido puede comprender 10 000 o más oligonucleótidos. El soporte sólido puede comprender 100 000 o más oligonucleótidos.

Además de un marcador universal, un marcador celular y un marcador molecular, un oligonucleótido puede comprender una región de unión a diana. Una región de unión a diana puede comprender una secuencia de ácido nucleico que puede unirse a un ácido nucleico diana (por ejemplo, un ácido nucleico celular a analizar). Una región de unión a diana puede ser una secuencia específica de gen. Por ejemplo, una región de unión a diana puede comprender una secuencia de ácido nucleico que puede adherirse (por ejemplo, hibridar) a una ubicación específica de un ácido nucleico diana, específico. Una región de unión a diana puede comprender una secuencia de ácido nucleico diana no específica. Una secuencia de ácido nucleico diana no específica puede referirse a una secuencia que puede unirse a múltiples ácidos nucleicos diana, independientemente de la secuencia específica del ácido nucleico diana. Por ejemplo, la región de unión a diana puede comprender una secuencia multimérica aleatoria o una secuencia de oligodT (por ejemplo, un tramo de nucleótidos de timidina que puede hibridar con una cola de poliadenilación en los ARNm). Una secuencia multimérica aleatoria puede ser, por ejemplo, un dímero, trímero, tetrámero, pentámero, hexámero, heptámero, octámero, nonámero, decámero aleatorio o secuencia multimérica superior de cualquier longitud. Una región de unión a diana puede ser de al menos aproximadamente 5, 10, 15, 20, 25, 30, 35, 40, 45, 50 o más nucleótidos de longitud. Una región de unión a diana puede ser como mucho de aproximadamente 5, 10, 15, 20, 25, 30, 35, 40, 45, 50 o más nucleótidos de longitud.

Un oligonucleótido puede comprender una pluralidad de marcadores. Por ejemplo, un oligonucleótido puede comprender al menos 1, 2, 3, 4, 5, 6, 7 u 8 o más marcadores universales. Un oligonucleótido puede comprender como mucho aproximadamente 1, 2, 3, 4, 5, 6, 7 u 8 o más marcadores universales. Un oligonucleótido puede comprender al menos aproximadamente 1, 2, 3, 4, 5, 6, 7 u 8 o más marcadores celulares. Un oligonucleótido puede comprender como mucho aproximadamente 1, 2, 3, 4, 5, 6, 7 u 8 o más marcadores celulares. Un oligonucleótido puede comprender al menos aproximadamente 1, 2, 3, 4, 5, 6, 7 u 8 o más marcadores moleculares. Un oligonucleótido puede comprender como mucho aproximadamente 1, 2, 3, 4, 5, 6, 7 u 8 o más marcadores moleculares. Un oligonucleótido puede comprender al menos aproximadamente 1, 2, 3, 4, 5, 6, 7 u 8 o más marcadores de muestra. Un oligonucleótido puede comprender como mucho aproximadamente 1, 2, 3, 4, 5, 6, 7 u 8 o más marcadores de muestra. Un oligonucleótido puede comprender al menos aproximadamente 1, 2, 3, 4, 5, 6, 7 u 8 o más regiones de unión a diana. Un oligonucleótido puede comprender como mucho aproximadamente 1, 2, 3, 4, 5, 6, 7 u 8 o más regiones de unión a diana.

Cuando un oligonucleótido comprende más de un tipo de marcador (por ejemplo, más de un marcador celular o más de un marcador molecular), los marcadores pueden intercalarse con una secuencia de marcador de conector. Una secuencia de marcador de conector puede ser de al menos aproximadamente 5, 10, 15, 20, 25, 30, 35, 40, 45, 50 o más nucleótidos de longitud. Una secuencia de marcador de conector puede ser como mucho de aproximadamente 5, 10, 15, 20, 25, 30, 35, 40, 45, 50 o más nucleótidos de longitud. En algunos casos, una secuencia de marcador de conector es de 12 nucleótidos de longitud. Una secuencia de marcador de conector puede usarse para facilitar la síntesis del oligonucleótido, tal como se esquematiza en la **FIG. 2A**.

El número de oligonucleótidos conjugados a un soporte sólido ser de 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10 veces más que el número de ácidos nucleicos diana en una célula. En algunos casos, al menos un 10, 20, 30, 40, 50, 60, 70, 80, 90 o 100 % de los oligonucleótidos se unen por un ácido nucleico diana. En algunos casos como mucho un 10, 20, 30, 40, 50, 60, 70, 80, 90 o 100 % de los oligonucleótidos se unen por un ácido nucleico diana. En algunos casos al menos 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90 o 100 o más diferentes ácidos nucleicos diana se capturan por los oligonucleótidos en un soporte sólido. En algunos casos, como mucho 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90 o 100 o más diferentes ácidos nucleicos diana se capturan por los oligonucleótidos en un soporte sólido.

Un polímero puede comprender soportes sólidos adicionales. Por ejemplo, un polímero puede estar salpicado de microesferas. Las microesferas pueden ubicarse espacialmente en diferentes regiones del polímero. Las microesferas o soportes que comprenden oligonucleótidos de la divulgación pueden abordarse espacialmente. Las microesferas o soportes pueden comprender un código de barras correspondiente a una dirección espacial en el polímero. Por ejemplo, cada microesfera o soporte de una pluralidad de microesferas o soportes puede comprender un código de barras que corresponde a una posición en un polímero, tal como una posición en una matriz o un micropocillo particular de una pluralidad de micropocillos. La dirección espacial puede descodificarse para determinar la ubicación desde la que se ubicó una microesfera o soporte. Por ejemplo, una dirección espacial, tal como un código de barras, puede descodificarse mediante hibridación de un oligonucleótido con el código de barras o por secuenciación del código de barras. Como alternativa, las microesferas o soporte pueden albergar otros tipos de códigos de barras, tales como elementos gráficos, grupos químicos, colores, fluorescencia o combinaciones de cualquier combinación de los mismos, con fines de descodificación de la dirección espacial.

Los métodos y kits divulgados en la presente memoria pueden comprender uno o más conjuntos de códigos de barras moleculares. Uno o más códigos de barras moleculares pueden comprender una región de índice de muestra y una región de marcador. Dos o más códigos de barras moleculares de un conjunto de códigos de barras moleculares pueden comprender la misma región de índice y dos o más regiones de marcador diferentes. Dos o más códigos de barras moleculares de dos o más conjuntos de códigos de barras moleculares pueden comprender dos o más regiones de índice de muestra diferentes. Dos o más códigos de barras moleculares de un conjunto de códigos de barras moleculares pueden comprender diferentes regiones de marcador. Dos o más códigos de barras moleculares de dos o más conjuntos de códigos de barras moleculares pueden comprender la misma región de marcador. Los códigos de barras moleculares de dos o más conjunto de códigos de barras moleculares pueden diferir en sus regiones de índice de muestra. Los códigos de barras moleculares de dos o más conjuntos de códigos de barras moleculares pueden ser similares basándose en sus regiones marcador.

ES 2 711 168 T3

Los códigos de barras moleculares pueden comprender además una región específica de diana, una región adaptadora, una región de PCR universal, una región específica de diana o cualquier combinación de las mismas. El código de barras molecular puede comprender una región de PCR universal y una región específica de diana. El código de barras molecular puede comprender una o más estructuras secundarias. El código de barras molecular puede comprender una estructura de horquilla. El código de barras molecular puede comprender una región específica de diana y un tallo escindible.

Los métodos y kits divulgados en la presente memoria pueden comprender uno o más conjuntos de marcas de muestra. Una o más marcas de muestra pueden comprender una región de índice de muestra. Una o más marcas de muestra pueden comprender una región de índice de muestra. Dos o más marcas de muestra de un conjunto de marcas de muestra pueden comprender la misma región de índice de muestra. Dos o más marcas de muestra de dos o más conjuntos de marcas de muestra pueden comprender dos o más regiones de índice de muestra diferentes.

Las marcas de muestra pueden comprender además una región específica de diana, una región adaptadora, una región de PCR universal, una región específica de diana o cualquier combinación de las mismas. La marca de muestra puede comprender una región de PCR universal y una región específica de diana. La marca de muestra puede comprender una o más estructuras secundarias. La marca de muestra puede comprender una estructura de horquilla. La marca de muestra puede comprender una región específica de diana y un tallo escindible.

Los métodos y kits divulgados en la presente memoria pueden comprender uno o más conjuntos de marcadores identificadores moleculares. Uno o más marcadores identificadores moleculares pueden comprender una región de marcador. Uno o más marcadores identificadores moleculares pueden comprender una región de marcador. Dos o más marcadores identificadores moleculares de un conjunto de marcadores identificadores moleculares pueden comprender dos o más regiones de marcador diferentes. Dos o más marcadores identificadores moleculares de dos o más conjuntos de marcadores identificadores moleculares pueden comprender dos o más regiones de marcador idénticas. Los marcadores identificadores moleculares pueden comprender además una región específica de diana, una región adaptadora, una región de PCR universal, una región específica de diana o cualquier combinación de las mismas. El marcador identificador molecular puede comprender una región de PCR universal y una región específica de diana. El marcador identificador molecular puede comprender una o más estructuras secundarias. El marcador identificador molecular puede comprender una estructura de horquilla. El marcador identificador molecular puede comprender una región específica de diana y un tallo escindible.

El código de barras molecular, marca de muestra o marcador identificador molecular 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 1000 nucleótidos o pares de bases. En otro ejemplo, la marca de muestra o marcador identificador molecular comprende al menos aproximadamente 1500, 2000; 2500, 3000; 3500, 4000; 4500, 5000; 5500, 6000; 6500, 7000; 7500, 8000; 8500, 9000; 9500 o 10 000 nucleótidos o pares de bases.

Los códigos de barras moleculares, marcas de muestra o marcadores identificadores moleculares pueden ser multímeros, por ejemplo, multímeros aleatorios. Una secuencia multimérica puede ser, por ejemplo, un dímero, trímero, tetramero, pentámero, hexámero, heptámero, octámero, nonámero, decámero aleatorio o no aleatorio o secuencia multimérica superior de cualquier longitud. Las marcas pueden generarse aleatoriamente a partir de un conjunto de mononucleótidos. Las marcas pueden ensamblarse por incorporación aleatoria de mononucleótidos.

Los códigos de barras moleculares, marcas de muestra o marcadores identificadores moleculares también pueden ensamblarse sin aleatoriedad, para generar una colección de marcas diferentes que no se generan aleatoriamente, sino que incluyen números suficientes de diferentes marcas para poner en práctica los métodos.

En algunas realizaciones, un código de barras molecular, marca de muestra o marcador identificador molecular puede comprender un recorte en un ácido nucleico diana. El recorte puede ser, por ejemplo, una digestión enzimática de uno o ambos extremos de un ácido nucleico diana. El recorte puede usarse junto con la adición de códigos de barras moleculares añadidos, marcas de muestra (por ejemplo, región de índice de muestra, marcador de muestra), marcador celular y marcadores identificadores moleculares (por ejemplo, marcador molecular). La combinación del recorte y las marcas añadidas puede contener información referida a la molécula de partida particular. Añadiendo un recorte aleatorio al código de barras molecular, marca de muestra o marcador identificador molecular, puede ser necesaria una diversidad menor de las marcas añadidas para contar el número de ácidos nucleicos diana cuando la detección permite una determinación tanto del recorte aleatorio como de los oligonucleótidos añadidos.

El código de barras molecular, marca de muestra o marcador identificador molecular puede comprender una región específica de diana. La región específica de diana puede comprender una secuencia que es complementaria a la molécula. En algunos casos, la molécula es una molécula de ARNm y la región específica de diana comprende una secuencia de oligodT que es complementaria a la cola de poliA de la molécula de ARNm. La región específica de diana también puede actuar como cebador para síntesis de ADN y/o ARN. Por ejemplo, la secuencia de oligodT de la región específica de diana puede actuar como cebador para la síntesis de la primera hebra de una copia de ADNc de la molécula de ARNm. Como alternativa, la región específica de diana comprende una secuencia que es complementaria a cualquier parte de la molécula. En otros casos, la región específica de diana comprende una secuencia aleatoria que puede hibridar o ligar con la molécula. La región específica de diana puede posibilitar la adhesión de la marca de muestra o marcador identificador molecular a la molécula. La adhesión de la marca de muestra o marcador identificador molecular puede producirse por cualquiera de los métodos divulgados en la

presente memoria. Por ejemplo, hibridación, ligamiento). En algunos casos, la región específica de diana comprende una secuencia que se reconoce por una o más enzimas de restricción. La región específica de 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, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900 o 1000 nucleótidos o pares de bases. En otro ejemplo, la región específica de diana comprende al menos aproximadamente 1500, 2000, 2500, 3000, 3500, 4000, 4500, 5000, 5500, 6000, 6500, 7000, 7500, 8000, 8500, 9000, 9500 o 10 000 nucleótidos o pares de bases. Preferiblemente, la región específica de diana comprende al menos aproximadamente 5-10, 10-15, 10-20, 10-30, 15-30 o 20-30 nucleótidos o pares de bases.

En algunos casos, la región específica de diana es específica para un gen o producto génico particular. Por ejemplo, la región específica de diana comprende una secuencia complementaria a una región de un gen o producto génico de p53. Por lo tanto, las marcas de muestra y marcadores identificadores moleculares pueden adherirse únicamente a moléculas que comprenden la secuencia específica de p53. Como alternativa, la región específica de diana es específica para una pluralidad de diferentes genes o productos génicos. Por ejemplo, la región específica de diana comprende una secuencia de oligodT. Por lo tanto, las marcas de muestra y marcadores identificadores moleculares pueden adherirse a cualquier molécula que comprenda una secuencia de poliA. En otro ejemplo, la región específica de diana comprende una secuencia aleatoria que es complementaria a una pluralidad de diferentes genes o productos génicos. Por tanto, la marca de muestra o marcador identificador molecular puede adherirse a cualquier molécula con una secuencia que sea complementaria a la región específica de diana. En otros casos, la región específica de diana comprende un saliente de sitio de restricción (por ejemplo, saliente de extremo adherente de EcoRI). La marca de muestra o marcador identificador molecular puede ligarse a cualquier molécula que comprenda una secuencia complementaria al saliente del sitio de restricción.

En algunos casos, la región específica de diana es específica para un microARN o producto de microARN particular. Por ejemplo, la región específica de diana comprende una secuencia complementaria a una región de un microARN o producto de microARN específico. Por ejemplo, las regiones específicas de diana comprenden secuencias complementarias a regiones de un panel específico de microARN o panel de productos de microARN. Por lo tanto, las marcas de muestra y marcadores identificadores moleculares pueden adherirse únicamente a moléculas que comprenden la secuencia específica de microARN. Como alternativa, la región específica de diana es específica para una pluralidad de diferentes microARN o productos de microARN. Por ejemplo, la región específica de diana comprende una secuencia complementaria a una región comprendida en dos o más microARN, tal como un panel de microARN que contienen una secuencia común. Por lo tanto, las marcas de muestra y marcadores identificadores moleculares pueden adherirse a cualquier molécula que comprenda la secuencia de microARN común. En otro ejemplo, la región específica de diana comprende una secuencia aleatoria que es complementaria a una pluralidad de diferentes microARN o productos de microARN. Por tanto, la marca de muestra o marcador identificador molecular puede adherirse a cualquier molécula de microARN con una secuencia que es complementaria a la región específica de diana. En otros casos, la región específica de diana comprende un saliente de sitio de restricción (por ejemplo, saliente de extremo adherente de EcoRI). La marca de muestra o marcador identificador molecular puede ligarse a cualquier molécula de microARN que comprenda una secuencia complementaria al saliente de sitio de restricción.

El código de barras molecular o marcador identificador molecular divulgado en la presente memoria a menudo comprende una región de marcador. La región de marcador puede usarse para identificar de forma unívoca la aparición de especies diana, haciendo de ese modo que cada especie con un identificador que puede usarse se distinga entre dos dianas por lo demás idénticas o casi idénticas. La región de marcador de la pluralidad de marcas de muestra y marcadores identificadores moleculares puede comprender una colección de diferentes nanocristales semiconductores, compuestos metálicos, péptidos, oligonucleótidos, anticuerpos, moléculas pequeñas, isótopos, partículas o estructuras que tienen diferentes formas, colores, códigos de barras o patrones de difracción asociados con los mismos o incluidos en los mismos, cadenas de números, fragmentos aleatorios de proteínas o ácidos nucleicos, diferentes isótopos o cualquier combinación de los mismos. La región de marcador puede comprender una secuencia degenerativa. La región de marcador 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 1000 nucleótidos o pares de bases. En otro ejemplo, la región de marcador comprende al menos aproximadamente 1500; 2000; 2500, 3000; 3500, 4000; 4500, 5000; 5500, 6000; 6500, 7000; 7500, 8000; 8500, 9000; 9500 o 10 000 nucleótidos o pares de bases. Preferiblemente, la región de marcador comprende al menos aproximadamente 10-30, 15-40 o 20-50 nucleótidos o pares de bases.

En algunos casos, el código de barras molecular, marca de muestra o marcador identificador molecular comprende un sitio de unión a cebador universal. El sitio de unión a cebador universal permite la adhesión de un cebador universal a la molécula marcada y/o amplicón marcado. Los cebadores universales son bien conocidos en la técnica e incluyen, aunque sin limitación, -47F (M13F), α MF, AOX3', AOX5', BGH_r, CMV_-30, CMV_-50, CVM_f, LACrmt, lamgda gt10F, lambda gt 10R, lambda gt11F, lambda gt11R, M13 rev, M13Forward(-20), M13Reverse, male, p10SEQP_pQE, pA_-120, pet 4, pGAP Forward, pGL_RVpr3, pGLpr2_R, pKLAC1_4, pQE_FS, pQE_RS, puc_U1, puc_U2, revers A, seq_IRES_tam, seq_IRES_zpet, seq_ori, seq_PCR, seq_pIRES-, seq_pIRES+, seq_pSecTag, seq_pSecTag+, seq_retro+PSI, SP6, T3-prom, T7-prom y T7-term_Inv. La adhesión del cebador universal al sitio de unión a cebador a cebador universal puede usarse para amplificación, detección y/o secuenciación de la molécula marcada y/o amplicón marcado. El sitio de unión a 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 1000 nucleótidos o pares de bases. En otro ejemplo, el sitio de unión a cebador universal comprende al menos aproximadamente 1500; 2000; 2500, 3000; 3500, 4000; 4500, 5000; 5500, 6000; 6500, 7000; 7500, 8000; 8500, 9000; 9500 o 10 000 nucleótidos o pares de bases.

ES 2 711 168 T3

Preferiblemente el sitio de unión a cebador universal comprende 10-30 nucleótidos o pares de bases.

El código de barras molecular, marca de muestra o marcador identificador molecular puede comprender una región adaptadora. La región adaptadora puede posibilitar la hibridación de una o más sondas. La región adaptadora puede posibilitar la hibridación de una o más sondas de HCR.

- 5 El código de barras molecular, marca de muestra o marcador identificador molecular puede comprender uno o más marcadores detectables.

10 El código de barras molecular, marca de muestra o marcador identificador molecular puede actuar como iniciador para una reacción en cadena de hibridación (HCR). La región adaptadora de la marca de muestra o marcador identificador molecular puede actuar como inicio para HCR. El sitio de unión a cebador universal puede actuar como iniciador para HCR.

15 En algunos casos, el código de barras molecular, marca de muestra o marcador identificador molecular es monocatenario. En otros casos, el código de barras molecular, marca de muestra o marcador identificador molecular es bicatenario. El código de barras molecular, marca de muestra o marcador identificador molecular puede ser lineal. Como alternativa, el código de barras molecular, marca de muestra o marcador identificador molecular comprende una estructura secundaria. Tal como se usa en la presente memoria, "estructura secundaria" incluye estructuras terciarias, cuaternarias, etc. En algunos casos, la estructura secundaria es una horquilla, una estructura de tallo-bucle, un bucle interno, un bucle abultado, una estructura ramificada o un seudonudo, múltiples estructuras de tallo y bucle, estructuras de tipo hoja de trébol o cualquier estructura tridimensional. En algunos casos, la estructura secundaria es una horquilla. La horquilla puede comprender una secuencia saliente. La secuencia saliente de la horquilla puede actuar como cebador para una reacción en cadena de la polimerasa y/o reacción de transcripción inversa. La secuencia saliente comprende una secuencia que es complementaria a la molécula a la que se adhiere la marca de muestra o marcador identificador molecular y la secuencia saliente hibrida con la molécula. La secuencia saliente puede ligarse a la molécula y actúa como molde para una reacción en cadena de la polimerasa y/o reacción de transcripción inversa. En algunas realizaciones, el código de barras molecular, la marca de muestra o marcador identificador molecular comprende ácidos nucleicos y/o ácidos nucleicos sintéticos y/o ácidos nucleicos modificados.

25 En algunos casos, la pluralidad de códigos de barras moleculares, marca de muestra (por ejemplo, región de índice de muestra, marcador de muestra), marcador celular y marcadores identificadores moleculares (por ejemplo, marcador molecular) comprende al menos aproximadamente 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 o 100 diferentes códigos de barras moleculares, marcas de muestra (por ejemplo, región de índice de muestra, marcador de muestra), marcador celular y marcadores identificadores moleculares (por ejemplo, marcador molecular). En otros casos, la pluralidad de códigos de barras moleculares, marcas de muestra (por ejemplo, región de índice de muestra, marcador de muestra), marcador celular y marcadores identificadores moleculares (por ejemplo, marcador molecular) comprende al menos aproximadamente 200; 300; 400; 500; 600; 700; 800; 900; 1000; 2000; 3000; 4000; 5000; 6000; 7000; 8000; 9000; o 10 000 diferentes códigos de barras moleculares, marcas de muestra (por ejemplo, región de índice de muestra, marcador de muestra), marcador celular y marcadores identificadores moleculares (por ejemplo, marcador molecular). Como alternativa, la pluralidad de códigos de barras moleculares, marcas de muestra (por ejemplo, región de índice de muestra, marcador de muestra), marcador celular y marcadores identificadores moleculares (por ejemplo, marcador molecular) comprende al menos aproximadamente 20 000; 30 000; 40 000; 50 000; 60 000; 70 000; 80 000; 90 000; o 100 000 diferentes códigos de barras moleculares, marcas de muestra (por ejemplo, región de índice de muestra, marcador de muestra), marcador celular y marcadores identificadores moleculares (por ejemplo, marcador molecular).

40 El número de códigos de barras moleculares, marcas de muestra (por ejemplo, región de índice de muestra, marcador de muestra), marcador celular y marcadores identificadores moleculares (por ejemplo, marcador molecular) en la pluralidad de códigos de barras moleculares, marcas de muestra (por ejemplo, región de índice de muestra, marcador de muestra), marcador celular y marcadores identificadores moleculares (por ejemplo, marcador molecular) a menudo está en exceso del número de moléculas a marcar. En algunos casos el número de códigos de barras moleculares, marcas de muestra (por ejemplo, región de índice de muestra, marcador de muestra), marcador celular y marcadores identificadores moleculares (por ejemplo, marcador molecular) en la pluralidad de códigos de barras moleculares, marcas de muestra (por ejemplo, región de índice de muestra, marcador de muestra), marcador celular y marcadores identificadores moleculares (por ejemplo, marcador molecular) es al menos aproximadamente 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30, 40, 50, 60, 70, 80, 90 o 100 veces mayor que el número de moléculas a marcar.

45 El número de diferentes códigos de barras moleculares, marcas de muestra (por ejemplo, región de índice de muestra, marcador de muestra), marcador celular y marcadores identificadores moleculares (por ejemplo, marcador molecular) en la pluralidad de códigos de barras moleculares, marcas de muestra (por ejemplo, región de índice de muestra, marcador de muestra), marcador celular y marcadores identificadores moleculares (por ejemplo, marcador molecular) a menudo está en exceso del número de diferentes moléculas a marcar. En algunos casos, el número de diferentes códigos de barras moleculares, marcas de muestra (por ejemplo, región de índice de muestra, marcador de muestra), marcador celular y marcadores identificadores moleculares (por ejemplo, marcador molecular) en la pluralidad de códigos de barras moleculares, marcas de muestra (por ejemplo, región de índice de muestra, marcador de muestra), marcador celular y marcadores identificadores moleculares (por ejemplo, marcador molecular) 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 mayor que el número de diferentes moléculas a marcar.

ES 2 711 168 T3

En algunos casos el marcaje estocástico de una molécula comprende una pluralidad de códigos de barras moleculares, marcas de muestra (por ejemplo, región de índice de muestra, marcador de muestra), marcador celular y marcadores identificadores moleculares (por ejemplo, marcador molecular), en la que la concentración de los diferentes códigos de barras moleculares, marcas de muestra (por ejemplo, región de índice de muestra, marcador de muestra), marcador celular y marcadores identificadores moleculares (por ejemplo, marcador molecular) es la misma. En dichos casos, la pluralidad de códigos de barras moleculares, marcas de muestra (por ejemplo, región de índice de muestra, marcador de muestra), marcador celular y marcadores identificadores moleculares (por ejemplo, marcador molecular) comprende números iguales de cada diferente código de barras molecular, marca de muestra o marcador identificador molecular.

En algunos casos, el marcaje estocástico de una molécula comprende una pluralidad de códigos de barras moleculares, marcas de muestra (por ejemplo, región de índice de muestra, marcador de muestra), marcador celular y marcadores identificadores moleculares (por ejemplo, marcador molecular), en la que la concentración de los diferentes códigos de barras moleculares, marcas de muestra (por ejemplo, región de índice de muestra, marcador de muestra), marcador celular y marcadores identificadores moleculares (por ejemplo, marcador molecular) es diferente. En dichos casos la pluralidad de códigos de barras moleculares, marcas de muestra (por ejemplo, región de índice de muestra, marcador de muestra), marcador celular y marcadores identificadores moleculares (por ejemplo, marcador molecular) comprende diferentes números de cada diferente código de barras molecular, marca de muestra o marcador identificador molecular.

En algunos casos, algunos códigos de barras moleculares, marcas de muestra (por ejemplo, región de índice de muestra, marcador de muestra), marcador celular y marcadores identificadores moleculares (por ejemplo, marcador molecular) están presentes a concentraciones mayores que otros códigos de barras moleculares, marcas de muestra (por ejemplo, región de índice de muestra, marcador de muestra), marcador celular y marcadores identificadores moleculares (por ejemplo, marcador molecular) en la pluralidad de códigos de barras moleculares, marcas de muestra (por ejemplo, región de índice de muestra, marcador de muestra), marcador celular y marcadores identificadores moleculares (por ejemplo, marcador molecular). En algunos casos, el marcaje estocástico con diferentes concentraciones de códigos de barras moleculares, marcas de muestra (por ejemplo, región de índice de muestra, marcador de muestra), marcador celular y marcadores identificadores moleculares (por ejemplo, marcador molecular) amplía el intervalo dinámico de medición de la muestra sin aumentar el número de diferentes marcadores usados. Por ejemplo, considérese el marcaje estocástico de 3 moléculas de muestra de ácido nucleico con 10 diferentes códigos de barras moleculares, marcas de muestra (por ejemplo, región de índice de muestra, marcador de muestra), marcador celular y marcadores identificadores moleculares (por ejemplo, marcador molecular) todos a igual concentración. Se espera observar 3 marcadores diferentes. Ahora en lugar de 3 moléculas de ácido nucleico considérense 30 moléculas de ácido nucleico y se espera observar los 10 marcadores. Por el contrario, si aún se usaran 10 marcadores estocásticos diferentes y se alterara las relaciones relativas de los marcadores a 1:2:3:4...10, entonces con 3 moléculas de ácido nucleico, se esperaría observar entre 1-3 marcadores, pero con 30 moléculas se esperaría observar únicamente 5 marcadores aproximadamente, ampliando de este modo el intervalo de medición con el mismo número de marcadores estocásticos.

Las relaciones relativas de los diferentes códigos de barras moleculares, marcas de muestra (por ejemplo, región de índice de muestra, marcador de muestra), marcador celular y marcadores identificadores moleculares (por ejemplo, marcador molecular) en la pluralidad de códigos de barras moleculares, marcas de muestra (por ejemplo, región de índice de muestra, marcador de muestra), marcador celular y marcadores identificadores moleculares (por ejemplo, marcador molecular) pueden ser 1:X, donde X es al menos aproximadamente 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 o 100. Como alternativa, las relaciones relativas de "n" diferentes códigos de barras moleculares, marcas de muestra (por ejemplo, región de índice de muestra, marcador de muestra), marcador celular y marcadores identificadores moleculares (por ejemplo, marcador molecular) en la pluralidad de códigos de barras moleculares, marcas de muestra (por ejemplo, región de índice de muestra, marcador de muestra), marcador celular y marcadores identificadores moleculares (por ejemplo, marcador molecular) son 1:A:B:C...Zn, donde A, B, C...Zn es al menos aproximadamente 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 o 100.

En algunos casos, la concentración de dos o más diferentes códigos de barras moleculares, marcas de muestra (por ejemplo, región de índice de muestra, marcador de muestra), marcador celular y marcadores identificadores moleculares (por ejemplo, marcador molecular) en la pluralidad de códigos de barras moleculares, marcas de muestra (por ejemplo, región de índice de muestra, marcador de muestra), marcador celular y marcadores identificadores moleculares (por ejemplo, marcador molecular) es igual. Para "n" diferentes códigos de barras moleculares, marcas de muestra (por ejemplo, región de índice de muestra, marcador de muestra), marcador celular y marcadores identificadores moleculares (por ejemplo, marcador molecular) la concentración de al menos 2, 3, 4, ...n diferentes códigos de barras moleculares, marcas de muestra (por ejemplo, región de índice de muestra, marcador de muestra), marcador celular y marcadores identificadores moleculares (por ejemplo, marcador molecular) es igual. Como alternativa la concentración de dos o más diferentes códigos de barras moleculares, marcas de muestra (por ejemplo, región de índice de muestra, marcador de muestra), marcador celular y marcadores identificadores moleculares (por ejemplo, marcador molecular) en la pluralidad de códigos de barras moleculares, marcas de muestra (por ejemplo, región de índice de muestra, marcador de muestra), marcador celular y marcadores identificadores moleculares (por ejemplo, marcador molecular) es diferente. Para "n" diferentes

códigos de barras moleculares, marcas de muestra (por ejemplo, región de índice de muestra, marcador de muestra), marcador celular y marcadores identificadores moleculares (por ejemplo, marcador molecular), la concentración de al menos 2, 3, 4, ...n diferentes códigos de barras moleculares, marcas de muestra (por ejemplo, región de índice de muestra, marcador de muestra), marcador celular y marcadores identificadores moleculares (por ejemplo, marcador molecular) es diferente. En algunos casos para "n" diferentes códigos de barras moleculares, marcas de muestra (por ejemplo, región de índice de muestra, marcador de muestra), marcador celular y marcadores identificadores moleculares (por ejemplo, marcador molecular), la diferencia en la concentración de al menos 2, 3, 4, ...n diferentes códigos de barras moleculares, marcas de muestra (por ejemplo, región de índice de muestra, marcador de muestra), marcador celular y marcadores identificadores moleculares (por ejemplo, marcador molecular) es al menos aproximadamente 0,1, 0,2, 0,3, 0,4, 0,5, 0,6, 0,7, 0,8, 0,9, 1, 1,25, 1,5, 1,75, 2, 2,25, 2,5, 2,75, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900 o 1000 veces.

En algunos casos, 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, marcas de muestra (por ejemplo, región de índice de muestra, marcador de muestra), marcador celular y marcadores identificadores moleculares (por ejemplo, marcador molecular) en la pluralidad de códigos de barras moleculares, marcas de muestra (por ejemplo, región de índice de muestra, marcador de muestra), marcador celular y marcadores identificadores moleculares (por ejemplo, marcador molecular) tienen la misma concentración. Como alternativa, 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, marcas de muestra (por ejemplo, región de índice de muestra, marcador de muestra), marcador celular y marcadores identificadores moleculares (por ejemplo, marcador molecular) en la pluralidad de códigos de barras moleculares, marcas de muestra (por ejemplo, región de índice de muestra, marcador de muestra), marcador celular y marcadores identificadores moleculares (por ejemplo, marcador molecular) tienen una concentración diferente.

Como se muestra en la FIG. 65, los códigos de barras moleculares (1004) pueden sintetizarse por separado. Los códigos de barras moleculares (1004) pueden comprender una región de PCR universal (1001), una o más regiones identificadoras (1002) y una región específica de diana. La una o más regiones identificadoras pueden comprender una región de índice de muestra, región de marcador o una combinación de las mismas. La una o más regiones identificadoras pueden estar adyacentes. La una o más regiones identificadoras pueden no estar adyacentes. El código de barras molecular individual puede combinarse para producir una pluralidad de códigos de barras moleculares (1005) que comprende una pluralidad de diferentes regiones identificadoras. Las marcas de muestra pueden sintetizarse de una manera similar a la representado en la FIG. 65, en la que la una o más regiones identificadoras comprenden una región de índice de muestra. Los marcadores identificadores moleculares pueden sintetizarse de una manera similar a la representada en la FIG. 65, en la que la una o más regiones identificadoras comprenden una región de marcador.

La región específica de diana puede ligarse a la región identificadora para producir un código de barras molecular que comprende una región específica de diana. Pueden añadirse exonucleasas 5' y 3' a la reacción para eliminar los productos no ligados. El código de barras molecular puede comprender el sitio de unión a cebador universal, región de marcador y región específica de diana y puede ser resistente a exonucleasa 5' y 3'. Tal como se usa en la presente memoria, las expresiones "sitio de unión a cebador universal" y "región de PCR universal" pueden usarse indistintamente y se refieren a una secuencia que puede usarse para cebar una reacción de amplificación. El grupo 3' fosfato de la región identificadora ligada puede retirarse para producir un código de barras molecular sin un grupo 3' fosfato. El grupo 3' fosfato puede retirarse enzimáticamente. Por ejemplo, puede usarse una polinucleótido cinasa T4 para retirar el grupo 3' fosfato.

Otro método de síntesis de códigos de barras moleculares se representa en la FIG. 66A. Como se muestra en la FIG. 66A, un código de barras molecular (1128) puede sintetizarse ligando dos o más fragmentos oligonucleótidos (1121 y 1127). Un fragmento oligonucleotídico (1121) puede comprender un sitio de unión a cebador universal (1122), región identificadora (1123) y un primer trozo (1123). El otro fragmento oligonucleotídico (1128) puede comprender un segundo trozo (1125) y una región específica de diana (1126). Puede usarse una ligasa (por ejemplo, ADN ligasa T4) para unir los dos fragmentos oligonucleotídicos (1121 y 1127) para producir un código de barras molecular (1128). El ligamiento bicatenario del primer trozo (1124) y el segundo trozo (1125) puede producir un código de barras molecular (1128) con un trozo de puente (1129).

Un método alternativo de síntesis de un código de barras molecular mediante el ligamiento de dos fragmentos oligonucleotídicos se representa en la FIG. 66B. Como se muestra en la FIG. 66B, un código de barras molecular (1158) se sintetiza ligando dos fragmentos oligonucleotídicos (1150 y 1158). Un fragmento oligonucleotídico (1150) puede comprender un sitio de unión a cebador universal (1151), una o más regiones identificadoras (1152) y una secuencia de ligamiento (1153). El otro fragmento oligonucleotídico (1158) puede comprender una secuencia de ligamiento (1154) que es complementaria a la secuencia de ligamiento (1153) del primer fragmento oligonucleotídico (1150), un complemento de una región específica de diana (1155) y un marcador (1156). El fragmento oligonucleotídico (1159) puede comprender también un 3' fosfato que evita la prolongación del fragmento oligonucleotídico. Como se muestra en la etapa 1 de la FIG. 66B, las secuencias de ligamiento (1153 y 1154) de los dos fragmentos oligonucleotídicos pueden hibridar y puede usarse una polimerasa para prolongar el extremo 3' del primer fragmento oligonucleotídico (1150) para producir el código de barras molecular (1158). El código de barras molecular (1158) puede comprender un sitio de unión a cebador universal (1151), una o más regiones identificadoras (1152), secuencia de ligamiento (1153) y una secuencia específica de diana (1157). La secuencia

específica de diana (1157) del código de barras molecular (1158) puede ser el complemento del complemento de la región específica de diana (1155) del segundo fragmento oligonucleotídico (1159). El fragmento oligonucleotídico que comprende el marcador (1156) puede retirarse del código de barras molecular (1158). Por ejemplo, el marcador (1156) puede comprender biotina y fragmentos oligonucleotídicos (1159) que comprenden el marcador biotina (1156) pueden retirarse mediante captura con estreptavidina. En otro ejemplo, el marcador (1156) puede comprender un 5' fosfato y fragmentos oligonucleotídicos (1159) que comprenden el 5' fosfato (1156) pueden retirarse mediante una exonucleasa (por ejemplo, exonucleasa Lambda).

Como se representa en la FIG. 66C, un primer fragmento oligonucleotídico (1170) que comprende un sitio de unión a cebador universal (1171), una o más regiones identificadoras (1172), una primera secuencia de ligamiento (1173) se hibrida a un segundo fragmento oligonucleotídico (1176) que comprende una segunda secuencia de ligamiento (1174) y un complemento de ARN o de la secuencia diana (1175). La etapa 1 puede comprender hibridar la primera y segunda secuencias de ligamiento (1173 y 1174) seguido de transcripción inversa del complemento de ARN de la secuencia diana (1175) para producir el código de barras molecular (1177) que comprende un sitio de unión a cebador universal (1171), una o más regiones identificadoras (1172), una primera secuencia de ligamiento (1173) y una región específica de diana (1178). Los fragmentos oligonucleotídicos que comprenden el complemento de ARN de la secuencia diana pueden degradarse selectivamente mediante el tratamiento con RNasa.

Las secuencias de los códigos de barras moleculares, marcas de muestra (por ejemplo, región de índice de muestra, marcador de muestra), marcador celular y marcadores identificadores moleculares (por ejemplo, marcador molecular) pueden optimizarse para minimizar la dimerización de códigos de barras moleculares, marcas de muestra (por ejemplo, región de índice de muestra, marcador de muestra), marcador celular y marcadores identificadores moleculares (por ejemplo, marcador molecular). El dímero de código de barras molecular, marca de muestra o marcador identificador molecular puede amplificarse y provocar la formación de un amplicón que comprende dos sitios de unión a cebador universal en cada extremo del amplicón y una región específica de diana y una región identificadora única. Como la concentración de los códigos de barras moleculares, marcas de muestra (por ejemplo, región de índice de muestra, marcador de muestra), marcador celular y marcadores identificadores moleculares (por ejemplo, marcador molecular) es mucho mayor que el número de moldes de ADN, estos dímeros de código de barras molecular, marca de muestra o marcador identificador molecular puede desplazar las moléculas de ADN marcadas en una reacción de amplificación. Los ADN no amplificados dan lugar a negativos falsos, y los dímeros amplificados de código de barras molecular, marca de muestra o marcador identificador molecular da lugar a positivos falsos elevados. Por tanto, los códigos de barras moleculares, marcas de muestra (por ejemplo, región de índice de muestra, marcador de muestra), marcador celular y marcadores identificadores moleculares (por ejemplo, marcador molecular) pueden optimizarse para minimizar la formación de dímeros de código de barras molecular, marca de muestra o marcador identificador molecular. Como alternativa, los códigos de barras moleculares, marcas de muestra (por ejemplo, región de índice de muestra, marcador de muestra), marcador celular y marcadores identificadores moleculares (por ejemplo, marcador molecular) que dimerizan se descartan, eliminando de ese modo la formación de dímeros de código de barras molecular, marca de muestra o marcador identificador molecular.

Como alternativa, la formación de dímeros de código de barras molecular, marca de muestra o marcador identificador molecular puede eliminarse o reducirse incorporando una o más modificaciones en la secuencia del código de barras molecular, marca de muestra o marcador identificador molecular. Un código de barras molecular, marca de muestra o marcador identificador molecular que comprende un sitio de unión a cebador universal, región identificadora única y región específica de diana que comprende uracilos y un grupo 3' fosfato se hibrida a un ácido nucleico diana. El ácido nucleico diana puede ser un fragmento digerido con endonucleasa de restricción. La endonucleasa de restricción puede reconocer el sitio de reconocimiento. La amplificación por PCR puede comprender uno o más cebadores directos y uno o más cebadores inversos. La amplificación por PCR puede comprender PCR con cebadores internos con un cebador directo específico para el sitio de unión a cebador universal del código de barras molecular, marca de muestra o marcador identificador molecular y un cebador directo específico para la región específica de diana del código de barras molecular, marca de muestra o marcador identificador molecular y cebadores inversos que son específicos para el ácido nucleico diana. El ácido nucleico diana puede amplificarse usando una ADN polimerasa Pfu, que no puede amplificar el molde que comprende uno o más uracilos. Por tanto, cualquier dímero de códigos de barras moleculares, marcas de muestra (por ejemplo, región de índice de muestra, marcador de muestra), marcador celular y marcadores identificadores moleculares (por ejemplo, marcador molecular) no puede amplificarse por la ADN polimerasa Pfu.

Métodos para sintetizar oligonucleótidos (por ejemplo, códigos de barras moleculares)

Puede sintetizarse un oligonucleótido. Un oligonucleótido puede sintetizarse, por ejemplo, acoplado (por ejemplo, mediante 1-etil-3-(3-dimetilaminopropil)carbodiimida) de un grupo 5' amino en el oligonucleótido al grupo carboxilo del soporte sólido funcionalizado.

Los oligonucleótidos no acoplados pueden retirarse de la mezcla de reacción mediante múltiples lavados. Los soportes sólidos pueden dividirse en pocillos (por ejemplo, 96 pocillos). Cada soporte sólido puede dividirse en un diferente pocillo. La síntesis de oligonucleótidos puede realizarse usando un método de síntesis de división-combinación. El método de división-combinación puede utilizar una combinación de soportes sólidos que comprenden restos reactivos (por ejemplo, oligonucleótidos a sintetizar). Esta combinación puede dividirse en varias combinaciones individuales de soportes sólidos. Cada combinación puede someterse a una primera reacción que puede provocar una modificación diferente a los soportes sólidos en cada una de las combinaciones (por ejemplo, una secuencia de ácido nucleico diferente añadida al oligonucleótido). Después de la reacción, las combinaciones de soportes sólidos pueden combinarse, mezclarse y dividirse de nuevo. Cada división-combinación puede

someterse a una segunda reacción o asignación aleatoria que de nuevo es diferente para cada una de las combinaciones. El proceso puede continuarse hasta que se forma una colección de compuestos diana.

5 Usando la síntesis de división-combinación, la secuencia de ácido nucleico añadida al oligonucleótido puede incorporarse por prolongación del cebador (por ejemplo, prolongación de Kenow). La secuencia de ácido nucleico a
añadir al oligonucleótido puede mencionarse como fragmento de cebador. Cada fragmento de cebador para cada
combinación individual puede comprender una secuencia diferente (por ejemplo, en el marcador celular, el marcador
molecular, el marcador de muestra o cualquier combinación de los mismos). El fragmento de cebador puede
comprender una secuencia que puede hibridar con la secuencia de marcador conector del oligonucleótido (por
ejemplo, el oligonucleótido acoplado al soporte sólido). El fragmento de cebador puede comprender además un
10 segundo marcador celular y una segunda secuencia de marcador conector. La prolongación del cebador puede
usarse para introducir la segunda secuencia de marcador celular y la segunda secuencia de marcador conector en el
oligonucleótido acoplado al soporte sólido (véase la **FIG. 2B**). Después de que la prolongación del cebador incorpore
las nuevas secuencias, los soportes sólidos pueden combinarse. Los soportes sólidos combinados pueden
calentarse para desnaturalizar la enzima. Los soportes sólidos combinados pueden calentarse para alterar la
15 hibridación. Los soportes sólidos combinados pueden dividirse en pocillos de nuevo. El proceso puede repetirse para
añadir secuencias adicionales al oligonucleótido conjugado al soporte sólido.

El proceso de división-combinación puede dar lugar a la creación de al menos aproximadamente 1000, 10 000,
100 000, 500 000 o 1 000 000 o más oligonucleótidos diferentes. El proceso puede dar lugar a la creación de como
mucho aproximadamente 1000, 10 000, 100 000, 500 000 o 1 000 000 o más oligonucleótidos diferentes.

20 La síntesis de división-combinación puede comprender síntesis química. Pueden sintetizarse diferentes
oligonucleótidos usando química de DMT en soportes sólidos en reacciones individuales, después pueden
combinarse en reacciones para la síntesis. El proceso de división-combinación puede repetirse 1, 2, 3, 4, 5, 6, 7, 8,
9, 10 o más veces. El proceso de división-combinación puede repetirse 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30,
35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100 o más veces. El proceso de división-combinación puede
repetirse 2 o más veces. El proceso de división-combinación puede repetirse 3 o más veces. El proceso de división-
combinación puede repetirse 5 o más veces. El proceso de división-combinación puede repetirse 10 o más veces.

En la presente memoria se divulgan además métodos de producción de uno o más conjuntos de microesferas
marcadas (por ejemplo, microesferas conjugadas a oligonucleótido). El método de producción de uno o más
conjunto de microesferas marcadas puede comprender adherir uno o más ácidos nucleicos a una o más
microesferas, produciendo de ese modo uno o más conjuntos de microesferas marcadas. El uno o más ácidos
nucleicos pueden comprender uno o más códigos de barras moleculares. El uno o más ácidos nucleicos pueden
comprender una o más marcas de muestra (por ejemplo, marcadores de muestra, regiones de índice de muestra). El
uno o más ácidos nucleicos pueden comprender uno o más marcadores celulares. El uno o más ácidos nucleicos
pueden comprender uno o más marcadores identificadores moleculares (por ejemplo, marcadores moleculares). El
35 uno o más ácidos nucleicos pueden comprender a) una región de cebador; b) una región de índice de muestra; y c)
una región de conector o adaptador. El uno o más ácidos nucleicos pueden comprender a) una región de cebador;
b) una región de marcador (por ejemplo, marcador molecular); y c) una región de conector o adaptador. El uno
o más ácidos nucleicos pueden comprender a) una región de índice de muestra (por ejemplo, marca de muestra); y b)
una región de marcador (por ejemplo, marcador molecular). El uno o más ácidos nucleicos pueden comprender a)
40 una región de índice de muestra; y b) un marcador celular. El uno o más ácidos nucleicos pueden comprender a) un
marcador celular; y b) un marcador molecular. El uno o más ácidos nucleicos pueden comprender a) una región de
índice de muestra; b) marcador celular; y c) un marcador molecular. El uno o más ácidos nucleicos pueden
comprender además una región de cebador. El uno o más ácidos nucleicos pueden comprender además una región
específica de diana. El uno o más ácidos nucleicos pueden comprender además una región de conector. El uno o
45 más ácidos nucleicos pueden comprender además una región de adaptador. El uno o más ácidos nucleicos pueden
comprender además una región de índice de muestra. El uno o más ácidos nucleicos pueden comprender además
una región de marcador.

Como alternativa, el método comprende: a) depositar una pluralidad de primeros ácidos nucleicos en una pluralidad
de pocillos, en la que dos o más pocillos diferentes de la pluralidad de pocillos pueden comprender dos o más ácidos
nucleicos diferentes de la pluralidad de ácidos nucleicos; b) poner en contacto uno o más pocillos de la pluralidad de
pocillos con una o menos microesferas para producir una pluralidad de microesferas de marcador individuales, en
las que una microesfera de marcador individual de la pluralidad de primeras microesferas marcadas comprende una
microesfera adherida a un ácido nucleico de la pluralidad de primeros ácidos nucleicos; c) combinar la pluralidad de
primeras microesferas marcadas de la pluralidad de pocillos para producir una combinación de primeras
55 microesferas marcadas; d) distribuir la combinación de primeras microesferas marcadas a una pluralidad posterior
de pocillos, en la que dos o más pocillos de la posterior pluralidad de pocillos comprenden dos o más ácidos
nucleicos diferentes de una pluralidad de posteriores ácidos nucleicos; y e) adherir uno o más ácidos nucleicos de la
pluralidad de posteriores ácidos nucleicos a una o más primeras microesferas marcadas para producir una pluralidad
de microesferas marcadas de forma unívoca.

60 Colecciones

En la presente memoria se divulgan métodos de producción de colecciones moleculares. El método puede
comprender: a) marcar de forma estocástica dos o más moléculas de dos o más muestras para producir moléculas
marcadas, en el que las moléculas marcadas comprenden (i) una región molecular basada en o derivada de las dos
o más moléculas, (ii) una región de índice de muestra para su uso en la diferencias de dos o más moléculas de dos

o más muestras; y (iii) una región de marcador para su uso en la diferenciación de dos o más moléculas de una única muestra. El marcaje estocástico puede comprender el uso de uno o más conjuntos de códigos de barras moleculares. El marcaje estocástico puede comprender el uso de uno o más conjuntos de marcas de muestra. El marcaje estocástico puede comprender el uso de uno o más conjuntos de marcadores identificadores moleculares.

- 5 El marcaje estocástico de las dos o más moléculas puede comprender poner en contacto las dos o más muestras con una pluralidad de marcas de muestra y la pluralidad de marcadores específicos de molécula para producir la pluralidad de ácidos nucleicos marcados. El contacto puede ser aleatorio. El método puede comprender además amplificar una o más de las moléculas marcadas, produciendo de ese modo una población enriquecidas de moléculas marcadas de la colección. El método puede comprender además realizar uno o más ensayos sobre las
10 dos o más moléculas de las dos o más muestras. El método puede comprender además realizar uno o más ensayos de arrastre.

El método de producción de una colección de ácidos nucleicos marcados puede comprender además añadir uno o más controles a las dos o más muestras. El uno o más controles pueden marcarse de forma estocástica para producir controles marcados. El uno o más controles pueden usarse para medir la eficacia de producción de las
15 moléculas marcadas.

Las colecciones divulgadas en la presente memoria pueden usarse en una diversidad de aplicaciones. Por ejemplo, la colección puede usarse para aplicaciones de secuenciación. La colección puede almacenarse y usarse múltiples veces para generar muestras para su análisis. Algunas aplicaciones incluyen, por ejemplo, genotipar polimorfismos, estudiar el procesamiento del ARN y seleccionar representantes clonales a secuenciar.

20 Preparación de muestras y aplicaciones

Los oligonucleótidos (por ejemplo, código de barras molecular, marca de muestra, marcador molecular, marcador celular) divulgados en la presente memoria pueden usarse en una diversidad de métodos. Los oligonucleótidos pueden estar en métodos para análisis de ácidos nucleicos. El análisis de ácidos nucleicos puede incluir, aunque sin limitación, genotipado, expresión génica, variación del número de copias y recuento molecular.

- 25 La divulgación proporciona métodos de análisis combinados de ácidos nucleicos. El método puede comprender (a) poner en contacto uno o más oligonucleótidos de una célula con uno o más oligonucleótidos adheridos a un soporte, en el que el uno o más oligonucleótidos adheridos al soporte comprenden (i) una región de marcador celular que comprende dos o más secuencias aleatorias conectadas mediante una secuencia no aleatoria; y (ii) una región de marcador molecular; y (b) realizar uno o más ensayos sobre el uno o más oligonucleótidos de la célula.

- 30 En la presente memoria se divulgan además métodos de producción de colecciones de ácidos nucleicos de células individuales. El método puede comprender (a) poner en contacto uno o más oligonucleótidos de una célula con uno o más oligonucleótidos adheridos a un soporte, en el que el uno o más oligonucleótidos adheridos al soporte comprenden (i) una región de marcador celular que comprende dos o más secuencias aleatorias conectadas mediante una secuencia no aleatoria; y (ii) una región de marcador molecular; y (b) realizar uno o más ensayos
35 sobre el uno o más oligonucleótidos de la célula.

En algunos casos, el método comprende añadir una o más células a una matriz de micropocillos. El número de células a añadir puede determinarse a partir del recuento. El exceso de células o las células no unidas pueden retirarse por lavado usando un tampón (por ejemplo, tampón de disolución salina tamponada con fosfato, HEPES, Tris). El número de células que puede capturarse por las células de la matriz de micropocillos puede estar
40 relacionado con el tamaño de la célula. Por ejemplo, dependiendo del diseño de la matriz, pueden capturarse más fácilmente células más grandes que células más pequeñas, como se representa en la **FIG. 6**. Pueden usarse diferentes micropocillos (por ejemplo, diferentes dimensiones) para capturar diferentes tipos celulares.

Los métodos descritos en esta ocasión permiten la adición de secuencias que pueden ser ácidos nucleicos para secuenciación u otros análisis moleculares. Estos métodos pueden permitir la detección de variantes, mutantes, polimorfismo, inversiones, eliminaciones, reversiones y otros eventos cualitativos de ácidos nucleicos encontrados en una población de moléculas de ARN o ADN. Por ejemplo, los métodos pueden permitir la identificación de frecuencias de dianas (por ejemplo, expresión génica o distribución alélica). Por ejemplo, los métodos también permiten la identificación de mutaciones o SNP en un genoma o transcriptoma, tal como de un sujeto enfermo o no enfermo. Los métodos también permiten la determinación de la presencia o ausencia de contaminación o
50 infecciones en una muestra biológica de un sujeto, tal como organismos o virus exógenos, tales como bacterias o un hongo.

Las células pueden añadirse a los micropocillos mediante cualquier método. En algunas realizaciones, las células se añaden a los micropocillos como una muestra celular diluida. En algunas realizaciones, las células se añaden a los micropocillos y se dejan sedimentar en los micropocillos por gravedad. En algunas realizaciones, las células se añaden a los micropocillos y se usa centrifugación para sedimentar las células en los micropocillos. En algunas realizaciones, las células se añaden a los micropocillos inyectando una o más células en uno o más micropocillos. Por ejemplo, puede añadirse una célula individual a un micropocillo inyectando la célula individual en un micropocillo. La inyección de una célula puede ser mediante el uso de cualquier dispositivo o método, tal como mediante el uso de un micromanipulador. En algunas realizaciones, la célula puede añadirse a los micropocillos usando un imán. Por ejemplo, las células pueden recubrirse en su superficie con partículas magnéticas, tales como micropartículas magnéticas o nanopartículas magnéticas y añadirse a los micropocillos usando un imán o un campo magnético.

La matriz de micropocillos que comprende células puede ponerse en contacto con un soporte sólido conjugado a oligonucleótido (por ejemplo, microesfera). Los soportes sólidos conjugados a oligonucleótido no capturados pueden retirarse (por ejemplo, retirarse por lavado con tampón). La **FIG. 5** representa una matriz de micropocillos con soporte sólidos capturados. Un micropocillo puede comprender al menos un soporte sólido. Un micropocillo puede comprender al menos dos soportes sólidos. Un micropocillo puede comprender como mucho un soporte sólido. Un micropocillo puede comprender como mucho dos soportes sólidos. Un micropocillo puede comprender al menos aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10 o más soportes sólidos. Un micropocillo puede comprender como mucho aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10 o más soportes sólidos. Algunos de los micropocillos de la matriz de micropocillos pueden comprender un soporte sólido y algunos de los micropocillos de la matriz de micropocillos pueden comprender dos o más soportes sólidos, como se muestra en la **FIG. 5**. Puede que el micropocillo no tenga que cubrirse para ninguno de los métodos de la divulgación. En otras palabras, puede que los micropocillos no tengan que sellarse durante el método. Cuando los micropocillos no se cubren (por ejemplo, sellan), los pocillos pueden espaciarse de modo que los contenidos de un micropocillo no puedan difundir en otro micropocillo.

Como alternativa o adicionalmente, las células pueden capturarse y/o purificarse antes de ponerse en contacto con un soporte conjugado a oligonucleótido. Los métodos para capturar y/o purificar células pueden comprender el uso de anticuerpos, estructuras moleculares y/o microesferas. Las células pueden purificarse por citometría de flujo. Pueden usarse kits disponibles en el mercado para capturar o purificar células. Por ejemplo, pueden usarse Dynabeads(R) para aislar células. Puede usarse aislamiento magnético para purificar células. Las células pueden purificarse por centrifugación.

Las células pueden ponerse en contacto con soportes conjugados a oligonucleótido creando una suspensión que comprende células y los soportes. La suspensión puede comprender un gel. Las células pueden inmovilizarse en un soporte o en una disolución antes del contacto con los soportes conjugados a oligonucleótido. Como alternativa las células pueden añadirse a una suspensión que comprende el soporte conjugado a oligonucleótido. Por ejemplo, las células pueden añadirse a un hidrogel que está incrustado con soportes conjugados a oligonucleótido.

Una célula individual puede ponerse en contacto con un único soporte sólido acoplado a oligonucleótido. Una célula individual puede ponerse en contacto con múltiples soportes sólidos conjugados a oligonucleótido. Múltiples células pueden interactuar con un único soporte sólido conjugado a oligonucleótido. Múltiples células pueden interactuar con múltiples soportes sólidos conjugados a oligonucleótido. Los soportes sólidos conjugados a oligonucleótido pueden ser específicos de tipo celular. Como alternativa, el soporte conjugado a oligonucleótido puede interactuar con dos o más tipos celulares diferentes.

Lisis

Las células en los micropocillos pueden lisarse. La lisis puede realizarse mediante lisis mecánica, lisis térmica, lisis óptica y/o lisis química. La lisis química puede incluir el uso de enzimas digestivas tales como proteinasa K, pepsina y tripsina. La lisis puede realizarse mediante la adición de un tampón de lisis a los micropocillos. Un tampón de lisis puede comprender Tris HCl. Un tampón de lisis puede comprender al menos aproximadamente 0,01, 0,05, 0,1, 0,5 o 1 M o más de Tris HCl. Un tampón de lisis puede comprender como mucho aproximadamente 0,01, 0,05, 0,1, 0,5 o 1 M o más de Tris HCl. Un tampón puede comprender aproximadamente 0,1 M de Tris HCl. El pH del tampón de lisis puede ser al menos aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10 o más. El pH del tampón de lisis puede ser como mucho aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10 o más. En algunos casos, el pH del tampón de lisis es aproximadamente 7,5. El tampón de lisis puede comprender una sal (por ejemplo, LiCl). La concentración de sal en el tampón de lisis puede ser al menos aproximadamente 0,1, 0,5 o 1 M o más. La concentración de sal en el tampón de lisis puede ser como mucho aproximadamente 0,1, 0,5 o 1 M o más. En algunos casos, la concentración de sal en el tampón de lisis es aproximadamente 0,5 M. El tampón de lisis puede comprender un detergente (por ejemplo, SDS, dodecilsulfato de Li, triton X, tween, NP-40). La concentración del detergente en el tampón de lisis puede ser al menos aproximadamente de un 0,0001, 0,0005, 0,001, 0,005, 0,01, 0,05, 0,1, 0,5, 1, 2, 3, 4, 5, 6 o 7 % o más. La concentración del detergente en el tampón de lisis puede ser como mucho aproximadamente de un 0,0001, 0,0005, 0,001, 0,005, 0,01, 0,05, 0,1, 0,5, 1, 2, 3, 4, 5, 6 o 7 % o más. En algunos casos la concentración del detergente en el tampón de lisis es de aproximadamente un 1 % de dodecilsulfato de Li. El tiempo usado en el método para la lisis puede depender de la cantidad de detergente usada. En algunos casos, cuanto más detergente se use, menor será el tiempo necesario para la lisis. El tampón de lisis puede comprender un agente quelante (por ejemplo, EDTA, EGTA). La concentración de un agente quelante en el tampón de lisis puede ser al menos aproximadamente 1, 5, 10, 15, 20, 25 o 30 mM o más. La concentración de un agente quelante en el tampón de lisis puede ser como mucho aproximadamente 1, 5, 10, 15, 20, 25 o 30 mM o más. En algunos casos, la concentración de agente quelante en el tampón de lisis es aproximadamente 10 mM. El tampón de lisis puede comprender un reactivo reductor (por ejemplo, betamercaptoetanol, DTT). La concentración del reactivo reductor en el tampón de lisis puede ser al menos aproximadamente 1, 5, 10, 15 o 20 mM o más. La concentración del reactivo reductor en el tampón de lisis puede ser como mucho aproximadamente 1, 5, 10, 15 o 20 mM o más. En algunos casos, la concentración de reactivo reductor en el tampón de lisis es aproximadamente 5 mM. En algunos casos, un tampón de lisis puede comprender aproximadamente 0,1 M de Tris HCl, aproximadamente pH 7,5, aproximadamente 0,5 M de LiCl, aproximadamente un 1 % de dodecilsulfato de litio, aproximadamente 10 mM de EDTA y aproximadamente 5 mM de DTT.

La lisis puede realizarse a una temperatura de aproximadamente 4, 10, 15, 20, 25 o 30 °C. La lisis puede realizarse durante aproximadamente 1, 5, 10, 15 o 20 o más minutos. Una célula lisada puede comprender al menos aproximadamente 100 000, 200 000, 300 000, 400 000, 500 000, 600 000 o 700 000 o más moléculas de ácido nucleico diana. Una célula lisada puede comprender como mucho aproximadamente 100 000, 200 000, 300 000,

ES 2 711 168 T3

400 000, 500 000, 600 000 o 700 000 o más moléculas de ácido nucleico diana. La **FIG. 7** ilustra la estadística ejemplar acerca de la concentración de ácido nucleico diana (es decir, ARNm) que puede obtenerse de la lisis.

Sellado

5 Los micropocillos de la matriz de micropocillos pueden sellarse durante la lisis. El sellado puede ser útil para evitar la hibridación cruzada de ácido nucleico diana entre micropocillos adyacentes. Un micropocillo puede sellarse usando una tapa como se muestra en la **FIG. 8A y B**. Una tapa puede ser un soporte sólido. Una tapa puede comprender una microesfera. El diámetro de la microesfera puede ser más grande que el diámetro del micropocillo. Por ejemplo, una tapa puede ser al menos aproximadamente un 10, 20, 30, 40, 50, 60, 70, 80 o 90 % más grande que el diámetro del micropocillo. Por ejemplo, una tapa puede ser como mucho aproximadamente un 10, 20, 30, 40, 50, 60, 70, 80 o 90 % más grande que el diámetro del micropocillo.

10 Una tapa puede comprender microesferas de dextrano reticulado (por ejemplo, Sephadex). El dextrano reticulado puede variar de aproximadamente 10 micrómetros a aproximadamente 80 micrómetros. El dextrano reticulado de la tapa puede ser de 20 micrómetros a aproximadamente 50 micrómetros. Una tapa puede comprender, por ejemplo, membranas inorgánicas de nanoporos (por ejemplo, óxidos de aluminio), membranas de diálisis, portaobjetos de vidrio, cubreobjetos y/o película de plástico hidrófilo (por ejemplo, película recubierta con una película delgada de agarosa hidratada con tampón de lisis).

15 La tapa puede permitir que el tampón pase a través al interior y a exterior del micropocillo, pero puede evitar que las macromoléculas (por ejemplo, ácido nucleico) migre al exterior del pocillo. Puede bloquearse que una macromolécula de al menos aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 13, 14, 15, 16, 17, 18, 19 o 20 o más nucleótidos migre al interior o el exterior del micropocillo mediante la tapa. Puede bloquearse que una macromolécula de como mucho aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 13, 14, 15, 16, 17, 18, 19 o 20 o más nucleótidos migre al interior o el exterior del micropocillo mediante la tapa.

20 Una matriz de micropocillos sellada puede comprender una única capa de microesferas en la parte superior de los micropocillos. Una matriz de micropocillos sellada puede comprender múltiples capas de microesferas en la parte superior de los micropocillos. La matriz de micropocillos sellada puede comprender aproximadamente 2, 3, 4, 5 o 6 o más capas de microesferas.

25 Depositar una microesfera, una pluralidad de microesferas, en un soporte sólido (por ejemplo, una matriz de micropocillos) puede ser aleatoria o no aleatoria. Por ejemplo, poner en contacto una microesfera con una matriz de micropocillos puede ser un contacto aleatorio o no aleatorio. En algunas realizaciones, la microesfera se pone en contacto con una matriz de micropocillos de forma aleatoria. En algunas realizaciones, la microesfera se pone en contacto con una matriz de micropocillos de forma no aleatoria. Depositar una pluralidad de microesferas en una matriz de micropocillos puede ser aleatoria o no aleatoria. Por ejemplo, el contacto de una pluralidad de microesferas con una matriz de micropocillos puede ser un contacto aleatorio o no aleatorio. En algunas realizaciones, la pluralidad de microesferas se pone en contacto con una matriz de micropocillos de forma aleatoria. En algunas realizaciones, la pluralidad de microesferas se pone en contacto con una matriz de micropocillos de forma no aleatoria.

Marcaje estocástico de moléculas

30 Cuando la marca de muestra o marcador identificador molecular es un oligonucleótido, la adhesión del oligonucleótido a un ácido nucleico puede producirse mediante una diversidad de métodos incluyendo, aunque sin limitación, hibridación del oligonucleótido con el ácido nucleico. En algunos casos, el oligonucleótido comprende una región específica de diana. La región específica de diana puede comprender una secuencia que es complementaria a al menos una parte de la molécula a marcar. La región específica de diana puede hibridar con la molécula, produciendo de ese modo un ácido nucleico marcado. La hibridación del oligonucleótido con el ácido nucleico puede estar seguida de una reacción de prolongación del ácido nucleico. La reacción de prolongación del ácido nucleico puede ser transcripción inversa.

35 La adhesión, como alternativa mencionada como contacto, de la pluralidad de ácidos nucleicos con la marca de muestra puede comprender la hibridación de la marca de muestra con uno o más de la pluralidad de ácidos nucleicos. El contacto de la pluralidad de ácidos nucleicos con la marca de muestra puede comprender la realización de una reacción de prolongación de ácido nucleico. La reacción de prolongación de ácido nucleico puede ser una reacción de transcripción inversa.

40 El contacto de la pluralidad de ácidos nucleicos con el marcador identificador molecular puede comprender la hibridación del marcador identificador molecular con uno o más de la pluralidad de ácidos nucleicos. El contacto de la pluralidad de ácidos nucleicos con el marcador identificador molecular puede comprender la realización de una reacción de prolongación de ácido nucleico. La reacción de prolongación de ácido nucleico puede comprender transcripción inversa.

45 El contacto de la pluralidad de ácidos nucleicos con el marcador identificador molecular puede comprender la hibridación de la marca de muestra con uno o más de la pluralidad de ácidos nucleicos. El contacto de la pluralidad de ácidos nucleicos con el marcador identificador molecular puede comprender la hibridación del marcador identificador molecular con la marca de muestra.

50 El contacto de la pluralidad de ácidos nucleicos con la marca de muestra puede comprender la hibridación del

marcador identificador molecular con uno o más de la pluralidad de ácidos nucleicos. El contacto de la pluralidad de ácidos nucleicos con la marca de muestra puede comprender la hibridación de la marca de muestra con el marcador identificador molecular.

5 La adhesión de la marca de muestra y/o el marcador identificador molecular con un ácido nucleico puede producirse por ligamiento. El contacto de la pluralidad de ácidos nucleicos con la marca de muestra puede comprender el ligamiento de marca de muestra con uno cualquiera de la pluralidad de ácidos nucleicos. El contacto de la pluralidad de ácidos nucleicos con el marcador identificador molecular puede comprender el ligamiento del marcador identificador molecular con uno o más de la pluralidad de ácidos nucleicos. El contacto de la pluralidad de ácidos nucleicos con la marca de muestra puede comprender el ligamiento del marcador identificador molecular con uno o más ácidos nucleicos. El contacto de la pluralidad de ácidos nucleicos con el marcador identificador molecular puede comprender ligamiento de la marca de muestra con uno o más de los ácidos nucleicos. Las técnicas de ligamiento comprenden ligamiento de extremos romos y ligamiento de extremos adherentes. Las reacciones de ligamiento pueden incluir ADN ligasas, tales como ADN ligasa I, ADN ligasa III, ADN ligasa IV y ADN ligasa T4. Las reacciones de ligamiento pueden incluir ARN ligasas tales como ARN ligasa T4 I y ARN ligase T4 II.

15 Se describen métodos de ligamiento, por ejemplo, en Sambrook *et al.* (2001) y el catálogo de New England BioLabs. Los métodos incluyen usar ADN ligasa T4 que cataliza la formación de un enlace fosfodiéster entre los extremos yuxtapuestos de 5' fosfato y 3' hidroxilo en ADN bicatenario o ARN con extremos romos y adhesivos; la ADN ligasa Taq que cataliza la formación de un enlace fosfodiéster entre extremos yuxtapuestos de 5' fosfato y 3' hidroxilo de dos oligonucleótidos adyacentes que hibridan con un ADN diana complementario; ADN ligasa de *E. coli* que cataliza la formación de un enlace fosfodiéster entre extremos yuxtapuestos de 5' fosfato y 3' hidroxilo en ADN bicatenario que contiene extremos cohesivos; y ARN ligasa T4 que cataliza el ligamiento de un donador de ácido nucleico terminado en 5' fosforilo con un aceptador de ácido nucleico terminado en 3' hidroxilo mediante la formación de un enlace fosfodiéster 3'→5', los sustratos incluyen ARN y ADN monocatenario, así como pirofosfatos dinucleosídicos; o cualquier otro método descrito en la técnica. El ADN fragmentado puede tratarse con una o más enzimas, por ejemplo, una endonucleasa, antes del ligamiento de los adaptadores a uno o ambos extremos para facilitar el ligamiento mediante la generación de extremos que son compatibles con el ligamiento.

En algunos casos, ambos extremos del oligonucleótido se adhieren a la molécula. Por ejemplos, ambos extremos del oligonucleótido pueden hibridarse y/o ligarse con uno o más extremos de la molécula. En algunos casos, la adhesión de ambos extremos del oligonucleótido a ambos extremos de la molécula provoca la formación de un ácido nucleico marcado circularizado. Ambos extremos del oligonucleótido también pueden adherirse al mismo extremo de la molécula. Por ejemplo, el extremo 5' del oligonucleótido se liga con el extremo 3' de la molécula y el extremo 3' del oligonucleótido se hibrida con el extremo 3' de la molécula, produciendo un ácido nucleico marcado con una estructura de horquilla en un extremo. En algunos casos, el oligonucleótido se adhiere al centro de la molécula.

En algunos casos, la adhesión del oligonucleótido al ácido nucleico comprende adherir uno o más conectores oligonucleotídicos a la pluralidad de ácidos nucleicos. El método puede comprender además la adhesión de uno o más conectores oligonucleotídicos a los ácidos nucleicos con marca de muestra. El método puede comprender además adherir uno o ms conectores oligonucleotídicos a los ácidos nucleicos marcados. La adhesión de uno o más conectores oligonucleotídicos a un ácido nucleico, marca de muestra o marcador identificador molecular puede comprender el ligamiento de uno o más conectores oligonucleotídicos a un ácido nucleico, marca de muestra o marcador identificador molecular. El uno o más conectores pueden comprender al menos aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 55, 60, 70, 80, 90, 100 nucleótidos. En algunos casos, el conector puede comprender al menos aproximadamente 1000 nucleótidos.

En algunos casos, la adhesión del código de barras molecular a la molécula comprende el uso de uno o más adaptadores. Tal como se usan en la presente memoria, las expresiones "adaptadores" y "regiones adaptadoras" pueden usarse indistintamente. Los adaptadores pueden comprender una región específica de diana, que permite la adhesión del adaptador a la molécula, y una región específica de oligonucleótido, que permite la adhesión del código de barras molecular al adaptador. Los adaptadores pueden comprender además un cebador universal. Los adaptadores pueden comprender además una región de PCR universal. Los adaptadores pueden adherirse a la molécula y/o códigos de barras moleculares mediante métodos que incluyen, aunque sin limitación, hibridación y/o ligamiento.

Los métodos para ligar adaptadores a fragmentos de ácido nucleico son bien conocidos. Los adaptadores pueden ser bicatenarios, monocatenarios o parcialmente monocatenarios. En algunos aspectos, se forman adaptadores a partir de dos oligonucleótidos que tienen una región de complementariedad, por ejemplo, de aproximadamente 10 a 30, o de aproximadamente 15 a 40 bases de complementariedad perfecta; de modo que cuando los dos oligonucleótidos se hibridan juntos forman una región bicatenaria. Opcionalmente, cualquiera o ambos oligonucleótidos pueden tener una región que no es complementaria al otro oligonucleótido y forma un saliente monocatenario en uno o ambos extremos del adaptador. Los salientes monocatenarios pueden ser de aproximadamente 1 a aproximadamente 8 bases, o de aproximadamente 2 a aproximadamente 4. El saliente puede ser complementario al saliente creado por escisión con una enzima de restricción para facilitar el ligamiento "de extremos adherentes". Los adaptadores pueden incluir otros elementos, tales como sitios de unión a cebador y sitios de restricción. En algunos aspectos, el sitio de restricción puede ser para una enzima de restricción de tipo IIS u otra enzima que corte fuera de su secuencia de reconocimiento, tal como EcoP151 (véase, Mucke *et al.* *J Mol Biol* 2001, 312(4):687-698 y el documento US 5710000).

En algunos casos, el recuento estocástico del número de copias de un ácido nucleico en una pluralidad de muestras

comprende detectar el adaptador, un complemento del adaptador, un complemento inverso del adaptador o una parte del mismo para determinar el número de ácidos nucleicos marcados diferentes. La detección del adaptador, un complemento del adaptador, un complemento inverso del adaptador o una parte del mismo puede comprender la secuenciación del adaptador, un complemento del adaptador, un complemento inverso del adaptador o una parte del mismo.

El código de barras molecular puede adherirse a cualquier región de una molécula. Por ejemplo, el código de barras molecular puede adherirse al extremo 5' o 3' de un polinucleótido (por ejemplo, ADN, ARN). Por ejemplo, la región específica de diana del código de barras molecular comprende una secuencia que es complementaria a una secuencia en la región 5' de la molécula. La región específica de diana del código de barras molecular también puede comprender una secuencia que es complementaria a una secuencia en la región 3' de la molécula. En algunos casos, el código de barras molecular se adhiere a una región dentro de un gen o producto génico. Por ejemplo, el ADN genómico se fragmenta y se adhiere una marca de muestra o marcador identificador molecular al ADN fragmentado. En otros casos, una molécula de ARN se corta y empalma de forma alternativa y el código de barras molecular se adhiere a las variantes de corte y empalme alternativo. En otro ejemplo, el polinucleótido se digiere y el código de barras molecular se adhiere al polinucleótido digerido. En otro ejemplo, la región específica de diana del código de barras molecular comprende una secuencia que es complementaria a una secuencia dentro de la molécula.

Un código de barras molecular, marca de muestra (por ejemplo, índice de muestra), marcador celular o marcador identificador molecular (por ejemplo, marcador molecular) que comprende una horquilla puede actuar como sonda para una reacción en cadena de hibridación (HCR) y, por tanto, puede mencionarse como sonda de HCR. La sonda de HCR puede comprender un código de barras molecular que comprende una estructura de horquilla. La sonda de HCR puede comprender una marca de muestra que comprende una estructura de horquilla. La sonda de HCR puede comprender un marcador identificador molecular que comprende una estructura de horquilla. En la presente memoria se divulga además un método de reacción en cadena de hibridación (HCR) basado en marcador estocástico que comprende el marcaje estocástico de una o más moléculas de ácido nucleico con una sonda de HCR, en el que la sonda de HCR comprende un código de barras molecular que comprende una horquilla y la una o más moléculas de ácido nucleico actúan como iniciadores para una reacción en cadena de hibridación. En la presente memoria se divulga además un método de reacción en cadena de hibridación (HCR) basado en marcador estocástico que comprende marcar de forma estocástica una o más moléculas de ácido nucleico con una sonda de HCR, en el que la sonda de HCR comprende una marca de muestra que comprende una horquilla y la una o más moléculas de ácido nucleico actúan como iniciadores para una reacción en cadena de hibridación. En la presente memoria se divulga además un método de reacción en cadena de hibridación (HCR) basado en marcador estocástico que comprende marcar de forma estocástica una o más moléculas de ácido nucleico con una sonda de HCR, en el que la sonda de HCR comprende un marcador identificador molecular que comprende una horquilla y la una o más moléculas de ácido nucleico actúan como iniciadores para una reacción en cadena de hibridación.

La sonda de HCR puede comprender una horquilla con una región saliente. La región saliente de la horquilla puede comprender una región específica de diana. La región saliente puede comprender una secuencia de oligodT. La muestra que comprende la una o más moléculas de ácido nucleico puede tratarse con una o más nucleasas de restricción antes del marcaje estocástico. La región saliente puede comprender una secuencia de reconocimiento por enzima de restricción. La muestra que comprende la una o más moléculas de ácido nucleico puede ponerse en contacto con uno o más adaptadores antes del marcaje estocástico para producir un híbrido de adaptador-molécula de ácido nucleico. La región saliente y el tallo pueden ser complementarios al uno o más adaptadores. La sonda de HCR puede comprender una horquilla con un bucle. El bucle de la sonda de HCR puede comprender una región de marcador y/o región de índice de muestra.

La hibridación de una primera sonda de HCR con las moléculas de ácido nucleico puede provocar la formación de un ácido nucleico marcado, en la que la sonda de HCR se linealiza para producir una primera sonda de HCR linealizada. La primera sonda de HCR linealizada del ácido nucleico marcado puede actuar como iniciador para la hibridación de una segunda sonda de HCR con el ácido nucleico marcado para producir un ácido nucleico marcado con dos sondas de HCR linealizadas. La segunda sonda de HCR linealizada puede actuar como iniciador para otra reacción de hibridación. Este proceso puede repetirse múltiples veces para producir un ácido nucleico marcado con múltiples sondas de HCR linealizadas. Los marcadores detectables en la sonda de HCR pueden posibilitar la detección del ácido nucleico marcado. Los marcadores detectables pueden ser cualquier tipo de marcador (por ejemplo, fluoróforo, cromóforo, molécula pequeña, nanopartícula, hapteno, enzima, anticuerpo, imán). Los marcadores detectables pueden comprender fragmentos de un único marcador. Los marcadores detectables pueden generar una señal detectable cuando están en cercana proximidad. Cuando la sonda de HCR es una horquilla, los marcadores detectables pueden estar demasiado lejos para producir una señal detectable. Cuando la sonda de HCR se linealiza y se hibridan juntas múltiples sondas de HCR linealizadas, los marcadores detectables pueden estar en proximidad suficientemente cercana para generar una señal detectable. Por ejemplo, una sonda de HCR puede comprender dos restos de pireno como marcadores detectables. Como alternativa, los marcadores detectables pueden ser nanopartículas. El método de HCR basado en marcador estocástico puede posibilitar la adhesión de múltiples sondas de HCR de horquilla con un ácido nucleico marcado, que puede provocar amplificación de la señal. La HCR basada en marcador estocástico puede aumentar la sensibilidad de detección, el análisis y/o la cuantificación de las moléculas de ácido nucleico. La HCR basada en marcador estocástico puede aumentar la precisión de la detección, el análisis y/o la cuantificación de una o más moléculas de ácido nucleico.

Después de la lisis, el ácido nucleico diana de las células puede hibridar con el oligonucleótido conjugado con el soporte sólido. El ácido nucleico diana puede hibridar con la región de unión a diana del oligonucleótido. El ácido

nucleico puede hibridar con cualquier región del oligonucleótido.

5 En algunos casos, no todos los oligonucleótidos pueden unirse a un ácido nucleico diana. Esto es porque en algunos casos el número de oligonucleótidos es mayor que el número de ácidos nucleicos diana. El número de oligonucleótidos conjugados a un soporte sólido puede ser 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10 veces mayor que el número de ácidos nucleicos diana en una célula. Al menos un 10, 20, 30, 40, 50, 60, 70, 80, 90 o 100 % de los oligonucleótidos pueden unirse por un ácido nucleico diana. Como mucho un 10, 20, 30, 40, 50, 60, 70, 80, 90 o 100 % de los oligonucleótidos pueden unirse por un ácido nucleico diana. En algunos casos, al menos 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90 o 100 o más ácidos nucleicos diana, diferentes pueden capturarse por los oligonucleótidos en un soporte sólido. En algunos casos, como mucho 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90 o 100 o más ácidos nucleicos diana, diferentes pueden capturarse por los oligonucleótidos en un soporte sólido.

15 En algunos casos, al menos aproximadamente un 40, 50, 60, 70, 80, 90, 95, 96, 97, 98, 99 o 100 % del número de copias de un ácido nucleico diana se unen a oligonucleótidos en un soporte sólido. En algunos casos, como mucho aproximadamente un 40, 50, 60, 70, 80, 90, 95, 96, 97, 98, 99 o 100 % del número de copias de un ácido nucleico diana se unen a oligonucleótidos en un soporte sólido.

Recuperación

20 Después de la lisis, los soportes sólidos pueden recuperarse. La recuperación de los soportes sólidos puede realizarse usando un imán. La recuperación de los soportes sólidos puede realizarse por fusión de la matriz de micropocillos y/o sonicación. La recuperación de los soportes sólidos puede comprender centrifugación. La recuperación de los soportes sólidos puede comprender exclusión molecular. En algunos casos, al menos aproximadamente un 50, 60, 70, 80, 90, 95 o 100 % de los soportes sólidos se recuperan de los micropocillos. En algunos casos, como mucho aproximadamente un 50, 60, 70, 80, 90, 95 o 100 % de los soportes sólidos se recuperan de los micropocillos.

Transcripción inversa

25 Los métodos divulgados en la presente memoria pueden comprender además la transcripción inversa de una molécula de ARN marcada para producir una molécula de ADNc marcada. En algunos casos, al menos una parte del oligonucleótido actúa como cebador para la reacción de transcripción inversa. La parte oligodT del oligonucleótido puede actuar como cebador para la síntesis de la primera hebra de la molécula de ADNc.

30 En algunos casos, la molécula de ADNc marcada puede usarse como una molécula para una nueva reacción de marcaje estocástico. El ADNc marcado puede tener una primera marca o conjunto de marcas de adhesión al ARN antes de la transcripción inversa y una segunda marca o conjunto de marcas adheridas a la molécula de ADNc. Estas múltiples reacciones de marcaje pueden usarse, por ejemplo, para determinar la eficacia de eventos que se producen entre la adhesión de la primera marca y la segunda marca, por ejemplo, una reacción de amplificación opcional o la reacción de transcripción inversa.

35 En otro ejemplo, un oligonucleótido se adhiere al extremo 5' de una molécula de ARN para producir una molécula de ARN marcada. La transcripción inversa de la molécula de ARN marcada puede producirse mediante la adición de un cebador de transcripción inversa. En algunos casos, el cebador de transcripción inversa es un cebador de oligodT, cebador hexanucleotídico aleatorio o un cebador oligonucleotídico específico de diana. En general, los cebadores de oligodT son de 12-18 nucleótidos de longitud y se unen a la cola poli(A)⁺ endógena en el extremo 3' del ARNm de mamífero. Los cebadores hexanucleotídicos aleatorios pueden unirse a ARNm en una diversidad de sitios complementarios. Los cebadores oligonucleotídicos específicos de diana típicamente ceban selectivamente el ARNm de interés.

45 En algunos casos, el método comprende transcribir de forma inversa repetidamente la molécula de ARN marcada para producir múltiples moléculas de ADNc marcadas. Los métodos divulgados en la presente memoria pueden comprender la realización de al menos aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 o 20 reacciones de transcripción inversa. El método puede comprender la realización de al menos aproximadamente 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95 o 100 reacciones de transcripción inversa.

50 La síntesis de ácidos nucleico (por ejemplo, síntesis de ADNc) puede realizarse sobre los soportes sólidos recuperados. La síntesis de ácido nucleico puede realizarse en un tubo y/o en un rotor para mantener los soportes sólidos suspendidos. El ácido nucleico sintetizado resultante puede usarse en posteriores tecnologías de amplificación y/o secuenciación de ácidos nucleicos. La síntesis de ácidos nucleico puede comprender la generación de copias de ADNc sobre un ARN adherido al oligonucleótido en el soporte sólido. La generación de copias de ADNc puede comprender el uso de transcriptasa inversa (RT) o ADN polimerasas que tienen actividad RT. Esto puede provocar la producción de moléculas de ADNc monocatenarias. Después de la síntesis de ácido nucleico, los oligonucleótidos no utilizados pueden retirarse del soporte sólido. La retirada de los oligonucleótidos puede producirse por tratamiento con exonucleasa (por ejemplo, por Exol).

60 En algunas realizaciones, los ácidos nucleicos pueden retirarse del soporte sólido usando escisión química. Por ejemplo, un grupo químico o una base modifica presente de un ácido nucleico puede usarse para facilitar su retirada de un soporte sólido. Por ejemplo, puede usarse una enzima para retirar un ácido nucleico de un soporte sólido. Por ejemplo, un ácido nucleico puede retirarse de un soporte sólido mediante digestión con una endonucleasa de restricción. Por ejemplo, el tratamiento de un ácido nucleico que contiene un dUTP o ddUTP con uracil-d-glucosilasa

(UDG) puede usarse para retirar un ácido nucleico de un soporte sólido. Por ejemplo, un ácido nucleico puede retirarse de un soporte sólido usando una enzima que realiza escisión de nucleótidos, tal como una enzima de reparación de escisión de bases, tal como una endonucleasaapurimica/apirimidinica (AP). En algunas realizaciones, un ácido nucleico puede retirarse de un soporte sólido usando un grupo fotoescindible y luz. En algunas realizaciones, puede usarse un conector escindible para retirar un ácido nucleico del soporte sólido. Por ejemplo, el conector escindible puede comprender al menos uno de biotina/avidina, biotina/estreptavidina, biotina/neutraavidina, Ig-proteína A, un conector fotoinestable, grupo conector inestable en ácido o base, o un aptámero.

En algunas realizaciones, los ácidos nucleicos no se amplifican. En algunas realizaciones, los ácidos nucleicos no se amplifican antes de la secuenciación de los ácidos nucleicos. En algunas realizaciones, los ácidos nucleicos no adheridos a un soporte sólido pueden secuenciarse directamente sin amplificación previa. En algunas realizaciones, los ácidos nucleicos pueden secuenciarse directamente sin realizar amplificación cuando están adheridos a un soporte sólido, por ejemplo, los ácidos nucleicos adheridos a un soporte sólido pueden secuenciarse directamente mientras están adheridos al soporte sólido. En algunas realizaciones, un ácido nucleico que se ha retirado de un soporte sólido puede secuenciarse directamente, por ejemplo, un ácido nucleico que se ha retirado de un soporte sólido puede secuenciarse directamente sin realizar amplificación. Cualquier plataforma de secuenciación que dé lugar a secuenciación sin amplificación puede usarse para realizar la secuenciación.

Amplificación

Después de haberse sintetizado el ácido nucleico (por ejemplo, transcribirse de forma inversa), puede amplificarse. La amplificación puede realizarse de una manera combinada, en la que múltiples secuencias de ácido nucleico diana se amplifican simultáneamente. La amplificación puede añadir adaptadores de secuenciación al ácido nucleico. La amplificación puede realizarse por reacción en cadena de la polimerasa (PCR). PCR puede referirse a una reacción para la amplificación *in vitro* de secuencias de ADN específicas por la prolongación simultánea de cebadores de hebras complementarias de ADN. La PCR puede abarcar formas derivadas de la reacción, incluyendo, aunque sin limitación, RT-PCR, PCR en tiempo real, PCR con cebadores internos, PCR cuantitativa, PCR combinada, PCR digital y PCR de ensamblaje.

El método puede comprender además realizar una o más reacciones de amplificación para producir amplicones de ácido nucleico marcados. Los ácidos nucleicos marcados pueden amplificarse antes de detectar los ácidos nucleicos marcados. El método puede comprender además combinar la primera y segunda muestra antes de realizar la una o más reacciones de amplificación.

Las reacciones de amplificación pueden comprender la amplificación de al menos una parte de la marca de muestra. Las reacciones de amplificación pueden comprender la amplificación de al menos una parte del marcador. Las reacciones de amplificación pueden comprender la amplificación de al menos una parte de la marca de muestra, marcador, ácido nucleico o una combinación de los mismos. Las reacciones de amplificación pueden comprender la amplificación de al menos 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 la pluralidad de ácidos nucleicos. El método puede comprender además la realización de una o más reacciones de síntesis de ADNc para producir una o más copias de ADNc de los ácidos nucleicos con marca de muestra o ácidos nucleicos con marcador identificador molecular.

La amplificación de los ácidos nucleicos marcados puede comprender métodos basados en PCR o métodos no basados en PCR. La amplificación de los ácidos nucleicos marcados puede comprender la amplificación exponencial de los ácidos nucleicos marcados. La amplificación de los ácidos nucleicos marcados puede comprender la amplificación lineal de los ácidos nucleicos marcados.

En algunos casos, la amplificación de los ácidos nucleicos marcados comprende métodos no basados en PCR. Los ejemplos de métodos no basados en PCR incluyen, aunque sin limitación, amplificación de desplazamiento múltiple (MDA), amplificación mediada por transcripción (TMA), amplificación basada en secuencia de ácido nucleico (NASBA), amplificación por desplazamiento de hebra (SDA), SDA en tiempo real, amplificación por círculo rodante o amplificación de círculo a círculo. Otros métodos de amplificación no basados en PCR incluyen múltiples ciclos de amplificación por transcripción de ARN dirigido por ARN polimerasa dependiente de ADN o síntesis de ADN dirigida por ARN y transcripción para amplificar dianas de ADN o ARN (documento WO 89/01050; documento WO 88/10315; y patentes de Estados Unidos n.º 5130238; 5409818; 5466586; 5514545; 5554517; 5888779; 6063603; y 6197554), una reacción en cadena de la ligasa (LCR), un método de replicasa Q β (Q β) como se describe en la patente de Estados Unidos n.º 4786600, uso de sondas palindrómicas, amplificación por desplazamiento de hebra, amplificación dirigida por oligonucleótido usando una endonucleasa de restricción, un método de amplificación en que se hibrida un cebador a una secuencia de ácido nucleico y el dúplex resultante se escinde antes de la reacción de prolongación y la amplificación, amplificación por desplazamiento de hebra usando una polimerasa de ácido nucleico que carece de actividad 5' exonucleasa (patente de Estados Unidos n.º 6214587), amplificación por círculo rodante y amplificación por prolongación de ramificación (RAM) (patente de Estados Unidos n.º 5942391).

La amplificación de los ácidos nucleicos marcados puede comprender métodos basados en reacción en cadena de hibridación (HCR) (Dirks y Pierce, PNAS, 2004; Zhang *et al.*, Anal Chem, 2012). Los métodos basados en HCR pueden comprender HCR basada en ADN. Los métodos basados en HCR pueden comprender una o más sondas macadas. La una o más sondas macadas pueden comprender una o más marcas de muestra o marcadores identificadores moleculares, o el complemento de los mismos, divulgados en la presente memoria.

En algunos casos, los métodos divulgados en la presente memoria comprenden además la realización de una

reacción en cadena de la polimerasa sobre el ácido nucleico marcado (por ejemplo, ARN marcado, ADN marcado, ADNc marcado) para producir un amplicón marcado. El amplicón marcado puede ser una molécula bicatenaria. La molécula bicatenaria puede comprender una molécula de ARN bicatenario, una molécula de ADN bicatenario o una molécula ARN hibridada con una molécula de ADN. Una o las dos hebras de la molécula bicatenaria pueden comprender la marca de muestra o marcador identificador molecular. Como alternativa, el amplicón marcado es una molécula monocatenaria. La molécula monocatenaria puede comprender ADN, ARN o una combinación de los mismos. Los ácidos nucleicos pueden comprender ácidos nucleicos sintéticos o alterados.

La reacción en cadena de la polimerasa puede realizarse por métodos tales como PCR, HD-PCR, PCR de siguiente generación, RTA digital o cualquier combinación de los mismos. Métodos de PCR adicionales incluyen, aunque sin limitación, PCR específica de alelo, Alu PCR, PCR de ensamblaje, PCR asimétrica, PCR de gotitas, PCR en emulsión, amplificación dependiente de helicasa HDA, PCR de inicio caliente, PCR inversa, PCR lineal después de exponencial (LATE), PCR larga, PCR combinada, PCR con cebadores internos, PCR con cebadores semiinternos, PCR cuantitativa, RT-PCR, PCR en tiempo real, PCR de células individuales, PCR de toma de contacto o combinaciones de los mismos.

Las reacciones de PCR combinada pueden comprender reacciones de PCR con cebadores internos. El método puede comprender un par de cebadores en el que un primer cebador que hibrida con uno cualquiera de la pluralidad de ácidos nucleicos al menos 300 a 400 nucleótidos desde el extremo 3' de uno cualquiera de la pluralidad de ácidos nucleicos y un segundo cebador que hibrida con uno cualquiera de la pluralidad de ácidos nucleicos al menos 200 a 300 nucleótidos desde el extremo 3' de uno cualquiera de la pluralidad de ácidos nucleicos, en el que el primer cebador y el segundo cebador generan síntesis de ADN complementario hacia el extremo 3' de uno cualquiera de la pluralidad de ácidos nucleicos.

En algunos casos, la realización de una reacción en cadena de la polimerasa comprende hibridar un primer cebador específico de diana al ácido nucleico marcado. Como alternativa o adicionalmente, la realización de una reacción en cadena de la polimerasa comprende además hibridar un cebador universal a una región de sitio de unión a cebador universal de la marca de muestra o marcador identificador molecular, en el que la marca de muestra o marcador identificador molecular es un ácido nucleico marcado o amplicón marcado. Los métodos divulgados en la presente memoria pueden comprender además la hibridación de un segundo cebador específico de diana al ácido nucleico marcado y/o amplicón marcado.

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

Otros métodos de amplificación adecuados incluyen la reacción en cadena de ligasa (LCR) (por ejemplo, Wu y Wallace, *Genomics* 4, 560 (1989), Landegren *et al.*, *Science* 241, 1077 (1988) y Barringer *et al.* *Gene* 89:117 (1990)), amplificación por transcripción (Kwoh *et al.*, *Proc. Natl. Acad. Sci. USA* 86, 1173 (1989) y documento WO 88/10315), replicación de secuencia automantenida (Guatelli *et al.*, *Proc. Nat. Acad. Sci. USA*, 87, 1874 (1990) y documento WO 90/06995). Ejemplos adicionales son amplificación selectiva de secuencia polinucleotídica diana (patente de Estados Unidos n.º 6410276), reacción en cadena de la polimerasa cebada por secuencia consenso (CP-PCR) (patente de Estados Unidos n.º 4437975), reacción en cadena de la polimerasa cebada arbitrariamente (APPCR) (patentes de Estados Unidos n.º 5413909, 5861245), amplificación por círculo rodante (RCA) (por ejemplo, Fire y Xu, *PNAS* 92:4641 (1995) y Liu *et al.*, *J. Am. Chem. Soc.* 118:1587 (1996)) y patente de Estados Unidos n.º 5648245, amplificación por desplazamiento de hebra (véase, Lasken y Egholm, *Trends Biotechnol.* 2003 21(12):531-5; Barker *et al.* *Genome Res.* 2004 May; 14(5):901-7; Dean *et al.* *Proc Natl Acad Sci USA.* 2002; 99(8):5261-6; Walker *et al.* 1992, *Nucleic Acids Res.* 20(7):1691-6, 1992 y Paez, *et al.* *Nucleic Acids Res.* 2004; 32(9):e71), replicasa Qbeta, descrita en la solicitud de patente PCT n.º PCT/US87/00880 y amplificación de secuencias basada en ácido nucleico (NABSA) (véanse, las patentes de Estados Unidos n.º 5409818, 5554517 y 6063603). Otros métodos de amplificación que pueden usarse se describen en las patentes de Estados Unidos n.º 6582938, 5242794, 5494810, 4988617 y la publicación de Estados Unidos n.º 20030143599. El ADN también puede amplificarse por PCR específica de locus combinada o usando PCR de ligamiento de adaptador y el cebador único (véase Kinzler y Vogelstein, *NAR* (1989) 17:3645-53. Otros métodos disponibles de amplificación, tales como PCR equilibrada (Makrigiorgos, *et al.* (2002), *Nat Biotechnol*, Vol. 20, pág. 936-9), también pueden usarse.

También pueden usarse sondas de inversión molecular ("MIP") para la amplificación de dianas seleccionada. Las MIP pueden generarse de modo que los extremos de la sonda precircular sean complementarios a regiones que flanquean la región a amplificar. El hueco puede cerrarse por prolongación del extremo de la sonda de modo que el complemento de la diana se incorpore en la MIP antes del ligamiento de los extremos para formar un círculo cerrado. El círculo cerrado puede amplificarse y detectarse por secuenciación o hibridación como se divulga previamente en Hardenbol *et al.*, *Genome Res.* 15:269-275 (2005) y patente de Estados Unidos n.º 6858412.

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

La amplificación puede comprender el uso de uno o más nucleótidos no naturales. Los nucleótidos no naturales

pueden comprender nucleótidos fotoinestables y/o activables. Los ejemplos de nucleótidos no naturales incluyen, aunque sin limitación, ácido péptido nucleico (PNA), ácido nucleico morfolino y bloqueado (LNA), así como ácido glicol nucleico (GNA) y ácido treosa nucleico (TNA). Pueden añadirse nucleótidos no naturales a uno o más ciclos de una reacción de amplificación. La adición de los nucleótidos no naturales puede usarse para identificar productos como ciclos específicos o puntos temporales en la reacción de amplificación.

La realización de la una o más reacciones de amplificación puede comprender el uso de uno o más cebadores. El uno o más cebadores pueden comprender uno o más oligonucleótidos. El uno o más oligonucleótidos pueden comprender menos de 12-15 nucleótidos. El uno o más cebadores pueden hibridar con al menos una parte de la pluralidad de ácidos nucleicos marcados. El uno o más cebadores pueden hibridar con el extremo 3' y/o extremo 5' de la pluralidad de ácidos nucleicos marcados. El uno o más cebadores pueden hibridar con una región interna de la pluralidad de ácidos nucleicos marcados. La región interna puede ser de al menos aproximadamente 50, 100, 150, 200, 220, 230, 240, 250, 260, 270, 280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 380, 390, 400, 410, 420, 430, 440, 450, 460, 470, 480, 490, 500, 510, 520, 530, 540, 550, 560, 570, 580, 590, 600, 650, 700, 750, 800, 850, 900 o 1000 nucleótidos desde los extremos 3' de la pluralidad de ácidos nucleicos marcados. 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 oligonucleótidos pueden comprender una secuencia seleccionada de un grupo que consiste en secuencias de la tabla 23. El uno o más cebadores pueden comprender un cebador universal. El cebador universal puede hibridar con un sitio de unión a cebador universal. El uno o más cebadores personalizados pueden hibridar con la primera marca de muestra, la segunda marca de muestra, el marcador identificador molecular, el ácido nucleico o un producto del mismo. El uno o más cebadores pueden comprender un cebador universal y un cebador personalizado. El cebador personalizado puede diseñarse para amplificar uno o más ácidos nucleicos diana. Los ácidos nucleicos diana pueden comprender un subconjunto de los ácidos nucleicos totales en una o más muestras. Los ácidos nucleicos diana pueden comprender un subconjunto de los ácidos nucleicos marcados totales en una o más muestras. El uno o más cebadores pueden comprender al menos 96 o más cebadores personalizados. El uno o más cebadores pueden comprender al menos 960 o más cebadores personalizados. El uno o más cebadores pueden comprender al menos 9600 o más cebadores personalizados. El uno o más cebadores personalizados pueden hibridar con dos o más ácidos nucleicos marcados diferentes. El uno o más ácidos nucleicos marcados diferentes pueden corresponder a uno o más genes.

En la presente memoria se divulga un método de selección de un cebador personalizado que comprende: a) un primer pase, en el que los cebadores elegidos pueden comprender: i) no más de tres guaninas secuenciales, no más de tres citosinas secuenciales, no más de cuatro adeninas secuenciales y no más cuatro timinas secuenciales; ii) al menos 3, 4, 5 o 6 nucleótidos que son guaninas o citosinas; y iii) una secuencia que no forma fácilmente una estructura de horquilla; b) un segundo pase, que comprende: i) una primera ronda de elección de una pluralidad de secuencias que tienen alta cobertura de todos los transcritos; y ii) una o más rondas posteriores, que seleccionan una secuencia que tiene la máxima cobertura de los transcritos restantes y una puntuación de complementariedad con otras secuencias elegidas de no más de 4; y c) añadir secuencias a un conjunto escogido hasta que se sature la cobertura o el número total de cebadores personalizados sea menor de o igual a aproximadamente 96.

El método de selección del cebador personalizado puede comprender además seleccionar el al menos un cebador común basado en uno o más transcritos de ARNm, transcritos no codificantes que incluyen ARN estructurales, seudogenes transcritos, ARNm de molde proporcionado por un proceso de anotación genómica, secuencias correspondientes al cóntigo genómico o cualquier combinación de los mismos.

El método de selección del cebador personalizado puede comprender además un método de selección de cebador que enriquece uno o más subconjuntos de ácidos nucleicos. El uno o más subconjuntos puede comprender ARNm de baja abundancia.

El método de selección del cebador personalizado puede comprender además un algoritmo informático. Los cebadores usados en el método pueden diseñarse con el uso de Primer 3, un programa informático que sugiere secuencias de cebador basadas en una secuencia de entrada definida por el usuario. También pueden usarse otros diseños de cebador, o pueden seleccionarse cebadores a simple vista sin la ayuda de programas informáticos. Hay muchas opciones disponibles con el programa para adaptar el diseño del cebador a la mayoría de aplicaciones. Primer 3 puede considerar muchos factores incluyendo, aunque sin limitación, la temperatura de fusión de oligo, la longitud, el contenido de GC, la estabilidad 3', la estructura secundaria estimada, la probabilidad de hibridación con o amplificación de secuencias indeseables (por ejemplo, repeticiones intercaladas) y la probabilidad de formación de dímeros de cebador entre dos copias del mismo cebador. En el diseño de pares de cebadores, Primer 3 puede considerar el tamaño del producto y la temperatura de fusión, la probabilidad de formación de dímeros de cebador entre los dos cebadores en el par, la diferencia entre las temperaturas de fusión del cebador y la ubicación del cebador respecto a regiones particulares de interés a evitar.

Los métodos, composiciones y kits divulgados en la presente memoria pueden comprender uno o más cebadores divulgados en las tablas 23-24.

Secuenciación

En algunos aspectos, la determinación del número de ácidos nucleicos marcados diferentes puede comprender la

determinación de la secuencia del ácido nucleico marcado o cualquier producto del mismo (por ejemplo, amplicones marcados, moléculas de ADNc marcadas). En algunos casos, un ácido nucleico diana amplificado puede someterse a secuenciación. Determinar la secuencia del ácido nucleico marcado o cualquier producto del mismo puede comprender la realización de una reacción de secuenciación para determinar la secuencia de al menos una parte de la marca de muestra, marcador identificador molecular, al menos una parte del ácido nucleico marcado, un complemento del mismo, un complemento inverso del mismo o cualquier combinación de los mismos. En algunos casos, únicamente la marca de muestra o únicamente la marca de muestra se secuencia. En algunos casos, únicamente el marcador identificador molecular o una parte del marcador identificador molecular se secuencia.

La determinación de la secuencia del ácido nucleico marcado o cualquier producto del mismo puede realizarse por métodos de secuenciación tales como secuenciación de molecular individual Helioscope®, secuenciación de ADN Nanopore, secuenciación masiva de características en paralelo (MPSS) de Lynx Therapeutics, pirosecuenciación 454, secuenciación en tiempo real de moléculas individuales (RNAP), secuenciación Illumina (Solexa), secuenciación SOLiD, Ion Torrent®, secuenciación de semiconductores iónicos, secuenciación de moléculas individuales SMRT®, secuenciación Polony, secuenciación de nanoesferas de ADN y estrategia de VisiGen Biotechnologies. Como alternativa, la determinación de la secuencia del ácido nucleico marcado o cualquier producto del mismo puede usar plataformas de secuenciación incluyendo, aunque sin limitación, Genome Analyzer IIX, HiSeq y MiSeq ofrecidos por Illumina, tecnología en tiempo real de moléculas individuales (SMRT®) tales como el sistema PacBio RS ofrecido por Pacific Biosciences (California) y el secuenciador Solexa, tecnología de secuenciación de moléculas individuales verdadera (tSMS®) tales como el secuenciador HeliScope® ofrecido por Helicos Inc. (Cambridge, MA).

En algunas realizaciones, los ácidos nucleicos marcados comprenden ácidos nucleicos que representa de aproximadamente un 0,01 % de los genes del genoma de un organismo a aproximadamente un 100 % de los genes del genoma de un organismo. Por ejemplo, aproximadamente un 0,01 % de los genes del genoma de un organismo a aproximadamente un 100 % de los genes del genoma de un organismo puede secuenciarse usando una región complementaria a diana que comprende una pluralidad de multímeros capturando los genes que contienen una secuencia complementaria de la muestra. En algunas realizaciones, los ácidos nucleicos marcados comprenden ácidos nucleicos que representan de aproximadamente un 0,01 % de los transcritos del transcriptoma de un organismo a aproximadamente un 100 % de los transcritos del transcriptoma de un organismo. Por ejemplo, aproximadamente un 0,501 % de los transcritos del transcriptoma de un organismo a aproximadamente un 100 % de los transcritos del transcriptoma de un organismo pueden secuenciarse usando una región complementaria a diana que comprende una cola de poli-T capturando los ARNm de la muestra.

En algunos casos, la determinación de la secuencia del ácido nucleico marcado o cualquier producto del mismo comprende secuenciación de extremos emparejados, secuenciación en nanoporos, secuenciación de alto rendimiento, secuenciación por disparo, secuenciación de tinte-terminador, secuenciación de ADN de múltiples cebadores, paseo con cebador, secuenciación con didesoxi de Sanger, secuenciación de Maxim-Gilbert, pirosecuenciación, secuenciación de moléculas individuales verdadera o cualquier combinación de las mismas. Como alternativa, la secuencia del ácido nucleico marcado o cualquier producto del mismo puede determinarse por microscopía electrónica o una matriz transistora de efecto de campo sensible a agentes químicos (chemFET).

La determinación de la secuencia de un ácido nucleico (por ejemplo, ácido nucleico amplificado, ácido nucleico marcado, copia de ADNc de un ácido nucleico marcado, etc.) puede realizarse usando una diversidad de métodos de secuenciación incluyendo, aunque sin limitación, secuenciación por hibridación (SBH), secuenciación por ligamiento (SBL), secuenciación por adición de nucleótidos fluorescentes en aumento cuantitativa (QIFNAS), ligamiento y escisión por etapas, transferencia de energía de resonancia de fluorescencia (FRET), balizas moleculares, digestión de sondas indicadoras TaqMan, pirosecuenciación, secuenciación *in situ* fluorescente (FISSEQ), microesferas FISSEQ, secuenciación de bamboleo, secuenciación combinada, secuenciación de colonias polimerizadas (POLONY); secuenciación por círculo rodante en nanorrejilla (ROLONY), ensayos de ligamiento de oligos específicos de alelo (por ejemplo, ensayo de ligamiento de oligo (OLA), OLA de moléculas de molde individuales usando una sonda lineal ligada y una lectura de amplificación de círculo rodante (RCA), sondas de candado ligadas y/o OLA de moléculas de molde individuales usando una sonda de candado circular ligada y una lectura de amplificación por círculo rodante (RCA)) y similares. También pueden utilizarse métodos de secuenciación de alto rendimiento, tales como secuenciación en matrices cíclicas usando plataformas tales como plataformas Roche 454, Illumina Solexa, ABI-SOLiD, ION Torrents, Complete Genomics, Pacific Bioscience, Helicos, Polonator. La secuenciación puede comprender secuenciación MiSeq. La secuenciación puede comprender secuenciación HiSeq. La secuenciación puede leer el marcador celular, el marcador molecular y/o el gen que estaba en el oligonucleótido original.

En otro ejemplo, la determinación de la secuencia de ácidos nucleicos marcados o cualquier producto de los mismos comprende secuenciación de secuencias de ARN o microARN. Como alternativa, la determinación de la secuencia de ácidos nucleicos marcados o cualquier producto de los mismos comprende técnicas de secuenciación de proteínas tales como degradación de Edman, huella molecular de masas de péptidos, espectrometría de masas o digestión con proteasa.

La reacción de secuenciación puede producirse, en determinadas realizaciones, en un soporte sólido o semisólido, en un gel, en una emulsión, en una superficie, en una microesfera, en una gota, en un flujo continuo, en una dilución o en uno o más volúmenes físicamente diferentes.

La secuenciación puede comprender la secuenciación de al menos aproximadamente 10, 20, 30, 40, 50, 60, 70, 80,

ES 2 711 168 T3

- 5 90, 100 o más nucleótidos o pares de bases del ácido nucleico marcado. En algunos casos, la secuenciación comprende secuenciar al menos aproximadamente 200, 300, 400, 500, 600, 700, 800, 900, 1000 o más nucleótidos o pares de bases del ácido nucleico marcado. En otros casos, la secuenciación comprende secuenciar a menos aproximadamente 1500; 2000; 3000; 4000; 5000; 6000; 7000; 8000; 9000; o 10 000 o más nucleótidos o pares de bases del ácido nucleico marcado.
- 10 La secuenciación puede comprender al menos aproximadamente 200, 300, 400, 500, 600, 700, 800, 900, 1000 o más lecturas de secuenciación por ejecución. En algunos casos, la secuenciación comprende secuenciar al menos aproximadamente 1500; 2000; 3000; 4000; 5000; 6000; 7000; 8000; 9000; o 10 000 o más lecturas de secuenciación por ejecución. La secuenciación puede comprender menos de o igual a aproximadamente 1 600 000 000 lecturas de secuenciación por ejecución. La secuenciación puede comprender menos de o igual a aproximadamente 200 000 000 lecturas por ejecución.
- La determinación del número de ácidos nucleicos marcados diferentes puede comprender una o más matrices.
- La determinación del número de ácidos nucleicos marcados diferentes puede comprender poner en contacto los ácidos nucleicos marcados con la una o más sondas.
- 15 Las sondas, tal como se describen en la presente memoria, pueden comprender una secuencia que es complementaria a al menos una parte del ácido nucleico marcado o amplicón marcado. La pluralidad de sondas puede disponerse sobre el soporte sólido en regiones concretas, en las que una región concreta en el soporte sólido comprende sondas de secuencias idénticas o casi idénticas. En algunos casos, dos o más regiones concretas en el soporte sólido comprenden dos sondas diferentes que comprenden secuencias complementarias a la secuencia de dos regiones identificadoras únicas diferentes de la marca oligonucleotídica.
- 20 En algunos casos, la pluralidad de sondas se hibrida con la matriz. La pluralidad de sondas puede permitir la hibridación de la molécula marcada con la matriz. La pluralidad de sondas puede comprender una secuencia que es complementaria al marcador estocástico oligodT. Como alternativa, o adicionalmente, la pluralidad de sondas comprende una secuencia que es complementaria a la molécula.
- 25 La determinación del número de ácidos nucleicos marcados diferentes puede comprender poner en contacto los ácidos nucleicos marcados con una matriz de una pluralidad de sondas. La determinación del número de ácidos nucleicos marcados diferentes puede comprender el contacto de los ácidos nucleicos marcados con un portaobjetos de vidrio de una pluralidad de sondas.
- 30 La determinación del número de ácidos nucleicos marcados diferentes puede comprender hibridación de sondas marcadas, amplificación específica de diana, secuenciación específica de diana, secuenciación con nucleótidos marcados específicos de polimorfismo nucleotídico pequeño diana, secuenciación con nucleótidos marcados específicos de patrones de digestión con enzima de restricción, secuenciación con nucleótidos marcados específicos de mutaciones o una combinación de los mismos.
- 35 La determinación del número de ácidos nucleicos marcados diferentes puede comprender clasificación por citometría de flujo de un marcador específico de secuencia. La determinación del número de ácidos nucleicos marcados diferentes puede comprender la detección de los ácidos nucleicos marcados adheridos a las microesferas. La detección de los ácidos nucleicos marcados adheridos a las microesferas puede comprender la detección de fluorescencia.
- 40 La determinación del número de ácidos nucleicos marcados diferentes puede comprender el recuento de la pluralidad de ácidos nucleicos marcados por transferencia de energía de resonancia de fluorescencia (FRET), entre una sonda específica de diana y un ácido nucleico marcado o una sonda marcada específica de diana.
- Detección de ácidos nucleicos marcados
- 45 Los métodos divulgados en la presente memoria pueden comprender además la detección de los ácidos nucleicos marcados y/o amplicones marcados. La detección de los ácidos nucleicos marcados y/o amplicones marcados puede comprender la hibridación de los ácidos nucleicos marcados con la superficie, por ejemplo, un soporte sólido. El método puede comprender además la inmunoprecipitación de una secuencia diana con una proteína de unión a ácido nucleico. La detección de los ácidos nucleicos marcados y/o amplicones marcados puede posibilitar o ayudar a determinar el número de ácidos nucleicos marcados diferentes.
- 50 En algunos casos, el método comprende además poner en contacto los ácidos nucleicos marcados y/o amplicones marcados con un marcador detectable para producir un ácido nucleico marcado conjugado a marcador detectable. Los métodos divulgados en la presente memoria pueden comprender además la detección del ácido nucleico marcado conjugado a marcador detectable. La detección de los ácidos nucleicos marcados o cualquier producto de los mismos (por ejemplo, amplicones marcados, ácido nucleico marcado conjugado a marcador detectable) puede comprender la detección de al menos una parte de la marca de muestra o marcador identificador molecular, molécula, marcador detectable, un complemento de la marca de muestra o marcador identificador molecular, o complemento de la molécula o cualquier combinación de los mismos.
- 55 La detección de los ácidos nucleicos marcados o cualquier producto de los mismos puede comprender una emulsión o una gota. Por ejemplo, los ácidos nucleicos marcados o cualquier producto de los mismos pueden estar en una emulsión o gota. Una gota puede ser un volumen pequeño de un primer líquido que está encapsulado por un

segundo líquido inmiscible, tal como una fase continua de una emulsión (y/o por una gota más grande). El volumen de una gota, y/o el promedio del volumen de gotas en una emulsión, por ejemplo, puede ser menor de aproximadamente un microlitro (o entre aproximadamente un microlitro y un nanolitro o entre aproximadamente un microlitro y un picolitro), menos de aproximadamente un nanolitro (o entre aproximadamente un nanolitro y un picolitro) o menos de aproximadamente un picolitro (o entre aproximadamente un picolitro y un femtolitro), entre otros. Una gota (o gotas de una emulsión) puede tener un diámetro (o un promedio de diámetro) de menos de aproximadamente 1000, 100 o 10 micrómetros, o aproximadamente de 1000 a 10 micrómetros, entre otros. Una gota puede ser esférica o no esférica. Pueden generarse gotas que tienen un promedio de diámetro de aproximadamente, de menos de aproximadamente o de más de 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 gotas pueden tener un promedio de diámetro de aproximadamente 0,001 a aproximadamente 500, de aproximadamente 0,01 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. Una gota puede ser una gota simple o una gota compuesta. El término emulsión, tal como se usa en la presente memoria, puede referirse a una mezcla de líquidos inmiscibles (tal como aceite y agua). Las emulsiones de fase oleosa y/o de agua en aceite permiten la compartimentación de las mezclas de reacción dentro de gotas acuosas. Las emulsiones pueden comprender gotas acuosas dentro de una fase oleosa continua. Las emulsiones proporcionadas en la presente memoria pueden ser emulsiones de aceite en agua, en las que las gotas son gotas de aceite dentro de una fase acuosa continua. Cuando se usa una emulsión o gota para aislar, por ejemplo, aislar espacialmente células individuales, puede no usarse un soporte sólido. Por tanto, los ácidos nucleicos a marcar y analizar pueden no unirse a un soporte sólido y, en dichos casos, un marcador celular puede corresponder a la célula individual o población de células presente en la emulsión o gota cuando se marca. La emulsión o gota, por tanto, puede aislar de forma eficaz las etapas de marcado o marcaje con una célula individual o pluralidad de células y el marcador celular puede usarse para identificar los ácidos nucleicos que provienen de la célula individual o pluralidad de células. En algunas realizaciones, pueden aplicarse gotas a micropocillos, por ejemplo, de forma similar a la aplicación de microesferas a matrices de micropocillos.

Como alternativa, la detección de los ácidos nucleicos marcados y cualquier producto de los mismos comprende una o más disoluciones. En otros casos, la detección de los ácidos nucleicos marcados comprende uno o más recipientes.

La detección de los ácidos nucleicos marcados o cualquier producto de los mismos (por ejemplo, amplicones marcados, ácido nucleico marcado conjugado con marcador detectable) puede comprender la detección de cada ácido nucleico marcado o productos de los mismos. Por ejemplo, los métodos divulgados en la presente memoria comprenden la secuenciación de al menos una parte de cada ácido nucleico marcado, detectado de ese modo cada ácido nucleico marcado.

En algunos casos, la detección de los ácidos nucleicos marcados y/o amplicones marcados comprende métodos de electroforesis, espectroscopia, microscopia, quimioluminiscencia, luminiscencia, fluorescencia, inmunofluorescencia, colorimetría o electroquimioluminiscencia. Por ejemplo, el método comprende la detección de un tinte fluorescente. La detección del ácido nucleico marcado o cualquier producto de los mismos puede comprender métodos colorimétricos. Por ejemplo, el método colorimétrico comprende el uso de un colorímetro o un lector colorimétrico. Una lista no limitante de colorímetros y lectores colorimétricos incluye el lector de imágenes de matrices colorimétricas de Sensovation (CLAIR), lector de inmunoensayo de flujo lateral ESEQuant, SpectraMax 340PC 38, SpectraMax Plus 384, SpectraMax 190, VersaMax, VMax, y EMax.

Métodos adicionales usados en solitario o en combinación con otros métodos para detectar ácidos nucleicos y/o amplicones marcados pueden comprender el uso de un detector de matriz, lector de fluorescencia, detector no fluorescente, lector de CR, luminómetro o escáner. En algunos casos, la detección de los ácidos nucleicos marcados y/o amplicones marcados comprende el uso de un detector de matriz. Ejemplos de detectores de matriz incluyen, aunque sin limitación, detectores de serie de diodos, detectores de serie de fotodiodos, detectores de serie de fotodiodos de HPLC, detectores de serie, detectores de serie de germanio, detectores de serie de CMOS y CCD, detectores de serie de CCD lineales sincronizados, sistemas de serie de fotodiodos InGaAs y sistemas de CCD refrigerados con TE. El detector de matriz puede ser un detector de micromatriz. Ejemplos no limitantes de detectores de micromatriz incluyen detectores de serie de microelectrodos, plataformas de detección de micromatriz de ADN óptica, detectores de micromatriz de ADN, detectores de micromatriz de ARN y detectores de micromatriz de proteínas.

En algunos casos, se usa un lector de fluorescencia para detectar el ácido nucleico marcado y/o amplicones marcados. El lector de fluorescencia puede leer 1, 2, 3, 4, 5 o más micromatrices de fluorescencia de color u otras estructuras en biochips, en portaobjetos o en microplacas. En algunos casos, el lector de fluorescencia es un lector de imágenes de matriz de fluorescencia Sensovation (FLAIR). Como alternativa, el lector de fluorescencia es un lector de microplaca de fluorescencia tal como el lector de microplaca de fluorescencia Gemini XPS, lector de microplaca de fluorescencia Gemini EM, lector de microplaca de fluorescencia basado en filtro Fluoroskan Finstruments®, lector de microplaca PHERAstar, lector de microplaca FIUOstar, lector de microplaca POLARstar Omega, lector de microplaca de multimodo FLUOstar OPTIMA y lector de microplaca de multimodo POLARstar ÓPTIMA. Ejemplos adicionales de lectores de fluorescencia incluyen los sistemas PharosFXTM y PharosFX Plus.

En algunos casos, la detección del ácido nucleico marcado y/o amplicón marcado comprende el uso de un lector de microplaca. En algunos casos, el lector de microplaca es un espectrofotómetro de absorbancia de microplaca xMarkTM, lector de absorbancia de microplaca iMark, lector de placa de multimodo EnSpire®, lector de placa de

ES 2 711 168 T3

multimarcador EnVision, lector de placa de multimarcador VICTOR X, FlexStation, SpectraMax Paradigm, SpectraMax M5e, SpectraMax M5, SpectraMax M4, SpectraMax M3, SpectraMax M2-M2e, serie FilterMax F, fluorímetro y luminómetro de microplaca Fluoroskan Ascent FL, fluorímetro de microplaca Fluoroskan Ascent, luminómetro de microplaca Luminoskan Ascent, fotómetro de microplaca Multiskan EX, fotómetro de microplaca Muliskan FC y fotómetro de microplaca Muliskan GO. En algunos casos, el lector de microplaca detecta la absorbancia, fluorescencia, luminiscencia, fluorescencia resuelta en el tiempo, dispersión de luz o cualquier combinación de los mismos. En algunas realizaciones, el lector de microplaca detecta la dispersión de luz dinámica. El lector de microplaca puede detectar, en algunos casos, la dispersión de luz estática. En algunos casos, la detección de los ácidos nucleicos marcados y/o amplicones marcados comprende el uso de un generador de imágenes de microplaca. En algunos casos, el generador de imágenes de microplaca comprende el generador de imágenes de microplaca ViewLux uHTS y el sistema de imágenes de microplaca BioRad.

La detección de ácidos nucleicos marcados y/o productos de los mismos puede comprender el uso de un luminómetro. Ejemplos de luminómetros incluyen, aunque sin limitación, SpectraMax L, luminómetro de microplaca GloMax®-96, luminómetro de un único tubo GloMax®-20/20, GloMax®-Multi+ con programa informático Instinct™, lector de multimodo de un único tubo GloMax®-Multi Jr, LUMIstar OPTIMA, luminómetro LEADER HC+, luminómetro LEADER 450i y luminómetro LEADER 50i.

En algunos casos, la detección de los ácidos nucleicos marcados y/o amplicones marcados comprende el uso de un escáner. Los escáneres incluyen escáneres de lecho plano tales como los proporcionados por Cannon, Epson, HP, Fujitsu y Xerox. Ejemplos adicionales de escáneres de lecho plano incluyen los escáneres de imágenes de fluorescencia FMBIO® (por ejemplo, sistemas FMBIO® II, III y III Plus). Los escáneres pueden incluir escáneres de microplaca tales como el escáner de microplaca de micromatriz Arrayit Pix™. En algunos casos, el escáner es un sistema Personal Molecular Imager™ (PMI) proporcionado por Bio-rad.

La detección del ácido nucleico marcado puede comprender el uso de una técnica analítica que mide la relación de masa a carga de partículas cargadas, por ejemplo, espectrometría de masas. En algunas realizaciones, la relación de masa a carga de partículas cargadas se mide en combinación con técnicas de separación cromatográfica. En algunas realizaciones, las reacciones de secuenciación se usan en combinación con la relación de masa a carga de mediciones de partículas cargadas. En algunas realizaciones, las marcas comprenden isótopos. En algunas realizaciones, el tipo de isótopo o relación se controla o manipula en la colección de marcas.

La detección de los ácidos nucleicos marcados o cualquier producto de los mismos, comprende el uso de partículas pequeñas y/o dispersión de luz. Por ejemplo, las moléculas amplificadas (por ejemplo, amplicones marcados) se adhieren a haptenos o directamente a partículas pequeñas y se hibridan a la matriz. Las partículas pequeñas pueden estar en el intervalo de tamaño nanométrico a micrométrico. Las partículas pueden detectarse cuando se dispersa luz de su superficie.

Puede usarse un ensayo colorimétrico donde las partículas pequeñas están coloreadas, o pueden teñirse haptenos con sistemas de detección colorimétrica. En algunos casos, puede usarse un escáner de lecho plano para detectar la luz dispersada de las partículas, o el desarrollo de materiales coloreados. Los métodos divulgados en la presente memoria pueden comprender además el uso de un material que absorbe luz. El material que absorbe luz puede usarse para bloquear la dispersión o el reflejo de luz indeseable. El material que absorbe luz puede ser un colorante alimenticio u otro material. En algunos casos, la detección del ácido nucleico marcado o cualquier producto del mismo comprende poner en contacto los ácidos nucleicos marcados con una luz blanca fuera de eje.

En algunas realizaciones, pueden detectarse dos o más tipos de diferentes de materiales biológicos de una muestra simultáneamente. Por ejemplo, pueden detectarse dos o más tipos diferentes de materiales biológicos seleccionados del grupo que consiste en ADN, ARN (por ejemplo, microARN, ARNm, etc.), nucleótido, proteína y carbohidrato, de una muestra simultáneamente. Por ejemplo, puede detectarse ADN y ARN de una muestra simultáneamente usando los métodos descritos en la presente memoria.

Análisis de datos

Los datos de secuenciación pueden usarse para contar el número de moléculas de ácido nucleico diana en una célula. Por ejemplo, una pluralidad de copias de un ácido nucleico diana en una célula puede unirse a un oligonucleótido diferente en el soporte sólido. Cuando la pluralidad de ácidos nucleicos diana se amplifica y secuencia, puede comprender diferentes marcadores moleculares. El número de marcadores moleculares para un mismo ácido nucleico diana puede ser indicativo del número de copias del ácido nucleico diana en la célula. La determinación del número de copias de un ácido nucleico diana puede ser útil para eliminar el sesgo de amplificación cuando se determina la concentración de un ácido nucleico diana en una célula.

Los datos de secuenciación pueden usarse para genotipar a un sujeto. Comparando los ácidos nucleicos diana con diferentes marcadores celulares, puede determinarse la variación del número de copias y/o la concentración del ácido nucleico diana. Comparando las concentraciones de ácidos nucleicos diana con diferentes marcadores celulares, los datos de secuenciación pueden usarse para determinar la heterogeneidad del genotipo celular. Por ejemplo, una primera célula de una muestra puede comprender un ácido nucleico diana a altas concentraciones, mientras que una segunda célula de la muestra puede no comprender el ácido nucleico diana o puede comprender el ácido nucleico diana a bajas concentraciones, indicando de ese modo la heterogeneidad de la muestra celular.

Determinar la heterogeneidad del genotipo celular puede ser útil para diagnosticar, pronosticar y determinar un ciclo de tratamiento de una enfermedad. Por ejemplo, si una primera célula de una muestra comprende el ácido nucleico

diana, pero una segunda célula de la muestra no comprende el ácido nucleico diana, pero comprende un segundo ácido nucleico diana, entonces un ciclo de tratamiento puede incluir un agente (por ejemplo, fármaco) para abordar el primer genotipo y un agente (por ejemplo, fármaco) para abordar el segundo genotipo.

5 En algunas realizaciones, determinados tipos de secuencia pueden ligarse a un perfil de ADN o ARN. Por ejemplo, las secuencias de receptor de linfocitos T y/o receptor de linfocitos B pueden ligarse a un perfil de transcripción, perfil de microARN o perfil de mutación genómica de una muestra, tal como una célula individual. En algunas realizaciones, determinados tipos de secuencia pueden ligarse a un perfil de antigenicidad o expresión de proteína. Por ejemplo, secuencias de receptor de linfocitos T y/o receptor de linfocitos B pueden ligarse a un perfil de antigenicidad o expresión de proteína mediante anticuerpos de unión a una superficie, tal como una superficie que comprende proteínas, tales como dianas proteínicas de anticuerpos que comprenden las secuencias de receptor de linfocitos T y/o receptor de linfocitos B.

10 En algunas realizaciones, la presencia o ausencia de una secuencia, tal como una secuencia vírica, puede ligarse a un perfil de ADN o ARN. Por ejemplo, la presencia o ausencia de una secuencia, tal como una secuencia vírica, puede ligarse a un perfil de transcripción, perfil de microARN o perfil de mutación genómica de una muestra, tal como una célula individual.

Kits

La presente divulgación proporciona kits para realizar los métodos de la divulgación. Un kit puede comprender uno o más de: una matriz de micropocillos, un oligonucleótido y un soporte sólido. Un kit puede comprender un reactivo para reconstituir y/o diluir los oligonucleótidos y/o soporte sólido. Un kit puede comprender reactivos para conjugar los oligonucleótidos al soporte sólido. Un kit puede comprender además uno o más reactivos adicionales, donde dichos reactivos adicionales pueden seleccionarse de: un tampón de lavado; un reactivo de control, un agente de amplificación para amplificar (por ejemplo, realizar síntesis de ADNc y PCR) un ácido nucleico diana y un agente de conjugación para conjugar un oligonucleótido al soporte sólido. Los componentes de un presente kit pueden estar en recipientes diferentes o pueden combinarse en un único recipiente.

20 Un kit puede comprender instrucciones para usar los componentes del kit para practicar los presente métodos. Las instrucciones para practicar los presentes métodos pueden registrarse en un medio de registro adecuado. Por ejemplo, las instrucciones pueden estar impresas en un sustrato, tal como papel o plástico, etc. Por tanto, las instrucciones pueden estar presentes en los kits como un prospecto, en la etiqueta del recipiente o del kit o componentes del mismo (es decir, asociadas con el envase o subenvase), etc. En algunas realizaciones, las instrucciones pueden estar presentes como un archivo de datos de almacenamiento electrónico presente en un medio de almacenamiento legible por ordenador adecuado, por ejemplo, CD-ROM, disquete, unidad flash, etc. En algunas realizaciones, las instrucciones reales pueden no estar presentes en el kit, sino que se proporciona un medio para obtener las instrucciones de una fuente remota, por ejemplo, mediante internet. Por ejemplo, un kit puede comprender una dirección web donde pueden verse las instrucciones y/o de la que pueden descargarse las instrucciones. Como con las instrucciones, este método para obtener las instrucciones está registrado en un sustrato adecuado.

25 En la presente memoria se divulgan además kits para su uso en el análisis de dos o más moléculas de dos o más muestras. Los kits divulgados en la presente memoria pueden comprender una pluralidad de microesferas, un cebador y agentes de amplificación suficientes para procesar al menos aproximadamente 384 muestras. Una cualquiera de las muestras puede comprender una célula individual. La amplificación de ácidos nucleicos puede producir una medición de aproximadamente 1, 2, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 200, 300, 400, 500, 600, 700, 800, 900 o 100 ácidos nucleicos diana en una muestra. La amplificación de ácidos nucleicos puede producir una medición de aproximadamente 1000 ácidos nucleicos diana en una muestra. La amplificación de ácidos nucleicos puede producir una medición de aproximadamente 100 ácidos nucleicos diana en una muestra. La amplificación de ácidos nucleicos puede producir una medición de aproximadamente un 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 100 % de ácidos nucleicos totales en células individuales. La amplificación de ácidos nucleicos puede producir una medición global de todas las secuencias de ácidos nucleico en células individuales. La amplificación de ácidos nucleicos puede producir una medición de secuencias de ácidos nucleico diana en células individuales por secuenciación. La amplificación de ácidos nucleicos puede producir una medición de secuencias de ácidos nucleico diana en células individuales mediante una matriz.

30 Los agentes de amplificación pueden comprender un panel fijo de cebadores. Los agentes de amplificación pueden comprender al menos un par de cebadores personalizados. Los agentes de amplificación pueden comprender al menos un par de cebadores de control. Los agentes de amplificación pueden comprender al menos un par de cebadores de genes constitutivos. Los agentes de amplificación adecuados pueden comprender una mezcla maestra de PCR. El kit puede comprender además instrucciones para el diseño y optimización de cebadores. El kit puede comprender además una placa de micropocillos, en el que la placa de micropocillos puede comprender al menos un pocillo en que se distribuye no más de una microesfera. El kit puede comprender además uno o más recipientes adicionales. El uno o más recipientes adicionales pueden comprender una o más pluralidades adicionales de marcas de muestra. La pluralidad de una o más marcas de muestra adicionales en el uno o más recipientes adicionales es diferentes de la primera pluralidad de marcas de muestra en el primer recipiente. El uno o más recipientes adicionales pueden comprender uno o más marcadores identificadores moleculares adicionales. El uno o más marcadores identificadores moleculares adicionales del uno o más recipientes adicionales son iguales al uno o más marcadores identificadores moleculares del segundo recipiente.

ES 2 711 168 T3

Los métodos y kits divulgados en la presente memoria pueden comprender el uso de una o más puntas de pipeta y/o recipientes (por ejemplo, tubos, viales, placas de multipocillos, placas de micropocillos, tubos Eppendorf, portaobjetos de vidrio, microesferas). En algunos casos, las puntas de pipetas son puntas de pipeta de baja unión. Como alternativa, o adicionalmente, los recipientes pueden ser recipientes de baja unión. Las puntas de pipeta de baja unión y los recipientes de baja unión pueden tener filtración y/o posterior degradación de la muestra reducidas asociadas con puntas basadas en silicona y recipientes que no son de baja unión. Las puntas de pipeta de baja unión y los recipientes de baja unión pueden tener unión a muestra reducida en comparación con las puntas de pipeta y recipientes que no son de baja unión. Los ejemplos de puntas de baja unión incluyen, aunque sin limitación, puntas de baja unión Corning® DeckWorks™ y puntas graduadas de baja unión Avant Premium. Una lista no limitante de recipientes de baja unión incluye tubos de microcentrifuga de baja unión Corning® Costar® y tubos de PCR y tubos de microcentrifuga de baja unión Cosmohand.

Cualquiera de los kits divulgados en la presente memoria puede comprender además un programa informático. Por ejemplo, un kit puede comprender un programa informático para analizar secuencias, tales como códigos de barras o secuencias diana. Por ejemplo, un kit puede comprender un programa informático para analizar secuencias, tales como códigos de barras o secuencias diana para contar moléculas diana únicas, tales como moléculas diana únicas de una célula individual. Por ejemplo, un kit puede comprender un programa informático para analizarse secuencias, tales como códigos de barras o secuencias diana para contar moléculas diana únicas, tales como moléculas diana únicas de un gen, tal como un gen de una célula individual.

Micropocillos y matrices de micropocillos

En algunos casos, los métodos de la divulgación proporcionan el contacto de un soporte sólido que comprende un oligonucleótido conjugado con una célula. La etapa de contacto puede realizarse en una superficie. Superficies ejemplares pueden incluir un micropocillo, un tubo, un matraz y chip. En algunos casos, la superficie comprende un micropocillo. En algunos casos, el micropocillo es parte de una matriz de micropocillos.

Los micropocillos de una matriz de micropocillos pueden ser de un tamaño y forma que pueden contener al menos 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10 o más células por micropocillo. Los micropocillos pueden ser de un tamaño y forma que puede contener como mucho 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10 o más células por micropocillo. Los micropocillos de una matriz de micropocillos pueden ser de un tamaño y forma que pueden contener al menos 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10 o más soportes sólidos por micropocillo. Los micropocillos pueden ser de un tamaño y forma que pueden contener como mucho 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10 o más soportes sólidos por micropocillo. Un micropocillo puede comprender como mucho una célula y un soporte sólido. Un micropocillo puede comprender como mucho una célula y dos soportes sólidos. Un micropocillo puede comprender al menos una célula y como mucho un soporte sólido. Un micropocillo puede comprender al menos una célula y como mucho dos soportes sólidos.

Los micropocillos en la matriz de micropocillos pueden estar dispuestos de forma horizontal. Los micropocillos pueden estar dispuestos de forma vertical. Los micropocillos pueden estar dispuestos con un espaciado igual o casi igual. La matriz de micropocillos puede tener marcadores asociados con uno o más micropocillos. Por ejemplo, los micropocillos de la matriz de micropocillos pueden estar divididos en grupos cada uno compuesto de un número prescrito de micropocillos. Estos grupos pueden proporcionarse en la superficie principal del sustrato. Pueden proporcionarse marcadores de modo que pueda determinarse la posición de cada grupo. Un marcador puede ser detectable a simple vista. Un marcador puede ser un marcador que requiere lentes para verlo (por ejemplo, marcador fluorescente, marcador de emisión, marcador de UV).

Una matriz de micropocillos puede comprender al menos aproximadamente 96, 384, 1000, 5000, 10 000, 15 000, 10 0000, 150 000, 500 000, 1 000 000 o 5 000 000 o más micropocillos. Una matriz de micropocillos pueden comprender como mucho aproximadamente 96, 384, 1000, 5000, 10 000, 15 000, 10 0000, 150 000, 500 000, 1 000 000 o 5 000 000 o más micropocillos.

La forma del micropocillo puede ser cilíndrica. La forma del micropocillo puede ser no cilíndrica, tal como un poliedro compuesto de múltiples caras (por ejemplo, un paralelepípedo, columna hexagonal o columna octogonal), un cono invertido, una pirámide invertida (pirámide triangular invertida, pirámide cuadrada invertida, pirámide pentagonal invertida, pirámide hexagonal invertida o una pirámide poligonal invertida con siete o más ángulos). El micropocillo puede comprender una forma que combina dos o más de estas formas. Por ejemplo, puede ser parcialmente cilíndrico, teniendo el resto la forma de un cono invertido. La forma del micropocillo puede ser una en que una parte de la parte superior de un cono invertido o pirámide invertida está cortada. La boca del micropocillo puede estar en la parte superior del micropocillo o el fondo del micropocillo. El fondo del micropocillo puede ser plano, pero también son posibles superficies curvadas (por ejemplo, convexa a cóncava). La forma y tamaño del micropocillo pueden determinarse considerando el tipo de célula y/o sustrato sólido (por ejemplo, forma, tamaño) a almacenar en el micropocillo.

El diámetro del micropocillo puede referirse al círculo más grande que puede inscribirse en la forma plana del micropocillo. El diámetro del micropocillo puede ser al menos aproximadamente 0,1, 0,5, 1, 2 o 3 veces o más el diámetro de la célula y/o soporte sólido a contener en el micropocillo. El diámetro del micropocillo puede ser como mucho aproximadamente 0,1, 0,5, 1, 2 o 3 veces o más el diámetro de la célula y/o soporte sólido a contener en el micropocillo. El diámetro del micropocillo puede ser al menos aproximadamente un 10, 20, 30, 40 o 50 % o más del diámetro del soporte sólido. El diámetro del micropocillo puede ser como mucho de aproximadamente un 10, 20, 30, 40 o 50 % o más del diámetro del soporte sólido. El diámetro del micropocillo puede ser de al menos aproximadamente 5, 10, 15, 20, 25, 30, 35, 40, 45 o 50 o más micrómetros. El diámetro del micropocillo puede ser

ES 2 711 168 T3

como mucho de aproximadamente 5, 10, 15, 20, 25, 30, 35, 40, 45 o 50 o más micrómetros. El diámetro del micropocillo es de aproximadamente 25 micrómetros. En algunos casos, el diámetro del micropocillo es de aproximadamente 30 micrómetros. En algunos casos, el diámetro del micropocillo es de aproximadamente 28 micrómetros.

- 5 La diferencia entre el volumen del micropocillo y el volumen del soporte sólido puede ser de al menos aproximadamente $1 \times 10^{-14} \text{ m}^3$, $1,5 \times 10^{-14} \text{ m}^3$, $1,7 \times 10^{-14} \text{ m}^3$, $2,0 \times 10^{-14} \text{ m}^3$, $2,5 \times 10^{-14} \text{ m}^3$, o $3,0 \times 10^{-14} \text{ m}^3$ o más. La diferencia entre el volumen del micropocillo y el volumen del soporte sólido puede ser como mucho de aproximadamente $1 \times 10^{-14} \text{ m}^3$, $1,5 \times 10^{-14} \text{ m}^3$, $1,7 \times 10^{-14} \text{ m}^3$, $2,0 \times 10^{-14} \text{ m}^3$, $2,5 \times 10^{-14} \text{ m}^3$, o $3,0 \times 10^{-14} \text{ m}^3$ o más. La diferencia entre el volumen del micropocillo y el volumen del soporte sólido puede ser de al menos aproximadamente $1 \times 10^{-11} \text{ l}$, $1,5 \times 10^{-11} \text{ l}$, $1,7 \times 10^{-11} \text{ l}$, $2,0 \times 10^{-11} \text{ l}$, $2,5 \times 10^{-11} \text{ l}$ o $3,0 \times 10^{-11} \text{ l}$ o más. La diferencia entre el volumen del micropocillo y el volumen del soporte sólido puede ser como mucho de aproximadamente $1 \times 10^{-11} \text{ l}$, $1,5 \times 10^{-11} \text{ l}$, $1,7 \times 10^{-11} \text{ l}$, $2,0 \times 10^{-11} \text{ l}$, $2,5 \times 10^{-11} \text{ l}$ o $3,0 \times 10^{-11} \text{ l}$ o más. La **FIG. 7** ilustra la estadística ejemplar acerca del volumen del micropocillo, el soporte sólido y las diferencias entre los volúmenes de micropocillo y el soporte sólido.
- 15 La profundidad del micropocillo puede ser al menos aproximadamente 0,1, 0,5, 1, 2, 3, 4 o 5 veces o más el diámetro de la célula y/o el soporte sólido a contener en el micropocillo. La profundidad del micropocillo puede ser como mucho aproximadamente 0,1, 0,5, 1, 2, 3, 4 o 5 veces o más el diámetro de la célula y/o el soporte sólido a contener en el micropocillo. La profundidad del micropocillo puede ser de al menos aproximadamente un 10, 20, 30, 40 o 50 % o más de la profundidad del soporte sólido. La profundidad del micropocillo puede ser como mucho de aproximadamente un 10, 20, 30, 40 o 50 % o más de la profundidad del soporte sólido. La profundidad del micropocillo puede ser de al menos aproximadamente 5, 10, 15, 20, 25, 30, 35, 40, 45 o 50 o más micrómetros. La profundidad del micropocillo puede ser como mucho de aproximadamente 5, 10, 15, 20, 25, 30, 35, 40, 45 o 50 o más micrómetros. La profundidad del micropocillo puede ser de aproximadamente 30 micrómetros. La profundidad del micropocillo puede ser de aproximadamente 28 micrómetros. El micropocillo puede ser plano o sustancialmente plano.

Una matriz de micropocillos puede comprender espaciado entre los pocillos. El espaciado entre los pocillos puede ser de al menos aproximadamente 5, 10, 25, 20, 25, 30, 35, 40, 45 o 50 o más micrómetros. El espaciado entre los pocillos puede ser como mucho de aproximadamente 5, 10, 25, 20, 25, 30, 35, 40, 45 o 50 o más micrómetros. El espaciado entre los pocillos puede ser de aproximadamente 15 micrómetros. El espaciado entre los pocillos puede ser de aproximadamente 25 micrómetros.

Puede haber diferencias en la altura de los valles y elevaciones en cualquier posición en la pared interior de un micropocillo. Creando valles y elevaciones en una parte de la pared interior de un pocillo que se ha tratado para que sea lisa, puede añadirse funcionalidad al pocillo. La pared interior de un micropocillo puede alisarse mediante grabado. El grado de vacío en el dispositivo de grabado, el tipo de gas de grabado, las etapas de grabado y similares pueden seleccionarse adecuadamente. Por ejemplo, el alisado de la pared interior de un micropocillo puede realizarse por grabado en húmedo o combinando una etapa de oxidación en caliente con grabado con película de óxido. La pared interior del micropocillo puede funcionalizarse (por ejemplo, funcionalizarse con un oligonucleótido, un grupo reactivo, un grupo funcional).

La matriz de micropocillos puede hacerse de silicio, metal (por ejemplo, aluminio, acero inoxidable, cobre, níquel, cromo y titanio), PDMS (elastómero), vidrio, polipropileno, agarosa, gelatina, pluronic (por ejemplo, pluronic F127), plástico (por ejemplo plásticos que son hidrófilos de forma natural, tales como PMMA), plástico (por ejemplo, PP, COP, COC) y elastómero (por ejemplo, PDMS) que son hidrófobos, pero pueden tratarse para que sean hidrófilos), hidrogeles (por ejemplo, poliacrilamida, alginato) o resina (por ejemplo, poliimida, polietileno, cloruro de vinilo, polipropileno, policarbonato, acrílico y tereftalato de polietileno). La matriz de micropocillos puede hacerse de un material que es hidrófobo. La matriz de micropocillos puede hacer de un material que es hidrófobo, pero se recubre para hacerlo hidrófilo (por ejemplo, por tratamiento con plasma de oxígeno). La matriz de micropocillos puede hacerse de un material que es hidrófilo, pero se recubre para hacerlo hidrófobo.

Una matriz de micropocillos puede ensamblarse. El ensamblaje de matriz de micropocillos puede comprender obtener una oblea de silicio con estampado (por ejemplo, postes estampados hechos con fotoprotector SU8) e incubándola con material de PDMS para crear matrices de pocillos mediante litografía suave (por ejemplo, a 80 °C durante unas pocas horas). Por ejemplo, el PDMS no curado puede ser líquido. El PDMS no curado puede rellenar huecos entre los postes. Cuando el PDMS se cura mediante calor, puede quedar sólido, generando de ese modo la matriz de pocillos. Puede aplicarse un adhesivo óptico (por ejemplo, NOA81/NOA63) al material de PDMS (por ejemplo, usando luz UV) para crear una matriz de postes (por ejemplo, una pluralidad de matrices). La aplicación puede realizarse durante al menos aproximadamente 1 segundo, 2 segundos, 3 segundos, 4 segundos, 5 segundos, 6 segundos, 7 segundos, 8 segundos, 9 segundos, 10 segundos o 1 minuto, 2 minutos, 3 minutos, 4 minutos o 5 o más minutos. Puede aplicarse una capa que comprende agarosa al adhesivo óptico. La capa de agarosa puede ser de al menos aproximadamente un 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 % o más de agarosa. La capa de agarosa puede ser como mucho de aproximadamente un 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 % o más de agarosa. La capa de agarosa puede ser de aproximadamente un 5 % de agarosa. La capa de agarosa puede establecerse en película Gelbond, o cualquier sustrato hidrófilo al que pueda adherirse la agarosa. La incubación de la capa de agarosa en la superficie óptica puede ser de al menos aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más minutos. La incubación de la capa de agarosa en la superficie óptica puede ser como mucho de aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más minutos.

En algunos casos, los métodos de la divulgación pueden usar una superficie que puede no comprender micropocillos. La superficie puede ser vidrio, plástico, metal. La superficie puede recubrirse con soportes sólidos, matriz extracelular, polímeros. La superficie puede no comprender pocillos. La superficie puede comprender soportes sólidos dispuestos de forma espacial para limitar la difusión molecular. Los métodos de la divulgación de captura de células y/o contenidos celulares pueden producirse sobre una superficie plana. Los métodos de la divulgación de captura de células y/o contenidos celulares puede producirse en una suspensión.

Células y muestras

La célula y de la divulgación puede ser una célula de un animal (por ejemplo, ser humano, rata, cerdo, caballo, vaca, perro, ratón). En algunos casos, la célula es una célula humana. La célula puede ser una célula humana fetal. La célula humana fetal puede obtenerse de una madre embarazada con el feto. La célula puede ser una célula de una madre embarazada. La célula puede ser una célula de un vertebrado, invertebrado, hongo, arqueobacteria o bacteria. La célula puede ser de un tejido multicelular (por ejemplo, un órgano (por ejemplo, cerebro, hígado, pulmón, riñón, próstata, ovario, bazo, ganglio linfático, tiroides, páncreas, corazón, musculo esquelético, intestino, laringe, esófago y estomago), un blastocisto). La célula puede ser una célula de un cultivo celular. La célula puede ser una célula HeLa, una célula K562, una célula Ramos, un hibridoma, una célula madre, una célula indiferenciada, una célula diferenciada, una célula en circulación, una célula CHO, una célula 3T3 y similares.

En algunos casos, la célula es una célula cancerosa. Ejemplos no limitantes de células cancerosas pueden incluir una célula de cáncer de próstata, una célula de cáncer de mama, una célula de cáncer de colon, una célula de cáncer pulmonar, una célula de cáncer cerebral y una célula de cáncer de ovario. En algunos casos, la célula es de un cáncer (por ejemplo, una célula tumoral en circulación). Ejemplos no limitantes de cánceres pueden incluir adenoma, adenocarcinoma, carcinoma escamocelular, carcinoma basocelular, carcinoma de células pequeñas, carcinoma indiferenciado de células grandes, condrosarcoma y fibrosarcoma.

En algunos casos, la célula es una célula infrecuente. Una célula infrecuente puede ser una célula tumoral en circulación (CTC), célula epitelial en circulación (CEC), célula madre en circulación (CSC), células madre, células madre indiferenciadas, células madre cancerosas, células de médula ósea, células progenitoras, macrófagos espumosos, células fetales, células mesenquimatosas, células endoteliales en circulación, células endometriales en circulación, trofoblastos, células del sistema inmunitario (hospedador o injerto), células de tejido conjuntivo, bacterias, hongos o patógenos (por ejemplo, bacterias o protozoos), micropartículas, fragmentos celulares, proteínas y ácidos nucleico, orgánulos celulares, otros componentes celulares (por ejemplo, mitocondrias y núcleos), y virus.

En algunos casos, la célula es de un tumor. En algunos casos, el tumor es benigno o maligno. La célula tumoral puede comprender una célula metastásica. En algunos casos, la célula es de un tejido sólido que comprende una pluralidad de diferentes tipos celulares (por ejemplo, diferentes genotipos).

La célula puede comprender un virus, bacteria, hongo y parásito. Los virus pueden incluir, aunque sin limitación, virus de animales de ADN o ARN (por ejemplo, *Picornaviridae* (por ejemplo, poliovirus), *Reoviridae* (por ejemplo, rotavirus), *Togaviridae* (por ejemplo, virus de la encefalitis, virus de la fiebre amarilla, virus de la rubéola), *Orthomyxoviridae* (por ejemplo, virus de la gripe), *Paramyxoviridae* (por ejemplo, virus sincitial respiratorio, virus del sarampión, virus de las paperas, virus de la parainfluenza), *Rhabdoviridae* (por ejemplo, virus de rabia), *Coronaviridae*, *Bunyaviridae*, *Flaviviridae*, *Filoviridae*, *Arenaviridae*, *Bunyaviridae* y *Retroviridae* (por ejemplo, virus linfotrópicos de linfocitos T humanos (HTLV), virus de la inmunodeficiencia humana (VIH), *Papovaviridae* (por ejemplo, virus del papiloma), *Adenoviridae* (por ejemplo, adenovirus), *Herpesviridae* (por ejemplo, virus del herpes simple) y *Poxviridae* (por ejemplo, virus de la viruela)).

Las bacterias ejemplares que pueden usarse en los métodos de la divulgación pueden incluir *Actinomeduræ*, *Actinomyces israelii*, *Bacillus anthracis*, *Bacillus cereus*, *Clostridium botulinum*, *Clostridium difficile*, *Clostridium perfringens*, *Clostridium tetani*, *Corynebacterium*, *Enterococcus faecalis*, *Listeria monocytogenes*, *Nocardia*, *Propionibacterium acnes*, *Staphylococcus aureus*, *Staphylococcus epiderm*, *Streptococcus mutans*, *Streptococcus pneumoniae* y similares. Las bacterias gramnegativas incluyen, aunque sin limitación, *Afpia felis*, *Bacterioides*, *Bartonella bacilliformis*, *Bordetella pertussis*, *Borrelia burgdorferi*, *Borrelia recurrentis*, *Brucella*, *Calymmatobacterium granulomatis*, *Campylobacter*, *Escherichia coli*, *Francisella tularensis*, *Gardnerella vaginalis*, *Haemophilus aegyptius*, *Haemophilus ducreyi*, *Haemophilus influenzae*, *Heliobacter pylori*, *Legionella pneumophila*, *Leptospira interrogans*, *Neisseria meningitidis*, *Porphyromonas gingivalis*, *Providencia sturti*, *Pseudomonas aeruginosa*, *Salmonella enteridis*, *Salmonella typhi*, *Serratia marcescens*, *Shigella boydii*, *Streptobacillus moniliformis*, *Streptococcus pyogenes*, *Treponema pallidum*, *Vibrio cholerae*, *Yersinia enterocolitica*, *Yersinia pestis* y similares. Otras bacterias pueden incluir *Myobacterium avium*, *Myobacterium leprae*, *Myobacterium tuberculosis*, *Bartonella henselae*, *Chlamydia psittaci*, *Chlamydia trachomatis*, *Coxiella burnetii*, *Mycoplasma pneumoniae*, *Rickettsia akari*, *Rickettsia prowazekii*, *Rickettsia rickettsii*, *Rickettsia tsutsugamushi*, *Rickettsia typhi*, *Ureaplasma urealyticum*, *Diplococcus pneumoniae*, *Ehrlichia chafensis*, *Enterococcus faecium*, meningococos y similares.

Los hongos ejemplares a usar en los métodos de la divulgación pueden incluir, aunque sin limitación, *Aspergilli*, *Candidae*, *Candida albicans*, *Coccidioides immitis*, *Cryptococci* y combinaciones de los mismos.

Parásitos ejemplares a usarse en los métodos de la divulgación pueden incluir, aunque sin limitación, *Balantidium coli*, *Cryptosporidium parvum*, *Cyclospora cayatanensis*, *Encephalitozoa*, *Entamoeba histolytica*, *Enterocytozoon bieneusi*, *Giardia lamblia*, *Leishmaniae*, *Plasmodii*, *Toxoplasma gondii*, *Trypanosomae*, amebas trapezoides, gusanos (por ejemplo, helmintos), particularmente gusanos parasitarios incluyendo, aunque sin limitación, nematodos (ascárides, por ejemplo, tricocéfalos, anquilostomas, oxiuros, ascáridos, filáridos y similares), cestodos

ES 2 711 168 T3

(por ejemplo, tenias).

- 5 La muestra de la divulgación puede ser una muestra de un animal (por ejemplo, ser humano, rata, cerdo, caballo, vaca, perro, gato). En algunas realizaciones, la muestra es una muestra humana. La muestra puede ser una muestra humana fetal. La muestra humana fetal puede obtenerse de una madre embarazada del feto. La muestra puede ser una muestra de una madre embarazada. La muestra puede ser una muestra de un vertebrado, invertebrado, hongo, arqueobacteria o bacteria. La muestra puede ser de un tejido multicelular (por ejemplo, un órgano (por ejemplo, cerebro, hígado, pulmón, riñón, próstata, ovario, bazo, ganglio linfático, tiroides, páncreas, corazón, músculo esquelético, intestino, laringe, esófago y estómago), un blastocisto). La muestra puede ser una célula de un cultivo celular.
- 10 La muestra puede comprender una pluralidad de células. La muestra puede comprender una pluralidad del mismo tipo de células. La muestra puede comprender una pluralidad de diferentes tipos de células. La muestra puede comprender una pluralidad de células en el mismo punto del ciclo celular y/o ruta de diferenciación. La muestra puede comprender una pluralidad de células en diferentes puntos del ciclo celular y/o ruta de diferenciación. Una muestra puede comprender una pluralidad de muestras.
- 15 La pluralidad de muestras puede comprender al menos 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 1000 o más muestras. La pluralidad de muestras puede comprender al menos aproximadamente 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000 muestras, 9000 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.
- 20 El uno o más ácidos nucleicos en la primera muestra pueden ser diferentes de uno o más ácidos nucleicos en la segunda muestra. El uno o más ácidos nucleicos en la primera muestra pueden ser diferentes de uno o más ácidos nucleicos en una pluralidad de muestras. El uno o más ácidos nucleicos pueden comprender una longitud de al menos aproximadamente 1 nucleótido, 2 nucleótidos, 5 nucleótidos, 10 nucleótidos, 20 nucleótidos, 50 nucleótidos, 100 nucleótidos, 200 nucleótidos, 300 nucleótidos, 500 nucleótidos, 1000 nucleótidos, 2000 nucleótidos, 3000 nucleótidos, 4000 nucleótidos, 5000 nucleótidos, 10 000 nucleótidos, 100 000 nucleótidos, 1 000 000 nucleótidos.
- 25 La primera muestra puede comprender una o más células y la segunda muestra puede comprender una o más células. La una o más células de la primera muestra pueden ser del mismo celular que la una o más células de la segunda muestra. La una o más células de la primera muestra pueden ser de un tipo celular diferente que una o más células diferentes de la pluralidad de muestras. El tipo celular puede ser condrocito, osteoclasto, adipocito, mioblasto, célula madre, célula endotelial o célula de músculo liso. El tipo celular puede ser un tipo celular inmunitario. El tipo celular inmunitario puede ser un linfocito T, linfocito B, trombocito, célula dendrítica, neutrófilo, macrófago o monocito.
- 30 La pluralidad de muestras puede comprender una o más células malignas. La una o más células malignas pueden obtenerse de un tumor, sarcoma o leucemia.
- 35 La pluralidad de muestras puede comprender al menos un líquido corporal. El líquido corporal puede comprender sangre, orina, líquido linfático, saliva. La pluralidad de muestras puede comprender al menos una muestra de sangre.
- 40 La pluralidad de muestras puede comprender al menos una célula de uno o más tejidos biológicos. El uno o más tejidos biológicos puede ser un hueso, corazón, timo, arteria, vaso sanguíneo, pulmón, músculo, estómago, intestino, hígado, páncreas, bazo, riñón, vesícula biliar, glándula tiroides, glándula suprarrenal, glándula mamaria, ovario, glándula prostática, testículo, piel, tejido adiposo, ojo o cerebro.
- El tejido biológico puede comprender un tejido infectado, tejido enfermo, tejido maligno, tejido calcificado o tejido sano.
- 45 La pluralidad de muestras puede ser de una o más fuentes. La pluralidad de muestras puede ser de dos o más fuentes. La pluralidad de muestras puede ser de uno o más sujetos. La pluralidad de muestras puede ser de dos o más sujetos. La pluralidad de muestras puede ser del mismo sujeto. El uno o más sujetos pueden ser de la misma especie. El uno o más sujetos pueden ser de diferentes especies. El uno o más sujetos pueden estar sanos. El uno o más sujetos pueden estar afectados por una enfermedad, trastorno o afección. La pluralidad de muestras puede comprender células de un origen seleccionado de un mamífero, bacteria, virus, hongo o planta. La una o más muestras pueden ser de un ser humano, caballo, vaca, pollo, cerdo, rata, ratón, mono, conejo, cobaya, oveja, cabra, perro, gato, pájaro, pez, rana y mosca de la fruta.
- 50 La pluralidad de muestras puede obtenerse simultáneamente. La pluralidad de muestras puede obtenerse al mismo tiempo. La pluralidad de muestras puede obtenerse secuencialmente. La pluralidad de muestras puede obtenerse durante el curso de años, 100 años, 10 años, 5 años, 4 años, 3 años, 2 años o 1 año de obtener una o más muestras diferentes. Una o más muestras pueden obtenerse en aproximadamente un año de obtener una o más muestras diferentes. Una o más muestras pueden obtenerse en 12 meses, 11 meses, 10 meses, 9 meses, 8 meses, 7 meses, 6 meses, 4 meses, 3 meses, 2 meses o 1 mes de obtener una o más muestras diferentes. Una o más muestras pueden obtenerse en 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 un día de obtener una o más muestras diferentes. Una o más muestras pueden obtenerse en 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
- 55
- 60

obtener una o más muestras diferentes. Una o más muestras pueden obtenerse en aproximadamente 60 segundos, 45 segundos, 30 segundos, 20 segundos, 10 segundos, 5 segundos, 2 segundos o 1 segundos de obtener una o más muestras diferentes. Una o más muestras pueden obtenerse en menos de un segundo de obtener una o más muestras diferentes.

5 Moléculas diana

Los métodos y kits divulgados en la presente memoria pueden usarse en el marcaje estocástico de moléculas. Dichas moléculas incluyen, aunque sin limitación, polinucleótidos y polipéptidos. Tal como se usan en la presente memoria, los términos "polinucleótido" y "molécula de ácido nucleico" se refieren a una forma polimérica de nucleótidos de cualquier longitud, ribonucleótidos, desoxirribonucleótidos, ácidos nucleicos bloqueados (LNA) o ácidos peptidonucleicos (PNA), que comprende bases de purina y pirimidina, u otras bases nucleotídicas naturales, modificadas química o bioquímicamente, no naturales o derivatizadas. Un "polinucleótido" o "molécula de ácido nucleico" puede consistir en un único nucleótido o par de bases. Como alternativa, el "polinucleótido" o "molécula de ácido nucleico" comprende dos o más nucleótidos o pares de bases. Por ejemplo, el "polinucleótido" o "molécula de ácido nucleico" comprende 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 1000 nucleótidos o pares de bases. En otro ejemplo, el polinucleótido comprende al menos aproximadamente 1500, 2000, 2500, 3000, 3500, 4000, 4500, 5000, 5500, 6000, 6500, 7000, 7500, 8000, 8500, 9000, 9500 o 10 000 nucleótidos o pares de bases. La cadena principal del polinucleótido puede comprender azúcares y grupos fosfato, que pueden encontrarse típicamente en ARN o ADN, o azúcares o grupos fosfato modificados o sustituidos. Un polinucleótido puede comprender nucleótidos modificados, tales como nucleótidos metilados o análogos nucleotídicos. La secuencia de nucleótidos puede estar interrumpida por componentes no nucleotídicos. Por tanto, los términos nucleósido, nucleótido, desoxinucleósido y desoxinucleótido en general incluyen análogos tales como los descritos en la presente memoria. Estos análogos son aquellas moléculas que tienen algunas características estructurales en común con un nucleósido o nucleótido de origen natural de modo que cuando se incorporan en un ácido nucleico o secuencia oligonucleosídica, permiten la hibridación con una secuencia de ácido nucleico de origen natural en disolución. Típicamente, estos análogos derivan de nucleósidos y nucleótidos de origen natural por remplazo y/o modificación de la base, la ribosa y el resto fosfodiéster. Los cambios pueden adaptarse para estabilizar o desestabilizar la formación de híbridos o potenciar la especificidad de hibridación con una secuencia de ácido nucleico complementaria según se desee. En algunos casos, las moléculas son ADN, ARN o híbridos de ADN-ARN. Las moléculas pueden ser monocatenarias o bicatenarias. En algunos casos, las moléculas son moléculas de ARN, tales como ARNm, ARNr, ARNt, ARNnc, ARNinc, ARNip, microARN o miARN. Las moléculas de ARN pueden estar poliadeniladas. Como alternativa, las moléculas de ARN no están poliadeniladas. Como alternativa, las moléculas son moléculas de ADN. Las moléculas de ADN pueden ser ADN genómico. Las moléculas de ADN pueden comprender exones, intrones, regiones no traducidas o cualquier combinación de las mismas. En algunos casos, las moléculas son un panel de moléculas.

Los métodos y kits divulgados en la presente memoria pueden usarse para marcar de forma estocástica acontecimientos individuales de moléculas idénticas o casi idénticas y/o moléculas diferentes. En algunos casos, los métodos y kits divulgados en la presente memoria pueden usarse para marcar de forma estocástica moléculas idénticas o casi idénticas (por ejemplo, moléculas que comprenden secuencias idénticas o casi idénticas). Por ejemplo, las moléculas a marcar comprenden al menos aproximadamente un 80 %, 85 %, 90 %, 95 %, 96 %, 97 %, 98 %, 99 % o un 100 % de identidad de secuencias. Las moléculas casi idénticas pueden diferir en menos de aproximadamente 100, 90, 80, 70, 60, 50, 40, 30, 25, 20, 25, 10, 9, 8, 7, 6, 5, 4, 3, 2 o 1 nucleótido o par de bases. La pluralidad de ácidos nucleicos en una o más muestras de la pluralidad de muestras puede comprender dos o más secuencias idénticas. 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 un 100 % de los ácidos nucleicos totales en una o más de la pluralidad de muestras puede comprender la misma secuencia. La pluralidad de ácidos nucleicos en una o más muestras de la pluralidad de muestras puede comprender al menos dos secuencias diferentes. Al menos aproximadamente un 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 %, 100 % de los ácidos nucleicos totales en una o más de la pluralidad de muestras puede comprender al menos dos secuencias diferentes. En algunos casos, las moléculas a marcar son variantes entre sí. Por ejemplo, las moléculas a marcar pueden contener polimorfismos mononucleotídicos u otros tipos de mutaciones. En otro ejemplo, las moléculas a marcar son variantes de corte y empalme. En algunos casos, al menos una molécula se marca de forma estocástica. En otros casos, al menos 2, 3, 4, 5, 6, 7, 8, 9 o 10 moléculas idénticas o casi idénticas se marcan de forma estocástica. Como alternativa, al menos 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900 o 1000 moléculas idénticas o casi idénticas se marcan de forma estocástica. En otros casos, al menos 1500; 2000; 2500; 3000; 3500; 4000; 4500; 5000; 6000; 7000; 8000; 9000 o 10 000 moléculas idénticas o casi idénticas se marcan de forma estocástica. En otros casos, al menos 15 000; 20 000; 25 000; 30 000; 35 000; 40 000; 45 000; 50 000; 60 000; 70 000; 80 000; 90 000 o 100 000 moléculas idénticas o casi idénticas se marcan de forma estocástica.

En otros casos, los métodos y kits divulgados en la presente memoria pueden usarse para marcar de forma estocásticas diferentes moléculas. Por ejemplo, las moléculas a marcar comprenden menos de un 75 %, 70 %, 65 %, 60 %, 55 %, 50 %, 45 %, 40 %, 35 %, 30 %, 25 %, 20 %, 15 %, 10 %, 5 %, 4 %, 3 %, 2 %, 1 % de identidad de secuencia. Las diferentes moléculas pueden diferir en al menos aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100 o más nucleótidos o pares de bases. En algunos casos, al menos una molécula se marca de forma estocástica. En otros casos, al menos 2, 3, 4, 5, 6, 7, 8, 9 o 10 moléculas diferentes se marcan de forma estocástica. Como alternativa, al menos 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900 o 1000 moléculas diferentes se marcan de forma estocástica. En otros casos, al menos 1500;

ES 2 711 168 T3

2000; 2500; 3000; 3500; 4000; 4500; 5000; 6000; 7000; 8000; 9000 o 10 000 moléculas diferentes se marcan de forma estocástica. En otros casos, al menos 15 000; 20 000; 25 000; 30 000; 35 000; 40 000; 45 000; 50 000; 60 000; 70 000; 80 000; 90 000 o 100 000 moléculas diferentes se marcan de forma estocástica.

5 Las diferentes moléculas a marcar pueden estar presentes en la muestra a diferentes concentraciones o cantidades. Por ejemplo, la concentración o cantidad de una molécula es mayor que la concentración o cantidad de otra molécula en la muestra. En algunos casos, la concentración o cantidad de al menos una molécula 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 o 100 o más veces mayor que la concentración o cantidad de al menos una molécula diferente en la muestra. En algunos casos, la concentración o cantidad de al menos una molécula en la muestra es al menos aproximadamente 1000 o más veces mayor que la concentración o cantidad de al menos una molécula diferente en la muestra. En otro ejemplo, la concentración o cantidad de una molécula es menor que la concentración o cantidad de otra molécula en la muestra. La concentración o cantidad de al menos una molécula en la muestra puede ser 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 o 100 o más menor que la concentración o cantidad de al menos una molécula diferente en la muestra. La concentración o cantidad de al menos una molécula en la muestra puede ser al menos aproximadamente 1000 o más veces menor que la concentración o cantidad de al menos una molécula diferente en la muestra.

20 En algunos casos, las moléculas a marcar están en una o más muestras. Las moléculas a marcar pueden estar en dos o más muestras. Las dos o más muestras pueden contener diferentes cantidades o concentraciones de las moléculas a marcar. En algunos casos, la concentración o cantidad de una molécula en una muestra puede ser mayor que la concentración o cantidad de la misma molécula en una muestra diferente. Por ejemplo, una muestra de sangre podría contener una cantidad mayor de una molécula particular que una muestra de orina. Como alternativa, una única muestra se divide en dos o más submuestras. Las submuestras pueden contener diferentes cantidades o concentraciones de la misma molécula. La concentración o cantidad de al menos una molécula en una muestra puede ser 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 o 100 o más veces mayor que la concentración o cantidad de la misma molécula en otra muestra. Como alternativa, la concentración o cantidad de una molécula en una muestra puede ser menor que la concentración o cantidad de la misma molécula en una muestra diferente. Por ejemplo, una muestra de tejido cardiaco podría contener una cantidad mayor de una molécula particular que una muestra de tejido pulmonar. La concentración o cantidad de al menos una molécula en una muestra puede ser 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 o 100 o más veces menor que la concentración o cantidad de la misma molécula en otra muestra. En algunos casos, las diferentes concentraciones o cantidades de una molécula en dos o más muestras diferentes se mencionan como sesgo muestral.

35 Los métodos y kits divulgados en la presente memoria pueden usarse para el análisis de dos o más moléculas de dos o más muestras. Las dos o más moléculas pueden comprender dos o más polipéptidos. El método puede comprender la determinación de la identidad de dos o más polipéptidos marcados. La determinación de la identidad de dos o más polipéptidos marcados puede comprender espectrometría de masas. El método puede comprender además combinar los polipéptidos marcados de la primera muestra con los polipéptidos marcados de la segunda muestra. Los polipéptidos marcados pueden combinarse antes de determinar el número de diferentes polipéptidos marcados. El método puede comprender además combinar los polipéptidos con marca de la primera muestra y los polipéptidos con marca de la segunda muestra. Los polipéptidos con marca de la primera muestra y los polipéptidos con marca de la segunda muestra pueden combinarse antes de entrar en contacto con la pluralidad de marcadores identificadores moleculares. Determinar el número de diferentes polipéptidos marcados puede comprender la detección de al menos una parte del polipéptido marcado. La detección de al menos una parte del polipéptido marcado puede comprender la detección de al menos una parte de la marca de la muestra, el marcador identificador molecular, el polipéptido o una combinación de los mismos.

50 Tal como se usa en la presente memoria, el término "polipéptido" se refiere a una molécula que comprende al menos un péptido. En algunos casos, el polipéptido consiste en un único péptido. Como alternativa, el polipéptido comprende dos o más péptidos. Por ejemplo, el polipéptido comprende 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 1000 péptidos. Los ejemplos de polipéptidos incluyen, aunque sin limitación, aminoácidos, proteínas, péptidos, hormonas, oligosacáridos, lípidos, glucolípidos, fosfolípidos, anticuerpos, enzimas, cinasas, receptores, factores de transcripción y ligandos.

Sujetos

55 Los métodos y kits divulgados en la presente memoria pueden comprender el uso de una célula o muestra de uno o más sujetos. Un sujeto puede ser un ser humano o un sujeto no humano. Un sujeto puede estar vivo. Un sujeto puede estar muerto. Un sujeto puede ser un ser humano que está bajo los cuidados de un cuidador (por ejemplo, profesional médico). Un sujeto puede ser sospechoso de tener una enfermedad. Un sujeto puede tener una enfermedad. Un sujeto puede tener síntomas de una enfermedad. Un sujeto puede ser un sujeto que proporciona una o más muestras. Un sujeto puede ser un mamífero, reptil, anfibio y/o ave. Un sujeto puede ser un primate no humano.

Enzimas

Los métodos y kits divulgados en la presente memoria pueden comprender una o más enzimas. Los ejemplos de enzimas incluyen, aunque sin limitación, ligasas, transcriptasas inversas, polimerasas y nucleasas de restricción. En

algunos casos, la adhesión de la marca oligonucleotídica a las moléculas comprende el uso de una o más ligasas. Los ejemplos de ligasas incluyen, aunque sin limitación, ADN ligasas tales como ADN ligasa I, ADN ligasas III, ADN ligasa IV y ADN ligasa T4, y ARN ligasas tales como ARN ligasa T4 I y ARN ligasa T4 II.

5 Los métodos y kits divulgados en la presente memoria pueden comprender además el uso de una o más transcriptasas inversas. En algunos casos, la transcriptasa inversa es una transcriptasa inversa de VIH-1, transcriptasa inversa de VLM-M, transcriptasa inversa de AMV y transcriptasa inversa telomerasa. En algunos casos, la transcriptasa inversa es la transcriptasa inversa de VLM-M.

10 En algunos casos, los métodos y kits divulgados en la presente memoria comprenden el uso de una o más polimerasas. Los ejemplos de polimerasas incluyen, aunque sin limitación, ADN polimerasas y ARN polimerasas. En algunos casos la ADN polimerasa es un ADN polimerasa I, ADN polimerasa II, holoenzima de 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 Bst 2.0 WarmStart™, ADN polimerasa Bst, ADN polimerasa IV de *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 II Therminator™, ADN polimerasa III Therminator™, ADN polimerasa VentR®, ADN polimerasa VentR® (exo-), ADN polimerasa Bsu, ADN polimerasa phi29, ADN polimerasa T4, ADN polimerasa T7, Terminal transferasa, polimerasa Taq Titanium®, ADN polimerasa Taq KAPA y ADN polimerasa Taq de inicio caliente KAPA.

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

25 En algunos casos, los métodos y kits divulgados en la presente memoria comprenden una o más enzimas de restricción. Las enzimas de restricción incluyen enzimas de restricción de tipo I, tipo II, tipo III y tipo IV. En algunos casos, las enzimas de tipo I son enzimas complejas, de múltiples subunidades, de combinación de restricción y modificación que cortan aleatoriamente ADN lejos de sus secuencias de reconocimiento. En general, las enzimas de tipo II cortan el ADN en posiciones definidas cerca de o dentro de sus secuencias de reconocimiento. Producen fragmentos de restricción concretos y distintos patrones de bandas en gel. Las enzimas de tipo III también son enzimas grandes de combinación de restricción y modificación. A menudo escinden fuera de sus secuencias de reconocimiento y pueden requerir dos de dichas secuencias en orientaciones opuestas dentro de la misma molécula de ADN para conseguir la escisión; infrecuentemente dan digestiones completas. En algunos casos, las enzimas de tipo IV reconocen ADN modificado, típicamente metilado y pueden ejemplificarse por los sistemas McrBC y Mrr de *E. coli*.

Reactivos adicionales

35 Los métodos y kits divulgados en la presente memoria pueden comprender el uso de uno o más reactivos. Los ejemplos de reactivos incluyen, aunque sin limitación, reactivos de PCR, reactivos de ligamiento, reactivos de transcripción inversa, reactivos enzimáticos, reactivos de hibridación, reactivos de preparación de muestras y reactivos para la purificación y/o aislamiento de ácidos nucleicos.

40 Los métodos y kits divulgados en la presente memoria pueden comprender el uso de uno o más tampones. Los ejemplos de tampones incluyen, aunque sin limitación, tampones de lavado, tampones de ligamiento, tampones de hibridación, tampones de amplificación y tampones de transcripción inversa. En algunos casos, el tampón de hibridación es un tampón disponible en el mercado, tal como la disolución TMAC Hyb, la disolución de hibridación de SSPE y el tampón de hibridación de ECONOTM. Los tampones divulgados en la presente memoria pueden comprender uno o más detergentes.

45 Los métodos y kits divulgados en la presente memoria pueden comprender el uso de uno o más vehículos. Los vehículos pueden potenciar o mejorar la eficacia de una o más reacciones divulgadas en la presente memoria (por ejemplo, reacción de ligamiento, transcripción inversa, amplificación, hibridación). Los vehículos pueden disminuir o evitar la pérdida no específica de las moléculas o cualquier producto de las mismas (por ejemplo, molécula marcada, molécula de ADNc marcada, amplicón marcado). Por ejemplo, el vehículo puede disminuir la pérdida no específica de una molécula marcada mediante absorción a las superficies. El vehículo puede disminuir la afinidad de la molécula, la molécula marcada o cualquier producto de las mismas por una superficie o sustrato (por ejemplo, recipiente, tubo Eppendorf, punta de pipeta). Como alternativa, el vehículo puede aumentar la afinidad de la molécula o cualquier producto de la misma por una superficie o sustrato (por ejemplo, microesfera, matriz, vidrio, portaobjetos, chip). Los vehículos pueden proteger la molécula o cualquier producto de la misma de la degradación.

50 Por ejemplo, los vehículos pueden proteger una molécula de ARN o cualquier producto de la misma de las ribonucleasas. Como alternativa, los vehículos pueden proteger una molécula de ADN o cualquier producto de la misma de una DNasa. Los ejemplos de vehículos incluyen, aunque sin limitación, moléculas de ácido nucleico tales como ADN y/o ARN, o polipéptidos. Los ejemplos de vehículos de ADN incluyen plásmidos, vectores, ADN poliadenilado y oligonucleótidos de ADN. Los ejemplos de vehículos de ARN incluyen ARN poliadenilado, ARN fágico, 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 oligonucleótidos de ARN. El vehículo de ARN puede ser un ARN poliadenilado. Como alternativa, el vehículo de ARN puede ser un ARN no poliadenilado. En algunos casos, el vehículo es de una bacteria, levadura o virus. Por ejemplo, el vehículo puede ser una molécula de ácido nucleico o

un polipéptido derivado de una bacteria, levadura o virus. Por ejemplo, el vehículo es una proteína de *Bacillus subtilis*. En otro ejemplo, el vehículo es una molécula de ácido nucleico de *Escherichia coli*. Como alternativa, el vehículo es una molécula de ácido nucleico o péptido de un mamífero (por ejemplo, ser humano, ratón, cabra, rata, vaca, oveja, cerdo, perro o conejo), ave, anfibio o reptil.

- 5 Los métodos y kits divulgados en la presente memoria pueden comprender el uso de uno o más agentes de control. Los agentes de control pueden incluir oligos de control, enzimas inactivas, competidores no específicos. Como alternativa, los agentes de control comprenden hibridación brillante, controles de sonda brillante, moldes de ácido nucleico, controles de adición, controles de amplificación por PCR. Los controles de amplificación por PCR pueden ser controles positivos. En otros casos, los controles de amplificación por PCR son controles negativos. Los
10 controles de molde de ácido nucleico pueden ser de concentraciones conocidas. Los agentes de control pueden comprender uno o más marcadores.

Los controles de adición pueden ser moldes que se añaden a una reacción o muestra. Por ejemplo, un molde de adición puede añadirse a una reacción de amplificación. El molde de adición puede añadirse a la reacción de amplificación cualquier momento después del primer ciclo de amplificación. En algunos casos, el molde de adición se añade a la reacción de amplificación después de la 2.^a, 3.^a, 4.^a, 5.^a, 6.^a, 7.^a, 8.^a, 9.^a, 10.^a, 11.^a, 12.^a, 13.^a, 14.^a,
15 15.^a, 20.^a, 25.^a, 30.^a, 35.^a, 40.^a, 45.^a o 50.^a ciclo de amplificación. El molde de adición puede añadirse a la reacción de amplificación 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.

20 Marcadores detectables

Los métodos, kits y composiciones divulgados en la presente memoria pueden comprender además un marcador detectable. Las expresiones "marcador detectable", o "marca" o "marcador" pueden usarse indistintamente y hacen referencia a cualquier resto químico adherido a una molécula (por ejemplo, nucleótido, polímero nucleotídico o factor de unión a ácido nucleico, código de barras molecular). El resto químico puede adherirse covalentemente a la
25 molécula. El resto químico puede adherirse no covalentemente a la molécula. Los códigos de barras moleculares, marcas de muestra y marcadores identificadores moleculares pueden comprender además un marcador detectable, marca o marcador. Preferiblemente, el marcador es detectable y hace que el nucleótido o polímero nucleotídico sea detectable para el facultativo de la invención. Los marcadores detectables en combinación con los métodos divulgados en la presente memoria incluyen, por ejemplo, un marcador fluorescente, un marcador quimioluminiscente, un inactivador, un marcador radiactivo, biotina, resto de pireno, oro o combinaciones de los mismos. Los ejemplos no limitantes de marcadores detectables incluyen moléculas luminiscentes, fluorocromos, agentes inactivadores fluorescentes, moléculas coloreadas, radioisótopos o agentes de centelleo.

En algunos casos, los métodos divulgados en la presente memoria comprenden además adherir uno o más marcadores detectables al código de barras molecular, marcador identificador molecular, la marca de muestra, el
35 ácido nucleico marcado o cualquier producto de los mismos (por ejemplo, amplicón marcado). Los métodos pueden comprender adherir dos o más marcadores detectables al código de barras molecular, marcador identificador molecular, la marca de muestra o el ácido nucleico marcado. Como alternativa, el método comprende adherir al menos aproximadamente 3, 4, 5, 6, 7, 8, 9 o 10 marcadores detectables al código de barras molecular, marcador identificador molecular, la marca de muestra o el ácido nucleico marcado. En algunos casos, el marcador detectable es un marcador CyTM. El marcador CyTM es un marcador Cy3. Como alternativa, o adicionalmente, el marcador detectable es biotina. En algunas realizaciones, el marcador detectable se adhiere a una sonda que se une al código de barras molecular, marcador identificador molecular, la marca de muestra o el ácido nucleico marcado. Esto puede producirse, por ejemplo, después de que el ácido nucleico o ácido nucleico marcado se haya hibridado a una matriz. En un ejemplo, el ácido nucleico o ácido nucleico marcado se une a compañeros en una matriz. Después de la
40 unión, una sonda que puede unirse al ácido nucleico marcado se une a las moléculas en la matriz. Este proceso puede repetirse con múltiples sondas y marcadores para disminuir la probabilidad de que una señal sea el resultado de unión no específica de un marcador o unión no específica de la molécula a la matriz.

Puede usarse un par de donador aceptador como los marcadores detectables. El donador o el aceptador pueden adherirse a una sonda que se une a un ácido nucleico. La sonda puede ser, por ejemplo, una sonda de ácido
50 nucleico que puede unirse a un ácido nucleico o el ácido nucleico marcado. El donador o aceptador correspondiente puede añadirse para causar una señal.

En algunos casos, el marcador detectable es un tinte Freedom, tinte Alexa Fluor®, tinte CyTM, tinte de fluoresceína o LI-COR IRDyes®. En algunos casos, el tinte Freedom es fluoresceína (6-FAMTM, 6-carboxifluoresceína), MAX (éster de NHS), TYETM 563, TEX 615, TYETM 665, TYE 705. El marcador detectable puede ser un tinte Alexa Fluor. Los ejemplos de tintes Alexa Fluor® incluyen Alexa Fluor® 488 (éster de NHS), Alexa Fluor® 532 (éster de NHS), Alexa Fluor® 546 (éster de NHS), Alexa Fluor® 594 (éster de NHS), Alexa Fluor® 647 (éster de NHS), Alexa Fluor® 660 (éster de NHS) o Alexa Fluor® 750 (éster de NHS). Como alternativa, el marcador detectable es un tinte CyTM. Los ejemplos de tintes CyTM incluyen, aunque sin limitación, Cy2, Cy3, Cy3B, Cy3.5, Cy5, Cy5.5 y Cy7. En algunos casos, el marcador detectable es un tinte de fluoresceína. Los ejemplos no limitantes de tintes de fluoresceína incluyen 6-FAMTM (azida), 6-FAMTM (éster de NHS), fluoresceína dT, JOE (éster de NHS), TETTM y HEXTM. En algunos casos, el marcador detectable es un LI-COR IRDyes®, tal como 5' IRDye® 700, 5' IRDye® 800 o IRDye® 800CW (éster de NHS). En algunos casos, el marcador detectable es TYETM 563. Como alternativa, el marcador detectable es Cy3.

El marcador detectable puede ser tinte de rodamina. Los ejemplos de tintes de rodamina incluyen, aunque sin limitación, verde de rodamina TM-X (éster de NHS), TAMRATM, TAMRATM (éster de NHS), rojo de rodamina TM-X (éster de NHS), ROXTM (éster de NHS) y 5'TAMRATM (azida). En otros casos, el marcador detectable es un tinte WellRED. Los tintes WellRED incluyen, aunque sin limitación, tinte WellRED D4, tinte WellRED D3 y tinte WellRED D2. En algunos casos, el marcador detectable es Texas Red®-X (éster de NHS), Lightcycler® 640 (éster de NHS) o Dy 750 (éster de NHS).

En algunos casos, los marcadores detectables incluyen una molécula conectora. Los ejemplos de moléculas conectoras incluyen, aunque sin limitación, biotina, avidina, estreptavidina, HRP, proteína A, proteína G, anticuerpos o fragmentos de los mismos, Grb2, polihistidina, Ni²⁺, marcas FLAG, marcas myc. Como alternativa, los marcadores detectables incluyen metales pesados, donadores/aceptadores de electrones, ésteres de acridinio, tintes y sustratos calorimétricos. En otros casos, los marcadores detectables incluyen enzimas tales como fosfatasa alcalina, peroxidasa y luciferasa.

Un cambio en la masa puede considerarse un marcador detectable, como es el caso de la detección de resonancia de plasmones superficiales. Los expertos en la materia reconocerían fácilmente los marcadores detectables útiles que no se mencionan en la presente memoria, que pueden emplearse en el funcionamiento de la presente invención.

En algunos casos, los marcadores detectables se usan con cebadores. Por ejemplo, el cebador universal se marca con el marcador detectable (por ejemplo, cebador universal marcado con Cy3, cebador universal marcado con fluoróforo). Como alternativa, el cebador específico de diana se marca con el marcador detectable (por ejemplo, cebador específico de diana marcado con TYE 563). En otros casos, los marcadores detectables se usan con las marcas de muestras o marcadores identificadores moleculares. Por ejemplo, la marca oligonucleotídica se marca con un marcador detectable (por ejemplo, marca oligonucleotídica marcada con biotina). En otros casos, los marcadores detectables se usan con la molécula de molde de ácido nucleico. Los marcadores detectables pueden usarse para detectar las moléculas marcadas o amplicones marcados. Como alternativa, los marcadores detectables se usan para detectar la molécula de molde de ácido nucleico.

En algunos casos, el marcador detectable se adhiere al cebador, código de barras molecular, marca de muestra, marcador identificador molecular, molécula marcada, amplicón marcado, sonda, sonda de HCR y/o ácido nucleico no marcado. Los métodos para adherir el marcador detectable al cebador, marca oligonucleotídica, molécula marcada, amplicón marcado y/o ácido nucleico no marcado incluyen, aunque sin limitación, marcaje químico y marcaje enzimático. En algunos casos, el marcador detectable se adhiere por marcaje químico. En algunas realizaciones, las técnicas de marcaje químico comprenden un grupo químicamente reactivo. Ejemplos no limitantes de grupos reactivos incluyen ésteres de succinimidilo de amina reactiva tales como NHS-fluoresceína o NHS-rodamina, derivados de isotiocianato de amina reactiva incluyendo FITC y fluoros activados con maleimida de sulfidrido reactivo tales como fluoresceína-5-maleimida. En algunas realizaciones, la reacción de cualquiera de estos tintes reactivos con otra molécula provoca un enlace covalente estable formado entre un fluoróforo y el conector y/o agente. En algunas realizaciones, el grupo reactivo es isotiocianatos. En algunas realizaciones, un marcador se adhiere a un agente mediante las aminas primarias de cadenas laterales de lisina. En algunas realizaciones, el marcaje químico comprende un método de química de NHS-éster.

Como alternativa, el marcador detectable se adhiere por marcaje enzimático. Los métodos de marcaje enzimático pueden incluir, aunque sin limitación, métodos de péptido aceptor de biotina/ligasa de biotina (AP/Bir A), proteína transportadora de acilo/fosfopanteteína transferasa (ACP/PPTasa), O6-alquilguanina transferasa humana (hAGT), marca-Q/transglutaminasa (TGasa), marca de aldehído/enzima generadora de formilglicina, deshalogenasa procarionota mutada (HaloTagTM) y motivo de farnesilación/proteína farnesiltransferasa (PFTasa). El marcaje por afinidad puede incluir, aunque sin limitación, métodos no covalentes que utilizan dihidrofolato reductasa (DHFR) y mutante Phe36Val de la proteína 12 de unión a FK506 (FKBP12(F36V)) y métodos de quelación de metales.

Pueden usarse reactivos de reticulación para adherir un marcador detectable al cebador, marca oligonucleotídica, molécula marcada, amplicón marcado y/o ácido nucleico no marcado. En algunos casos, el reactivo de reticulación es glutaraldehído. El glutaraldehído puede reaccionar con grupos amina para crear reticulaciones mediante varias rutas. Por ejemplo, en condiciones reductoras, los aldehídos en ambos extremos del glutaraldehído se acoplan con aminas para formar enlaces amina secundarios.

En algunos casos, la adhesión del marcador detectable al cebador, marca oligonucleotídica, molécula marcada, amplicón marcado y/o ácido nucleico no marcado comprende activación con peryodato seguido de aminación reductora. En algunos casos, se usan reticulantes de sulfo-SMCC u otros reticulantes heterobifuncionales para conjugar el marcador detectable al cebador, marca oligonucleotídica, molécula marcada, amplicón marcado y/o ácido nucleico no marcado. Por ejemplo, se usa sulfo-SMCC para conjugar una enzima a un fármaco. En algunas realizaciones, la enzima se activa y purifica en una etapa y después se conjuga con el fármaco en una segunda etapa. En algunas realizaciones, la direccionalidad de la reticulación está limitada a una orientación específica (por ejemplo, aminas en la enzima con grupos sulfidrido en el anticuerpo).

Enfermedades/afecciones

En la presente memoria se divulgan métodos, kits y composiciones para diagnosticar, controlar y/o pronosticar un estado o resultado de una enfermedad o afección en un sujeto. En general, el método comprende (a) marcar de forma estocástica dos o más moléculas de dos o más muestras para producir dos o más ácidos nucleicos marcados; (b) detectar y/o cuantificar los dos o más ácidos nucleicos marcados; y (c) diagnosticar, controlar y/o pronosticar un estado o resultado de una enfermedad o afección en un sujeto basándose en la detección y/o cuantificación de los

dos o más ácidos nucleicos marcados. El método puede comprender además determinar un régimen terapéutico. Las dos o más muestras pueden comprender una o más muestras de un sujeto que padece una enfermedad o afección. Las dos o más muestras pueden comprender una o más muestras de un sujeto sano. Las dos o más muestras pueden comprender una o más muestras de un control.

- 5 El control de una enfermedad o afección puede comprender además el control de un régimen terapéutico. El control de un régimen terapéutico puede comprender determinar la eficacia de un régimen terapéutico. En algunos casos, el control de un régimen terapéutico comprende administrar, terminar, añadir o alterar un régimen terapéutico. La alteración de un régimen terapéutico puede comprender aumentar o reducir la dosificación, frecuencia de dosis o modo de administración de un régimen terapéutico. Un régimen terapéutico puede comprender uno o más fármacos terapéuticos. Los fármacos terapéuticos pueden ser un fármaco antineoplásico, fármaco antivírico, fármaco antibacteriano, fármaco antipatógeno o cualquier combinación de los mismos.

Cáncer

- 15 En algunos casos, la enfermedad o afección es un cáncer. Las moléculas a marcar de forma estocástica pueden ser de una célula o tejido canceroso. En algunos casos, el cáncer es un sarcoma, carcinoma, linfoma o leucemia. Los sarcomas son cánceres del hueso, cartílago, grasa, músculo, vasos sanguíneos u otro tejido conjuntivo o de soporte. Los sarcomas incluyen, aunque sin limitación, cáncer de huesos, fibrosarcoma, condrosarcoma, sarcoma de Ewing, hemangioendotelioma maligno, schwannoma maligno, schwannoma vestibular bilateral, osteosarcoma, sarcomas de tejido blando (por ejemplo, sarcoma de la parte blanda alveolar, angiosarcoma, cistosarcoma filoides, dermatofibrosarcoma, tumor desmoide, sarcoma epiteliode, osteosarcoma extraesquelético, fibrosarcoma hemangiopericitoma, hemangiosarcoma, sarcoma de Kaposi, leiomiomasarcoma, liposarcoma, linfangiosarcoma, linfosarcoma, histiocitoma fibroso maligno, neurofibrosarcoma, rbdomiosarcoma y sarcoma sinovial).

- 25 Los carcinomas son cánceres que empiezan en las células epiteliales, que son células que cubren la superficie del organismo, producen hormonas y componen las glándulas. A modo de ejemplo no limitante, los carcinomas incluyen cáncer de mama, cáncer pancreático, cáncer pulmonar, cáncer de colon, cáncer colorrectal, cáncer rectal, cáncer renal, cáncer de vejiga, cáncer de estómago, cáncer de próstata, cáncer hepático, cáncer de ovario, cáncer cerebral, cáncer vaginal, cáncer vulvar, cáncer uterino, cáncer bucal, cáncer de pene, cáncer testicular, cáncer esofágico, cáncer de piel, cáncer de las trompas de Falopio, cáncer de cabeza y cuello, cáncer del estroma gastrointestinal, adenocarcinoma, melanoma cutáneo o intraocular, cáncer de la región anal, cáncer del intestino delgado, cáncer del sistema endocrino, cáncer de glándula tiroidea, cáncer de la glándula paratiroidea, cáncer de la glándula suprarrenal, cáncer de la uretra, cáncer de la pelvis renal, cáncer del uréter, cáncer de endometrio, cáncer del cuello uterino, cáncer de la glándula pituitaria, neoplasias del sistema nervioso central (SNC), linfoma primario del SNC, glioma del tronco cerebral y tumores del eje medular. En algunos casos, el cáncer es un cáncer de piel, tal como un carcinoma basocelular, carcinoma escamocelular, melanoma, no melanoma o queratosis actínica (solar).

- 35 En algunos casos, el cáncer es un cáncer pulmonar. El cáncer pulmonar puede empezar en las vías respiratorias que se ramifican desde la tráquea para suministrar a los pulmones (bronquios) o a los sacos aéreos pequeños del pulmón (los alvéolos). Los cánceres pulmonares incluyen carcinoma broncopulmonar no microcítico (NSCLC), carcinoma broncopulmonar microcítico y mesotelioma. Los ejemplos de NSCLC incluyen carcinoma escamocelular, adenocarcinoma y carcinoma de células grandes. El mesotelioma puede ser un tumor canceroso del revestimiento del pulmón y la cavidad torácica (pleura) o revestimiento del abdomen (peritoneo). El mesotelioma puede deberse a exposición a asbestos. El cáncer puede ser un cáncer cerebral, tal como un glioblastoma.

- 40 Como alternativa, el cáncer puede ser un tumor del sistema nervioso central (SNC). Los tumores del SNC pueden clasificarse como gliomas o no gliomas. El glioma puede ser glioma maligno, glioma de alto grado, glioma pontino intrínseco difuso. Los ejemplos de gliomas incluyen astrocitomas, oligodendrogliomas (o mezclas de elemento de oligodendroglioma y astrocitoma) y ependimomas. Los astrocitos incluyen, aunque sin limitación, astrocitomas de grado bajo, astrocitomas anaplásicos, glioblastoma multiforme, astrocitoma pilocítico, xantastrocitoma pleomórfico y astrocitoma de gigantocitos subependimarias. Los oligodendrogliomas incluyen oligodendrogliomas de grado bajo (u oligoastrocitomas) y oligodendrogliomas anaplásicos. Los no gliomas incluyen meningiomas, adenomas de la pituitaria, linfomas primarios del SNC y meduloblastomas. En algunos casos, el cáncer es un meningioma.

- 50 La leucemia puede ser una leucemia linfocítica aguda, leucemia mielocítica aguda, leucemia linfocítica crónica o leucemia mielocítica crónica. Tipos adicionales de leucemia incluyen leucemia de células pilosas, leucemia mielomonocítica crónica y leucemia mielomonocítica juvenil.

- 55 Los linfomas son cánceres de los linfocitos y pueden desarrollarse a partir de linfocitos B o T. Los dos tipos principales de linfoma son linfoma de Hodgkin, previamente conocido como enfermedad de Hodgkin y linfoma no hodgkiniano. El linfoma de Hodgkin está evidenciado por la presencia de la célula de Reed-Sternberg. Los linfomas no hodgkinianos son todos linfomas que no son linfomas de Hodgkin. Los linfomas no hodgkinianos pueden ser linfomas indolentes y linfomas agresivos. Los linfomas no hodgkinianos incluyen, aunque sin limitación, linfoma de linfocitos B grande difuso, linfoma folicular, linfoma de tejido linfático asociado a la mucosa (MALT), linfoma linfocítico de células pequeñas, linfoma de células del manto, linfoma de Burkitt, linfoma de linfocitos B grande mediastinal, macroglobulinemia de Waldenstrom, linfoma de linfocitos B de la zona marginal nodal, linfoma de la zona marginal esplénica (SMZL), linfoma de linfocitos B de la zona marginal extranodal, linfoma de linfocitos B grande intravascular, linfoma de efusión primaria y granulomatosis linfomatoide.

Infección patógena

En algunos casos, la enfermedad o afección es una infección patógena. Las moléculas a marcar de forma estocástica pueden ser de un patógeno. El patógeno puede ser un virus, bacteria, hongo o protozoo. En algunos casos, el patógeno puede ser un protozoo, tal como *Acanthamoeba* (por ejemplo, *A. astronyxis*, *A. castellanii*, *A. culbertsoni*, *A. hatchetti*, *A. polyphaga*, *A. rhyssodes*, *A. healyi*, *A. divionensis*), *Brachiola* (por ejemplo, *B. connori*, *B. vesicularum*), *Cryptosporidium* (por ejemplo, *C. parvum*), *Cyclospora* (por ejemplo, *C. cayetanensis*), *Encephalitozoon* (por ejemplo, *E. cuniculi*, *E. hellem*, *E. intestinalis*), *Entamoeba* (por ejemplo, *E. histolytica*), *Enterocytozoon* (por ejemplo, *E. bienesi*), *Giardia* (por ejemplo, *G. lamblia*), *Isospora* (por ejemplo, *I. belli*), *Microsporidium* (por ejemplo, *M. africanum*, *M. ceylonensis*), *Naegleria* (por ejemplo, *N. fowleri*), *Nosema* (por ejemplo, *N. algerae*, *N. oculorum*), *Pleistophora*, *Trachipleistophora* (por ejemplo, *T. anthropophthera*, *T. hominis*) y *Vittaforma* (por ejemplo, *V. corneae*). El patógeno puede ser un hongo tal como *Candida*, *Aspergillus*, *Cryptococcus*, *Histoplasma*, *Pneumocystis* y *Stachybotrys*.

El patógeno puede ser una bacteria. Las bacterias ejemplares incluyen, aunque sin limitación, *Bordetella*, *Borrelia*, *Brucella*, *Campylobacter*, *Chlamydia*, *Chlamydomydia*, *Clostridium*, *Corynebacterium*, *Enterococcus*, *Escherichia*, *Francisella*, *Haemophilus*, *Helicobacter*, *Legionella*, *Leptospira*, *Listeria*, *Mycobacterium*, *Mycoplasma*, *Neisseria*, *Pseudomonas*, *Rickettsia*, *Salmonella*, *Shigella*, *Staphylococcus*, *Streptococcus*, *Treponema*, *Vibrio*, o *Yersinia*.

El virus puede ser un virus de transcripción inversa. Los ejemplos de virus de transcripción inversa incluyen, aunque sin limitación, virus de ARN monocatenario-RT (ARNmc-RT) y virus de ADN bicatenario-RT (ADNbc-RT). Los ejemplos no limitantes de virus ARNmc-RT incluyen retrovirus, alfarretrovirus, betarretrovirus, gammaretrovirus, deltarretrovirus, epsilonretrovirus, lentivirus, espumavirus, metavirus y pseudovirus. Los ejemplos no limitantes de virus ADNbc-RT incluyen hepadenovirus y caulimovirus. El virus puede ser un virus de ADN. El virus puede ser un virus de ADN. El virus de ADN puede ser un virus de ADN bicatenario (ADNbc). En algunos casos, el virus de ADNbc es un adenovirus, herpesvirus o poxvirus. Los ejemplos de adenovirus incluyen, aunque sin limitación, adenovirus y virus de la hepatitis canina infecciosa. Los ejemplos de herpesvirus incluyen, aunque sin limitación, virus del herpes simple, virus de varicella-zoster, citomegalovirus y virus de Epstein-Barr. Una lista no limitante de poxvirus incluye virus de la viruela, virus de la viruela bovina, virus de la viruela ovina, virus de la viruela del simio y virus *vaccinia*. El virus de ADN puede ser un virus de ADN monocatenario (ADNmc). El virus de ADNmc puede ser un parvovirus. Los ejemplos de parvovirus incluyen, aunque sin limitación, parvovirus B19, parvovirus canino, parvovirus de ratón, parvovirus porcino, virus de panleucopenia felina y de la enteritis visón.

El virus puede ser un virus de ARN. El virus de ARN puede ser un virus de ARN bicatenario (ARNbc), virus de ARN monocatenario codificante (+) (ARNc(+)), o virus de ARN monocatenario no codificante (-) (ARNc(-)). Una lista no limitante de virus de ARNbc incluye reovirus, ortorreovirus, cipovirus, rotavirus, virus de la lengua azul y fitorreovirus. Los ejemplos de virus de ARNmc (+) incluyen, aunque sin limitación, picornavirus y togavirus. Los ejemplos de picornavirus incluyen, aunque sin limitación, enterovirus, rinovirus, hepatovirus, cardiovirus, aftovirus, poliovirus, parecovirus, erbovirus, cobuvirus, tescovirus y coxsackie. En algunos casos, el togavirus es un virus de la rubéola, virus Sindbis, virus de la encefalitis equina oriental, virus de la encefalitis equina occidental, virus de la encefalitis equina venezolana, virus del río Ross, virus de la selva O'nyong'nyong, Chikungunya o Semliki. Una lista no limitante de virus de ARNmc (-) incluye ortomixovirus y rabdovirus. Los ejemplos de ortomixovirus incluyen, aunque sin limitación, virus de la gripe A, virus de la gripe B, virus de la gripe C, isavirus y togotovirus. Los ejemplos de rabdovirus incluyen, aunque sin limitación, citorabdovirus, dicorabdovirus, efemerovirus, lisavirus, novirabdovirus y vesiculovirus.

Trastornos fetales

En algunos casos, la enfermedad o trastorno es en el embarazo. Los métodos y kits divulgados en la presente memoria pueden comprender el diagnóstico de una afección fetal en una mujer embarazada. Los métodos y kits divulgados en la presente memoria pueden comprender la identificación de mutaciones o anomalías genéticas fetales. Las moléculas a marcar de forma estocástica pueden ser de una célula o tejido fetal. Como alternativa, o adicionalmente, las moléculas a marcar pueden ser de la mujer embarazada.

Los métodos y kits divulgados en la presente memoria pueden usarse en el diagnóstico, predicción o control de trisomía autosómica (por ejemplo, trisomía del 13, 15, 16, 18, 21 o 22). En algunos casos, la trisomía puede estar asociada con una posibilidad aumentada de aborto (por ejemplo, trisomía del 15, 16 o 22). En otros casos, la trisomía que se detecta es una trisomía en el recién nacido que puede indicar que un bebé nacerá con defectos al nacer (por ejemplo, trisomía del 13 (síndrome de Patau), trisomía del 18 (síndrome de Edwards) y trisomía del 21 (síndrome de Down). La anomalía también puede ser de un cromosoma sexual (por ejemplo, XXY (síndrome de Klinefelter), XYY (síndrome de Jacobs) o XXX (trisomía del X). La molécula o moléculas a marcar pueden ser en uno o más de los siguientes cromosomas: 13, 18, 21, X o Y. Por ejemplo, la molécula está en el cromosoma 21 y/o en el cromosoma 18 y/o en el cromosoma 13.

Afecciones fetales adicionales que pueden determinarse basándose en los métodos y kits divulgados en la presente memoria incluyen monosomía de uno o más cromosomas (monosomía del cromosoma X, también conocido como síndrome de Turner), trisomía de uno o más cromosomas (13, 18, 21 y X), tetrasomía y pentasomía de uno o más cromosomas (que en seres humanos se observa más habitualmente en los cromosomas sexuales, por ejemplo, XXXX, XYYY, XXXY, XYYY, XXXXX, XXXXY, XXXYY, XYYYY y XYYYY), monoploidía, triploidía (tres de cada cromosoma, por ejemplo, 69 cromosomas en seres humanos), tetraploidía (cuatro de cada cromosoma, por ejemplo, 92 cromosomas en seres humanos), pentaploidía y multiploidía.

En la presente memoria se divulga además un método de análisis forense, que comprende cualquiera de los

métodos descritos anteriormente. La ciencia forense puede usar ácidos nucleicos de diversas muestras (por ejemplo, sangre, semen, piel, saliva, cabello) encontradas en una escena del crimen para identificar la presencia de un individuo en la escena, tal como un autor del crimen. Este proceso se denomina formalmente perfil de ADN, pero también puede llamarse "huella genética". Por ejemplo, el perfil de ADN comprende medir y comparar las longitudes de secciones variables de ADN repetitivo, tales como repeticiones en tándem corto y minisatélites, en diversas muestras y personas. Este método es habitualmente una técnica extremadamente fiable para emparejar una muestra de ADN de una persona con ADN en una muestra encontrada en la escena de un crimen. Sin embargo, la identificación puede complicarse si la escena está contaminada con ADN de varias personas. En este caso, así como en otras aplicaciones forenses, puede ser ventajoso obtener la cuantificación absoluta de ácidos nucleicos de una célula individual o pequeñas cantidades de células.

En algunos casos, la enfermedad o afección es un trastorno inmunitario. Un trastorno inmunitario puede ser un trastorno inflamatorio, un trastorno autoinmunitario, síndrome irritable del intestino, o colitis ulcerante. Los ejemplos de enfermedades autoinmunitarias incluyen enfermedad de Crohn, lupus y enfermedad de Graves.

En algunos casos, la enfermedad o trastorno es una afección o trastorno neurológico. Una afección o trastorno neurológico puede ser afasia epileptiforme adquirida, encefalitis diseminada aguda, adrenoleucodistrofia, agénesis del cuerpo caloso, agnosia, síndrome de Aicardi, enfermedad de Alexander, enfermedad de Alper, hemiplejia alternante, enfermedad de Alzheimer, esclerosis lateral amiotrófica (véase enfermedad de neuronas motoras), anencefalia, síndrome de Angelman, angiomatosis, anoxia, afasia, apraxia, quistes aracnoideos, aracnoiditis, malformación de Arnold-Chiari, malformación arteriovenosa, síndrome de Asperger, ataxia telangiectasia, trastorno de hiperactividad con déficit de atención, autismo, trastorno del procesamiento auditivo, disfunción autonómica, dolor de espalda, enfermedad de Batten, enfermedad de Behcet, parálisis de Bell, blefaroespasma esencial benigno, amiotrofia focal benigna, hipertensión intracraneal benigna, polimicroguría frontoparietal bilateral, enfermedad de Binswanger, blefaroespasma, síndrome de Bloch-Sulzberger, lesión del plexo braquial, abscesos cerebrales, daños cerebrales, lesión cerebral, tumor cerebral, síndrome de Brown-Sequard, enfermedad de Canavan, síndrome del túnel carpiano (CTS), causalgia, síndrome de dolor central, mielínolisis pontina central, miopatía centronuclear, trastorno cefálico, aneurisma cerebral, arteriosclerosis cerebral, atrofia cerebral, gigantismo cerebral, parálisis cerebral, enfermedad de Charcot-Marie-Tooth, malformación de Chiari, corea, polineuropatía desmielinizante inflamatoria crónica (CIDP), dolor crónico, síndrome del dolor regional crónico, síndrome de Coffin Lowry, coma, incluyendo estado vegetativo persistente, síndrome de deficiencia del complejo I, síndrome de deficiencia de complejo I, síndrome de deficiencia de complejo II, síndrome de deficiencia de complejo III, síndrome de deficiencia de deficiencia de complejo IV/COX, síndrome de deficiencia de complejo V, diplejía facial congénita, degeneración corticobasal, arteritis craneal, craneosinostosis, enfermedad de Creutzfeldt-Jakob, trastornos de traumatismo acumulado, síndrome de Cushing, enfermedad por cuerpos de inclusión citomegálicos (CIBD), infección por citomegalovirus, síndrome de Dandy-Walker, enfermedad de Dawson, deficiencia del componente NADH deshidrogenasa mitocondrial del complejo I, síndrome de Morsier, parálisis de Dejerine-Klumpke, enfermedad de Dejerine-Sottas, síndrome de la fase del sueño retardada, demencia, dermatomiositis, dispraxia neurológica, neuropatía diabética, esclerosis difusa, disautonomía, discalculia, disgrafía, dislexia, distonía, encefalopatía epiléptica infantil temprana, síndrome de Empty sella, encefalitis, encefalocele, angiomatosis encefalotrigeminal, encopresis, epilepsia, parálisis de Erb, eritromelalgia, temblores esenciales, enfermedad de Fabry, síndrome de Fahr, desmayo, parálisis espástica familiar, ataques febriles, síndrome de Fisher, ataxia de Friedrich, Síndrome de FART, enfermedad de Gaucher, síndrome de Gerstmann, arteritis de gigantocitos, enfermedad por inclusión de gigantocitos, leucodistrofia de células globoides, heterotopía de la materia gris, síndrome de Guillain-Barre, mielopatía asociada a HTLV-1, enfermedad de Hallervorden-Spatz, traumatismo craneal, cefalea, espasmo hemifacial, paraplejia espástica hereditaria, heredopatía atáctica polineurítica, herpes zoster ótico, herpes zoster, síndrome de Hirayama, holoprosencefalia, enfermedad de Huntington, hidranencefalia, hidrocefalo, hipercortisolismo, hipoxia, encefalomielitis mediada por el sistema inmunitario, miositis con cuerpos de inclusión, incontinencia pigmentaria, enfermedad por almacenamiento de ácido fitánico infantil, enfermedad de Refsum infantil, espasmos infantiles, miopatía inflamatoria, quiste intracraneal, hipertensión intracraneal, síndrome de Joubert, síndrome de Kearns-Sayre, enfermedad de Kennedy, síndrome de Kinsbourne, síndrome de Klippel Feil, enfermedad de Krabbe, síndrome de Kufor-Rakeb, enfermedad de Kugelberg-Welander, Kuru, enfermedad de Lafora, síndrome de miasténico de Lambert-Eaton, síndrome de Landau-Kleffner, síndrome medular lateral (de Wallenberg), dificultades de aprendizaje, enfermedad de Leigh, síndrome de Lennox-Gastaut, síndrome de Lesch-Nyhan, leucodistrofia, demencia con cuerpos de Lewy, lisencefalia, síndrome de enclaustramiento, enfermedad de Lou Gehrig, enfermedad del disco lumbar, secuelas neurológicas por enfermedad de Lyme, enfermedad de Machado-Joseph (ataxia medulocerebral de tipo 3), macrencefalia, enfermedad de orina de jarabe de arce, megalencefalia, síndrome de Melkersson-Rosenthal, enfermedad de Menieres, enfermedad de Menkes, leucodistrofia metacromática, microcefalia, migraña, síndrome de Miller Fisher, miniapoplejías, enfermedad mitocondrial, disfunción mitocondrial, miopatías mitocondriales, deficiencia de la cadena respiratoria mitocondrial del complejo I, síndrome de Mobius, amiotrofia monomérica, enfermedad de neuronas motoras, trastorno de habilidades motoras, enfermedad de Moyamoya, mucopolisacaridosis, demencia por infarto múltiple, neuropatía motora multifocal, esclerosis múltiple, atrofia de múltiples sistemas con hipotensión postural, distrofia, encefalomielitis mialgia, miastenia grave, esclerosis difusa mielínoclastica, encefalopatía mioclónica infantil, mioclono, miopatía, miopatía miotubular, miotonía congénita, deficiencia de NADH-coenzima Q reductasa, deficiencia de NADH:Q(1) oxidoreductasa, narcolepsia, neurofibromatosis, síndrome maligno neuroléptico, manifestaciones neurológicas del SIDA, secuelas neurológicas de lupus, neuromiotonía, lipofuscinosis cerioide neuronal, trastornos de migración neuronal, enfermedad de Niemann-Pick, síndrome sin sueño/vigilia en 24 horas, trastorno de aprendizaje no verbal, síndrome de O'Sullivan-McLeod, neuralgia occipital, secuencia de disrafia espinal oculta, síndrome de Ohtahara, atrofia olivopontocerebelar, síndrome de mioclono obsoclon, neuritis óptica, hipotensión ortostática, síndrome

abusivo, trastornos de fosforilación oxidativa, palinopsia, parestesia, enfermedad de Parkinson, paramiotonía congénita, enfermedades paraneoplásicas, ataques paroximales, síndrome de Parry-Romberg (también conocido como síndrome de Rombergs), enfermedad de Pelizaeus-Merzbacher, parálisis periódicas, neuropatía periférica, estado vegetativo persistente, trastornos neurológicos dominantes, reflejo de estornudo por luz brillante, enfermedad por almacenamiento de ácido fitánico, enfermedad de Pick, tumores nervio pinzado, tumores de la pituitaria, PMG, poliomielitis, polimicroguría, polimiositis, pencefalia, síndrome pospoliomielitis, neuralgia posherpética (PHN), encefalomielitis posinfecciosa, hipotensión postural, síndrome de Prader-Willi, esclerosis lateral primaria, enfermedades por priones, atrofia hemifacial progresiva también conocida como síndrome de Rombergs, leucoencefalopatía focal progresiva, polidistrofias esclerosante progresiva, parálisis supranuclear progresiva, pseudotumor cerebri, síndrome de Ramsay-Hunt (de tipo I y tipo II), encefalitis de Rasmussen, síndrome de distrofia del sistema simpático reflejo, enfermedad de Refsum, trastornos de movimiento repetitivo, lesión por estrés repetitivo, síndrome de piernas inquietas, mielopatía asociada a retrovirus, síndrome de Rett, síndrome de Reye, síndrome de Rombergs, rabia, baile de San Vito, enfermedad de Sandhoff, esquizofrenia, enfermedad de Schilder, esquizencefalia, disfunción de integración sensitiva, displasia septoóptica, síndrome del bebé zarandeado, herpes, síndrome de Shy-Drager, síndrome de Sjogren, apnea del sueño, enfermedad del sueño, Sniation, síndrome de Sotos, espasticidad, espina bífida, lesión de la médula espinal, tumores de la médula espinal, atrofia muscular espinal, estenosis espinal, síndrome de Steele-Richardson-Olszewski, véase parálisis supranuclear progresiva, ataxia espinocerebral, síndrome de la persona rígida, subplejía, síndrome de Sturge-Weber, panencefalitis esclerosante subaguda, encefalopatía arterioesclerótica subcortical, siderosis superficial, corea de Sydenham, síncope, sinestesia, siringomielia, disquinesia tardía, enfermedad de Tay-Sachs, arteritis temporal, síndrome de médula espinal anclada, enfermedad de Thomsen, síndrome del opérculo torácico, tic doloroso, parálisis de Todd, síndrome de Tourette, ataque isquémico transitorio, encefalopatías espongiiformes transmisibles, mielitis transversa, lesión cerebral traumática, temblores, neuralgia del trigémino, paraparesia espástica tropical, tripanosomiasis, esclerosis tuberosa, vasculitis incluyendo arteritis temporal, enfermedad de Von Hippel-Lindau (VHL), encefalomielitis de Viliuisk (VE), síndrome de Wallenberg, enfermedad de Werdnig-Hoffman, síndrome de West, hiperextensión cervical, síndrome de Williams, enfermedad de Wilson, atrofia espinal y muscular bulbar ligada al X o síndrome de Zellweger.

Definiciones

Cuando se proporciona un intervalo de valores, se entiende que cada valor intermedio, hasta la décima de la unidad del límite inferior salvo que el contexto indique claramente lo contrario, entre el límite superior e inferior de ese intervalo también se divulga específicamente. Cada intervalo más pequeño entre cualquier valor indicado o valor intermedio en un intervalo indicado o cualquier otro valor indicado o intermedio en ese valor indicado está abarcado dentro de la invención. El límite superior e inferior de estos intervalos más pequeños puede incluirse o excluirse independientemente en el intervalo, y cada intervalo en el que se incluya alguno, ninguno o ambos límites en los intervalos más pequeños también está abarcado dentro de la invención, sometido a cualquier límite específicamente excluido en el intervalo indicado. Cuando el intervalo indicado incluye uno o ambos límites, los intervalos que excluyen alguno o ambos límites incluidos también se incluyen en la invención.

Salvo que se definan de otro modo, todos los términos técnicos y científicos usados en la presente memoria tienen el mismo significado que el comprendido habitualmente por un experto en la materia a la que pertenece esta invención. Aunque puede usarse cualquier método y material similar o equivalente a los descritos en la presente memoria en la práctica o ensayo de la presente invención, ahora se describen algunos métodos y materiales potenciales y preferidos.

Como será evidente para los expertos en la materia tras leer esta divulgación, cada una de las realizaciones individuales descrita e ilustrada en la presente memoria tiene características y componentes concretos que pueden separarse fácilmente de o combinarse con las características de cualquiera de las otras varias realizaciones sin alejarse del alcance de la presente invención. Cualquier método indicado puede realizarse en el orden de eventos indicado o en cualquier otro orden que sea lógicamente posible.

Debe apreciarse que tal como se usa en la presente memoria y las reivindicaciones anejas, las formas singulares "un/o", "una" y "el/la" incluyen referentes en plural salvo que el contexto indique claramente lo contrario. Por tanto, por ejemplo, una referencia a "una célula" incluye una pluralidad de dichas células y una referencia a "el péptido" incluye referencias a uno o más péptidos y equivalentes de los mismos, por ejemplo, polipéptidos, conocidos por los expertos en la materia, y así sucesivamente.

Tal como se usa en la presente memoria, el término "marcador" puede referirse a una secuencia oligonucleotídica única que puede permitir que una base de ácido nucleico correspondiente y/o secuencia de ácido nucleico se identifique. En algunas realizaciones, la base de ácido nucleico y/o secuencia de ácido nucleico puede estar ubicada en una posición específica en una secuencia polinucleotídica más grande (por ejemplo, un polinucleótido adherido a una microesfera).

Tal como se usa en la presente memoria, el término "hibridación" puede referirse al proceso en que dos polinucleótidos monocatenarios se unen no covalentemente para formar un polinucleótido bicatenario estable. El término "hibridación" también se refiere a hibridación de tres hebras. El polinucleótido bicatenario resultante (habitualmente) es un "híbrido" o "dúplex".

Tal como se usa en la presente memoria, "nucleósido" puede incluir nucleósidos naturales, tal como las formas 2'-desoxi y 2'-hidroxilo. "Análogos" en referencia a nucleósidos pueden incluir nucleósidos sintéticos que comprenden

restos de base modificados y/o restos de azúcar modificados, o similares. los análogos pueden tener capacidad de hibridación. Los análogos pueden incluir nucleósidos sintéticos diseñados para potenciar las propiedades de unión, reducir la complejidad, aumentar la especificidad y similares. Tipos ejemplares de análogos pueden incluir fosforamidatos oligonucleotídicos (mencionados en la presente memoria como "amidatos"), ácidos peptidonucleicos (mencionados en la presente memoria como "PNA"), oligo-2'-O-alkuilribonucleótidos, polinucleótidos que contienen propinilpirimidinas C-5 y ácidos nucleicos bloqueados (LNA).

Tal como se usan en la presente memoria, las expresiones "molécula de ácido nucleico", "secuencia de ácido nucleico", "fragmento de ácido nucleico", "oligonucleótido", "fragmento oligonucleotídico" y "polinucleótido" pueden usarse indistintamente y pueden estar destinadas a incluir, aunque sin limitación, formas poliméricas de nucleótidos que pueden tener diversas longitudes, desoxirribonucleótidos o ribonucleótidos, o análogos de los mismos. Las moléculas de ácido nucleico pueden incluir ADN monocatenario (ADNmc), ADN bicatenario (ADNbc), ARN monocatenario (ARNmc) y ARN bicatenario (ARNbc). Diferentes moléculas de ácido nucleico pueden tener diferentes estructuras tridimensionales, y pueden realizar diversas funciones. Ejemplos no limitantes de moléculas de ácido nucleico pueden incluir un gen, un fragmento génico, un hueco genómico, un exón, un intrón, ADN intergénico (incluyendo, sin limitación, ADN heterocromático), ARN mensajero (ARNm), ARN transferente, ARN ribosómico, ribozimas, ARN interferente pequeño (ARNip), miARN, ARN nucleolar pequeño (ARNnp), ADNc, polinucleótidos recombinantes, polinucleótidos ramificados, plásmidos, vectores, ADN aislado de una secuencia, ARN aislado de una secuencia, sondas de ácido nucleico y cebadores.

Los oligonucleótidos pueden referirse a un polímero lineal de monómeros nucleosídicos naturales o modificados unidos mediante enlaces fosfodiéster o análogos de los mismos. Un "fragmento oligonucleotídico" se refiere a una secuencia oligonucleotídica que se ha escindido en dos o más secuencias oligonucleotídicas más pequeñas. Los oligonucleótidos pueden ser naturales o sintéticos. Los oligonucleótidos pueden incluir desoxirribonucleósidos, ribonucleósidos y análogos no naturales de los mismos, tales como formas anoméricas de los mismos, ácidos peptidonucleicos (PNA) y similares. Los oligonucleótidos pueden tener capacidad de unión específica a un genoma diana mediante un patrón regular de interacciones monómero a monómero, tales como el emparejamiento de bases de Watson-Crick, el apilamiento de bases, el emparejamiento de bases de Hoogsteen o Hoogsteen inverso o similares. Los oligonucleótidos y el término "polinucleótidos" pueden usarse indistintamente en la presente memoria.

Siempre que un oligonucleótido se represente mediante una secuencia de letras, tal como "ATGCCTG", puede entenderse que los nucleótidos están en orden de 5' a 3' de izquierda a derecha y que "A" indica desoxiadenosina, "C" indica desoxicitidina, "G" indica desoxiguanina, "T" indica desoxitimidina y "U" indica el ribonucleósido uridina, salvo que se indique lo contrario.

Los oligonucleótidos pueden incluir uno o más oligonucleótidos no convencionales, análogos nucleotídicos y/o nucleótidos modificados. Ejemplos de nucleótidos modificados pueden incluir, aunque sin limitación, diaminopurina, 5-fluorouracilo, 5-bromouracilo, 5-clorouracilo, 5-yodouracilo, hipoxantina, xantina, 4-acetilcitosina, 5-(carboxihidroximetil)uracilo, 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-metilguanina, 5-metilaminometiluracilo, 5-metoxiametil-2-tiouracilo, beta-D-manosilqueosina, 5'-metoxicarboximetiluracilo, 5'-metoxiuracilo, 2-metil-D46-isopenteniladenina, ácido uracil-5-oxiacético (v), wibutoxosina, pseudouracilo, queosina, 2-tiocitosina, 5-metil-2-tiouracilo, 2-tiouracilo, 4-tiouracilo, 5-metiluracilo, éster metílico de ácido uracil-5-oxiacético, ácido uracil-5-oxiacético (v), 5-metil-2-tiouracilo, 3-(3-amino-3-N-2-carboxipropil) uracilo, (acp3)w, fosforamidatos oligonucleotídicos (mencionados en la presente memoria como "amidatos"), ácidos peptidonucleicos (mencionados en la presente memoria como "PNA"), oligo-2'-O-alkuilribonucleótidos, polinucleótidos que contienen propinilpirimidinas C-5, ácidos nucleicos bloqueados (LNA), 2,6-diaminopurina y similares. Las moléculas de ácido nucleico también pueden modificarse en el resto de base (por ejemplo, en uno o más átomos que están típicamente disponibles para formar un enlace de hidrógeno con un nucleótido complementario y/o en uno o más átomos que típicamente no pueden formar un enlace de hidrógeno con un nucleótido complementario, el resto de azúcar o la cadena principal de fosfato).

Tal como se usa en la presente memoria, una "muestra" puede referirse a una célula individual o muchas células. Las moléculas de ácido nucleico pueden obtenerse de una o más muestras. Una muestra puede comprender un único tipo celular o una combinación de dos o más tipos celulares. Una muestra puede incluir una colección de células que realiza una función similar tal como las encontradas, por ejemplo, en un tejido. Una muestra puede comprender uno o más tejidos. Los ejemplos de tejidos pueden incluir, aunque sin limitación, tejido epitelial (por ejemplo, piel, el revestimiento de las glándulas, intestino, piel y órganos tales como hígado, pulmón, riñón), endotelio (por ejemplo, el revestimiento de los vasos sanguíneos y linfáticos), mesotelio (por ejemplo, el revestimiento de los espacios pleural, peritoneal y pericárdico), mesénquima (por ejemplo, células que rellenan los espacios entre los órganos, incluyendo células de grasa, músculo, hueso, cartilago y tendón), células sanguíneas (por ejemplo, eritrocitos, granulocitos, neutrófilos, eosinófilos, basófilos, monocitos, linfocitos T (también conocidos como células T), linfocitos B (también conocidos como células B), células plasmáticas, megacariocitos y similares), neuronas, células germinales (por ejemplo, espermatozoides, ovocitos), células de líquido amniótico, placenta, células madre y similares. Una muestra puede obtenerse de una o más células individuales en cultivo, muestras metagenómicas, células madre embrionarias, células madre pluripotentes inducidas, muestras cancerosas, secciones tisulares y biopsias, o cualquier combinación de las mismas.

Tal como se usa en la presente memoria, el término "organismo" puede incluir, aunque sin limitación, un ser humano, un primate no humano, una vaca, un caballo, una oveja, una cabra, un cerdo, un perro, un gato, un conejo,

un ratón, una rata, un jerbo, una rana, un sapo, un pez (por ejemplo, *Danio rerio*) una ascáride (por ejemplo, *C. elegans*) y cualquier especie transgénica de los mismos. El término "organismo" también puede incluir, aunque sin limitación, una célula de levadura (por ejemplo, *S. cerevisiae*), una tetrada de levadura, una colonia de levadura, una bacteria, una colonia bacteriana, un virión, virosoma, partícula de tipo virus y/o cultivos de los mismos, y similares.

- 5 Tal como se usa en la presente memoria, el término "adherir", "conjugar" y "acoplar" pueden usarse indistintamente y pueden referirse a tanto interacciones covalentes como a interacciones no covalentes.

Aunque se han mostrado y descrito en la presente memoria realizaciones preferidas de la presente invención, será obvio para los expertos en la materia que dichas realizaciones se proporcionan a modo de ejemplo únicamente. A los expertos en la materia se les ocurrirán numerosas variaciones, cambios y sustituciones sin alejarse de la invención. Debe entenderse que pueden emplearse diversas alternativas a las realizaciones de la invención descrita en la presente memoria en la práctica de la invención. Se pretende que las siguientes reivindicaciones definan el alcance de la invención y que los métodos y estructuras dentro del alcance de estas reivindicaciones y sus equivalentes estén cubiertos por la misma.

Ejemplos

15 **Ejemplo 1: Síntesis de división-combinación enzimática**

En este ejemplo, se usó un método de síntesis de división-combinación enzimática para producir microesferas acopladas al oligonucleótido. Como se muestra en la **FIG. 2A**, se añadió un conjunto de oligonucleótidos a cada pocillo de una primera placa. Un oligonucleótido en un conjunto de oligonucleótidos comprende una 5' amina, secuencia universal, marcador celular y un conector. La 5' amina, secuencia universal y el conector son iguales para cada conjunto de oligonucleótidos. La secuencia universal y el conector son diferentes entre sí. Sin embargo, el marcador celular es diferente para cada conjunto de oligonucleótidos. Por tanto, cada pocillo tiene un marcador celular diferente. En la etapa 1 de la síntesis de división-combinación enzimática, se sintetizaron microesferas acopladas a oligonucleótido añadiendo una única microesfera a cada pocillo y realizando reacciones de acoplamiento de 1-Etil-3-(3-dimetilaminopropil)carbodiimida (EDC). Las microesferas con oligonucleótido resultantes de la etapa 1 comprenden una microesfera acoplada a múltiples oligonucleótidos. El oligonucleótido comprende una 5' amina, secuencia universal, marcador celular 1 y conector 1 (véase la **FIG. 2A**). Los oligonucleótidos en la misma microesfera son iguales. Sin embargo, los oligonucleótidos en una primera microesfera son diferentes de los oligonucleótidos en una segunda microesfera.

En la etapa 2 de la síntesis de división-combinación enzimática, se realizaron múltiples lavados para eliminar los oligonucleótidos no acoplados. Una vez se retiraron los oligonucleótidos no acoplados, se combinaron las microesferas acopladas a oligonucleótido (véase la **FIG. 2A**). Las microesferas acopladas a oligonucleótido resultantes de la etapa 2 comprenden una microesfera acoplada a múltiples oligonucleótidos monocatenarios. El oligonucleótido monocatenario comprende una 5' amina, secuencia universal, marcador celular 1 y conector 1. Cada oligonucleótido en una microesfera es idéntico. Sin embargo, cada microesfera comprende un oligonucleótido diferente. Los oligonucleótidos acoplados a las diferentes microesferas difieren en la secuencia de marcador celular 1.

Como se muestra en la **FIG. 2B**, se añadió un conjunto de oligonucleótidos a cada pocillo de una segunda placa. Un oligonucleótido en un conjunto de oligonucleótidos comprende un primer conector, marcador celular y un segundo conector. El primer y segundo conector son iguales para cada conjunto de oligonucleótidos. El primer y segundo conector son diferentes entre sí. Sin embargo, el marcador celular es diferente para cada conjunto de oligonucleótidos. Por tanto, cada pocillo tiene un marcador celular diferente.

En la etapa 3 de la síntesis de división-combinación enzimática, las microesferas acopladas a oligonucleótido que se combinaron en la etapa 2 se dividieron en los pocillos de la segunda placa. Como el primer conector de los oligonucleótidos en los pocillos de la segunda placa son complementarios al conector de los oligonucleótidos acoplados a las microesferas, se realizó extensión de cebador usando el fragmento grande Klenow para acoplar los oligonucleótidos de la segunda placa con las microesferas acopladas a oligonucleótido de la etapa 2. Las microesferas acopladas a oligonucleótido resultantes de la etapa 3 comprenden una microesfera acoplada a múltiples oligonucleótidos bicatenarios. El oligonucleótido bicatenario comprende una 5' amina, secuencia universal, marcador celular 1, conector 1, marcador celular 2 y conector 2 (véase la **FIG. 2B**).

En la etapa 4 de la síntesis de división-combinación enzimática, se realizaron múltiples lavados para retirar los oligonucleótidos no acoplados y las enzimas de fragmento grande Klenow. La segunda placa se calentó para desnaturalizar los oligonucleótidos bicatenarios y se combinaron las microesferas acopladas a oligonucleótido (véase la **FIG. 2B**). Las microesferas acopladas a oligonucleótido resultante de la etapa 4, comprenden una microesfera acoplada a múltiples oligonucleótidos monocatenarios. El oligonucleótido monocatenario comprende una 5' amina, secuencia universal, marcador celular 1, conector 1, marcador celular 2 y conector 2. Cada oligonucleótido en una microesfera es idéntico. Sin embargo, cada microesfera comprende un oligonucleótido diferente. Los oligonucleótidos acoplados a las diferentes microesferas difieren por las secuencias de marcador celular combinadas. Por ejemplo, una primera microesfera puede comprender oligonucleótidos que comprenden un primer marcador celular de marcador celular A y un segundo marcador celular de marcador celular C y una segunda microesfera puede comprender oligonucleótidos que comprenden un primer marcador celular de marcador celular C y un segundo marcador celular de marcador celular D. Por tanto, la primera microesfera y la segunda microesfera pueden comprender el mismo marcador celular (en este caso, marcador celular C), sin embargo, las secuencias de marcador celular combinadas de la primera esfera y la segunda microesfera son diferentes (por ejemplo, para la

primera microesfera, la secuencia de marcador celular combinada es marcador celular A + marcador celular C; para la segunda microesfera, la secuencia de marcador celular combinada es marcador celular C + marcador celular A). En otros casos, dos microesferas pueden comprender oligonucleótidos que comprenden diferentes marcadores celulares. Por ejemplo, una primera microesfera puede comprender oligonucleótidos que comprenden marcador celular A y marcador celular B y una segunda microesfera puede comprender oligonucleótidos que comprenden marcador celular C y marcador celular D. En este caso, ambos marcadores celulares de la primera microesfera son diferentes de ambos marcadores celulares de la segunda microesfera.

Como se muestra en la **FIG. 2C**, se añade un conjunto de oligonucleótidos a cada pocillo de una tercera placa. Un oligonucleótido en un conjunto de oligonucleótidos comprende un conector, marcador celular, marcador molecular y un oligodT. El conector y las secuencias de oligodT son iguales para cada conjunto de oligonucleótidos. Sin embargo, el marcador celular es diferente para cada conjunto de oligonucleótidos. Por tanto, cada pocillo tiene un marcador celular diferente. Además, el marcador molecular es diferente para oligonucleótidos dentro de un conjunto. Por tanto, un único pocillo contiene una pluralidad de oligonucleótidos con el mismo marcador celular, pero diferentes marcadores moleculares. Los oligonucleótidos de diferentes pocillos pueden contener el mismo marcador molecular.

En la etapa 5 de la síntesis de división-combinación enzimática, las microesferas acopladas a oligonucleótido que se combinaron en la etapa 4 se dividieron en los pocillos de la tercera placa. Como el conector de los oligonucleótidos en los pocillos de la segunda placa son complementarios al segundo conector de los oligonucleótidos acoplados a las microesferas, se realizó extensión de cebador usando el fragmento grande Klenow para acoplar los oligonucleótidos de la tercera placa con las microesferas acopladas a oligonucleótido de la etapa 4. Las microesferas acopladas a oligonucleótido resultantes de la etapa 5 comprenden una microesfera acoplada a múltiples oligonucleótidos bicatenarios. El oligonucleótido bicatenario comprende una 5' amina, secuencia universal, marcador celular 1, conector 1, marcador celular 2, conector 2, marcador celular 3, marcador molecular y oligodT (véase la **FIG. 2C**).

En la etapa 6 de la síntesis de división-combinación enzimática, se realizaron múltiples lavados para retirar los oligonucleótidos no acoplados y las enzimas de fragmento grande Klenow. La tercera placa se calentó para desnaturalizar los oligonucleótidos bicatenarios y las microesferas acopladas a oligonucleótido se combinaron (véase la **FIG. 2C**). Las microesferas acopladas a oligonucleótido resultantes de la etapa 4 comprenden una microesfera acoplada a múltiples oligonucleótidos monocatenarios. El oligonucleótido monocatenario comprende una 5' amina, secuencia universal, marcador celular 1, conector 1, marcador celular 2, conector 2, marcador celular 3, marcador molecular y oligodT. Los múltiples oligonucleótidos monocatenarios en una única microesfera pueden diferenciarse por el marcador molecular. Las partes de marcador celular de los múltiples oligonucleótidos en una única microesfera son idénticas. Cada microesfera comprende diferentes oligonucleótidos. Los oligonucleótidos acoplados a las diferentes microesferas difieren en las secuencias de marcador celular. El marcador molecular en los oligonucleótidos de diferentes microesferas puede ser igual. El marcador molecular en los oligonucleótidos de diferentes microesferas puede ser diferente. Dos o más microesferas pueden diferir en las secuencias de marcador celular combinadas. Por ejemplo, una primera microesfera puede comprender un oligonucleótido que comprende marcador celular A, marcador celular B y marcador celular C y una segunda microesfera puede comprender un oligonucleótido que comprende marcador celular B, marcador celular D y marcador celular A. En este caso, la primera y segunda microesfera contienen ambas marcadores celulares B, sin embargo, los otros dos marcadores celulares son diferentes. Por tanto, dos o más microesferas pueden comprender oligonucleótidos que difieren en al menos un marcador celular. Dos o más microesferas pueden comprender oligonucleótidos que difieren en al menos dos marcadores celulares. Dos o más microesferas pueden comprender oligonucleótidos que difieren en al menos tres marcadores celulares. Sin embargo, una microesfera puede comprender un oligonucleótido que comprende dos o más marcadores celulares idénticos. Por ejemplo, una microesfera puede comprender un oligonucleótido que comprende marcador celular A, marcador celular A y marcador celular D. Una microesfera puede comprender oligonucleótidos que comprenden al menos tres marcadores celulares idénticos. Por ejemplo, una microesfera puede comprender un oligonucleótido que comprende marcador celular A, marcador celular A y marcador celular A. Una microesfera puede comprender oligonucleótidos que comprenden tres marcadores celulares no idénticos. Por ejemplo, una microesfera puede comprender un oligonucleótido que comprende marcador celular A, marcador celular D y marcador celular E. Una microesfera puede comprender al menos dos oligonucleótidos que comprenden al menos dos marcadores moleculares diferentes. Por ejemplo, una microesfera puede comprender un primer oligonucleótido que comprende marcador molecular A y un segundo oligonucleótido que comprende marcador molecular D. Sin embargo, una microesfera puede comprender múltiples copias de un oligonucleótido que comprende un primer marcador molecular. Por tanto, una microesfera puede comprender al menos dos oligonucleótidos que comprenden el mismo marcador molecular. Por ejemplo, una microesfera puede comprender un primer oligonucleótido que comprende marcador molecular A y un segundo oligonucleótido que comprende marcador molecular A. Al menos un 30 % de los oligonucleótidos en una microesfera pueden comprender diferentes marcadores moleculares. Al menos un 40 % de los oligonucleótidos en una microesfera pueden comprender diferentes marcadores moleculares. Al menos un 50 % de los oligonucleótidos en una microesfera pueden comprender diferentes marcadores moleculares. Al menos un 60 % de los oligonucleótidos en una microesfera pueden comprender diferentes marcadores moleculares. Menos de un 30 % de los oligonucleótidos en una microesfera pueden comprender el mismo marcador molecular. Menos de un 20 % de los oligonucleótidos en una microesfera pueden comprender el mismo marcador molecular. Menos de un 15 % de los oligonucleótidos en una microesfera pueden comprender el mismo marcador molecular. Menos de un 10 % de los oligonucleótidos en una microesfera pueden comprender el mismo marcador molecular. Menos de un 5 % de los oligonucleótidos en una microesfera pueden comprender el mismo marcador molecular.

La técnica de síntesis de división-combinación enzimática puede realizarse en múltiples placas o placas con un mayor número de pocillos para producir una cantidad mayor de microesferas acopladas a oligonucleótido. El uso de tres partes de marcador celular diferentes puede aumentar la diversidad de las partes de marcador celular totales en las microesferas. Con 96 opciones de secuencia diferentes para cada parte de marcador celular, pueden crearse 884 736 diferentes combinaciones de marcador celular.

Ejemplo 2: Comparación de amplificación en tubo y micropocillo

La divulgación proporciona un método para capturar células. Aproximadamente 5000 células Ramos se capturaron en una matriz de micropocillos que comprende micropocillos de aproximadamente 30 micrómetros de diámetro. Algunas células no se capturaron. El control para el experimento fue un número equivalente de células capturadas en un tubo. Tanto las células en el tubo como las células en la matriz de micropocillos se lisaron. Se permitió que el ácido nucleico hibridara con una microesfera conjugada. Se realizó PCR en tiempo real de los genes de GAPDH y RPL19.

La FIG. 9 muestra los resultados de la amplificación por PCR en tiempo real. El rendimiento del micropocillo fue mayor que el rendimiento del ácido nucleico en el tubo, lo que indica que la hibridación del ácido nucleico con el oligonucleótido fue más eficaz en el micropocillo que en el tubo (compárese la barra gris y la barra blanca, respectivamente).

Ejemplo 3: Comparación de amplificación de la segunda hebra sintetizada y la síntesis en la microesfera

Se obtuvieron células y se lisaron como se describe en el ejemplo 1. Se amplificaron RPL19, TUBB y GAPDH de la segunda hebra sintetizada de los soportes sólidos, o directamente en los soportes sólidos usando un cebador universal. La FIG. 10 muestra que la amplificación directamente en los soportes sólidos (FIG. 10) producía menos amplificación inespecífica que la amplificación no directamente de un soporte sólido. Las amplificaciones de GAPDH y TUBB produjeron productos del tamaño correcto independientemente del método (el carril de la izquierda de cada triplete en la FIG. 10 corresponde a soporte sólido más lisado en formato de tubo, el carril central de cada triplete corresponde a soportes sólidos del micropocillo y el carril de la derecha de cada triplete corresponde a soportes sólidos más ácido nucleico purificado. El producto RPL19 tenía productos de amplificación inespecíficos mínimos, pero únicamente producía un producto fuerte cuando se usaba ácido nucleico purificado con el soporte sólido. Estos experimentos indican que la amplificación directamente sobre las microesferas produce menos productos de amplificación inespecíficos que la amplificación usando una segunda hebra sintetizada del soporte sólido.

Ejemplo 4: Análisis combinado de ácidos nucleicos diana

Se obtuvieron células y se lisaron como se describe en el ejemplo 1. Los ácidos nucleicos diana se hibridan al soporte sólido que comprende oligonucleótidos. Se hibridan una pluralidad de copias del ácido nucleico diana a una región de unión a diana que comprende una secuencia de oligodT. La pluralidad de copias del ácido nucleico diana se transcriben de forma inversa usando transcriptasa inversa. La transcripción inversa incorpora las características del oligonucleótido al que se hibridó la copia del ácido nucleico diana (por ejemplo, el marcador molecular, el marcador celular y el marcador universal). La pluralidad de copias del ácido nucleico diana se amplifica usando PCR. Las copias amplificadas del ácido nucleico diana se secuencian. Los ácidos nucleicos diana secuenciados se cuentan para determinar el número de copias del ácido nucleico diana en la célula. El recuento se realiza contando el número de diferentes marcadores moleculares para cada una de las mismas lecturas de secuencia de ácido nucleico diana. De esta manera, puede disminuirse el sesgo de amplificación.

Ejemplo 5: Evaluación de la eficacia de la síntesis de división-combinación para producir microesferas con copias clonales de una combinación de marcador celular

En este ejemplo, se evaluó la eficacia de la síntesis de división-combinación para producir microesferas con copias clonales de una combinación de marcador celular. Se sintetizaron microesferas acopladas a oligonucleótido mediante el método de síntesis de división-combinación enzimática como se describe en el ejemplo 1. Se purificaron 250 ng de ARN total de células Ramos, que es equivalente al ARN de 25 000 células. El ARN total se puso en contacto con 35 000 microesferas acopladas a oligonucleótido, que provoca hibridación del ARNm a las microesferas acopladas a oligonucleótido. Se realizó síntesis de ADNc en el ARNm hibridado a las microesferas acopladas a oligonucleótido. Se usaron muestras que comprenden 18, 175 y 1750 microesferas para análisis adicional. Se realizaron reacciones de amplificación por PCR usando cebadores específicos de GAPDH y cebadores específicos de IGJ sobre el ADNc unido a las microesferas de las muestras de 18, 175 y 1750 microesferas. Las moléculas de ADNc adheridas a las microesferas se secuenciaron. La FIG. 11A-I muestra representaciones gráficas de los resultados de secuenciación. Para la FIG. 11A-C, se representa el número de lecturas por microesfera en el eje de ordenadas y el código de barras único (por ejemplo, combinación de marcador celular) se representa en el eje de abscisas para las muestras de 18 perlas, 175 perlas y 1750 perlas, respectivamente. Para la FIG. 11D-F, se representa el número de moléculas únicas por microesfera en el eje de ordenadas y el código de barras único (por ejemplo, combinación de marcador celular) se representa en el eje de abscisas para las muestras de 18 perlas, 175 perlas y 1750 perlas, respectivamente. Para la FIG. 11G-I, se representa el número de moléculas únicas por microesfera en el eje de ordenadas y el código de barras único se representa en el eje de abscisas para las muestras de 18 perlas, 175 perlas y 1750 perlas, respectivamente. Los resultados para la FIG. 11G-I se clasifican mediante el número total de moléculas. La mediana del número de moléculas únicas por microesfera para las diversas muestras se muestra en la tabla 1. Los valores numéricos para los resultados de secuenciación se muestran en la tabla 2. Para la FIG. 11J-L, el número de combinación de código de barras único (bc) usando el índice se representa en el eje de ordenadas y el índice de segmento de código de barras (bc) se representa en el eje

ES 2 711 168 T3

5 de abscisas para el marcador celular 1, marcador celular 2 y marcador celular 3 para la muestra de 1750 microesferas, respectivamente. El código de barras (bc) se refiere al marcador celular (por ejemplo, bc segmento 1 = parte 1 de marcador celular). Como se muestra en la FIG. 11J-L, la presencia de casi todos los 96 códigos de barras dentro de cada segmento se detectó por secuenciación. Estos resultados demuestran el éxito del método de síntesis de división-combinación enzimática para producir microesferas con copias clonales de una combinación de marcador celular.

Tabla 1

	Mediana del número de moléculas únicas		
	18 microesferas	175 microesferas	1750 microesferas
IGJ	78	85	40
GAPDH	22	45	25

Tabla 2

N.º esperado de microesferas	17,5	175	1750
Número total de lecturas	58 321	60 308	133 043
> = 8 coincidencia en constante 1	56 385	57 615	123 349
> = 8 coincidencia en constante 2	54 117	55 187	115 126
> = 8 coincidencia en constante 1 y 2	54 114	55 185	115 107
Coincidencia perfecta en los 3 subcódigos de barras	38 585	46 066	95 217
Coincidencia perfecta en el gen (40 pb)	29 968	33 775	72 260
Número total de combinación de código de barras único	239	407	1654
% de lecturas útiles	51,388 %	56,00 %	54,31 %
Número de combinaciones de códigos de barras únicos > 20 lecturas	5	26	288

Ejemplo 6. Marcaje de ARN de célula individual usando microesferas acopladas a oligonucleótido

10 En este ejemplo, se evaluó la eficacia del marcaje de ARN de célula individual usando microesferas acopladas a oligonucleótido. Se prepararon tres muestras celulares del siguiente modo:

	Muestra 1: K562 únicamente	Muestra 2: Ramos únicamente	Muestra 3: mezcla Ramos + K562
N.º de micropocillos	~10 000	~10 000	~10 000
N.º de células Ramos	0	5000	3750
N.º de células K562	1000	0	2500

15 La suspensión celular de las muestras se añadió a la parte superior de un micropocillo y se permitió que las células sedimentaran en los pocillos de la matriz de micropocillos. Las células no capturadas por la matriz de micropocillos se retiraron por lavado en un baño de disolución salina tamponada con fosfato (PBS). Las microesferas acopladas a oligonucleótido, preparadas por el método de síntesis de división-combinación enzimática descrito en el ejemplo 1 se añadieron a la matriz de micropocillos. La microesfera acoplada a oligonucleótido comprende una microesfera magnética con una pluralidad de oligonucleótidos. Cada oligonucleótido en la microesfera comprende una 5' amina, secuencia universal, marcador celular 1, conector 1, marcador celular 2, conector 2, marcador celular 3, marcador molecular y oligodT. Para cada oligonucleótido en la misma microesfera, las secuencias de los oligonucleótidos son idénticas excepto por el marcador molecular. Para oligonucleótidos en diferentes microesferas, las combinaciones de marcador celular 1, 2 y 3 son diferentes. Se añadieron aproximadamente 5-6 microesferas por pocillo de la matriz de micropocillos. En algunos casos, por cada 10 pocillos, pueden depositarse 50 microesferas en la matriz, cayendo 0-2 microesferas en cada pocillo. Se permitió que las microesferas sedimentaran en los pocillos y las microesferas no capturadas se retiraron por lavado en un baño de PBS. Se colocó un imán por debajo de la matriz de micropocillos. Las células se lisaron mediante la adición de tampón de lisis frío. La matriz y el imán se colocaron en un bloque de aluminio frío durante 5 minutos. El ARNm de las células lisadas se hibridó con los oligonucleótidos acoplados a las microesferas. La matriz se lavó con exceso de tampón de lisis para retirar el ARNm no unido. Las microesferas se recuperaron de los pocillos colocando un imán en la parte superior de la matriz de micropocillos. Las

ES 2 711 168 T3

microesferas recuperadas se lavaron. Se realizó la síntesis de ADNc en las microesferas usando Superscript III a 50 °C durante 50 minutos en un rotor. Se retiró el oligodT no prolongado de los oligonucleótidos en las microesferas por tratamiento con Exol realizado a 37 °C durante 30 minutos en un rotor. Se realizó amplificación por PCR específica de gen en el ADNc. Los genes seleccionados para la PCR específica de gen eran específicos de tipo celular y se muestran en la tabla 3. Los productos amplificados por PCR se secuenciaron. La estadística de secuenciación se muestra en la tabla 4. Las figuras 12A-C muestran un histograma de los resultados de secuenciación para la muestra de K562 únicamente, muestra de Ramos únicamente y muestra de mezcla de K562 + Ramos, respectivamente. Para la FIG. 12A-C, la molécula única por código de barras se representó en el eje de ordenadas y el índice de combinación de código de barras único, clasificado por lectura por código de barras, se representó en el eje de abscisas.

Tabla 3

Número	Gen	Tipo celular
1	CD74	Específico de Ramos
2	CD79a	Específico de Ramos
3	IGJ	Específico de Ramos
4	TCL1A	Específico de Ramos
5	SEPT9	Específico de Ramos
6	CD27	Específico de Ramos
7	CD41	Específico de K562
8	GYPA	Específico de K562
9	GATA1	Específico de K562
10	GATA2	Específico de K562
11	HBG1	Específico de K562
12	GAPDH	Común

Tabla 4

	Muestra 1: K562 únicamente	Muestra 2: Ramos únicamente	Muestra 3: mezcla Ramos + K562
Número de células Ramos	0	5000	3750
Número de células K562	1000	0	2500
Número total de lecturas	717 718	1 329 189	2 399 025
> = 8 coincidencia en constante 1	657 911	1 201 081	2 026 726
> = 8 coincidencia en constante 2	581 581	1 071 364	1 513 466
> = 8 coincidencia en constante 1 y 2	581 508	1 071 153	1 513 102
Coincidencia perfecta en los 3 subcódigos de barras	481 564	862 348	1 248 073
Coincidencia perfecta en el gen (40 pb)	283 463	575 713	1 004 338
% de lecturas útiles	39,50 %	43,31 %	41,86 %
Número total de combinación de código de barras único	8501	29 647	28 783
Número de combinaciones de códigos de barras únicos > 30 moléculas	145	1072	768
Eficacia de captura	0,145	0,2144	0,12288

ES 2 711 168 T3

Se usó el marcaje de célula individual para determinar el número de copias para las muestras de tipo de célula individual (por ejemplo, muestra de K562 únicamente, muestra de Ramos únicamente). La FIG. 12D-E muestra un gráfico del número de copias para genes enumerados en la tabla 3 para la muestra de células Ramos únicamente y la muestra de células K562 únicamente, respectivamente. Para la FIG. 12D-E, el número de moléculas por combinación de código de barras (bc) se representa en el eje de ordenadas y la combinación de código de barras única, clasificada por el número total de moléculas por combinación de bc se representa en el eje de abscisas. Los resultados mostrados en las figuras 12D-E se basaban en los datos de secuenciación de microesferas con > 30 de número total de moléculas únicas. Estos resultados demuestran que la proporción de moléculas por amplicón por microesfera cumple las expectativas para el tipo celular. Para la muestra de células únicamente, el desajuste del número de moléculas es más importante y parece que HBG1, que es muy abundante en este tipo celular, tiene un número de copias variables. Sin embargo, el número de copias de GAPDH parece ser constante aunque el número total de moléculas por microesfera se desajusta. El número de copias para los genes individuales se muestra en la FIG. 12F-I. Para la FIG. 12F-G, el número de copias se representa como copia por microesfera o célula individual para células Ramos únicamente y células K562 únicamente, respectivamente. Para la FIG. 12H-I, el número de copias se representa como la abundancia relativa por microesfera o célula individual para células Ramos únicamente y células K562 únicamente, respectivamente.

Se usó el marcaje de célula individual para determinar el tipo de célula de células individuales en la muestra de mezcla de K562+Ramos. Los resultados de secuenciación de 100 combinaciones de código de barras únicas con las moléculas más abundantes se analizaron para evaluar la eficacia del marcaje de célula individual para determinar el tipo de célula de células individuales en la muestra de mezcla de K562+Ramos. La FIG. 12J-M muestra gráficos del número de moléculas únicas por gen (eje de ordenadas) para las microesferas con las 100 combinaciones de códigos de barras únicas. Los números en el eje de ordenadas se refieren al gen (véase la tabla 3). La FIG. 12J-M representa claramente los patrones de expresión génica generales para las células K562 y Ramos. La FIG. 12N-O muestra gráficos ampliados de dos microesferas que representan el patrón general de perfiles de expresión génica para los dos tipos celulares. La FIG. 12N muestra el patrón general de perfil de expresión génica para células de tipo K562 y la FIG. 12O muestra el patrón general de perfil de expresión génica para células de tipo Ramos. La FIG. 12P muestra un diagrama de dispersión de los resultados basados en el análisis de componentes principales del perfil de expresión génica de 768 microesferas con > 30 moléculas por microesfera de la muestra de mezcla de K562+Ramos. El componente 1, que se representa en el eje de abscisas, separa los dos tipos celulares. El componente 2, que se representa en el eje de ordenada, separa células K562 con un número de copias de HBG1 alto y bajo. Cada punto en el diagrama de dispersión representa una combinación de código de barras única, que es equivalente a una microesfera o una célula. Basándose en el análisis de componentes principales, 409 microesferas correspondían a células K562 y 347 microesferas correspondían a células Ramos. El número de copias de los genes de la tabla 3 se determinó para los tipos de células de tipo K562 y de tipo Ramos. La FIG. 12Q-R muestra histogramas del número de copias por amplicón por microesfera para las células de tipo K562 (microesferas a la izquierda del componente principal basado en la FIG. 12P) y células de tipo Ramos (microesferas a la derecha del primer componente principal basado en la FIG. 12P), respectivamente. Para la FIG. 12Q-R, el número por combinación de bc está en el eje de ordenadas y la combinación de bc única está, clasificada por el número total de moléculas por combinación de bc está en el eje de abscisas. La FIG. 12S-T muestra el número de copias por microesfera o célula individual de los genes individuales para las células de tipo K562 (microesferas a la izquierda del primer componente principal basado en la FIG. 12P) y células de tipo Ramos (microesferas a la derecha del primer componente principal basado en la FIG. 12P), respectivamente. La tabla 5 muestra la media del número de copias por microesfera para muestras de células individuales y de mezcla.

Tabla 5

Gen	Muestra de tipo celular individual		Muestra de mezcla de K562+Ramos	
	K562 únicamente	Ramos únicamente	Tipo K562	Tipo Ramos
CD74	0,00	39,95	0,10	7,50
CD79a	0,02	30,97	0,84	18,88
IGJ	0,03	42,43	0,81	27,76
TCL1A	0,01	31,78	0,71	19,44
SEPT9	0,88	3,89	1,35	1,52
CD27	0,00	5,31	0,03	1,30
CD41	0,61	0,00	0,47	0,01
GYPA	1,92	0,00	0,73	0,02
GATA2	1,38	0,00	0,60	0,04
GATA1	0,94	0,00	1,04	0,04

ES 2 711 168 T3

HBG1	201,09	0,00	72,27	1,37
GAPDH	51,77	39,13	44,94	13,53
Redundancia de lectura de GAPDH	2,04	1,47	7,67	7,22

Ejemplo 7. Evaluación de interferencia entre microesferas

En este ejemplo, se evaluó la interferencia entre microesferas. Se prepararon muestras que comprenden mezclas de células EL4 de ratón y células Ramos de la siguiente manera:

	Alta densidad	Baja densidad
Número de micropocillos	~10 000	~10 000
Número de células EL4 de ratón	2500	1500
Número de células Ramos	3750	1500

5

La suspensión celular de las muestras se añadió a la parte superior de un micropocillo y se permitió que las células sedimentaran en los pocillos de la matriz de micropocillos. Las células no capturadas por la matriz de micropocillos se retiraron por lavado en un baño de disolución salina tamponada con fosfato (PBS). Las microesferas acopladas a oligonucleótido, preparadas por el método de síntesis de división-combinación enzimática descrito en el ejemplo 1 se añadieron a la matriz de micropocillos. La microesfera acoplada a oligonucleótido comprende una microesfera magnética con una pluralidad de oligonucleótidos. Cada oligonucleótido en la microesfera comprende 5' amina, secuencia universal, marcador celular 1, conector 1, marcador celular 2, conector 2, marcador celular 3, marcador molecular y oligodT. Para cada oligonucleótido en la misma microesfera, las secuencias de los oligonucleótidos son idénticas excepto por el marcador molecular. Para oligonucleótidos en diferentes microesferas, las combinaciones de marcador celular 1, 2 y 3 son diferentes. Se añadieron aproximadamente 5-6 microesferas por pocillo de la matriz de micropocillos. Se permitió que las microesferas sedimentaran en los pocillos y las microesferas no capturadas se retiraron por lavado en un baño de PBS. Se colocó un imán por debajo de la matriz de micropocillos. Las células se lisaron mediante la adición de tampón de lisis frío. La matriz y el imán se colocaron en un bloque de aluminio frío durante 5 minutos. El ARNm de las células lisadas se hibridó con los oligonucleótidos acoplados a las microesferas. La matriz se lavó con exceso de tampón de lisis para retirar el ARNm no unido. Las microesferas se recuperaron de los pocillos colocando un imán en la parte superior de la matriz de micropocillos. Las microesferas recuperadas se lavaron. Se realizó la síntesis de ADNc en las microesferas usando Superscript III a 50 °C durante 50 minutos en un rotor. Se retiró el oligodT no prolongado de los oligonucleótidos en las microesferas por tratamiento con Exol realizado a 37 °C durante 30 minutos en un rotor. Se realizó amplificación por PCR específica de gen en el ADNc. Los genes seleccionados para la PCR específica de gen eran específicos de tipo celular y se muestran en la tabla 6.

10

15

20

25

Tabla 6

Número	Gen	Tipo celular
1	HS_CD74	Humano
2	HS_CD79a	Humano
3	HS_IGJ	Humano
4	HS_TCL1A	Humano
5	HS_SEPT9	Humano
6	HS_CD27	Humano
7	HS_CD27	Humano
8	MM_B2M	Ratón
9	MM_ACTM	Ratón
10	MM_HPRT	Ratón
11	MM_SHDA	Ratón

Los productos amplificados por PCR se secuenciaron. La estadística de secuenciación se muestra en la tabla 7.

Tabla 7

	Baja densidad	Alta densidad
Número de células Ramos	15 000	3750
Número de células de ratón	1500	2500

ES 2 711 168 T3

Tabla 7

Número total de lecturas	2 391 780	4 038 217
> = 8 coincidencia en constante 1	2 162 945	3 651 643
> = 8 coincidencia en constante 2	1 981 835	3 356 493
> = 8 coincidencia en constante 1 y 2	1 981 626	3 355 787
Coincidencia perfecta en los 3 subcódigos de barras	1 645 994	2 790 879
Coincidencia perfecta en el gen (40 pb)	1 083 013	2 171 930
% de lecturas útiles	45 %	54 %
Número total de combinación de código de barras único	16 695	36 595
Número de combinaciones de códigos de barras únicos > 30 moléculas	80	281
Eficacia de captura	0,03	0,04

Se determinaron los perfiles de expresión génica para 100 combinaciones de código de barras único con las moléculas más abundantes para las muestras de alta densidad y baja densidad. Los perfiles de expresión génica se generaron basándose en los resultados de secuenciación. La FIG. 13A muestra gráficos del perfil de expresión génica para 35 de las 100 combinaciones de código de barras único de la muestra de alta densidad. Para la FIG. 13A, el número de moléculas única está en el eje de ordenadas y el número de referencia de gen está en el eje de abscisas (véase la tabla 6 para genes correspondientes al número de referencia de gen). La FIG. 13A representa claramente los patrones de expresión génica generales para las células de ratón y Ramos. La FIG. 13B-C muestra diagramas de dispersión de resultados basados en el análisis de componentes principales del perfil de expresión génica de la muestra de alta densidad y la muestra de baja densidad, respectivamente. El componente 1, que se representa en el eje de abscisas, separa los dos tipos celulares. El componente 2, que se representa en el eje de ordenadas, indica la variabilidad en la expresión génica dentro de la población de células Ramos. Cada punto en el diagrama de dispersión representa una combinación de código de barras único, que es equivalente a una microesfera o una célula. Basándose en el análisis de componentes principales de la muestra de alta densidad, 144 microesferas correspondían a las células de ratón y 132 microesferas correspondían a células Ramos. Basándose en el análisis de componentes principales de la muestra de baja densidad, 52 microesferas correspondían a las células de ratón y 27 microesferas correspondían a células Ramos.

Una vez determinados los tipos celulares, se evaluó la interferencia entre las microesferas detectando los genes de la tabla 6 en los diferentes tipos celulares. La FIG. 13D-E representa gráficos de la lectura por combinación de código de barras (bc) (eje de abscisas) frente a la combinación de código de barras único, clasificado por el número total de moléculas por combinación de código de barras (eje de abscisas) para células de tipo Ramos y células de tipo ratón de la muestra de alta densidad, respectivamente. La FIG. 13F-G representa gráficos del número de moléculas por combinación de código de barras (eje de abscisas) frente a la combinación de código de barras único, clasificado por el número total de moléculas por combinación de código de barras (eje de ordenadas) para células de tipo Ramos y células de tipo ratón de la muestra de alta densidad, respectivamente. La FIG. 13H-I representa gráficos de la lectura por combinación de código de barras (bc) (eje de ordenadas) frente a la combinación de código de barras único, clasificado por el número total de moléculas por combinación de código de barras (eje de abscisas) para células de tipo Ramos y células de tipo ratón de la muestra de baja densidad, respectivamente. La FIG. 13J-K representa gráficos del número de moléculas por combinación de código de barras (eje de abscisas) frente a la combinación de código de barras único, clasificado por el número total de moléculas por combinación de código de barras (eje de ordenadas) para células de tipo Ramos y células de tipo ratón de la muestra de baja densidad, respectivamente. La tabla 8 muestra el promedio factorial de cobertura o redundancia de lecturas por molécula única para las muestras de baja y alta densidad.

Tabla 8

Gen	Baja densidad		Alta densidad	
	Células tipo Ramos	Células tipo ratón	Células tipo Ramos	Células tipo ratón
HS_CD74	29,75	3,17	23,75	2,15
HS_CD79a	47,2	4,09	42,30	2,67
HS_IGJ	29,65	1,39	30,23	2,4
HS_TCL1A	45,74	2,26	39,00	4,13

ES 2 711 168 T3

Tabla 8

HS_SEPT9	11,85	1,00	12,75	1,18
HS_CD27	37,99	1,00	32,12	1,10
HS_GAPDH	19,97	1,55	17,37	2,57
MM_B2M	1,21	31,98	3,05	31,48
MM_ACTM	1,05	29,08	1,90	28,38
MM_HPRT	1,02	39,96	1,03	43,65
MM_SHDA	1,00	39,60	1,02	29,60

Los resultados en la tabla 8 muestran que el promedio factorial de cobertura por molécula única era mucho mayor para genes humanos que para genes de ratón en células Ramos y viceversa.

5 Como control, se lisó una mezcla de células de ratón y humanas en un tubo, se convirtió para la síntesis de ADNc con las microesferas y se secuenció el ADNc. La FIG. 4XL muestra una representación gráfica de los resultados de secuenciación. Como se esperaba, se observó una gran cantidad de combinaciones de códigos de barras (bc) únicos, y la mayoría de las microesferas tenían únicamente de una a dos copias en total.

Estos resultados demuestran que hubo interferencia mínima entre las microesferas y que la interferencia puede identificarse de forma bioinformática.

10 Ejemplo 8. Producción de colección de ácidos nucleicos de células individuales

Los soportes conjugados a oligonucleótido divulgados en la presente memoria pueden usarse para producir colecciones de ácidos nucleicos de células individuales. En este ejemplo, se producen colecciones de ácidos nucleicos de células individuales añadiendo una muestra celular a una superficie (por ejemplo, rejilla) que tiene los soportes conjugados a oligonucleótido. Un soporte conjugado a oligonucleótido comprende una pluralidad de oligonucleótidos conjugados a una microesfera. Un oligonucleótido comprende (a) una región de marcador celular que comprende al menos dos regiones distintas conectadas por un conector; y (b) una región de marcador molecular. Dos o más oligonucleótidos en una microesfera comprenden regiones de marcador celular idénticas. Dos o más oligonucleótidos en una microesfera comprenden dos o más regiones de marcador molecular diferentes. Dos o más oligonucleótidos en dos o más microesferas diferentes comprenden dos o más regiones de marcador celular diferentes. Por tanto, cada célula asociada con un soporte conjugado a oligonucleótido tiene una región de marcador celular diferente. La concentración de células en la muestra de células está suficientemente diluida para posibilitar la asociación de una o menos células a soporte conjugado a oligonucleótido en la superficie. Las células se lisan usando un tampón de lisis. Los ARNm de una célula se hibridan con los oligonucleótidos del soporte conjugado a oligonucleótido. Por tanto, todos los ARNm de una célula se marcan con oligonucleótidos que comprenden regiones de marcador celular idénticas. Dos o más ARNm de una célula se marcan con dos o más oligonucleótidos que comprenden dos o más regiones de marcador molecular diferentes. Se aplica un imán a la superficie para purificar los soportes sólidos conjugados a oligonucleótido de la superficie. Los soportes sólidos conjugados a oligonucleótido pueden purificarse individualmente de la superficie. Los ARNm hibridados con los oligonucleótidos en el soporte sólido conjugado a oligonucleótido se transcriben de forma inversa para producir ADNc marcado. El ADNc marcado comprende un complemento inverso del ARNm y una copia del oligonucleótido con el que hibridó el ARNm. El ADNc marcado se amplifica por PCR para producir amplicones marcados. El ADNc marcado y/o los amplicones marcados pueden retirarse de la microesfera por digestión con enzimas de restricción. Se produce una colección de ácidos nucleicos de la célula individual a partir de los amplicones marcados.

35 Como alternativa, los soportes sólidos conjugados a oligonucleótido se purifican juntos. Puede realizarse transcripción inversa del ARNm en los soportes sólidos conjugados a oligonucleótido combinados. Como los ARNm de diferentes células se marcan con oligonucleótidos que comprenden diferentes regiones de marcador celular, las regiones de marcador celular pueden usarse para determinar la célula de la que se originó el ADNc marcado o los amplicones marcados. Por tanto, puede producirse una colección de ácidos nucleicos a partir de una pluralidad de células, en la que la identidad de la célula de la que se originó el amplicón marcado puede determinarse mediante la región de marcador celular.

45 También pueden producirse colecciones de ácidos nucleicos de células individuales poniendo en contacto las células con un agente antes de la lisis de la célula. El agente puede ser un antígeno, fármaco, célula, toxina, etc. por tanto, pueden producirse colecciones de ácidos nucleicos de células individuales especializadas. Puede usarse análisis de las colecciones de ácidos nucleicos para generar perfiles de expresión de fármacos de células individuales. También pueden determinarse las rutas de transducción de señales a nivel de célula individual a partir de estas colecciones de ácidos nucleicos. Las colecciones de ácidos nucleicos también pueden usarse para determinar los efectos de antígenos sobre tipos celulares específicos.

Ejemplo 9. Perfilado de expresión de célula individual

Los soportes conjugados a oligonucleótido divulgados en la presente memoria pueden usarse para determinar el perfil de expresión de células individuales. En este ejemplo, una muestra de células que comprende una mezcla de células se pone en contacto con una pluralidad de anticuerpos. Un subconjunto de las células se purifica usando citometría de flujo. El subconjunto de células se añade a una matriz de micropocillos. Una pluralidad de soportes conjugados a oligonucleótido se añade a la matriz de micropocillos. Un soporte conjugado a oligonucleótido comprende una pluralidad de oligonucleótidos acoplados a una nanopartícula. Un oligonucleótido comprende (a) una región de marcador celular que comprende tres secuencias distintas conectadas a dos secuencias predeterminadas; y (b) una región de marcador molecular. Dos o más oligonucleótidos en una nanopartícula comprenden regiones de marcador celular idénticas. Dos o más oligonucleótidos en una nanopartícula comprenden dos o más regiones de marcador molecular diferentes. Dos o más oligonucleótidos en dos o más nanopartículas diferentes comprenden dos o más regiones de marcador celular diferentes. Por tanto, cada célula asociada con un soporte conjugado a oligonucleótido tiene una región de marcador celular diferente.

Se aplica un imán a la matriz de micropocillos y las células que no están asociadas con un soporte conjugado a oligonucleótido se retiran por lavado. Una esponja que comprende un tampón de lisis se coloca en la parte superior de la matriz de micropocillos, lisando de ese modo las células.

Los ARNm de las células lisadas hibridan con los oligonucleótidos en la microesfera. Los ARNm se transcriben de forma inversa para producir ADNc marcado. El ADNc marcado comprende un complemento inverso del ARNm y una copia del oligonucleótido con el que hibridó el ARNm. El ADNc marcado se amplifica por PCR para producir amplicones marcados. Los amplicones marcados se secuencian. Como cada ARNm de una célula se marca con el mismo marcador celular y los ARNm de diferentes células se marcan con diferentes marcadores celulares, la información de secuencia de los amplicones marcados se usa para generar perfiles de expresión de célula individual.

Ejemplo 10: Inmunofenotipado por secuenciación de célula individual

Se recogió una muestra de sangre de un sujeto y se aislaron las células mononucleares de sangre periférica (PBMC) de la muestra de sangre. Las PBMC se cultivaron en medio RPMI 1640 y se colocaron en una incubadora durante la noche. Las PBMC se lavaron múltiples veces en PBS para retirar el suero. Se depositaron aproximadamente 7000 PBMC en una matriz de micropocillos con 32 400 pocillos. Por tanto, la mayoría de los pocillos en la matriz de micropocillos no contenían células y algunos pocillos en la celda contenían únicamente 1 célula. Las microesferas conjugadas a oligonucleótido se aplicaron a la matriz de micropocillos. Cada microesfera conjugada a oligonucleótido contenía aproximadamente 1000 millones de oligonucleótidos adheridos a una microesfera. Cada oligonucleótido adherido a la microesfera contenía una 5' amina, secuencia universal, marcador celular de tres partes (por ejemplo, tres secciones de marcador celular conectadas por dos conectores), marcador molecular y oligodT. Cada microesfera contenía un marcador celular de tres partes único, que es un resultado de la combinación única de las tres secciones de marcador celular. Todos los oligonucleótidos en una única microesfera contenían el mismo marcador celular de tres partes. Los oligonucleótidos de diferentes microesferas contenían diferentes marcadores celulares de tres partes. Cada micropocillo contenía uno o menos microesfera conjugada a oligonucleótido. Se aplicó un reactivo de lisis celular a la matriz de micropocillos, provocando la lisis de las células. Las moléculas poliadeniladas (por ejemplo, ARNm) de la célula hibridaron con la secuencia de oligodT de los oligonucleótidos de las microesferas conjugadas a oligonucleótido. Las moléculas poliadeniladas que hibridaron con los oligonucleótidos de la microesfera conjugada a oligonucleótido se transcribieron de forma inversa con SuperScript II a 42 °C durante 90 minutos en un rotor. El oligonucleótido de la microesfera conjugada a oligonucleótido sirvió como cebador para la síntesis de ADNc de primera hebra. Se incorporó un oligo SMART en la síntesis de ADNc de modo que SuperScript II pueda añadir el complemento de la secuencia del oligo SMART al extremo 3' del ADNc cuando alcanza el extremo. La reacción de síntesis de ADNc produce una microesfera conjugada a oligonucleótidos no prolongados (por ejemplo, oligonucleótidos que no se adhirieron a la molécula poliadenilada de la célula) y los oligonucleótidos prolongados (por ejemplo, oligonucleótidos que se adhirieron a la molécula poliadenilada y comprenden un híbrido de molécula poliadenilada/ADNc).

Las microesferas se combinan y los oligonucleótidos que comprenden el híbrido de molécula poliadenilada/ADNc se amplificaron. Se realizó PCR combinada para amplificar un panel de 98 genes (véase la tabla 9) del ADNc en las microesferas. Los cebadores para la PCR combinada comprendían un primer cebador específico de gen que se diseñó para situarse aproximadamente 500 pares de bases desde el extremo 3' del ARNm y un cebador específico de gen interno que se diseñó para situarse aproximadamente 300 pares de bases desde el extremo 3' del ARNm. Los cebadores para la PCR combinada se diseñaron para que no requirieran complementariedad significativa en las últimas 6 bases de los cebadores en el panel. Si se detectaba complementariedad en los cebadores de PCR combinada, entonces los cebadores se reemplazaban manualmente. La reacción de PCR combinada comprendía las siguientes etapas: 1) 15 ciclos de primera PCR específica de gen (mezcla combinada KAPA, 50 nM de cada cebador - primer cebador específico de gen y cebador universal que es complementario a la secuencia universal de la microesfera conjugada a oligonucleótido), limpieza Ampure (relación de microesfera a molde 0,7 x), 15 ciclos de PCR específica de gen con cebadores internos (mezcla combinada KAPA, 50 nM de cada cebador - cebador específico de gen interno y cebador universal que es complementario a la secuencia universal de la microesfera conjugada a oligonucleótido), limpieza Ampure (relación de microesfera a molde 0,7 x), 8 ciclos de PCR final para añadir el adaptador Illumina de longitud completa (KAPA HiFi ReadyMix) y limpieza Ampure (relación de microesfera a molde 1 x).

Tabla 9. Panel de genes

Tipo de célula	Gen	Tipo de célula	Gen	Tipo de célula	Gen	Tipo de célula	Gen
linfocito B	PAX5	monocitos	CD14	virgen	CD62L (SELL)	Th17	IL17A
	CD19	monocitos clásicos	S100A12	Th virgen	CD45RA		IL17F
	CD20		CCR2		THPOK/ZBTB7B		IL21
	BCMA/ TNFRSF17		SELL/CD82L (L-selectina)	Tc virgen	RUNX3		IL22
	BAFF	monocitos no clásicos	CD16/FCGR3B	de memoria	CD45RO/PTPRC		CCL20
TCL1A	monocitos no clásicos	CX3CR1	CD8+/CD4+ de memoria central	CD44	IL23R		
TACI		ITGAL		CCR7	RORα/RORA		
IGHD		células dendríticas mieloides convencionales	CD1b	CD8+/CD4+ de memoria efector	TXK	RORgammat/RORC	
IGHM	FOXQ1		MBD2		OX40L/TNFSF4/ CD252		
CD27	CD209/ DC-SIGN		BCL6	CXCR5			
CD38	CD1e		BLIMP1	T auxiliar folicular	SLAMF1/SLAMF1		
CD24	CCL17		CXCR3		ICOS		
B de memoria	AICDA	DTNA	Th1	IFNGR1	SAP/SH2D1A		
B de transición	CD95	célula dendrítica plasmocitoide (infrecuente)	CLEC4C/ CD303	T y B activados	IL12RB2	CD69	
	CD10	célula dendrítica mielocide infrecuente (0.02 %)	CD141/ITM		IFN gamma	CD30	

Tabla 9. Panel de genes

Tipo de célula	Gen	Tipo de célula	Gen	Tipo de célula	Gen	Tipo de célula	Gen
Breg	IL10				IL33R/IL1RL1		TLR1
	RASD1	NKT	PLZF/ZBTB16		IL4R		TLR2
	AMPD1		SLAMF1		CCR4		TLR3
plasmática	SDC1 (CD138)			Th2	CRTH2/PTGDR2		TLR4
	OSBPL5	linfocito T	CD3 (CD3D)		IL4		TLR5
	CD56/NCAM 1		CD3 (CD3E)		IL5		TLR6
NK	IGFBP7		CD8 (CD8A)		CD25		TLR7
	KIR2DS5	T citotóxico	CD8 (CD8B)		FOXP1		TLR8
	KIR2DS2		PRF1 (perforina)	Treg	TGFbeta		TLR9
	RAB4B	T auxiliar	EOMES		IL10		TLR10
			CD4				
							receptores de tipo Toll (innatos)

5 Los productos amplificados se secuenciaron. Las lecturas de secuencia con 150 pb se alinearon con las secuencias de ARNm completas de los 98 genes enumerados en la tabla 9 usando Bowtie2. Los resultados de la alineación de secuencias (véase la tabla 10) demuestran que la reacción combinada de PCR produjo productos altamente específicos. La FIG. 14 muestra un gráfico que representa los genes en el eje de ordenadas y el log10 de los números de lecturas. De los 98 genes, 16 genes no estaban presentes. La ausencia de estos genes puede deberse al hecho de que algunos de los genes abordan células infrecuentes que pueden no estar presentes en esta muestra de sangre. Globalmente, se detectó aproximadamente un 84 % de los genes del panel de 98 genes.

Tabla 10

total	6 357 075
alineadas 0 veces	703 616
alineadas exactamente 1 vez	5 584 201
alineadas > 1 vez	69 258
% de alineadas exactamente una vez	88 %

10 La tabla 11 muestra los resultados de la estadística global de secuenciación. Para la lectura1, los criterios de coincidencia de lectura1 total requirieron una coincidencia perfecta con el marcador celular de tres partes (por ejemplo, código de barras celular) y como mucho 1 emparejamiento incorrecto con los conectores.

Tabla 11

número total de lecturas	6 357 075
criterios de coincidencia de lectura1 total	4 384 245
lectura2 también alineación	3 943 667
% de alineación de lectura2	89,95 %
número de bc celular único	31 129
recuento de lecturas por bc único > 100	3228
recuento de lecturas por bc único > 50	3721
% de lecturas útiles	62,04 %

15 La FIG. 15A muestra un gráfico de la distribución de genes detectados por marcador celular de tres partes (por ejemplo, código de barras celular). La FIG. 15B muestra un gráfico de la distribución de moléculas únicas detectadas por microesfera (que expresa el panel de genes).

20 Se realizó análisis de agrupación celular para determinar si los resultados de secuenciación podían usarse para analizar poblaciones celulares basándose en el código de barras celular. Se usó SPADE (un algoritmo de árbol de cobertura mínimo desarrollado por el laboratorio Nolan para datos de CyTOF) para agrupar células basándose en la presencia/ausencia de 17 genes. Para que un gen se considere presente, el promedio de redundancia de secuenciación para el gen tiene que ser mayor de 5 veces. Después de filtrar las secuencias, había aproximadamente 500 códigos de barras celulares únicos (por ejemplo, marcadores celulares) asociados con más de 20 moléculas únicas. Cada código de barras celular único corresponde a una célula individual. Basándose en los genes que estaban asociados con un código de barras celular único, las células se agruparon en tipos celulares. La tabla 12 muestra una lista de genes que puede usarse para identificar de forma definitiva un tipo celular. Por tanto, 25 los códigos de barras que están asociados con CD20, IGHM, TCL1A y CD24 se designaron como linfocitos B, mientras que los códigos de barras celulares que están asociados con CD8A, CD3D, CD3E, CD4 y CD62L se designaron como linfocitos T. Los genes restantes de la tabla 9 se cartografiaron a los grupos celulares. La FIG. 16 representa los grupos celulares basándose en los genes asociados con un código de barras celular. El tamaño del grupo es proporcional al número de células que se asignaron al grupo. Los resultados mostrados en la FIG. 16 demuestran que la combinación de código de barras celular y molecular puede usarse para marcar de forma unívoca copias de moléculas de una célula individual, lo que puede posibilitar el inmunofenotipado por secuenciación de célula individual. Además de la agrupación de PBMC en los tipos celulares principales basándose en los genes enumerados en la tabla 12, el panel de 98 genes también puede usarse para identificar grupos de subtipos de los tipos celulares principales. La tabla 13 muestra la frecuencia de cada tipo celular principal detectado por 30

secuenciación de célula individual. Como se muestra en la tabla 13, con la excepción de los linfocitos T CD8+, el porcentaje de cada tipo celular correspondía al intervalo de porcentaje normal de células. Se observó un porcentaje ligeramente mayor de linfocitos T CD8+ en la muestra de PBMC. Usando los grupos celulares basados en la FIG. 16, se usaron los perfiles de expresión de genes adicionales del panel de 98 genes para analizar adicionalmente los grupos celulares.

5

Tabla 12

Tipos celulares principales	Genes
Linfocitos B	CD20, IGHM, TCL1A, CD24
Linfocitos T	CD8A, CD3D, CD3E, CD4, CD62L
Linfocitos NKT	ZBTB16
Células dendríticas	CD209
Linfocitos citolíticos naturales	KIR2DS5, KIR2DS2, CD16
Monocitos	CD16, CD14, CCR2, S100A12, CD62L

Tabla 13

Tipo de célula	n.º de células	porcentaje	intervalo normal
monocitos	67	13,3 %	10-30 %
NK	85	16,9 %	hasta un 15 %
B	47	9,3 %	hasta un 15%
CD8	210	41,7 %	5-30 %
CD4	94	18,7 %	25-60 %
grupo asignado total	503	100,0 %	

La FIG. 17A-D muestra el análisis de marcadores específicos de monocitos. La FIG. 17E muestra el grupo celular representado en la FIG. 16. La FIG. 17A muestra el perfil de expresión celular para CD14, que es un marcador específico de monocitos. Los "colores calientes" (por ejemplo, rojo) representan alta expresión génica y los "colores fríos" (por ejemplo, azul) representan baja expresión génica. Como se muestra en la FIG. 17A, CD14 se expresa altamente en la población de monocitos y tenía de baja a ninguna expresión en los otros tipos celulares. El perfil de expresión celular para CD16 que sabe que está presente tanto en monocitos como en NK, se muestra en la FIG. 17B. Como se muestra en la FIG. 17B, los grupos celulares de monocitos y NK tenían alta expresión de CD16, mientras que los otros tipos celulares tenían de baja a ninguna expresión. Se sabe que CCR2 y S100A12 se expresan altamente en monocitos. La expresión específica de monocitos de CCR2 y S100A12 también se demostró en los perfiles de expresión celular mostrados en la FIG. 17C y D, respectivamente. Sin embargo, la expresión de CCR2 y S100A12 se separaba en dos ramas de células monocitos. Los otros tipos celulares mostraban de baja a ninguna expresión de CCR2 y S100A12.

10

15

La FIG. 18A-B muestra el análisis de los marcadores específicos de linfocitos T. La FIG. 18C muestra el grupo celular representado en la FIG. 16. La FIG. 18A muestra el perfil de expresión celular para CD3D que es una cadena de la molécula CD3. CD3 es un marcador de linfocitos T global. La FIG. 18A muestra que CD3D se expresa altamente en dos ramas de linfocitos T CD8+ T y se expresa moderadamente en una tercera rama de linfocitos T CD8+. Sin embargo, CD3D no se expresa altamente en linfocitos T CD4+. Además, los otros tipos celulares tienen de baja a ninguna expresión de CD3D. La FIG. 18B muestra el perfil de expresión celular para CD3E que es una cadena de la molécula CD3. La FIG. 18B muestra que CD3D se expresa altamente en linfocitos T CD4. Diferentes ramas de linfocitos T CD8+ muestran de alta a moderada expresión de CD3D. Se observa de poca a ninguna expresión de CD3D en los otros tipos celulares.

20

25

La FIG. 19A-B muestra el análisis de los marcadores específicos de linfocitos T CD8+. La FIG. 19C muestra el grupo celular representado en la FIG. 16. La FIG. 19A muestra el perfil de expresión celular para CD8A que es una cadena de la molécula CD8. Como se muestra en la FIG. 19A, diferentes ramas de linfocitos T CD8+ tienen diversos niveles de expresión de CD8A, con algunas ramas que tienen alta expresión, otras ramas que tienen expresión moderada y

30

una rama que muestra de baja a ninguna expresión de CD8A. Se observó alta expresión de CD8A en una rama de los linfocitos T NK CD16+. Se ha informado en la bibliografía de que hasta un 80 % de los linfocitos NK expresan CD8. Se observó de poca a ninguna expresión de CD8A en los otros tipos celulares. La FIG. 19B muestra el perfil de expresión celular para CD8B que es una cadena del modelo CD8. Como se muestra en la FIG. 19B, diferentes ramas de linfocitos T CD8+ tienen diversos niveles de expresión de CD8B, con una rama que tiene alta expresión, algunas ramas que tienen expresión moderada y dos ramas que muestran de baja a ninguna expresión de CD8B. También se observó alta expresión de CD8B en una rama de los linfocitos NK CD16+. Se observó de poca a ninguna expresión de CD8B en los otros tipos celulares.

La FIG. 20A muestra el análisis de marcadores específicos de linfocitos T CD4+ T. La FIG. 20B muestra el grupo celular representado en la FIG. 16. La FIG. 20A muestra el perfil de expresión para CD4. Se observó expresión moderada de CD4 en un subconjunto de células en el grupo de linfocitos T CD4+ y se observó alta expresión de CD4 en una rama del grupo de monocitos. Se ha documentado previamente en la bibliografía que los monocitos también expresan CD4. Se observó de moderada a baja expresión de CD4 en una rama de linfocitos T CD8+ y en linfocitos NK. Se observó de baja a ninguna expresión de CD4 en los otros tipos celulares.

La FIG. 21A-D muestra el análisis de marcadores específicos de linfocitos citolíticos naturales (NK). La FIG. 20E muestra el grupo celular representado en la FIG. 16. La FIG. 20A muestra el perfil de expresión para KIR2DS2. Todos los tipos celulares mostraron de poca a ninguna expresión de KIR2DS2. La FIG. 20B muestra el perfil de expresión para KIR2DS5. Se sabe que los receptores citolíticos de inmunoglobulina (KIR) se expresan en linfocitos NK y un subconjunto de linfocitos T. Se observó alta expresión de KIR2DS5 en 2 ramas de linfocitos NK y se observó de moderada a baja expresión de KIR2DS5 en una rama de linfocitos NK. Se observó de moderada a alta expresión de KIR2DS5 en 2 ramas de linfocitos T CD8+. Se observó de baja a ninguna expresión de KIR2DS5 en todos los demás tipos celulares. Se sabe que OSBPL5 e IGFBP7 se expresan altamente en linfocitos NK. La FIG. 20C muestra el perfil de expresión para OSBPL5. OSBPL5 se expresaba altamente en una rama de linfocitos NK. Se observó de moderada a baja expresión de OSBPL5 en una rama de linfocitos B. Se observó de baja a ninguna expresión de OSBPL5 en todos los demás tipos celulares. La FIG. 20D muestra el perfil de expresión para IGFBP7. Se observó alta expresión de IGFBP7 en dos ramas de linfocitos NK y una rama de monocitos. Se observó expresión moderada de IGFBP7 en una rama de linfocitos B. Se observó de baja a ninguna expresión de IGFBP7 en todos los demás tipos celulares.

La FIG. 22A-E muestra el análisis de marcadores específicos de linfocitos B. La FIG. 22F muestra el grupo celular representado en la FIG. 16. La FIG. 22A muestra el perfil de expresión para IGHM CH4. IGHM CH4 se expresaba altamente en una rama de linfocitos B y se expresaba moderadamente en la segunda rama de linfocitos B. Se observó de baja a ninguna expresión de IGHM CH4 en todos los demás tipos celulares. La FIG. 22B muestra el perfil de expresión para PAX5. PAX5 se expresaba altamente en una rama de linfocitos B. Se observó de baja a ninguna expresión de PAX5 en todos los demás tipos celulares. La FIG. 22C muestra el perfil de expresión para CD20. CD20 se expresaba altamente en una rama de linfocitos B. Se observó de baja a ninguna expresión de CD20 en todos los demás tipos celulares. La FIG. 22D muestra el perfil de expresión para TCL1A. Se observó de baja a ninguna expresión de TCL1A en todos los demás tipos celulares. La FIG. 22E muestra el perfil de expresión para IGHD CH2. IGHD CH2 se expresaba altamente en una rama de linfocitos B. Se observó de baja a ninguna expresión de IGHD CH2 en todos los demás tipos celulares.

La FIG. 23A-F muestra el análisis de receptores de tipo Toll. Los receptores de tipo Toll se expresan principalmente por monocitos y algunos linfocitos B. La FIG. 23G muestra el grupo celular representado en la FIG. 16. La FIG. 23A muestra el perfil de expresión para TLR1. Una rama de monocitos mostraba alta expresión de TLR1 y dos ramas de monocitos mostraban expresión moderada de TLR1. Se observó de baja a ninguna expresión de TLR1 en todos los demás tipos celulares. La FIG. 23B muestra el perfil de expresión para TLR4. Una rama de monocitos mostraba alta expresión de TLR4. Se observó expresión moderada de TLR4 en dos ramas de monocitos y una rama de linfocitos NK. Se observó de baja a ninguna expresión de TLR4 en todos los demás tipos celulares. La FIG. 23C muestra el perfil de expresión para TLR7. Se observó alta expresión de TLR7 en una rama de monocitos y se observó expresión moderada de TLR7 en una rama de linfocitos NK. Se observó de baja a ninguna expresión de TLR7 en todos los demás tipos celulares. La FIG. 23D muestra el perfil de expresión para TLR2. Se observó alta expresión de TLR2 en una rama de linfocitos B. Se observó de baja a ninguna expresión de TLR2 en todos los demás tipos celulares. La FIG. 23E muestra el perfil de expresión para TLR3. Se observó alta expresión de TLR3 en una rama de linfocitos B. Se observó de baja a ninguna expresión de TLR3 en todos los demás tipos celulares. La FIG. 23F muestra el perfil de expresión de TLR8. Se observó alta expresión de TLR8 en tres ramas de monocitos. Se observó de moderada a baja expresión de TLR8 en dos ramas de monocitos y una rama de linfocitos NK. Se observó de baja a ninguna expresión de TLR8 en todos los demás tipos celulares.

Estos resultados demuestran que la secuenciación masiva de células individuales en paralelo puede identificar de forma satisfactoria tipos celulares principales en PBMC. Los resultados de secuenciación también determinaron que algunos marcadores celulares que se usan en FACS para identificar tipos celulares no tienen alta expresión de ARNm (por ejemplo, CD56 para linfocitos NK, CD19 para linfocitos B). Además, muchos de los genes en el panel de genes se expresaban entre múltiples tipos celulares. Estos perfiles de expresión pueden usarse para subtipar células dentro de un tipo celular principal (por ejemplo, célula activada frente a célula en reposo, etc.).

Ejemplo 11: Identificación de células infrecuentes en una población

En este experimento, se usa secuenciación masiva de células individuales en paralelo para identificar células cancerosas de una muestra de células cancerosas y no cancerosas. Se añadieron células Ramos (linfoma de Burkitt) a una población de linfocitos B CD19+ que se aislaron de un individuo sano. La concentración de las células Ramos en la población mixta era de aproximadamente un 4-5 %. Se depositaron aproximadamente 7000 linfocitos B normales y 300 células Ramos en una matriz de micropocillos con 25 200 pocillos. Por tanto, la mayoría de los pocillos en la matriz de micropocillos no contenía células y algunos pocillos en la celda contenían únicamente una célula. Se aplicaron microesferas conjugadas a oligonucleótido a la matriz de micropocillos. Cada microesfera conjugada a oligonucleótido contenía aproximadamente 1000 millones de oligonucleótidos unidos a una microesfera. Cada oligonucleótido adherido a la microesfera contenía una 5' amina, secuencia universal, marcador celular de tres partes (por ejemplo, tres secciones de marcador celular conectadas por dos conectores), marcador molecular y oligodT. Cada microesfera contenía un marcador celular de tres partes único, que es un resultado de la combinación única de las tres secciones de marcador celular. Todos los oligonucleótidos en una única microesfera contenían el mismo marcador celular de tres partes. Los oligonucleótidos de diferentes microesferas contenían diferentes marcadores celulares de tres partes. Cada pocillo contenía 1 o menos microesferas conjugadas a oligonucleótido. Se aplicó un reactivo de lisis celular a la matriz de micropocillos, provocando la lisis de las células. Las moléculas poliadeniladas (por ejemplo, ARNm) de la célula hibridaron con la secuencia de oligodT de los oligonucleótidos de las microesferas conjugadas a oligonucleótido. Las moléculas poliadeniladas que hibridaron con los oligonucleótidos de la microesfera conjugada a oligonucleótido se transcribieron de forma inversa con SuperScript II a 42 °C durante 90 minutos en un rotor. El oligonucleótido de la microesfera conjugada a oligonucleótido sirvió como cebador para la síntesis de ADNc de primera hebra. Se incorporó un oligo SMART en la síntesis de ADNc de modo que SuperScript II puede añadir el complemento de la secuencia del oligo SMART al extremo 3' del ADNc cuando alcanza el extremo. La reacción de síntesis de ADNc produce una microesfera conjugada a oligonucleótidos no prolongados (por ejemplo, oligonucleótidos que no se adhirieron a la molécula poliadenilada de la célula) y los oligonucleótidos prolongados (por ejemplo, oligonucleótidos que se adhirieron a la molécula poliadenilada y comprenden un híbrido de molécula poliadenilada/ADNc).

Las microesferas se combinan y los oligonucleótidos que comprenden el híbrido de molécula poliadenilada/ADNc se amplificaron. Se realizó PCR combinada para amplificar un panel de 111 genes a partir del ADNc de las microesferas. Los 111 genes representan marcadores para diferentes subconjuntos de linfocitos B. Los cebadores para la PCR combinada comprendían un primer cebador específico de gen que se diseñó para situarse aproximadamente 500 pares de bases desde el extremo 3' del ARNm y un cebador específico de gen interno que se diseñó para situarse aproximadamente 300 pares de bases desde el extremo 3' del ARNm. Los cebadores para PCR combinada se diseñaron para que no requieran complementariedad significativa en las últimas 6 bases de los cebadores en el panel. Si se detectaba complementariedad en los cebadores de la PCR combinada, entonces los cebadores se reemplazaban manualmente. La reacción de PCR combinada comprendía las siguientes etapas: 1) 15 ciclos de primera PCR específica de gen (mezcla combinada KAPA, 50 nM de cada cebador - primer cebador específico de gen y cebador universal que es complementario a la secuencia universal de la microesfera conjugada a oligonucleótido), limpieza Ampure (relación de microesfera a molde 0,7 x), 15 ciclos de PCR específica de gen con cebadores internos (mezcla combinada KAPA, 50 nM de cada cebador - cebador específico de gen interno y cebador universal que es complementario a la secuencia universal de la microesfera conjugada a oligonucleótido), limpieza Ampure (relación de microesfera a molde 0,7 x), 8 ciclos de PCR final para añadir el adaptador Illumina de longitud completa (KAPA HiFi ReadyMix) y limpieza Ampure (relación de microesfera a molde 1 x).

Los productos amplificados se secuenciaron. Las lecturas de secuencia que comprendían 150 pb se alinearon con las secuencias de ARNm completas de los 111 genes (tabla 17) usando Bowtie2. Los resultados de la alineación de secuencias (véase la tabla 14) demuestran que la reacción de PCR combinada provocaba productos altamente específicos. La FIG. 24 representa un gráfico de los genes frente al log10 del número de lecturas. De los 111 genes, 24 no estaban presentes. Al menos dos de los genes, RAG1 y RAG2 que están implicados en la recombinación VDJ y deben estar presentes únicamente en linfocitos pre-B, no deben estar presentes. Unos pocos de los genes ausentes eran específicos de células plasmáticas, que se conservan muy raramente en células congeladas.

Tabla 17

CD19	AURKB	FOXP1	CCND3	TLR1	FOXP3	CXCL12	GNAI2
CD27	CD81	MCL1	IL12A	TLR2	LAG3	CCL3	RGS1
CD138	CD80	IFNB1	IFNG	TLR3	CD73	CCL14	CD5
CD38	CD23a	BLNK	TNFA	TLR4	CD70	CCL20	CD22
CD24	CD44	CD40LG	IL2	TLR5	CCR7	CCL18	PIK3CD
CD10	LEF1	IGBP1	IL4	TLR6	CD45RA	TCL1A	DOCK8

Tabla 17

CD95	CXCR5	IRF4	IL6	TLR7	PDCD1	TACI	CD11b
CD21	PRKCB	CD79a	BAFF	TLR8	MYC	AICDA	FCGR2B
CXCR3	PRKCD	LTA	IGHE	TLR9	CD25	FCRL4	CD72
CD40	CD20	HDAC5	IGHD	TLR10	FCAMR	BCL2	BCL11B
CD69	CD30	RAG1	IGHM	GAPDH	CCND2	FASLG	CD86
CD1c	CD30L	RAG2	IGHA	CD9	MKI67	BCL6	TBX21
IL10	BAFFR	CD1d	IGHG1	CD11c	IL21R	IGHG2	PRDM1
IL4R	CMRF-35H	TGFB1	IGHG4	IL6R	HLA-DRA	IGHG3	

Tabla 14

total	5 711 013
alineadas 0 veces	504 775
alineadas exactamente 1 vez	5 203 308
alineadas > 1 vez	2930
% de alineadas exactamente una vez	91,6 %

5 La tabla 15 muestra los resultados de la estadística de secuenciación global. Para la lectura1, los criterios de coincidencia de lectura1 total requirieron una coincidencia perfecta con el marcador celular de tres partes (por ejemplo, código de barras celular) y como mucho 1 emparejamiento incorrecto con los conectores.

Tabla 15

número total de lecturas	5 711 013
criterios de coincidencia de lectura1 total	3 795 915
lectura2 también alineación	3 495 392
% de alineación de lectura2	92 %
número de bc celular único	40 764
recuento de lecturas por bc único > 100	3313
recuento de lecturas por bc único > 50	4154
% de lecturas útiles	61 %

10 La FIG. 25A-D muestra gráficos del código de barras molecular frente al número de lecturas o log10 del número de lecturas para dos genes. La FIG. 25A muestra un gráfico de código de barras molecular (clasificado por abundancia) frente al número de lecturas para CD79. La FIG. 25B muestra un gráfico del código de barras molecular (clasificado por abundancia) frente al log10 del número de lecturas para CD79. La FIG. 25C muestra un gráfico del código de barras molecular (clasificado por abundancia) frente al número de lecturas para GAPDH. La FIG. 25D muestra un gráfico del código de barras molecular (clasificado por abundancia) frente al log 10 del número de lecturas para GAPDH.

15 Se retuvieron 856 células para el análisis. La FIG. 26A muestra un gráfico del número de genes en el panel expresados por código de barras celular frente al número de códigos de barras celulares únicos/célula individual. La FIG. 26B muestra un histograma del número de moléculas únicas detectadas por microesfera frente a la frecuencia del número de células por código de barras único que porta un número dado de moléculas. Un pequeño subconjunto

de células mostró un número distintivo mayor de moléculas de ARNm y un número distintivo de genes expresados a partir del panel de 111 genes (véanse se las secciones rodeadas con un círculo en la FIG. 26A-B). La FIG. 26C muestra un histograma del número de moléculas GAPDH únicas detectadas por microesfera frente a la frecuencia del número de células/código de barras celular único que porta un número dado de moléculas.

5 Se usó análisis de componentes principales (PCA) para generar un diagrama de dispersión de células. La FIG. 27 muestra un diagrama de dispersión de las 856 células. El PCA identificó el pequeño subconjunto de células con un patrón de expresión génica diferente a la mayoría de células. El subconjunto de células contenía 18 células, que es aproximadamente un 2 % del total de las células analizadas. Este porcentaje es similar al porcentaje de células Ramos que se añadió a la población.

10 Las células Ramos se obtienen de linfocitos B foliculares y expresan fuertemente los marcadores de diferenciación de linfocitos B CD20, CD22, CD19, CD10 y BCL6. Las células Ramos también expresan IgM y sobreexpresan c-myc. La FIG. 28 muestra un mapa térmico de expresión de los 100 principales (en términos de número total de moléculas detectadas). El subconjunto de células (18 células) que expresan niveles mucho mayores de ARNm también expresan fuertemente genes que se sabe que son marcadores para células Ramos (por ejemplo, CD10, Bcl-6, CD22, C-myc e IgM).

15 Estos resultados demuestran que la secuenciación masiva de células individuales en paralelo identificó satisfactoriamente pequeños subconjuntos (tan poco como un 2 %) de tipos celulares anómalos en una suspensión celular. La secuenciación masiva de células individuales en paralelo puede usarse en el diagnóstico del cáncer (por ejemplo, biopsia/células tumorales en circulación). Como las células cancerosas son de mayor tamaño y portan más ARNm, se pueden diferenciar fácilmente de las células normales.

20 Ejemplo 12: Genoma completo masivo de células individuales en paralelo y amplificación múltiple de dianas de ADN_g usando RESOLVE

La FIG. 29 muestra un flujo de trabajo para este ejemplo. Como se muestra en la FIG. 29, se aplica una suspensión de células a una matriz de pocillos (2901). El número de células en la suspensión de células es menor que la cantidad de pocillos en la matriz de pocillos, de modo que la aplicación de la suspensión de células a la matriz de micropocillos provoca que un pocillo en la matriz de pocillos contenga 1 o menos células. Se aplican microesferas conjugadas a oligonucleótido (2905) a la matriz de pocillos. Una microesfera conjugada a oligonucleótido (2905) contiene una microesfera (2910) adherida a un oligonucleótido que comprende una 5' amina (2915), secuencia de cebador universal (2920), marcador celular (2925), marcador molecular (2930) y oligómero aleatorio (2935). La microesfera conjugada a oligonucleótido contiene aproximadamente 1000 millones de oligonucleótidos. Un oligonucleótido contiene una 5' amina, secuencia de cebador universal, marcador celular, marcador molecular y oligómero aleatorio. Cada oligonucleótido en una única microesfera contiene el mismo marcador celular. Sin embargo, dos o más oligonucleótidos en una única microesfera pueden contener dos o más marcadores moleculares diferentes. Una microesfera puede contener múltiples copias de oligonucleótidos que contienen el mismo marcador molecular.

Después de añadir las microesferas conjugadas a oligonucleótido a la matriz de pocillos, se aplica un tampón de lisis celular a la superficie de la matriz. Como se muestra en la FIG. 29, el ADN genómico (2945) de la célula híbrida con las secuencias de oligómero aleatorio (2935) de las microesferas conjugadas a oligonucleótido (2940). Se añade un tampón de neutralización a la superficie de la matriz. Se añade una ADN polimerasa (por ejemplo, Phi29) y dNTP a la superficie de la matriz. La secuencia del oligómero aleatorio (2935) actúa como un cebador para la amplificación del ADN genómico, produciendo de ese modo una microesfera conjugada a ADN_g (2955). La microesfera conjugada a ADN_g (2955) contiene un oligonucleótido que comprende una 5' amine (2915), secuencia de cebador universal (2920), marcador celular (2925), marcador molecular (2925), oligómero aleatorio (2935) y copia del ADN genómico (2955). El ADN genómico original (2945) se hibrida con el oligómero aleatorio (2935) y la copia del ADN genómico (2955). Para una única microesfera, hay múltiples moléculas de ADN genómico diferentes adheridas a los oligonucleótidos.

Como se muestra en la FIG. 29, las microesferas conjugadas a ADN_g (2950) de los pocillos se combinan en un tubo Eppendorf (2960). El ADN genómico en la mezcla MDA de ADN_g que contiene oligómeros aleatorios, dNTP y un ADN polimerasa (por ejemplo, Phi29) se añade al tubo Eppendorf que contienen las microesferas conjugadas a ADN_g combinadas. El ADN genómico marcado se amplifica adicionalmente para producir amplicones marcados (2965) en disolución. Un amplicón marcado (2965) comprende una secuencia de cebador universal (2920), un marcador celular (2925), marcador molecular (2930), oligómero aleatorio (2935) y copia del ADN genómico (2955). Los amplicones marcados se cortan en trozos más pequeños de aproximadamente 1 kb o menos. Como alternativa, los amplicones marcados pueden fragmentarse mediante Tagmentation (Nextera). El corte o fragmentación de los amplicones marcados produce fragmentos marcados (2980) y fragmentos no marcados (2985). El fragmento marcado (2980) contiene la secuencia de cebador universal (2920), marcador celular (2925), marcador molecular (2930), oligómero aleatorio (2935) y fragmento de la copia del ADN genómico (2955). Se añaden adaptadores (2970, 2975) a los fragmentos. La secuencia de cebador universal puede usarse para seleccionar fragmentos marcados (2980) mediante arrastre por hibridación o PCR usando la secuencia de cebador universal y un cebador contra uno de los adaptadores (2970, 2975).

Los fragmentos marcados pueden secuenciarse. Las lecturas de secuencia que comprenden una secuencia del marcador celular tal como marcador molecular y fragmento genómico pueden usarse para identificar poblaciones celulares de la suspensión celular. Puede usarse análisis de componentes principales para generar diagramas de dispersión de las células basados en marcadores celulares conocidos. Como alternativa, o adicionalmente, puede usarse SPADE para producir diagramas de grupos celulares. Puede usarse un programa informático para generar una lista que comprende un marcador celular y los marcadores moleculares y fragmentos genómicos asociados con el marcador celular.

Ejemplo 13: Secuenciación masiva en paralelo para identificar células en una población heterogénea

El flujo de trabajo experimental para este ejemplo se muestra en la FIG. 30. Como se muestra en la FIG. 30, una población mixta de células se dispersó de forma estocástica en una matriz de pocillos. En este ejemplo, la población mixta de células comprende una mezcla de células Ramos y células K562. La suspensión celular comprende una baja concentración de células de modo que cada pocillo en la matriz contiene una o menos células. Después de aplicar las células a la matriz de pocillos, se dispersó de forma estocástica una pluralidad de microesferas conjugadas a oligonucleótido en la matriz de pocillos. La microesfera de oligonucleótido contiene una pluralidad de oligonucleótidos que comprenden una 5' amina, secuencia de cebador universal, marcador celular, marcador molecular y oligodT. Los marcadores celulares de la pluralidad de oligonucleótidos de una única microesfera son idénticos. Una única microesfera puede comprender múltiples oligonucleótidos que comprenden el mismo marcador molecular. Además, una única microesfera puede comprender múltiples oligonucleótidos que comprenden diferentes marcadores moleculares. Un marcador celular de un oligonucleótido conjugado a una primera microesfera es diferente de un marcador celular de un oligonucleótido conjugado con una segunda microesfera. Por tanto, el marcador celular puede usarse para diferenciar dos o más microesferas conjugadas a oligonucleótido. Las células se lisaron y las moléculas de ARN de una célula individual se adhirieron a las microesferas conjugadas a oligonucleótido en el mismo pocillo. La FIG. 30 muestra la adhesión de la secuencia de poliA de un ARN a la secuencia de oligodT del oligonucleótido. Después de la adhesión de las moléculas de ARN de las células individuales a las microesferas conjugadas a oligonucleótido en el mismo pocillo, las microesferas se combinaron en una única muestra. Se realizó una reacción de síntesis de ADNc en las microesferas en la muestra única. La FIG. 30 muestra que el producto de la síntesis de ADNc comprende una microesfera adherida a un oligonucleótido, comprendiendo el oligonucleótido la 5' amina, secuencia de cebador universal, marcador celular, marcador molecular, oligodT y una copia de la molécula de ARN. Por motivos de simplicidad, se representa únicamente un oligonucleótido en la FIG. 30, sin embargo, en este ejemplo cada microesfera conjugada a oligonucleótido comprende aproximadamente 1000 millones de oligonucleótidos. Como se muestra en la FIG. 30, se realizó PCR combinada con las microesferas en la muestra única usando un cebador universal que hibridó con la secuencia de cebador universal y un cebador específico de gen que hibridó con la copia de la molécula de ARN. Se diseñaron cebadores específicos de gen para que se unieran a genes específicos de Ramos o genes específicos de K562 del panel de genes mostrado en la tabla 16. Como control, también se usó un cebador específico del gen de GAPDH en la reacción de PCR combinada. Finalmente, se usó secuenciación de la siguiente generación para secuenciar los productos amplificados. Las lecturas de secuenciación incluían información perteneciente al marcador celular, marcador molecular y el gen. Usando análisis de componentes principales, se construyó un diagrama de dispersión de las células basado en la información de secuenciación perteneciente al marcador celular, marcador molecular y el gen. Análogo a la forma en que se usa FACS para clasificar células y la forma en que se usan los diagramas de dispersión basados en los marcadores superficiales para agrupar células, el marcador celular se usa para identificar genes de una célula individual y el marcador molecular se usa para determinar la cantidad de los genes. Esta información combinada se usa entonces para relacionar el perfil de expresión génica de células individuales. Como se muestra en la FIG. 31A, la secuenciación masiva de células individuales en paralelo con marcadores celulares y moleculares pudo identificar satisfactoriamente las dos poblaciones celulares (células K562 y Ramos) en la población de células mixta.

Tabla 16

Gen	Célula	Gen	Célula
CD74	Específica de Ramos	CD41	Específica de K562
CD79a	Específica de Ramos	GYPA	Específica de K562
IGJ	Específica de Ramos	GATA2	Específica de K562
TCL1A	Específica de Ramos	GATA1	Específica de K562
SEPT9	Específica de Ramos	HBG1	Específica de K562
CD27	Específica de Ramos	GAPDH	Común

Ejemplo 14: Secuenciación masiva de células individuales en paralelo con análisis de componentes principales

- 5 En este ejemplo, se marcaron de forma estocástica moléculas de ARNm de células individuales con microesferas conjugadas a oligonucleótido en paralelo. Se aislaron las PBMC de sangre y se congelaron a -80 °C en RPMI1640 más FBS y DMSO. Las PBMC se descongelaron y lavaron tres veces con PBS. Una muestra de PBMC que comprende una mezcla de tipos celulares (4000 células en total) se aplicó de forma estocástica a una matriz de pocillos de agarosa. La matriz de pocillos de agarosa contenía 37 500 células. Una mezcla de 150 000 microesferas conjugadas a oligonucleótido se aplicó de forma estocástica a la matriz de pocillos mediante una junta de PDMS que rodeaba la matriz de pocillos. La microesfera conjugada a oligonucleótido se representa en la FIG. 1. Por motivos de simplicidad, se muestra únicamente un oligonucleótido adherido a la microesfera, sin embargo, las microesferas conjugadas a oligonucleótido contenían aproximadamente 1000 millones de oligonucleótidos.
- 10 Las células se lisaron colocando la matriz de pocillos en un bloque frío durante 10 minutos y aplicando tampón de lisis a la superficie de la matriz. Una vez que las células en los pocillos se lisaron, las moléculas de ARNm de las células individuales se adhirieron a la microesfera conjugada a oligonucleótido mediante la secuencia de oligodT. Se aplicó un imán a la matriz y la matriz se lavó dos veces con tampón de lavado.
- 15 Las microesferas con las moléculas de ARNm adheridas se combinaron en un tubo Eppendorf. Las moléculas de ARNm adheridas a las microesferas se transcribieron de forma inversa para producir ADNc. La siguiente mezcla de síntesis de ADNc se preparó de la siguiente manera:

Componente	Volumen (ul)
Agua	8
dNTP (10 mM)	2
Tampón de primera hebra 5 x	4
MgCl ₂	2,4
SuperRase In	1
oligo SMART (50 uM)	0,4
DTT 0,1 M	1
BSA 100 x	0,2
SSII	1
Total	20

- 20 La mezcla de síntesis de ADNc se añadió al tubo Eppendorf que contenía las microesferas con las moléculas de ARNm adheridas. El tubo Eppendorf se incubó a 40 °C durante 90 minutos en un rotor. La reacción de síntesis de ADNc se produjo sobre las microesferas. Después de 90 minutos, se aplicó un imán al tubo y se retiró la mezcla de ADNc y se reemplazó con la siguiente mezcla de reacción de Exol:

Componente	Volumen (ul)
Tampón de Exol	2
Agua	17
Exol	1

- 25 Se incubaron los tubos a 37 °C durante 30 minutos en un rotor. Los tubos después se transfirieron a un termociclador durante 15 minutos a 80 °C. Después de incubar el tubo a 80 °C durante 15 minutos, se añadieron 70 microlitros de TE+Tween20 al tubo. Se aplicó un imán al tubo y se retiró el tampón. Las microesferas entonces se resuspendieron en 50 microlitros de TE+Tween20.

El ADNc adherido a las microesferas se amplificó por PCR en tiempo real usando la siguiente mezcla de amplificación:

Componente	Volumen (ul)
Mezcla iTaq 2 x	10
GAPDH ILMN (10 uM)	0,6
ILR2 (10 uM)	0,6
Microesfera	2
Agua	6,8
Total	20

- Los amplicones de ADNc marcados se secuenciaron para detectar el marcador celular, índice molecular y gen. Las lecturas de secuenciación se alinearon para el marcador celular, después el gen y finalmente el marcador molecular. Un marcador celular asociado con 4 o más genes o asociado con 10 o más moléculas de transcrito únicas, con cada molécula de transcrito única secuenciada más de una vez, se denominó una célula. El análisis de componentes principales con todos los genes de la tabla 9 detectados se usó para identificar el conjunto de genes que tenía la mayor contribución en la variación en los datos. Se usaron 632 células individuales en el análisis de componentes principales. Basándose en los resultados de secuenciación, se detectaron 81 de los 98 genes.
- 5 La FIG. 32 muestra un diagrama de análisis de componentes principales para la expresión de GAPDH. Como se muestra en la FIG. 32, se observaron dos grupos celulares basándose en la ubicación del espacio de componentes principales.
- 10 La FIG. 33A-F muestra el análisis de componentes principales (PCA) para genes asociados a monocitos. La FIG. 33A muestra el PCA para CD16. La FIG. 33B muestra el PCA para CCRvarA. La FIG. 33C muestra el PCA para CD14. La FIG. 33D muestra el PCA para S100A12. La FIG. 33E muestra el PCA para CD209. La FIG. 33F muestra el PCA para IFNGR1.
- 15 La FIG. 34A-B muestra el análisis de componentes principales (PCA) para marcadores de linfocitos T globales (CD3). La FIG. 34A muestra el PCA para CD3D y la FIG. 34B muestra el PCA para CD3E.
- La FIG. 35A-E muestra el análisis de componentes principales (PCA) para genes asociados a linfocitos T CD8. La FIG. 35A muestra el PCA para CD8A. La FIG. 35B muestra el PCA para EOMES.
- 20 La FIG. 35C muestra el PCA para CD8B. La FIG. 35D muestra el PCA para PRF1. La FIG. 35E muestra el PCA para RUNX3.
- La FIG. 36A-C muestra el análisis de componentes principales (PCA) para genes asociados a linfocitos T CD4. La FIG. 36A muestra el PCA para CD4. La FIG. 36B muestra el PCA para CCR7. La FIG. 36C muestra el PCA para CD62L.
- 25 La FIG. 37A-F muestra el análisis de componentes principales (PCA) para genes asociados a linfocitos B. La FIG. 37A muestra el PCA para CD20. La FIG. 37B muestra el PCA para IGHD. La FIG. 37C muestra el PCA para PAX5. La FIG. 37D muestra el PCA para TCL1A. La FIG. 37E muestra el PCA para IGHM. La FIG. 37F muestra el PCA para CD24.
- 30 La FIG. 38A-C muestra el análisis de componentes principales (PCA) para genes asociados a linfocitos citolíticos naturales. La FIG. 38A muestra el PCA para KIR2DS5. La FIG. 38B muestra el PCA para CD16. La FIG. 38C muestra el PCA para CD62L.
- Basándose en los análisis de componentes principales, los monocitos y los linfocitos formaban dos grupos distintos en PC1. Los linfocitos B, T y NK formaban otro grupo que residía como un continuo en el grupo a lo largo de PC2. La FIG. 39 muestra el análisis PCA de la expresión de GAPDH con anotaciones para los tipos celulares y subtipos celulares. La FIG. 40 representa un mapa térmico que muestra la correlación en el perfil de expresión génica entre las células. A lo largo de la diagonal que parte desde la esquina superior izquierda, las células son monocitos, linfocitos T CD4 vírgenes, linfocitos T CD8 vírgenes, linfocitos T CD8 citotóxicos, linfocitos NK y linfocitos B. La FIG. 41 muestra otra versión de un mapa térmico que demuestra la correlación entre la expresión génica y el tipo celular. La FIG. 42 muestra un mapa térmico que demuestra la correlación en el perfil de expresión génica entre genes.
- 35
- 40 Ejemplo 15: Descubrimiento de heterogeneidad celular por citometría de expresión génica digital
- Se presenta una estrategia para citometría de expresión génica combinando secuenciación de la siguiente generación con códigos de barras estocásticos de células individuales. Se depositaron miles de células aleatoriamente en una matriz de aproximadamente 150 000 micropocillos. Se añadió una colección de microesferas

que albergan sondas de captura de código de barras celular y de transcrito de modo que cada célula se reparte a lo largo de una microesfera con un código de barras celular único. Después de la lisis celular, los ARNm se hibridaron con las microesferas y se combinaron para transcripción inversa, amplificación y secuenciación. El perfil de expresión génica digital para cada célula se reconstruyó cuando se contaron los transcriptos con código de barras y se asignaron a la célula de origen. Se aplicó la tecnología para examinar el sistema hematopoyético humano en subpoblaciones celulares, y para caracterizar la respuesta heterogénea de células inmunitarias a estimulación *in vitro*. Además, se demostró la alta sensibilidad del método mediante la detección de células infrecuentes, tales como linfocitos T específicos de antígeno y células tumorales en un elevado fondo de células normales.

Introducción

La comprensión de la diversidad celular y la función en una gran colección de células requiere la medición de genes específicos o proteínas expresados por células individuales. La citometría de flujo está bien establecida para medir la expresión de proteínas de células individuales, aunque las mediciones de expresión de ARNm se realizan típicamente en muestras voluminosas, lo que obstaculiza las contribuciones de las células individuales. Aunque se ha informado recientemente de mediciones de expresión de ARNm en células individuales usando placas de microvaloración o chips de microfluidos comerciales (1-5), estas estrategias son de un rendimiento extremadamente bajo y difíciles de cambiar de escala. A causa de estas limitaciones, la mayoría de los estudios hasta la fecha están restringidos tanto en la cantidad de células consultadas como en la cantidad de condiciones exploradas.

En esta ocasión, se ha desarrollado una estrategia de alta escalabilidad que posibilita un perfilado rutinario de expresión génica digital de miles de células individuales entre un número arbitrario de genes. Se usó genomanipulación a microescala y química combinatoria para marcar todas las moléculas de ARNm en una célula con un código de barras celular único de una manera masiva en paralelo. Además, cada copia de transcrito dentro de una célula se marcó con un código de barras molecular, lo que permite mediciones de expresión génica digital absolutas (6). Las moléculas ARNm con marca de todas las células se combinaron, amplificaron y secuenciaron. El perfil de expresión génica digital de cada célula se reconstruyó usando los códigos de barras celulares y moleculares en cada secuencia. Esta tecnología de alta escalabilidad posibilita la citometría de expresión génica, que se denomina CytoSeq. Se ha aplicado la técnica para la clasificación genética de múltiples parámetros del sistema hematopoyético y se demostró su uso para estudiar la heterogeneidad celular y detectar células infrecuentes en una población.

Resultados

CytoSeq

El procedimiento se resumió en la **FIG. 43A**. En primer lugar, se cargó una suspensión celular en una superficie microfabricada con hasta 150 000 micropocillos. Cada micropocillo de 30 micrómetros de diámetro tiene un volumen de ~20 picolitros. El número de células se ajustó de modo que únicamente ~1 de 10 o más pocillos reciba una célula. Las células sedimentaron dentro de los pocillos por gravedad.

Se cargaron microesferas magnéticas en la matriz de micropocillos hasta saturación, de modo que una microesfera se asentara parcialmente en la parte superior de, o adyacente a, cada célula dentro de un pocillo. La dimensión de la microesfera se eligió de modo que cada micropocillo pueda albergar únicamente una microesfera. Cada microesfera magnética portaba aproximadamente mil millones de moldes oligonucleotídicos con la estructura resumida en la **FIG. 43B**. Cada oligonucleótido presentaba un sitio de cebador universal, seguido por un marcador celular, un marcador molecular y una secuencia de captura de oligo(dT). Todos los oligonucleótidos en cada microesfera tienen el mismo marcador celular, pero contienen una diversidad de marcadores moleculares. Se ideó un método de división-combinación combinatorio para sintetizar microesferas con una diversidad de cerca de un millón. La probabilidad de tener dos células individuales marcadas con el mismo marcador celular era baja (del orden de 10^{-4}) porque únicamente ~10 % de los pocillos estaban ocupados por una célula individual. Asimismo, la diversidad de los marcadores moleculares en una única microesfera era del orden de 10^4 , y la probabilidad de que dos moléculas de transcrito del mismo gen en la misma célula que estén marcadas con el mismo marcador molecular también era baja.

Se aplicó tampón de lisis a la superficie de la matriz de pocillos y difunde a los pocillos. Las moléculas de ARNm con cola de poli(dA) liberadas de una célula hibridan con el oligo(dT) en el extremo 3' de los oligonucleótidos en la microesfera. Como la célula estaba adyacente a la microesfera, en las condiciones de alta salinidad del tampón de lisis y la alta concentración local de ARNm (decenas de nanomolar), se capturaron moléculas de ARNm en la microesfera.

Después de la lisis y la hibridación, todas las microesferas se recogieron de la matriz de pocillos en un tubo usando un imán. Desde este punto en adelante, todas las reacciones se realizaron en un único tubo. La síntesis de ADNc se realizó en las microesferas usando protocolos convencionales (métodos). Las moléculas de ADNc derivadas de cada célula se adhirieron covalentemente a su microesfera correspondiente, cada una marcada en el extremo 5' con un marcador celular y un marcador molecular. Se realizaron PCR combinadas con cebadores internos para amplificar genes de interés (**FIG. 55**). Como el ARNm de cada célula se había copiado en una microesfera como

ADNc, las microesferas pueden amplificarse y analizarse repetidamente, por ejemplo, para un conjunto diferente de genes.

5 La secuenciación de los amplicones reveló el marcador celular, el marcador molecular y la identidad génica (**FIG. 55**). El análisis informático agrupó las lecturas basándose en el marcador celular y colapsó las lecturas con el mismo marcador molecular y secuencia génica en una única entrada para suprimir cualquier sesgo de amplificación. El uso de marcador molecular permitió medir el número absoluto de moléculas por gen por célula y, por lo tanto, permitió la comparación directa de niveles de expresión celular entre muestras biológicas que pueden haber experimentado diferentes profundidades de secuenciación.

Identificación de distintos tipos celulares en mezclas de células controladas

10 Para medir la capacidad del método de separar dos tipos celulares, se cargó una mezcla ~1:1 de células K562 y Ramos en la matriz de pocillos con 10 000 pocillos. Se usaron aproximadamente 6000 células para capturar ~1000 células. Se seleccionó un panel de 12 genes y se amplificó a partir de las microesferas. El panel consiste en 5 genes específicos para células K562 (leucemia mielógena), 6 genes específicos para células Ramos (linfoma folicular) y el gen constitutivo GAPDH (**tabla 18**). Con aproximadamente 1000 células capturadas en una matriz de 10 000, cada uno con una microesfera única, únicamente un 10 % de las microesferas debe portar ARNm y una en teoría se debe observar únicamente un máximo de 1000 marcadores celulares únicos en los datos de secuenciación. De hecho, se encontraron 768 marcadores celulares que estaban asociados con un número significativo de lecturas después del filtrado de los datos (véanse los métodos para los criterios de filtrado). Como comparación, se realizó lisis celular a granel y captura de ARNm en un tubo de microcentrífuga con números similares de células y microesferas, y se observó un gran número de marcadores celulares con principalmente una lectura únicamente asociada con cada marcador celular. Esto demuestra que la matriz de pocillos era eficaz para confinar la hibridación de ARNm de una célula individual a la microesfera en el mismo pocillo.

25 El perfil de expresión génica de cada una de las 768 células individuales se agrupó usando análisis de componentes principales (PCA) (**FIG. 31A**). El primer componente principal (PC) separó claramente las células individuales en dos grupos principales basados en el tipo celular. Los genes que contribuían al lado positivo del primer componente principal eran los que son específicos de Ramos, mientras que los genes que contribuían al lado negativo del mismo componente principal eran los que son específicos para K562. Esta agrupación satisfactoria de las células en grupos basados en su expresión específica mostró que la contaminación entre pocillos, si la hay, era insignificante. El segundo componente principal resaltó el alto grado de variabilidad en hemoglobina fetal (HBG1) dentro de las células K562, que se había observado previamente (7).

35

40

45

ES 2 711 168 T3

Tabla 18		
Gen	Cebador externo	Cebador interno con secuencia flanqueante 5' común
CD41	CCCCTGGAAGAAGATGATG A	CAGACGTGTGCTCTTCCGATCTTTCTCCAACAAGTTGCCTC C
GYPD	GAGGAAATGAAGCCAAACA CA	CAGACGTGTGCTCTTCCGATCTAATCGTGACCTTAAAGGC CC
GATA 1	TTAGCCACCTCATGCCTTT C	CAGACGTGTGCTCTTCCGATCTCTACTGTGGTGGCTCCGC T
GATA 2	GGAGGAGGATTGTGCTGA TG	CAGACGTGTGCTCTTCCGATCTGTGTCCGCATAAGAAAAAGAA TC
HBG1	GCAAGAAGGTGCTGACTTC C	CAGACGTGTGCTCTTCCGATCTCTGCATGTGGATCCTGAG AA
CD27	CTGCAGTCCCATCCTTTG T	CAGACGTGTGCTCTTCCGATCTGATGAGGTGGAGAGTGGG AA
IGJ	GGACATAACAGACTTGGAA GCA	CAGACGTGTGCTCTTCCGATCTCAATCCATTTTGTAACTGAACC TT
TCL1 A	AAGCCTCTGGGTCAGTGGT	CAGACGTGTGCTCTTCCGATCTTGAAAAGGGATAGAGGT TGG
CD74	TAGACAGATCCCCGTTCT G	CAGACGTGTGCTCTTCCGATCTACAGGGAGAAGGGATAAC CC
SEPT 9	CAGCATCCCAGCCTTGAG	CAGACGTGTGCTCTTCCGATCTCCTCAATGGCCTTTTGCTA C
CD79 a	CCTCTAAACTGCCCCACCT C	CAGACGTGTGCTCTTCCGATCTCCTTAATCGCTGCCTCTAG G
GAPD H	CACATGGCCUCCAAGGAGU AA	CAGACGTGTGCTCTTCCGATCTCAGCAAGAGCACAAGAGG AA

En otro experimento, se añadieron células Ramos (linfoma de Burkitt) a un pequeño porcentaje a linfocitos B primarios de un individuo sano. Se diseñó un panel de 111 genes (**tabla 22**) para que representaran diferentes estados de linfocitos B. Se analizaron 1198 células individuales. Se encontró que un pequeño grupo de la población, que constituye 18 células individuales (~1,5 % de la población) tiene un patrón de expresión génica distinto en comparación con el resto (**FIG. 31B**). Se sabe que los genes que se expresaban preferentemente por este grupo están asociados con linfoma de Burkitt, tal como MYC e IgM, así como marcadores de diferenciación de linfocitos B (CD10, CD20, CD22, BCL6) que se expresan específicamente por linfocitos B foliculares, que son el subconjunto de linfocitos B que origina el linfoma de Burkitt (**FIG. 31C y 31D**). Además, este grupo de células portaba un mayor nivel de CCND3 y GAPDH, así como un contenido global mayor de ARNm, determinado por el número total de moléculas ARNm únicas detectadas basándose en el análisis de índices moleculares (**FIG. 31B**). Este hallazgo fue coherente con el hecho de que las células de linfoma son físicamente más grandes que los linfocitos B primarios en individuos normales, y que son de rápida proliferación y producen una cantidad mayor de transcritos.

Identificación simultánea de múltiples tipos celulares en PBMC humanas

Aunque los experimentos controlados implicaban mezclas artificiales de dos tipos distintos de células, la mayoría de las muestras biológicas de origen natural contienen diversas poblaciones con numerosos tipos celulares y estados con diferencias más sutiles en el perfil de expresión génica. Un ejemplo prominente es la sangre. Se realizó un experimento en que el objetivo era identificar simultáneamente todos los tipos celulares principales en células mononucleares de sangre periférica (PBMC) humanas, incluyendo monocitos, linfocitos NK y los diferentes subconjuntos de linfocitos T y B, midiendo el perfil de expresión de un panel de 98 genes (**tabla 19**) que son específicos para cada tipo celular principal. A diferencia del inmunofenotipado tradicional que se limita principalmente a los marcadores proteínicos de superficie, se incluyeron genes que codifican citocinas, factores de transcripción y proteínas intracelulares de diversas funciones celulares además de proteínas superficiales. Se analizó con PCA el perfil de expresión génica digital de 632 PBMC individuales usando 81 genes presentes (**FIG. 32-39**). El primer componente principal separó claramente los monocitos y los linfocitos en dos grupos ortogonales, como se evidencia por la expresión de CD16a, CD14, S100A12 y CCR2 en un grupo, y genes asociados a linfocitos en el otro. Los diferentes subtipos de linfocitos recaen en un continuo a lo largo del segundo componente principal, con linfocitos B (que expresan IgM, IgD, TCL1A, CD20, CD24, PAX5) en un extremo, linfocitos T vírgenes (que expresan CD4, CCR7, CD62L) en el centro y linfocitos T citotóxicos (que expresan CD8A, CD8B, EOMES, PRF1) en el otro extremo. Los linfocitos citolíticos naturales que expresan el receptor de inmunoglobulina de tipo citolítico, CD16a, y perforina (PRF1) recaen en el espacio entre los monocitos y los linfocitos T citotóxicos. También se observó que GAPDH, un indicador de metabolismo celular, se expresaba a los niveles más altos en monocitos y los más bajos en linfocitos B, que supuestamente están principalmente en reposo. El análisis de correlación del perfil de expresión génica entre las células reiteró las observaciones con PCA y reveló subconjuntos más pequeños adicionales de células dentro de cada tipo celular principal (**FIG. 40A-B**). Un experimento replicado de la misma muestra de PBMC con 731 células produjo una segregación y frecuencia de tipos celulares en gran medida similares (**FIG. 41**).

40

45

50

ES 2 711 168 T3

Tabla 22		
Gen	Cebador externo	Cebador interno con secuencia flanqueante 5' común
CD19	GCAGGGTCCCAGTCCTATG	CAGACGTGTGCTCTTCCGATCTCCAATCATGAGGAAGAT GCA
CD27	TCCAGGAGGATTACCGAAAA	CAGACGTGTGCTCTTCCGATCTCCATCCAAGGGAGAGT GAGA
CD138	AATGGCAAAGGAAGGTGGAT	CAGACGTGTGCTCTTCCGATCTGCAGACACCTTGGACAT CCT
CD38	AGATCTGAGCCAGTCGCTGT	CAGACGTGTGCTCTTCCGATCTTGGTGCAGAGCTGAAG ATTTT
CD24	AAAAGTGGGCTTGATTCTGC	CAGACGTGTGCTCTTCCGATCTTTTTGTTGCGATGGTCA CAC
CD10	ATATTCCTTTGGGCCTCTGC	CAGACGTGTGCTCTTCCGATCTTCAAGTTTGGGTCTGTG CTG
CD95	CCCCCGAAAATGTTCAATAA	CAGACGTGTGCTCTTCCGATCTTGCTCTTGTCATACCCC CA
CD21	TAGCTTCCTCCTCTGGTGGT	CAGACGTGTGCTCTTCCGATCTTTTGCCTTTCCATAATCA CTCA
CXCR3	CTGGCTCTCCCAATATCCT	CAGACGTGTGCTCTTCCGATCTGCTCTGAGGACTGCAC CATT
CD40	GTGGTGTGGGGTATGGTTT	CAGACGTGTGCTCTTCCGATCTATACACAGATGCCCAT GCA
CD69	AGACAGGTCCTTTTCGATGG	CAGACGTGTGCTCTTCCGATCTTGCAATATGTGATGT GGC
CD1c	TTGAGACAGGCACATACAGC TT	CAGACGTGTGCTCTTCCGATCTTTGCTTCTCAATCTGT CCA
IL10	CCCCAACCCTTCA TTCTTG	CAGACGTGTGCTCTTCCGA TCTTTCAA TTCCTCTGGGAA TGTT
IL4R	TGCCTAGAGGTGCTCATTCA	CAGACGTGTGCTCTTCCGATCTGTTGATGCTGGAGGCA GAAT
IL21R	AGCCTGGGTCACAGATCAAG	CAGACGTGTGCTCTTCCGATCTAGGTAGGAGGGTGGAT GGAG
IL6R	CCAGCACCAGGGAGTTTCTA	CAGACGTGTGCTCTTCCGATCTAGGAAAGGATTGGAACA GCA
CXCL1 2	GGGTTTCAGGTTCCAATCAG	CAGACGTGTGCTCTTCCGATCTTTTGTAACTTTTTGCAAG GCA
CCL3	GTGAGGAGTGGTCCAGAA A	CAGACGTGTGCTCTTCCGATCTAGTGGGAGGAGCAGG AG
CCL14	CCATTCCCTTCTTCCTCCTC	CAGACGTGTGCTCTTCCGATCTTACCTACAAGATCCCCG GTC
CCL20	TTGGACATAGCCCAAGAACA	CAGACGTGTGCTCTTCCGATCTTGTGCCTCACTGGACTT GTC
CCL18	ACCTGAAGCTGAATGCCTGA	CAGACGTGTGCTCTTCCGATCTCTGGAGGCCACCTCTTC TAA
TCL1A	GGTAAACACGCCTGCAAAC	CAGACGTGTGCTCTTCCGATCTCAGGACTCAGAAGCCT CTGG

ES 2 711 168 T3

Tabla 22		
Gen	Cebador externo	Cebador interno con secuencia flanqueante 5' común
TACI	CAACAAAGCACAGTGTTAAA TGAA	CAGACGTGTGCTCTTCCGATCTTGTGTCAGCTACTGCCG AAA
AICDA	TGAGCAGA TCCACAGGAAAA	CAGACGTGTGCTCTTCCGA TCTGAAA TGGAGTCTCAAAGCTTCA
FCRL4	TCCCAACTACGCTGATTTGA	CAGACGTGTGCTCTTCCGATCTGACCAAAGGAATGTGT GGG
BCL2	TGCAAGAGTGACAGTGGATT G	CAGACGTGTGCTCTTCCGATCTTCAACCAAGGTTTGCTT TTGT
FASLG	AGAGGCTGAAAGAGGCCAA T	CAGACGTGTGCTCTTCCGATCTAATATGGGTTGCATTTG GTCA
BCL6	AAATCTGCAGAAGGAAAAAT GTG	CAGACGTGTGCTCTTCCGATCTAGTTTTCAATGATGGGC GAG
AURKB	GCTCAAGGGAGAGCTGAAG A	CAGACGTGTGCTCTTCCGATCTGACTACCTGCCCCAG AGAT
CD81	GTGGCGTGTATGAGTGGAG A	CAGACGTGTGCTCTTCCGATCTCACTCGCCCAGAGACT CAG
CD80	GCACATCTCATGGCAGCTAA	CAGACGTGTGCTCTTCCGATCTGCTTCACAAACCTTGCT CCT
CD23a	ACATTTTCTGCCACCCAAAC	CAGACGTGTGCTCTTCCGATCTAACAGCACCCCTCTCCAG ATG
CD44	GCCTGGTAGAATTGGCTTTT C	CAGACGTGTGCTCTTCCGATCTTTTTGTAGCCAACATTC ATTCAA
LEF1	CAATTGGCAGCCCTATTTCA	CAGACGTGTGCTCTTCCGATCTGTTCAGCAGACTGGTTT GCA
CXCR5	CCGTGAGGATGTCACTCAGA	CAGACGTGTGCTCTTCCGATCTACGAGGAAGCCCTAAG ACGT
PRKCB	TTGAGCCTGGGGTGTAAAGAC	CAGACGTGTGCTCTTCCGATCTGTCTTCCAGGATTCACG GTG
PRKC D	GAGCACCTCCTGGAAGATTG	CAGACGTGTGCTCTTCCGATCTTAAGCACCAAGTGGGACT GTG
CD20	TAGGAGCAGGCCTGAGAAA A	CAGACGTGTGCTCTTCCGATCTGATTCCTCTCCAAACCC ATG
CD30	TGTTTTGGGAAAGTTGGAG	CAGACGTGTGCTCTTCCGATCTCTGTTTGCCAGTGTTT GTG
CD30L	TGCAACCCAAGTGTGTGTTA	CAGACGTGTGCTCTTCCGATCTTTTACCAACTGTTCTC TGAGC
BAFFR	GCCCTGAGCAACAATAGCAG	CAGACGTGTGCTCTTCCGATCTTTCAGCTCTTCACTCCA GCA
CMRF- 35H	AGGAAAAGATGTGGCTCACG	CAGACGTGTGCTCTTCCGATCTGGAGTTGGGGAGAACT GTCA
PRDM 1	TCGAATAATCCAGGGAAACC	CAGACGTGTGCTCTTCCGATCTACCAAAGCATCACGTTG ACA
HLA- DRA	GGCTTTACAAAGCTGGCAAT	CAGACGTGTGCTCTTCCGATCTTATGCCTCTTCGATTGC TCC
GNAI2	CCTTGAGTGTGTCTGCGTGT	CAGACGTGTGCTCTTCCGATCTCCACAGAATTGGGTTCC AAG

ES 2 711 168 T3

Tabla 22		
Gen	Cebador externo	Cebador interno con secuencia flanqueante 5' común
RGS1	AACTGGGAAGGCCAGGTAA C	CAGACGTGTGCTCTTCCGATCTTGTTTTCAAATTGCCATT GC
CD5	CTTTCTCCACGCCATTTGAT	CAGACGTGTGCTCTTCCGATCTACTAGGATATGGGGTG GGCT
CD22	GGGATCTGCTCGTCATCATT	CAGACGTGTGCTCTTCCGATCTGTTTCTGCCTCTGAGGG AAA
PIK3C D	GCGTGCGCGTTATTTATTTA	CAGACGTGTGCTCTTCCGATCTTGCTGGGGAAGGCAA GTTA
DOCK8	GCAGTCAGCCAGAAATCACA	CAGACGTGTGCTCTTCCGATCTTTTTCTCCTCTCTGGGA CCA
CD11b	TGAAAAGTCTCCCTTTCCAG A	CAGACGTGTGCTCTTCCGATCTCCTTCAGACAGATTCCA GGC
FCGR2 B	GGAGAGGAGAGATGGGGAT T	CAGACGTGTGCTCTTCCGATCTGAGTGAGTGCCCTTTT CTT
CD72	CTCATGCCAACAAGAACCTG	CAGACGTGTGCTCTTCCGATCTTGACCCACACCTGACAC TTC
BCL11 B	TCGTGGAACACAGGCAAAC	CAGACGTGTGCTCTTCCGATCTTGCATTTGTAAGGCA AGG
CD86	TCAAGGCAACCAGAGGAAAC	CAGACGTGTGCTCTTCCGATCTACTAAGGGATGGGGCA GTCT
TBX21	ACCTTTTCGTTGGCATGTGT	CAGACGTGTGCTCTTCCGATCTTCAGGGAAAGGACTCA CCTG
FOXP1	ATGCTGAAGGCATTTCTTGG	CAGACGTGTGCTCTTCCGATCTCTGTGAGCATGGTGCTT CAT
MCL1	GAGGGGAGTGGTGGGTTTA T	CAGACGTGTGCTCTTCCGATCTCAAAGGGAAAGGGAG GATT
IFNB1	AGGGGAAAACATGAGCA G	CAGACGTGTGCTCTTCCGATCTTCACTGTGCCTGGACCA TAG
BLNK	TTGGGCAGAAAGAAAATGG	CAGACGTGTGCTCTTCCGATCTCAAAGATTCCACCAGA CTGAA
CD40L G	CCTCCCCAGTCTCTCTTCT	CAGACGTGTGCTCTTCCGATCTGAGTCAGGCCGTTGCT AGTC
IGBP1	GGCTGATCTTCCACAACAC	CAGACGTGTGCTCTTCCGATCTACGAGGGCAAAGATGC TAAA
IRF4	ATTCCCGTGTGCTTCAAAC	CAGACGTGTGCTCTTCCGATCTAGAACTGCCAGCAGGT AGGA
CD79a	CACTTCCCTGGGACATTCTC	CAGACGTGTGCTCTTCCGATCTTCACTCTTCTCCAGGC CAG
LTA	TGATGTCTGTCTGGCTGAGG	CAGACGTGTGCTCTTCCGATCTCCACACACAGAGGAAG AGCA
HDAC5	CCAGCCTGT AGGAAACCAA	CAGACGTGTGCTCTTCCGA TCTCTCCTTCT A TCTCCAGGGCC
RAG1	GGATGCAGGTGGTTTTTATG	CAGACGTGTGCTCTTCCGATCTCATTGTACCCATTTTAC ATTTTCTT
RAG2	CAAACCTTAAACACCCAGAA GC	CAGACGTGTGCTCTTCCGATCTATAACAATTCGGCAGTT GGC

ES 2 711 168 T3

Tabla 22		
Gen	Cebador externo	Cebador interno con secuencia flanqueante 5' común
CD1d	GAACCAGTTTCCTCCTGTGC	CAGACGTGTGCTCTTCCGATCTAAGATGTGGAGGCTGTT GCT
TGFB1	GACTGCGGATCTCTGTGTCA	CAGACGTGTGCTCTTCCGATCTTCTGCACTATTCCTTTG CCC
CD9	TCAGTATGATCTTGTGCTGT GCT	CAGACGTGTGCTCTTCCGATCTTACCCATGAAGATTGGT GGG
CD11c	CACAGCATGAGAGGCTCTGT	CAGACGTGTGCTCTTCCGATCTTCTCAGTTCGGATTTCC CAG
FOXP3	TCAGGATCTGAGGTCCCAAC	CAGACGTGTGCTCTTCCGATCTTACCTGTGTATCTCAC GCA
LAG3	AGAGCTGTCTAGCCCAGGT G	CAGACGTGTGCTCTTCCGATCTTGGTGTCTTTCTCTGC TCC
CD73	CTTAACGTGGGAGTGGAACC	CAGACGTGTGCTCTTCCGATCTGTGTGCAAATGGCAGCT AGA
CD70	TCTCAGCTTCCACCAAGGTT	CAGACGTGTGCTCTTCCGATCTTCACTGGGACACTTTTG CCT
CCR7	CAGGGGAGAGTGTGGTGT T	CAGACGTGTGCTCTTCCGATCTGACATGCACTCAGCTCT TGG
CD45R A	TGCATAGTTCCCATGTAAAT CC	CAGACGTGTGCTCTTCCGATCTTACCAGGAATGGATGTC GCT
PDCD1	ACATCCTACGGTCCCAAGGT	CAGACGTGTGCTCTTCCGATCTGCAGAAGTGCAGGCAC CTA
MYC	TGCATGATCAAATGCAACCT	CAGACGTGTGCTCTTCCGATCTTTGGACTTTGGGCATAA AAGA
CD25	AAATCACGGCAGTTTTCAGC	CAGACGTGTGCTCTTCCGATCTCTCATCTGTGCACTCTC CCC
FCAM R	GTGGGAAGAGAAGCTGATG C	CAGACGTGTGCTCTTCCGATCTTCAAGCATTATCCACGT CCA
CCND2	TGTGATGCCATATCAAGTCC A	CAGACGTGTGCTCTTCCGATCTTCAAGTGTATGCGAAAAG GTTTTT
MKI67	AGCCTCTCTTGGGCTTTCTT	CAGACGTGTGCTCTTCCGATCTGTTTTCCCTGCCTGGAA CTT
CCND3	CTTTGCTGCTGAAGGCTCAT	CAGACGTGTGCTCTTCCGATCTACAAGTGGTGGTAACCC TGG
IL12A	TGCTTCCTAAAAAGCGAGGT	CAGACGTGTGCTCTTCCGATCTGAACTAGGGAGGGGGA AAGA
IFNG	GCAGCCAACCTAAGCAAGAT	CAGACGTGTGCTCTTCCGATCTATCCAGTTACTGCCGGT TTG
TNFA	GAATGCTGCAGGACTTGAGA	CAGACGTGTGCTCTTCCGATCTACTTCCTTGAGACACGG AGC
IL2	ACCCAGGGACTTAATCAGCA	CAGACGTGTGCTCTTCCGATCTGCTGATGAGACAGCAA CCATT
IL4	GACATCTTTGCTGCCTCCA	CAGACGTGTGCTCTTCCGATCTATGAGAAGGACACTCG CTGC
IL6	TTAAGGAGTTCCTGCAGTCC A	CAGACGTGTGCTCTTCCGATCTTCCACTGGGCACAGAA CTTA

ES 2 711 168 T3

Tabla 22		
Gen	Cebador externo	Cebador interno con secuencia flanqueante 5' común
BAFF	TCCTTCGCTTTGCTTGTCTT	CAGACGTGTGCTCTTCCGATCTAGGTGGAAAAATAGATGCCAGTC
IGHE	CCCGGAAGTCTATGCGTTT	CAGACGTGTGCTCTTCCGATCTAGGACATCTCGGTGCA GTG
IGHD	TGTGTGAGGTGTCTGGCTTC	CAGACGTGTGCTCTTCCGATCTAGGAGCACCACGTTCT GG
IGHM	CCCGGAGAAGTATGTGACCA	CAGACGTGTGCTCTTCCGATCTGTA CTTCGCCACAGCA TC
IGHA	CTGAACGAGCTGGTGACG	CAGACGTGTGCTCTTCCGATCTAGTACCTGACTTGGGCA TCC
IGHG1	CAAGGGCCCATCGGTCTT	CAGACGTGTGCTCTTCCGATCTTTGTGACAAA ACTCACACATGC
IGHG4	CAAGGGCCCA TCGGTCTT	CAGACGTGTGCTCTTCCGA TCTCAA TA TGGTCCCCCA TGC
IGHG2	CAAGGGCCCATCGGTCTT	CAGACGTGTGCTCTTCCGATCTGCAAATGTTGTGTGCGAG TGC
IGHG3	CAAGGGCCCATCGGTCTT	CAGACGTGTGCTCTTCCGATCTACCCCACTTGGTGACAC AAC
TLR1	CCATTCCGCAGTACTCCATT	CAGACGTGTGCTCTTCCGATCTAAGGAAAAGAGCAAAC GTGG
TLR2	TTGGTTGACTTCATGGATGC	CAGACGTGTGCTCTTCCGATCTGGAAACAGCACAAATGA ACTTAA
TLR3	CATCATGCAGTTCAACAAGC	CAGACGTGTGCTCTTCCGATCTATGCACTCTGTTTGCGA AGA
TLR4	GGGTGTGTTTCCATGTCTCA	CAGACGTGTGCTCTTCCGATCTTTGAAAGTGTGTGTGTC CGC
TLR5	TCAGGCTGTTGCATGAAGAA	CAGACGTGTGCTCTTCCGATCTGTATGCCCTTGCTGGAC CTA
TLR6	ATGCGCAGTAAAACTCGTG	CAGACGTGTGCTCTTCCGATCTTACAGTTCCACGCTGAG CTG
TLR7	GCCTGTACTTTCAGCTGGGT A	CAGACGTGTGCTCTTCCGATCTAAGGTGTTTGTGCCATT TGG
TLR8	GGTGAGCTCTGATTGCTTCA	CAGACGTGTGCTCTTCCGATCTTATCAGGAGGCAGGGA TCAC
TLR9	GACCGGGTCAGTGGTCTCT	CAGACGTGTGCTCTTCCGATCTGGTGATCCTGAGCCCT GAC
TLR10	TGCAGTGAGCTGAGATCGA G	CAGACGTGTGCTCTTCCGATCTATGGAAAACATCCTCAT GGC
GAPDH	CACATGGCCUCCAAGGAGU AA	CAGACGTGTGCTCTTCCGATCTCAGCAAGAGCACAAAGA GGAA

ES 2 711 168 T3

Tabla 19		
Gen	Cebador externo	Cebador interno con secuencia flanqueante 5' común
CD19	GCAGGGTCCCAGTCCTATG	CAGACGTGTGCTCTTCCGATCTCCAATCATGAGGAA GATGCA
CD20	TAGGAGCAGGCCTGAGAA AA	CAGACGTGTGCTCTTCCGATCTGATTCTCTCCAAAC CCATG
BAFF	TCCTTCGCTTTGCTTGTCTT	CAGACGTGTGCTCTTCCGATCTAGGTGGAAAAATAG ATGCCAGTC
TCL1A	GGTAAACACGCCTGCAAAC	CAGACGTGTGCTCTTCCGATCTCAGGACTCAGAAGC CTCTGG
TACI	CAACAAAGCACAGTGTTAA ATGAA	CAGACGTGTGCTCTTCCGATCTTGTGTGCTCAGCTACTGC GGAAA
IGHD	TGTGTGAGGTGTCTGGCTT C	CAGACGTGTGCTCTTCCGATCTAGGAGCACCACGTT CTGG
IGHM	CCCGGAGAAGTATGTGACC A	CAGACGTGTGCTCTTCCGATCTGTACTTCGCCACA GCATC
CD27	TCCAGGAGGATTACCGAAA A	CAGACGTGTGCTCTTCCGATCTCCATCCAAGGGAGA GTGAGA
CD38	AGATCTGAGCCAGTCGCTG T	CAGACGTGTGCTCTTCCGATCTTGGTGCAGAGCTGA AGATTTT
CD24	AAAAGTGGGCTTGATTCTG C	CAGACGTGTGCTCTTCCGATCTTTTTGTTCGCATGGT CACAC
AICDA	TGAGCAGA TCCACAGGAAAA	CAGACGTGTGCTCTTCCGA TCTGAAA TGGAGTCTCAAAGCTTCA
CD95	CCCCCGAAAATGTTCAATA A	CAGACGTGTGCTCTTCCGATCTTGTCTTGTGCATACC CCCA
CD10	ATATTCTTTGGGCCTCTG C	CAGACGTGTGCTCTTCCGATCTTCAAGTTTGGGTCTG TGCTG
IL10	CCCCAACCATTCA TTCTTG	CAGACGTGTGCTCTTCCGA TCTTTCAA TTCCTCTGGGAA TGTT
CD138	AATGGCAAAGGAAGGTGG AT	CAGACGTGTGCTCTTCCGATCTGCAGACACCTTGGA CATCCT
CD45RA	TGCATAGTTCCCATGTTAA ATCC	CAGACGTGTGCTCTTCCGATCTTACCAGGAATGGAT GTCGCT
BCL6	AAATCTGCAGAAGGAAAA TGTG	CAGACGTGTGCTCTTCCGATCTAGTTTTCAATGATGG GCGAG
PRDM1	TCGAATAATCCAGGGAAAC C	CAGACGTGTGCTCTTCCGATCTACCAAAGCATCACGT TGACA
CXCR3	CTGGCTCTCCCAATATCC T	CAGACGTGTGCTCTTCCGATCTGCTCTGAGGACTGC ACCATT
IFNG	GCAGCCAACCTAAGCAAGA T	CAGACGTGTGCTCTTCCGATCTATCCAGTTACTGCCG GTTTG
IL4R	TGCCTAGAGGTGCTCATT A	CAGACGTGTGCTCTTCCGATCTGTTGATGCTGGAGG CAGAAT
IL4	GACATCTTTGCTGCCTCCA	CAGACGTGTGCTCTTCCGATCTATGAGAAGGACACT CGCTGC
CCL20	TTGGACATAGCCCAAGAAC A	CAGACGTGTGCTCTTCCGATCTTGTGCCTCACTGGA CTTGTC

ES 2 711 168 T3

Tabla 19		
Gen	Cebador externo	Cebador interno con secuencia flanqueante 5' común
CD25	AAATCACGGCAGTTTTCCAG C	CAGACGTGTGCTCTTCCGATCTCTCATCTGTGCACTC TCCCC
FOXP1	ATGCTGAAGGCATTTCTTG G	CAGACGTGTGCTCTTCCGATCTCTGTGAGCATGGTG CTTCAT
TGFB1	GACTGCGGATCTCTGTGTG A	CAGACGTGTGCTCTTCCGATCTTCTGCACTATTCCCT TGCCC
CXCR5	CCGTGAGGATGTCACTCAG A	CAGACGTGTGCTCTTCCGATCTACGAGGAAGCCCTA AGACGT
CD69	AGACAGGTCTTTTCGATG G	CAGACGTGTGCTCTTCCGATCTTGTGCAATATGTGAT GTGGC
CD30	TGTTTTGGGGAAAGTTGGA G	CAGACGTGTGCTCTTCCGATCTCTGTTTGCCAGTGT TTGTG
PDCD1	ACATCCTACGGTCCCAAGG T	CAGACGTGTGCTCTTCCGATCTGCAGAAGTGCAGGC ACCTA
LAG3	AGAGCTGTCTAGCCCAGGT G	CAGACGTGTGCTCTTCCGATCTTGGTGTCTTTCTCT GCTCC
PAX5	TGACGTGTGTTGCTTTTGT G	CAGACGTGTGCTCTTCCGATCTACTTGGGAGAAAAC AGGGGT
TNFRSF17	GCTTTCCACTCCCAGCTAT G	CAGACGTGTGCTCTTCCGATCTTGGCTTTGAGTGCTAC GGAGA
RASD1	GGGGGAGGGA TGTGAAGTT A	CAGACGTGTGCTCTTCCGA TCTA TCTTGTCTGTGA TTCCGGG
AMPD1	ACAGATGACCCAATGCAAT TC	CAGACGTGTGCTCTTCCGATCTGAGCACCTGTGATAT GTGCG
OSBPL5	AGACCGATGCACAGTCTTC C	CAGACGTGTGCTCTTCCGATCTCTTCACGTCTGGCCT CAGTC
CD56	GGAGCACTCAAGTGTGAC GA	CAGACGTGTGCTCTTCCGATCTTTTTCTATGGAGCCT TCCGA
IGFBP7	CATCCAATTCCCAAGGACA G	CAGACGTGTGCTCTTCCGATCTGGTGAAGGTGCCGA GCTATA
KIR2DS5	GCTCTTCTCAAACCACGA A	CAGACGTGTGCTCTTCCGATCTCACACTCCTTTGCTT AGCCC
KIR2DS2	TCCTCACACCACGAATCTG A	CAGACGTGTGCTCTTCCGATCTCACTCCTTTGCTTAG CCCAC
RAB4B	CCAGCTCACCTGTTCTCCA G	CAGACGTGTGCTCTTCCGATCTGAATCCCGTACCTG CTGCT
CD14	CTAAAGGACTGCCAGCCAA G	CAGACGTGTGCTCTTCCGATCTATAACCTGACACTGG ACGGG
S100A12	CACATTCCTGTGCATTGAG G	CAGACGTGTGCTCTTCCGATCTATACTCAGTTCGGAA GGGGC
CCR2varian teB	GAAGGAGGGAGACATGAG CA	CAGACGTGTGCTCTTCCGATCTACTGGTCCTTAGCC CCATCT
CD62L	TCAGTTGGCTGACTTCCAC A	CAGACGTGTGCTCTTCCGATCTTTAGTTTGGGGGTTT TGCTG
CD16	TCTTGGCCAGGGTAGTAAG AA	CAGACGTGTGCTCTTCCGATCTGTGAGTTCCAATGAG GTGGG

ES 2 711 168 T3

Tabla 19		
Gen	Cebador externo	Cebador interno con secuencia flanqueante 5' común
CX3CR1	CGTCCAGACCTTGTTCACAC	CAGACGTGTGCTCTTCCGATCTCCACAAATAGTGCTCGCTTTC
CD1b	TAGAGGGCCAGGACATCATC	CAGACGTGTGCTCTTCCGATCTTTGCTCCTTTTGCTATGTCCT
FOXQ1	TGCTATTGACCGATGCTTCA	CAGACGTGTGCTCTTCCGATCTGCAACGGGCTACAGCTTTAT
CD209	GCTCTTGTTCTTGCCGTTTT	CAGACGTGTGCTCTTCCGATCTGAGTCCCTCAGTGGAGCAAG
CD1e	CACAAGCACATTCATCTCTTCC	CAGACGTGTGCTCTTCCGATCTATTCAGGGCCAGCTTCATAA
CCL17	TACTTCAAGGGAGCCATTCAC	CAGACGTGTGCTCTTCCGATCTTTTGTAACTGTGCAGGGCAG
DTNA	AGCAACGTGGAGTCAGTCTGT	CAGACGTGTGCTCTTCCGATCTCTCACCTTCTCTTGCCTTGG
CLEC4C	TTATTTTCTGGGGCTGTCAAGA	CAGACGTGTGCTCTTCCGATCTCATTCTGGCACTCAGGTGAA
ZBTB16	TGATCAAGCACCTGAGAACG	CAGACGTGTGCTCTTCCGATCTTACCAGTGCACCATCTGCAC
SLAMF1	TGCAAAACCCAGAAGCTAAAA	CAGACGTGTGCTCTTCCGATCTGTTCTGTGCAAATGGCATTTC
CD3D	AGAGCTGTGTGGAGCTGGAT	CAGACGTGTGCTCTTCCGATCTGGAGTCTTCTGCTTTGCTGG
CD3E	GCCCTCTTGCCAGGATATTT	CAGACGTGTGCTCTTCCGATCTGCATGTAAGTTGTCCCCCAT
CD8A	CTGGCCTCTGCTCAACTAGAC	CAGACGTGTGCTCTTCCGATCTATGGTACAAGCAATGCCTGC
CD8B	CAGCCTCAAGGGGAAGGTAT	CAGACGTGTGCTCTTCCGATCTTGCTTAACCCATGGATCCTG
PRF1	CCCTGCAGTCACAGCTACAC	CAGACGTGTGCTCTTCCGATCTTCAGGGCTGGTCTTTTAGGA
EOMES	TGGGATAATGTAAAAGTGGTGCT	CAGACGTGTGCTCTTCCGATCTCATCCCCATGATATTGGGA
CD4	AGCTAGCCTGAGAGGGAAAC	CAGACGTGTGCTCTTCCGATCTTCTCCAGACCATTCAGGAC
THPOK	GGCTCTGCCTTGCACTATTT	CAGACGTGTGCTCTTCCGATCTCTTCTCCTCCCTTCCATGC
RUNX3	TAAGGCCCAAAGTGGGTACA	CAGACGTGTGCTCTTCCGATCTTAGGAAGCACGAGGAAAGGA
CD45RO	ACCCTCTCTCCCTCCCTTTCC	CAGACGTGTGCTCTTCCGATCTTAGTTGGCTATGCTGGCATG
CD44	GCCTGGTAGAATTGGCTTTTC	CAGACGTGTGCTCTTCCGATCTTTTTGTAGCCAACATTCATTCAA
CCR7	CAGGGGAGAGTGTGGTGTTT	CAGACGTGTGCTCTTCCGATCTACTCAGCTCTTGGCTCCTACT
TXK	ACATCAAGCTCCATTGTTTCG	CAGACGTGTGCTCTTCCGATCTTTTGCCTGCACTCTTTGTAGG

ES 2 711 168 T3

Tabla 19		
Gen	Cebador externo	Cebador interno con secuencia flanqueante 5' común
MBD2	GCCTGGCACGTAATAGCTT G	CAGACGTGTGCTCTTCCGATCTAGGAAAGAAATGCC CTTGGT
IFNGR1	GAGGATGTGTGGCATTTC A	CAGACGTGTGCTCTTCCGATCTGGTTCTAGGTGAG CAGGTG
IL12RB2	AGCAGGCTGTACACAGCA GA	CAGACGTGTGCTCTTCCGATCTGACACTAGGCACATT GGCTG
IL33R	ACTGTGCCCTCATCCAGAA C	CAGACGTGTGCTCTTCCGATCTAACGACGCCAAGGT GATACT
CCR4	TGGTGAATGCAGAGTCAA TG	CAGACGTGTGCTCTTCCGATCTTCAGGAGGAAGGCT TACACC
CRTN2	TGAATTTTGCTTGGTGGAT G	CAGACGTGTGCTCTTCCGATCTTGTCAGTGAAGAA GCAGATG
IL5	CAGTGAGAATGAGGGCCA AG	CAGACGTGTGCTCTTCCGATCTGAATGAGGGCCAAG AAAGAG
IL17A	AAAATGAAACCCTCCCAA A	CAGACGTGTGCTCTTCCGATCTTCCTTTGGAGATTAA GGCCC
IL17F	CTGCA TCAA TGCTCAAGGAA	CAGACGTGTGCTCTTCCGA TCTCCAAGGCTGCTCTGTTTCTT
IL21	AAA TCAAGCTCCCAAGGTCA	CAGACGTGTGCTCTTCCGA TCTTGTGAA TGACTTGGTCCCTG
IL22	ATGCCCAAAGCGATTTTT	CAGACGTGTGCTCTTCCGATCTCAAAGGAAACCAAT GCCACT
IL23R	TCCCTCATTGAAAGATGCA A	CAGACGTGTGCTCTTCCGATCTTAGAATCATTAGGCC AGGCG
RORA	TGCAAGCCATTTATGGGAA T	CAGACGTGTGCTCTTCCGATCTCCTTGGGTTTTCTTT TCAATTC
RORC	ATTTCCATGGTGTCCAGT C	CAGACGTGTGCTCTTCCGATCTAGAGAAGCAGAAGT CGCTCG
OX40L	CTGCTGGCCCTGTACCTG	CAGACGTGTGCTCTTCCGATCTCTCCACCCTGGCCA AGAT
ICOS	TTCAGCTGACTTGGACAAC CT	CAGACGTGTGCTCTTCCGATCTGGACAACCTGACTG GCTTTG
SH2D1A	GGGTGTTGGTGAACCTGGT T	CAGACGTGTGCTCTTCCGATCTTTAATATGGATGCC GTGGG
CCR2varian teA	GTTGCCCAGTGTGTTTCTG A	CAGACGTGTGCTCTTCCGATCTAACCAGGCAACTTG GGAACT
TLR1	CCATTCCGCAGTACTCCAT T	CAGACGTGTGCTCTTCCGATCTAAGGAAAAGAGCAA ACGTGG
TLR2	TTGGTTGACTTCATGGATG C	CAGACGTGTGCTCTTCCGATCTGGAAACAGCACAAA TGAACCTAA
TLR3	CATCATGCAGTTCAACAAG C	CAGACGTGTGCTCTTCCGATCTATGCACTCTGTTTGC GAAGA
TLR4	GGGTGTGTTTCCATGTCTC A	CAGACGTGTGCTCTTCCGATCTTTGAAAGTGTGTGTG TCCGC
TLR5	TCAGGCTGTTGCATGAAGA A	CAGACGTGTGCTCTTCCGATCTGTATGCCCTTGCTG GACCTA

ES 2 711 168 T3

Tabla 19		
Gen	Cebador externo	Cebador interno con secuencia flanqueante 5' común
TLR6	ATGCGCAGTAAAACTCGT G	CAGACGTGTGCTCTTCCGATCTTACAGTTCCACGCTG AGCTG
TLR7	GCCTGTACTTTCAGCTGGG TA	CAGACGTGTGCTCTTCCGATCTAAGGTGTTTGTGCCA TTTGG
TLR8	GGTGAGCTCTGATTGCTTC A	CAGACGTGTGCTCTTCCGATCTTATCAGGAGGCAGG GATCAC
TLR9	GACCGGGTCAGTGGTCTCT	CAGACGTGTGCTCTTCCGATCTGGTGATCCTGAGCC CTGAC
TLR10	TGCAGTGAGCTGAGATCGA G	CAGACGTGTGCTCTTCCGATCTATGGAAAACATCCTC ATGGC
GAPDH	CACATGGCCUCCAAGGAG UAA	CAGACGTGTGCTCTTCCGATCTCAGCAAGAGCACAA GAGGAA

Estudio de la diversidad de la respuesta de linfocitos T humanos a estímulo in vitro

Cuando se examina el patrón de expresión génica de una muestra masiva, al patrón observado contribuía tanto la composición celular de la muestra como el nivel de expresión de gen en cada tipo o subtipo celular. Estos dos efectos no pueden desconvolucionarse por análisis masivo, sino únicamente con análisis de células individuales a gran escala. Para ilustrarlo, se utilizó nuestra plataforma para estudiar la variabilidad de respuesta de linfocitos T humanos a un estímulo *in vitro*.

Se purificaron linfocitos T CD3+ por selección negativa de un donador de sangre y se estimularon con microesferas anti-CD28/anti-CD3 durante 6 horas, y se realizaron experimentos con una alícuota estimulada y una diferente de células no estimuladas. Se diseñó un panel de 93 genes (**tabla 20**) que abarcaba proteínas superficiales, citocinas, quimiocinas y moléculas efectoras expresadas por los diferentes subconjuntos de linfocitos T. Se analizó un total de 3517 y 1478 células individuales para las muestras estimuladas y no estimuladas, respectivamente.

5

10

15

20

25

30

35

40

ES 2 711 168 T3

Tabla 20		
Gen	Cebador externo	Cebador interno con secuencia flanqueante 5' común
GAPDH	GACTTCAACAGCGACACCC A	CAGACGTGTGCTCTTCCGATCTGCCCTCAACGACCACTT TGT
CD3D	GAAAACGCATCCTGGACC CA	CAGACGTGTGCTCTTCCGATCTTGATGTCATTGCCACTCT GCT
CD3E	AAGTTGTCCCCATCCCAA A	CAGACGTGTGCTCTTCCGATCTCTGGGGATGGACTGGGT AAAT
CD8A	ACTGCTGTCCCAAACATGC A	CAGACGTGTGCTCTTCCGATCTATGCCTGCCATTGGAG AGAA
CD8B	CCACCATCTTTGCAGGTTG C	CAGACGTGTGCTCTTCCGATCTGCTGTCCAGTTCCCA AGG
CD4	CTGGGAGAGGGGGTAGCT AG	CAGACGTGTGCTCTTCCGATCTACCACTTCCCTCAGTCC CAA
FOXP3	ACAGAAGCAGCGTCAGTA CC	CAGACGTGTGCTCTTCCGATCTGGGTCTCTTGAGTCCCG TG
CCR7	GGGGAGAGTGTGGTGT CC	CAGACGTGTGCTCTTCCGATCTCTTGGCTCCACTGGG ATG
CD5	ATCAATGGTCCAAGCCGCA T	CAGACGTGTGCTCTTCCGATCTAGGTACAGATCTTCCC CCG
IL32	CTTCCAGTCTACGGAGC C	CAGACGTGTGCTCTTCCGATCTTGCTCTGAACCCCAATC CTC
CD28	ACCATCACAGGCATGTTCC T	CAGACGTGTGCTCTTCCGATCTTGATAGATGACCTGGCTT GCC
SELL	GCATCTCATGAGTGCCAAG C	CAGACGTGTGCTCTTCCGATCTCCTGCCCCAGACCTTT TATC
CD27	TGCAGAGCCTTGTCGTTAC A	CAGACGTGTGCTCTTCCGATCTCGTGACAGAGTGCCTTT TCG
GZMB	AGGTGAAGATGACAGTGC AGG	CAGACGTGTGCTCTTCCGATCTAGGCCCTCTTGTGTGTA ACA
GZMA	GGAACCATGTGCCAAGTTG C	CAGACGTGTGCTCTTCCGATCTCCTTTGTTGTGCGAGGG TGT
GZMH	AGTGTGCTGACAGTGCAG A	CAGACGTGTGCTCTTCCGATCTCAAAGAAGACACAGAC CGGT
GZMK	TTGCCACAAAGCCTGGAAT C	CAGACGTGTGCTCTTCCGATCTAAAGCAACCTTGTCCTG CCT
PRF1	GGAGTCCAGCGAATGACG TC	CAGACGTGTGCTCTTCCGATCTCATGGCCACGTTGTCAT TGT
NKG2D	CAACACCCAGGGGATCAG TG	CAGACGTGTGCTCTTCCGATCTCCACCCTCCACAGGAAA TTG
LAG3	AGCTGTACCAGGGGGAGA G	CAGACGTGTGCTCTTCCGATCTCTTTGGAGAAGACAGTG GCGA
CD160	GGAAGACAGCCAGATCCA GTG	CAGACGTGTGCTCTTCCGATCTTTGTGCAGACCAAGAGC ACC
CD244	GGGCTGAGAATGAGGCAG TT	CAGACGTGTGCTCTTCCGATCTGGAAAGCGACAAGGGTG AAC
EOMES	ACTTAACAGCTGCAGGGG C	CAGACGTGTGCTCTTCCGATCTACTAACTTGAACCGTGT TAAGG

ES 2 711 168 T3

TBX21	TTATAACCATCAGCCCGCC A	CAGACGTGTGCTCTTCCGATCTAGAAAAGGGGCTGGAAA GGG
PRDM1	ACCAAAGCATCACGTTGAC AT	CAGACGTGTGCTCTTCCGATCTACATGTGAATGTTGAGC CCA
IRF4	CTCTTCAGCATCCCCCGTA C	CAGACGTGTGCTCTTCCGATCTGCCCCAAATGAAAGCT TGA
ZNF683	GGAGAGCGTCCATTCCAG TG	CAGACGTGTGCTCTTCCGATCTATCCACCTGAAGCTGCA CC
ZBED2	AATGTACCAGCCAGTCAGC G	CAGACGTGTGCTCTTCCGATCTGGTTTTGGTGGAGCTGA CGA
CD30	TTTACTCATCGGGCAGCCA C	CAGACGTGTGCTCTTCCGATCTTGTGGCCAGTGTGGT GC
CD69	GCTGTAGACAGGTCCTTTT CG	CAGACGTGTGCTCTTCCGA TCTAGTGTGGAAAA TGTGCAA TA TGTG
HLA-DRA	GGGTCTGGTGGGCATCAT TA	CAGACGTGTGCTCTTCCGATCTGCCTCTTCGATTGCTCC GTA
CD38	AGGTCAATGCCAGAGACG GA	CAGACGTGTGCTCTTCCGATCTATCAGCATACCTTTATTG TGATCTATC
TNFRSF9	TGGCATGTGAGTCATTGCT C	CAGACGTGTGCTCTTCCGATCTTTTGTGTGAGGGGCG GAT
MKI67	TACTTTTTGCCTCCCAGG G	CAGACGTGTGCTCTTCCGATCTTCTGCCCCACCAAGAT CAT
BIRC5	TGCCACGGCCTTTCCTTAA A	CAGACGTGTGCTCTTCCGATCTTGTCTAAGTGCAACCG CCT
FOSL1	CTCCTGACAGAAGGTGCC AC	CAGACGTGTGCTCTTCCGATCTGGTGATTGGACCAGGCC ATT
MCL1	GACTGGCTACGTAGTTG GG	CAGACGTGTGCTCTTCCGATCTTTGCTTAGAAGGATGG CGC
MYC	AGCTACGGAACTCTTGTGC G	CAGACGTGTGCTCTTCCGATCTCAACCTTGGCTGAGTCT TGA
TYMS	TCAGTCTTTAGGGGTTGGG C	CAGACGTGTGCTCTTCCGATCTATGTGCATTTCAATCCCA CGTAC
CDCA7	CCAGTCTAGTTTCTGGGCA GG	CAGACGTGTGCTCTTCCGATCTATGTAAACCATTGCTGTG CCATT
UHRF1	CCAGTTCTTCTGACACCG G	CAGACGTGTGCTCTTCCGATCTCCAAAGTTTGCAGCCTAT ACC
SAP30	ACCAACCAGACCAGGACTT A	CAGACGTGTGCTCTTCCGATCTTCACTAGGAGACGTGGA ATTG
CX3CR1	CACCCGTCCAGACCTTGTT T	CAGACGTGTGCTCTTCCGATCTTGTTCCTCTTAACGTT AGACCAC
BCL2	TGCAAGAGTGACAGTGGAT TG	CAGACGTGTGCTCTTCCGATCTGCTGATATTCTGCAACAC TGTA
BCL6	TGTCCTCACGGTGCCTTTT T	CAGACGTGTGCTCTTCCGATCTGTAGGCAGACACAGGGA CTT
FASLG	CCTCAAGGGGGACTGTCTT TC	CAGACGTGTGCTCTTCCGATCTGCATATCCTGAGCCATC GGT
FAS	ATTGCTGGTAGAGACCCCC A	CAGACGTGTGCTCTTCCGATCTCCCCATTTCCCGATG T

ES 2 711 168 T3

CCL4	CCCAGCCAGCTGTGGTATT C	CAGACGTGTGCTCTTCCGATCTTGGAACTGAACTGAGCT GCT
IFNG	CTAGGCAGCCAACCTAAG CA	CAGACGTGTGCTCTTCCGATCTCCTGCAATCTGAGCCAG TGC
TNF	AGTGGACCTTAGGCCTTCC T	CAGACGTGTGCTCTTCCGATCTGGCTCAGACATGTTTTTC CGTG
IL2	TCACTTAAGACCCAGGGAC TT	CAGACGTGTGCTCTTCCGATCTAAGCATCATCTCAACACT GACTT
IL4	ACCATGAGAAGGACACTC GC	CAGACGTGTGCTCTTCCGATCTCGGGCTTGAATTCCTGT CCT
IL6	CGGCAATGTAGCATGGG C	CAGACGTGTGCTCTTCCGATCTGGAAAGTGGCTATGCAG TTTG
IL1A	GGCATCCTCCACAATAGCA GA	CAGACGTGTGCTCTTCCGATCTGCATTTTGGTCCAAGTTG TGC
IL1B	CTTAAAGCCCGCCTGACAG A	CAGACGTGTGCTCTTCCGATCTACATTCTGATGAGCAAC CGC
IL3	ACAGACGACTTTGAGCCTC G	CAGACGTGTGCTCTTCCGATCTATTTACCTTTTCCTGCG GC
IL13	GGAGCCAAGGGTTCAGAG AC	CAGACGTGTGCTCTTCCGATCTTGCTACCTCACTGGGGT CCT
IL31	GGCCATCTCTTCCTTTCGG A	CAGACGTGTGCTCTTCCGATCTGTGTGGAACTCTGCCG TG
IL24	CTCACCCCA TCA TCCCTTTCC	CAGACGTGTGCTCTTCCGA TCTGCCCAGTGAGACTGTGTTGT
IL26	TACTGACGGCATGTTAGGT G	CAGACGTGTGCTCTTCCGATCTTGTGTGTGGAGTGGGAT GTG
LTA	AGGCAGGGAGGGGACTAT TT	CAGACGTGTGCTCTTCCGATCTGGAGAAACAGAGACAGG CCC
IL5	GCAGTGAGAATGAGGGCC A	CAGACGTGTGCTCTTCCGATCTAGGCATACTGACACTTT GCC
CSF2	AGCCAGTCCAGGAGTGAG AC	CAGACGTGTGCTCTTCCGATCTGGCCACACTGACCCTGA TAC
IL21	CCCAAGGTCAAGA TCGCCAC	CAGACGTGTGCTCTTCCGA TCTCTGCCAGCTCCAGAAGA TGT
IL22	TGGGAAGCCAACTCCATC AT	CAGACGTGTGCTCTTCCGATCTGGAAACCAATGCCACTT TTGT
IL17 A	GCCTTCAAGACTGAACACC GA	CAGACGTGTGCTCTTCCGA TCTGCCCTCAGAGA TCAACAGAC
IL17F	TTGGAGAAGGTGCTGGTG AC	CAGACGTGTGCTCTTCCGATCTCTTACCCAGTGCTCTGC AAC
TGFB1	TATTCCTTTGCCCGGCATC A	CAGACGTGTGCTCTTCCGATCTACCTTGGGCACTGTTGA AGT
CCL20	ACTTGCACA TCA TGGAGGGT	CAGACGTGTGCTCTTCCGA TCTTCCA TAAGCTA TTTTGGTTTAGTGC
IL12A	GGTCCCTCCAAACCGTTGT C	CAGACGTGTGCTCTTCCGATCTGAACTAGGGAGGGGGAA AGAAG
CXCL12	TGGGAGTTGA TCGCCTTTCC	CAGACGTGTGCTCTTCCGA TCTCTCA TTCTGAAGGAGCCCCA T

En la muestra no estimulada, el análisis PCA reveló dos subconjuntos principales de células. Un examen más detenido de los genes enriquecidos en cada subconjunto mostró que un subconjunto representaba linfocitos CD8+ con expresión de CD8A, CD8B, NKG2D, GZMA, GZMH, GZMK y EOMES, y el otro subconjunto representaba linfocitos CD4+ con expresión de CD4, CCR7 y SELL (**FIG. 44A** y **FIG. 45**).

5 En la muestra estimulada, había dos ramas de células inmediatamente claras en el diagrama de PCA (**FIG. 44B** y **FIG. 46A-D**). El primer componente principal representaba el grado de respuesta de células individuales a estimulación en término de nivel variable de expresión de IFNG, TNF, CD69 y GAPDH. La expresión de CCL3, CCL4 y GZMB, que son citocinas y moléculas efectoras asociadas con linfocitos citotóxicos, y LAG3, un marcador asociado con células agotadas, estaba localizada en células en la rama superior. La expresión de IL2, LTA, CD40LG y CCL20, que son citocinas asociadas con linfocitos T auxiliares, estaba localizada en la rama inferior. Otros genes que se ha sabido que están regulados por aumento en linfocitos T activados, incluyendo ZBED2, IL4R, PRDM1, TBX21, MYC, FOSL1, CSF2, TNFRSF9, BCL2 y FASLG, se expresaban en diversos grados en un número más pequeño de células (**FIG. 46A-D**). La mayoría de estas citocinas, moléculas efectoras y factores de transcripción no se expresaban o se expresaban a niveles muy bajos por células en la muestra no estimulada. Aunque la mayoría de las células que respondían dentro de este corto periodo de estimulación eran supuestamente células de memoria, se observó una pequeña población de células que producían un nivel menor de IL2 y no producían otras citocinas ni moléculas efectoras, y pueden representar células vírgenes (**FIG. 44B**, **flecha**).

Para apreciar completamente la heterogeneidad en la respuesta, se agruparon las células basándose en un coeficiente de correlación por pares. Aunque los dos grupos principales de linfocitos CD4 y CD8 eran obvios, había una diversidad considerable dentro de cada conjunto en términos de la combinación y el nivel de genes activados expresados (**FIG. 47** y **FIG. 48**).

Se observó que había unas pocas citocinas, concretamente IL4, IL5, IL13, IL17F, IL22, LIF, IL3 e IL21, que estaban reguladas por aumento en unos pocos cientos o más veces en la muestra estimulada como un conjunto en comparación con la no estimulada, pero a la que contribuían únicamente unas pocas células en la muestra (**FIG. 44C**). Se expresaron subconjuntos de estas citocinas por las mismas células (**FIG. 49A-C**). Por ejemplo, la misma célula individual contribuía a la mayoría de los recuentos de IL17F e IL22, que son características distintivas de linfocitos Th17. Otras 7 células expresaban diversas combinaciones de IL4, IL5, IL13, que eran características distintivas de linfocitos Th2, y expresaban diversas combinaciones de los mismos. Dicha observación resalta la importancia del análisis de células individuales a gran escala, especialmente cuando la contribución a los cambios de expresión global se obtiene de una subpoblación infrecuente.

Se repitió el mismo experimento de estimulación con linfocitos T de un segundo donador de sangre y se analizó el perfil de 669 y 595 células individuales en la muestra estimulada y no estimulada, respectivamente. Aunque el nivel global de activación era inferior (magnitud más pequeña en términos de cambio en la expresión) en este individuo (lo que posiblemente indica variabilidad entre individuos a la estimulación), se observaron las mismas tendencias en análisis PCA, así como heterogeneidad en la respuesta de la célula individual a estímulo (**FIG. 48**).

Identificación de linfocitos T específicos de antígeno infrecuentes

Se demostró la utilidad de nuestra plataforma para identificar células infrecuentes usando el modelo de células específicas de antígeno en población de linfocitos T CD8+. Se expuso sangre fresca de los mismos dos donadores de sangre que eran seropositivos para citomegalovirus (CMV) a la combinación de péptido pp65 de CMV. Una alícuota de sangre no tratada diferente de cada donador sirvió como control negativo. Posteriormente se aislaron linfocitos T CD8+ y se analizó la respuesta de células estimuladas y no estimuladas en nuestra plataforma. Se obtuvieron datos de 2274, 2337, 581 y 253 células en las muestras estimulada y no estimulada por CMV del donador 2, y de las muestras estimulada y no estimulada por CMV del donador 1, respectivamente.

45 Excepto por el control negativo del donador 1 que produjo números relativamente pequeños de células para formar grupos obvios en análisis de agrupación, todas las demás muestras mostraron dos grupos principales de células (**FIG. 50A**, **51** y **52**). Las células en un grupo expresaban marcadores asociados a células vírgenes y a memoria central SELL, CCR7 y CD27, mientras que las células en el otro grupo expresaban genes asociados a células de memoria efectoras (CCL4, CX3CR1, CXCR3) y células efectoras (EOMES, GZMA, GZMB, GZMH, TBX21, ZNF683). Hubo un subconjunto pequeño de células que ocupaba espacio entre las dos ramas y expresa granzima K (GZMK), así como otro subconjunto de células que expresan HLA-DRA. La expresión diferencial de los diferentes tipos de granzimas se ha presentado previamente (8). Los resultados recapitulaban los observados en los experimentos CyTOF previos con linfocitos T CD8+ (9).

Aunque una proporción considerable de células parecía responder a la exposición a un antígeno mediante expresión de CD69 y MYC (**FIG. 52**), se encontraron únicamente unas pocas células que expresaban IFNG, una citocina característica de célula específica de antígeno activada. La mayoría de las células que expresan IFNG estaban también entre aquellas células que portaban la mayoría de moléculas de transcrito detectadas totales en el panel de genes, una indicación del estado activo, y pertenecen al grupo de células de memoria efectoras/efectoras (**FIG. 50B** y **53**). Se identificaron 5 de 581 (0,86 %) y 2 de 2274 (0,09 %) células en los donadores 1 y 2 respectivamente que probablemente eran específicas de CMV basándose en la expresión de IFNG y el nivel de transcripción global. Entre

esas células, había una cantidad sustancial de heterogeneidad en términos de combinaciones y niveles de moléculas efectoras (por ejemplo, granzimas) y citocinas (por ejemplo, IFNG, IL2, CCL3, CCL4, TNF, CSF2, IL4) expresadas (**FIG. 54**). De forma interesante, la célula individual que expresaba la mayoría de transcritos en el donador 2, expresaba tanto IL6 como IL1B, pero no IFNG.

5 *Discusión*

En este ejemplo, se presentó citometría de ARNm de alta escalabilidad que usó una estrategia de Poisson recursiva para aislar células individuales, para poner un código de barras único en el contenido celular y para poner un código de barras de moléculas individuales para análisis cuantitativos. Se mostró que se pueden identificar y contar simultáneamente moléculas de transcrito que pertenecen a cada célula en una muestra que contiene unos pocos miles de células. Además, se ha demostrado el uso de esta técnica para caracterizar células individuales basándose en sus perfiles de expresión en sistemas heterogéneos de origen natural, y la detección de células infrecuentes en una gran población de fondo.

El rendimiento y simplicidad de CytoSeq presenta un avance importante sobre las estrategias existentes que implican placas de microvaloración o chips de microfluidos para la medición basada en secuenciación de la expresión génica de células individuales. Como el procedimiento experimental es simple y el consumo de reactivo por célula es bajo (en el intervalo de nanolitros), posibilita realizar fácilmente el análisis de células individuales para números grandes de células entre múltiples condiciones. En este estudio en solitario, se realizó perfilado de expresión génica de un total de ~14 600 células individuales entre 12 experimentos, que sería muy costoso y llevaría mucho tiempo si se realizara por las estrategias existentes. El número de células medidas por CytoSeq puede aumentarse de escala adicionalmente simplemente aumentando el tamaño de la matriz de micropocillos y el tamaño de la colección de las microesferas con código de barras, lo que se consigue fácilmente por síntesis combinatoria. Además, no hay restricciones sobre la uniformidad de los tamaños celulares, lo que permite por tanto el análisis directo de muestras complejas que contienen células con una diversidad de tamaños y formas celulares, tales como las PBMC mostradas en este ejemplo, sin ninguna preclasificación.

Los datos de CytoSeq se parecían a los de citometría de flujo (FC), pero con diferencias importantes. En primer lugar, CytoSeq ofrece mayor versatilidad en términos del número y tipo de productos génicos estudiados. A diferencia de la citometría de flujo que está confinada principalmente a un manejo de proteínas superficiales y requiere anticuerpos de unión óptima, CytoSeq permitía la medición de cualquier ARNm transcrito mediante técnicas de amplificación de ácido nucleico. El diseño de cebadores óptimos y las condiciones de ensayo posibilitan conseguir de forma rutinaria una tasa de cartografía de ~88 % mediante PCR combinada para un panel elegido de forma arbitraria de 100 o más genes (**tabla 21**). Adicionalmente, el transcriptoma completo de cada célula individual en la muestra también puede medirse mediante amplificación universal del ADNc unido a la microesfera, aunque hay que ser cuidadoso con la eficacia relativamente baja de las técnicas de amplificación universales usadas habitualmente (7) y la alta profundidad de secuenciación requerida para medir el transcriptoma completo entre miles de células.

En segundo lugar, en contraste con la citometría de flujo que depende de la cinética de la unión del anticuerpo, CytoSeq proporciona una lectura absoluta digital del nivel de expresión génica mediante indización molecular. Tiene mayor sensibilidad y especificidad para un evento de célula infrecuente individual porque la detección se conseguía mediante la coexpresión de un número grande de genes específicos para las células infrecuentes. Por lo tanto, consume una cantidad mucho más pequeña de muestra en comparación con la citometría de flujo, que requiere un determinado número de eventos para formar grupos fiables para la sincronización.

Los datos ilustran la importancia del análisis de células individuales frente a masivo. Por ejemplo, se mostraron escenarios donde los genes más altamente expresados en una muestra de miles de células como conjunto, los aportaban únicamente una o unas pocas células. De forma más importante, los experimentos ilustran la importancia de examinar tanto un número grande de células como un número grande de genes en estudios de expresión génica de células individuales, una capacidad que es extremadamente limitante en las estrategias anteriores. La disponibilidad de dicha herramienta para la medición rutinaria de expresión entre miles de células individuales en una muestra biológica puede ayudar a acelerar la comprensión de los sistemas biológicos completos e impulsar aplicaciones novedosas en diagnóstico clínico, tal como análisis de células tumorales en circulación y control de respuestas inmunitarias. Se prevé que el régimen de código de barras de células individuales masivo en paralelo puede también adoptarse para medir el genoma, así como el genoma y el transcrito más simultáneamente, para estudiar la inestabilidad del genoma de células individuales en áreas tales como biología del cáncer y neurociencia.

Tabla 21

Experimento	Número total de lecturas	Número de lecturas con exactamente 1 coincidencia con gen en el panel	% de lecturas alineadas con un gen en el panel	Número de lecturas con coincidencia exacta con un código de barras celular y alineación con un gen	% de lecturas después de alineación del gen y el código de barras	Número de códigos de barras celulares únicos que satisfacen los criterios de filtrado	Número de lecturas asociadas con esos códigos de barras celulares
K562 + Ramos	2 399 025	2 154 454	90 %	1 175 715	49 %	768	859 470
B primarios + Ramos	5 711 013	5 203 308	91 %	3 495 392	61 %	1198	2 868 577
PBMC	1 270 214	1 105 687	87 %	803 151	63 %	632	670 576
Réplica de PBMC	3 927 672	3 468 538	88 %	2 459 367	63 %	731	1 920 956
Estimulado con antiCD3/antiCD28 de donador 1	3 529 898	3 249 998	92 %	2 122 416	60 %	3517	1 466 000
Control negativo antiCD3/antiCD28 de donador 1	1 557 996	1 292 211	83 %	939 094	60 %	1478	719 351
Estimulado con antiCD3/antiCD28 de donador 2	606 865	552 877	91 %	403 943	67 %	669	246 234
Control negativo antiCD3/antiCD28 de donador 2	332 951	283 723	85 %	205 762	62 %	595	86 866
Estimulado con CMV de donador 1	1 064 648	958 410	90 %	697 057	65 %	581	401 629
Control negativo de CMV de donador 1	619 957	547 259	88 %	406 801	66 %	253	192 605
Estimulado con CMV de donador 2	1 902 977	1 692 734	89 %	1 229 667	65 %	2274	688 296
Control negativo de CMV de donador 2	1 671 419	1 346 637	81 %	977 344	58 %	2337	715 453

Síntesis de colección de microesferas

- 5 Las microesferas se fabricaron por Cellular Research, Inc. usando una estrategia combinatoria de división-combinación. En resumen, se distribuyeron microesferas magnéticas de veinte micrómetros funcionalizadas con grupos carboxilo en 96 tubos que contenían oligos con 5' amina, seguida de una secuencia universal, primera parte del marcador celular que es diferente para diferentes tubos, y una secuencia de conector. Los oligos se acoplaron covalentemente en las microesferas por química de carbodiimida. Las microesferas se combinaron y se dividieron en un segundo conjunto de 96 tubos que contenían oligos con una segunda secuencia de conector en el extremo 5',
- 10 seguido de la segunda parte del marcador celular que es diferente para diferentes tubos, y secuencia

complementaria al primer conector. Los oligos en las microesferas se prolongaron por ADN polimerasa tras hibridación con los oligos en disolución mediante el primer conector. Las microesferas se combinaron y dividieron en un tercer conjunto de 96 tubos que contenían oligos con oligo(dA) en el extremo 5', seguido de una secuencia de oligómero aleatorio que sirve como marcador molecular, la tercera parte del marcador celular y una secuencia complementaria al segundo conector. Los oligos en las microesferas se prolongaron por ADN polimerasa tras la hibridación con los oligos en disolución mediante el segundo conector. La colección de microesferas final tiene un tamaño de 96 x 96 x 96 (884 736) marcadores celulares.

Fabricación de matriz de micropocillo

Las matrices de micropocillo se fabricaron usando fotolitografía convencional. Se formaron patrones de series de pilares en oblea de silicio fotorresistente. Se vertió PDMS en la oblea para crear matrices de micropocillos. Se hicieron réplicas de la oblea con adhesivo óptico NOA63 usando la matriz de micropocillos de PDMS como molde. Se moldearon matrices de micropocillos de agarosa (5 %, tipo IX-A, Sigma) a partir de las réplicas de NAO63 antes de cada experimento.

Preparación de muestra

Se cultivaron células K562 y Ramos en RPMI-1640 con FBS al 10 % y antibiótico-antimicótico 1 x. Se adquirieron linfocitos B primarios de un donador sano de Sanguine Biosciences. Las PBMC de un donador sano se aislaron de sangre completa fresca en tubo con heparina de sodio adquirido del Stanford Blood Center usando disolución Lymphoprep (StemCell).

Estimulación de linfocitos T

La sangre completa heparinizada de dos donadores de sangre seropositivos para CMV se obtuvo del Stanford Blood Center. Para la estimulación con CMV, se estimuló 1 ml de sangre completa con combinación de péptido pp65 de CMV diluida en PBS (Miltenyi Biotec) a una concentración final de 1,81 µg/ml durante 6 horas a 37 °C. Una alícuota separada de sangre completa de cada donador se incubó con PBS como control negativo. Se aislaron linfocitos T CD8+ usando el cóctel RosetteSep (StemCell) y posteriormente se depositó en matrices de micropocillos. Para la estimulación con anti-CD3/anti-CD28, los linfocitos T de los dos mismos donadores se aislaron de sangre completa usando cóctel de enriquecimiento de linfocitos T RosetteSep y se resuspendieron en RPMI-1640 con FBS al 10 % y antibiótico-antimicótico 1 x. Una alícuota de células de cada donador se incubó con Dynabeads Human T-Activator CD3/CD28 (Life Technologies) a relación de microesferas a célula ~1:1 a 37 °C durante 6 horas. Una alícuota separada de células de cada donador se puso en la incubadora sin estimulación y sirvió como control negativo.

Captura de células individuales

Se pipeteó suspensión de células individuales en la matriz de micropocillos a una densidad de ~1 célula por 10 micropocillos. Después de lavar para retirar las células no capturadas, se cargaron microesferas magnéticas a una densidad de ~5 microesferas por pocillo para saturar la matriz de micropocillos. Después de lavar para retirar el exceso de microesferas, se pipeteó tampón de lisis frío (Tris-HCl 0,1 M, pH 7,5, LiCl 0,5 M, LiSDS al 1 %, EDTA 10 mM, DTT 5 mM) sobre la superficie de la matriz de micropocillos. Después de 10 minutos de incubación en un portaobjetos imantado, las microesferas se recuperaron de la matriz de micropocillos. Las microesferas se recogieron en un tubo de microcentrífuga, y se lavaron dos veces con tampón de lavado A (Tris-HCl 0,1 M, LiCl 0,5 M, EDTA 1 mM) y una vez con tampón de lavado B (Tris-HCl 20 mM, pH 7,5, KCl 50 mM, MgCl₂ 3 mM). Desde este punto en adelante, todas las reacciones se realizaron en un único tubo.

Síntesis de ADNc

Las microesferas lavadas se resuspendieron en 40 µl de mezcla de RT (tampón de primera hebra 1 x, 1 µl de inhibidor SuperRase, 1 µl de SuperScript II o SuperScript III, MgCl₂ adicional 3 mM, dNTP 1 mM, 0,2 µg/µl de BSA) en un tubo de microcentrífuga colocado en un rotor en un horno de hibridación a temperaturas de 50 °C durante 50 minutos (cuando se usa SuperScript III para el primer experimento con células K562 y Ramos) o a 42 °C durante 90 minutos (cuando se usa Superscript II para el resto de los experimentos). Las microesferas se trataron con 1 µl de Exol (NEB) en 20 µl de tampón Exol 1 x a 37 °C durante 30 minutos, y a 80 °C durante 15 minutos.

PCR combinada y secuenciación

Cada panel de genes contenía dos conjuntos de cebadores específicos de gen diseñados por Primer3. Se escribió una secuencia de comandos MATLAB personalizada para seleccionar cebadores de PCR de modo que hubiera complementariedad mínima en el extremo 3' entre los cebadores dentro del conjunto. Los cebadores en cada panel se enumeran en la **tabla 21**. El esquema de amplificación se muestra en la **FIG. 55**. Se realizó PCR con las microesferas con kit KAPA Fast Multiplex, con 50 nM de cada cebador específico de gen en el primer conjunto de cebadores y 400 nM de cebador universal, en un volumen de 100 µl o 200 µl, con el siguiente protocolo de ciclado: 3 min a 95 °C; 15 ciclos de 15 s a 95 °C, 60 s a 60 °C, 90 s a 72 °C; 5 min a 72 °C. Las microesferas magnéticas se recuperaron y los productos de PCR se purificaron con Ampure XP 0,7 x. La mitad de los productos purificados se usaron para la siguiente ronda de PCR con cebadores internos con el segundo conjunto de cebadores usando el

mismo kit KAPA y protocolo de ciclado. Después de limpiar con Ampure XP 0,7, se introdujo 1/10.^a parte del producto en una reacción de PCR final por la que se adjuntaron los adaptadores Illumina de longitud completa (mezcla KAPA HiFi Ready 1 x, 200 nM de P5, 200 nM de P7. 95 °C 5 min; 8 ciclos de 98 °C 15 s, 60 °C 30 s, 72 °C 30 s; 72 °C 5 min).

5 Análisis de datos

Se realizó secuenciación de la colección en instrumento Illumina MiSeq con química de 150 x 2 pb a una mediana de profundidad de 1,6 millones de lecturas por muestra. La secuenciación reveló el marcador celular, el marcador molecular y el gen de cada lectura (**FIG. 55**). La asignación de gen de cada lectura se hizo con el programa informático de alineación "bowtie" (ref). Los marcadores celulares y moleculares de cada lectura se analizaron usando secuencias de comandos MATLAB personalizadas. Las lecturas se agruparon en primer lugar por marcador celular, después por gen y marcador molecular. Para calcular el número de moléculas únicas por gen por célula, los marcadores moleculares de las lecturas con la misma asignación de marcador celular y gen se agruparon. Una distancia de edición mayor de 1 base se consideró un grupo único y, por tanto, una molécula de transcrito única. Una tabla que contiene la información de expresión génica digital de cada célula se construyó para cada muestra, cada fila en la tabla representaba un marcador celular único, cada columna representaba un gen y cada entrada en la tabla representaba el recuento de moléculas únicas dentro del gen por marcador celular. La tabla se filtró de modo que se retiraron las moléculas únicas que se secuenciaron únicamente una vez (es decir, redundancia = 1). Posteriormente, las células con una suma de moléculas únicas menor de 10 o con coexpresión de 4 o menos genes en el panel se retiraron. La tabla filtrada entonces se usó para el análisis de agrupación. El análisis de componentes principales y la agrupación jerárquica se realizaron sobre el recuento de transcritos transformado por logaritmo (con pseudo recuento de 1 añadido) con funciones integradas en MATLAB.

Referencias citadas en el ejemplo 15, todas las cuales se incorporan por referencia en sus totalidades:

1. A. K. Shalek et al., Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature* 498, 236 (13 de junio de 2013).
2. S. C. Bendall et al., Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. *Science* 332, 687 (6 de mayo de 2011).
3. A. R. Wu et al., Quantitative assessment of single-cell RNA-sequencing methods. *Nature methods* 11, 41 (enero de 2014).
4. B. Treutlein et al., Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. *Nature* 509, 371 (15 de mayo de 2014).
5. S. Islam et al., Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. *Genome research* 21, 1160 (julio de 2011).
6. G. K. Fu, J. Hu, P. H. Wang, S. P. Fodor, Counting individual DNA molecules by the stochastic attachment of diverse labels. *Proceedings of the National Academy of Sciences of the United States of America* 108, 9026 (31 de mayo de 2011).
7. G. K. Fu, J. Wilhelmy, D. Stern, H. C. Fan, S. P. Fodor, Digital encoding of cellular mRNAs enabling precise and absolute gene expression measurement by single-molecule counting. *Analytical chemistry* 86, 2867 (18 de mayo de 2014).
8. K. Bratke, M. Kuepper, B. Bade, J. C. Virchow, Jr., W. Luttmann, Differential expression of human granzymes A, B, and K in natural killer cells and during CD8⁺ T cell differentiation in peripheral blood. *European journal of immunology* 35, 2608 (septiembre de 2005).
9. E. W. Newell, N. Sigal, S. C. Bendall, G. P. Nolan, M. M. Davis, Cytometry by time-of-flight shows combinatorial cytokine expression and virus-specific cell niches within a continuum of CD8⁺ T cell phenotypes. *Immunity* 36, 142 (27 de enero de 2012).

45 Ejemplo 16: Desarrollo de protocolo de cuantificación de células individuales

La FIG. 56 representa un flujo de trabajo general para la cuantificación de moléculas de ARN en una muestra. En este ejemplo, el número total de moléculas de ARN en la muestra era equivalente al número total de moléculas de ARN en una célula individual. Como se muestra en la etapa 1 de la FIG. 56, las moléculas de ARN (110) se transcribieron de forma inversa para producir moléculas de ADNc (105) por la hibridación estocástica de un conjunto de marcadores identificadores moleculares (115) con la región de cola de poliA de las moléculas de ARN. Los marcadores identificadores moleculares (115) comprendían una región de oligodT (120), región de marcador (125) y región de PCR universal (130). El conjunto de marcadores identificadores moleculares contenía 960 tipos diferentes de regiones de marcador.

Parte I. Transcripción inversa y marcaje de moléculas de ARN

Se preparó una muestra de ARN mezclando lo siguiente:

Genes	Número de moléculas de ARN
Lys (control de adición)	456
Phe (control de adición)	912
Thr (control de adición)	1824
Dap (control de adición)	6840
Kan (control de adición)	7352
ARN de línea celular de linfocitos	10 pg (1 equivalente de célula)
Vehículo MS2 (sin poliA)	6×10^{11}

- 5 Las moléculas de ARN se marcaron preparando en un tubo Eppendorf una mezcla de marcaje de la siguiente manera:

	Cantidad (µl)
Muestra de ARN	2
ARN de ms2 1 µg/µl	1
dNTP 10 mM	1
Combinación de 960 oligodT (conjunto n.º 4) 10 µM	0,4
Agua	9,1

Nota: la combinación de oligodT (conjunto n.º 4) se refiere al conjunto de marcadores identificadores moleculares.

Los marcadores identificadores moleculares se hibridaron con las moléculas de ARN por incubación a 65 °C durante 5 minutos. La mezcla de marcaje se almacenó en hielo durante al menos 1 minuto.

- 10 Las moléculas de ARN marcadas se transcribieron de forma inversa mediante la adición de la mezcla de transcripción inversa como se describe a continuación:

	Cantidad (µl)
Tampón de primera hebra 5 X	4
DTT 0,1 M	1
superase-in 20 u/µl	0,5
RT superscript III	1

- 15 Una vez se añadió la mezcla de transcripción inversa al tubo Eppendorf que contenía la reacción de mezcla de marcaje, se realizó la reacción de transcripción inversa incubando la muestra a 37 °C durante 5 minutos, seguida de incubación a 50 °C durante 30 minutos y finalmente incubación a 75 °C durante 15 minutos. La transcripción inversa de las moléculas de ARN marcadas produjo moléculas de ADNc marcadas (170).

- 20 Una vez que las moléculas de ARN se transcribieron de forma inversa y se marcaron, se retiró el exceso de oligos de la muestra por purificación con microesferas Ampure (etapa 2 de la FIG. 1). La purificación con microesferas Ampure se realizó añadiendo 20 µl de microesferas Ampure al tubo Eppendorf que contenía las moléculas de ARN transcritas de forma inversa y marcadas e incubando el tubo a temperatura ambiente durante 5 minutos. Las microesferas se lavaron dos veces con etanol al 70 % para retirar el exceso de oligos. Una vez se retiró el exceso de oligos mediante los lavados con etanol, se añadieron 20 µl de Tris 10 mM al tubo que contenía las moléculas de

ADNc marcadas unidas a microesfera.

- 5 Como se muestra en la etapa 3 de la FIG. 56, las moléculas de ADNc marcadas (170) se amplificaron por PCR combinada. Se realizó amplificación personalizada de las moléculas de ADNc marcadas usando un cebador directo personalizado (F1, 135 en la FIG. 1) y un cebador de PCR universal (140). La tabla 23 enumera los 96 cebadores directos personalizados diferentes que se usaron para amplificar 96 genes diferentes para producir amplicones marcados (140) en un único volumen de reacción.

Para optimizar las reacciones de PCR combinada, se prepararon 3 mezclas de reacción de PCR combinada. La reacción 1 de PCR combinada se preparó del siguiente modo:

	Reacción 1
	Cantidad (µl)
Titanium 10 X	5
dNTP 10 mM	1,5
Agua	35,5
Combinación de cebadores F1 1 µM cada uno	5
PCR004 10 µM	1
ADNc purificado	1
Polimerasa titanium	1

- 10 Las condiciones de reacción para la reacción de PCR combinada fue 1 ciclo a 94 °C durante 2 minutos, seguido de 25 ciclos a 94 °C durante 30 segundos, 57 °C durante 60 segundos y 68 °C durante 1 minuto, después 1 ciclo de 68 °C durante 7 minutos y 1 ciclo de mantenimiento a 4 °C.

Las reacciones 2 y 3 de PCR combinada se prepararon de la siguiente manera:

	Reacción 2	Reacción 3
	Cantidad (µl)	Cantidad (µl)
Mezcla Qiagen Multiplex 2 X	25	25
Combinación de cebadores F1 1 µM cada uno	5	5
PCR004 10 µM	1	1
Disolución Q		5
Agua	18	13
ADNc purificado	1	1

- 15 Las condiciones de reacción de PCR combinada para las reacciones 2 y 3 fueron 1 ciclo a 95 °C durante 15 minutos, seguido de 25 ciclos de 94 °C durante 30 segundos, 57 °C durante 90 segundos y 72 °C durante 1 minuto, después 1 ciclo de 68 °C durante 7 minutos y 1 ciclo de mantenimiento a 4 °C.

Las combinaciones de cebadores F1 contenían los siguientes cebadores:

Cebadores de PCR F1	Secuencia
100611 KanF2	CTGCCTCGGTGAGTTTTCTC
Lys_L_269	CTTCCCGTTACGGTTTTGAC
phe_L_177	AAAACCGGATTAGGCCATTA

Cebadores de PCR F1	Secuencia
thr_L_332	TCTCGTCATGACCGAAAAAG
dap_L_276	CAACGCCTACAAAAGCCAGT

Los genes de control Kan, Phe y Dap se amplificaron selectivamente por PCR con cebadores internos. Las reacciones de amplificación por PCR con cebadores internos se prepararon de la siguiente manera:

N.º de reacción de PCR combinada →	1	2	3	1	2	3	1	2	3
N.º de reacción de PCR →	1	2	3	4	5	6	7	8	9
	µl	µl	µl	µl	µl	µl	µl	µl	µl
Taq 10 x	2,5	2,5	2,5	2,5	2,5	2,5	2,5	2,5	2,5
dNTP 10 mM	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5
Agua	22,25	22,25	22,25	22,25	22,25	22,25	22,25	22,25	22,25
Cy3 PCR004 10 µM	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5
KanF3_5P 5 µM	1	1	1						
Phe_L_215 5 µM				1	1	1			
Dap_L_290 5 µM							1	1	1
Reacción de PCR combinada	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5
USB taq	0,25	0,25	0,25	0,25	0,25	0,25	0,25	0,25	0,25

- 5 Nota: La reacción de PCR combinada usada para las reacciones de PCR 1, 4 y 7 fue la reacción de PCR combinada n.º 1. La reacción de PCR combinada usada para las reacciones de PCR 2, 5 y 8 fue la reacción de PCR combinada n.º 2. La reacción de PCR combinada usada para las reacciones de PCR 3, 6 y 9 fue la reacción de PCR combinada n.º 3.

Los cebadores usados para la PCR con cebadores internos se divulgan a continuación:

Cebador de PCR con cebadores internos	Secuencia
KanF4_5P	/5Phos/GTGGCAAAGCAAAAGTTCAA
Phe_L_215	TGAGAAAGCGTTTGATGATGTA
Dap_L_290	GCCAGTTTATCCCGTCAAAG

10

Las condiciones de reacción para amplificación por PCR para las reacciones 1-9 fue 1 ciclo de 94 °C durante 2 minutos, 30 ciclos de 94 °C durante 20 segundos, 55 °C durante 20 segundos y 72 °C durante 20 segundos, después 1 ciclo a 72 °C durante 4 minutos y 1 ciclo de mantenimiento a 4 °C.

15

Los 4 µl de productos de PCR de las reacciones de amplificación por PCR 1-9 se ejecutaron en un gel de agarosa. Como se muestra en la FIG. 57A, las reacciones 1-3 mostraron la presencia del gen de control Kan, las reacciones 4-6 mostraron la presencia del gen de control Phe y las reacciones 7-9 mostraron la presencia del gen de control Dap.

Los productos de PCR de las reacciones de PCR 1-9 se prepararon para hibridación con una matriz Applied Microarray Inc. (AMI). Las mezclas de hibridación se prepararon de la siguiente manera:

20

ES 2 711 168 T3

	μl
Producto de PCR	20
Lavado A (SSPE 6 X + Triton X-100 al 0,01 %)	55
Cy3 Oligo (760 pM)	1

5 Las mezclas de hibridación 1-9, correspondientes a las mezclas que contienen los productos de PCR de las reacciones de PCR 1-9, respectivamente, se desnaturalizaron a 95 °C durante 5 minutos y después se colocaron a 4 °C. Las mezclas de hibridación se transfirieron a un portaobjetos de matriz AMI y se incubaron durante la noche a 37 °C.

Después de la hibridación durante la noche, el portaobjetos de matriz AMI se lavó y después se exploró. Las mediciones teóricas y reales y el porcentaje de precisión se representan a continuación.

10

15

20

25

30

N.º de mezclas de hibridación →	1	2	3	4	5	6	7	8	9
Condición de PCR combinada (n.º de reacción)	Titanium (1)	Qiagen (2)	Qiagen +Q (3)	Titanium (1)	Qiagen (2)	Qiagen +Q (3)	Titanium (1)	Qiagen (2)	Qiagen +Q (3)
Medición teórica	3676 (bioanalizador)								
Medición real	1826	1740	2116	299	235	251	1165	1077	172
% de detección	49,7	47,3	57,6	32,8	25,8	27,5	17	15,7	2,5
Nota: La medición teórica se basa en la detección de un 100 % de los genes de control Kan, Phe y Dap.									

Los productos de PCR de la reacción 2 se purificaron por purificación Ampure. La purificación Ampure se realizó de la siguiente manera:

	µl
Productos de PCR F1 de la muestra 2 X01	30
Microesferas Ampure	30

5 Las reacciones de purificación Ampure se incubaron a temperatura ambiente durante 5 minutos y después se lavaron en etanol al 70 %. Los productos de PCR purificados se eluyeron de las microesferas en 30 µl de agua. La concentración de los productos de PCR fue de 6 ng/µl determinada por un espectrómetro Nanodrop.

Parte II: Protocolo de preparación de colecciones

Los productos de PCR purificados de la muestra 2 X01 (véase el ejemplo 1) se usaron para preparar una colección de ADN. Se creó una combinación de cebadores F2 mezclando los siguientes cebadores:

Cebadores de PCR F2	Secuencia
Lys_L_269	CTTCCCGTTACGGTTTTGAC
phe_L_177	AAAACCGGATTAGGCCATTA
thr_L_332	TCTCGTCATGACCGAAAAAG
dap_L_276	CAACGCCTACAAAAGCCAGT

10

Se preparó una mezcla de cebadores F2 mezclando lo siguiente

Mezcla de cebadores F2	µl
Agua	750
Combinación de cebadores F2 1 uM cada uno/100 uM en total	100

15

La mezcla de cebadores F2 se incubó a 95 °C durante 30 minutos y después se almacenó en hielo. Se añadió la siguiente mezcla de ligamiento a la mezcla de cebadores F2 para producir una mezcla de ligamiento de cebadores F2:

Mezcla de ligamiento	µl
Tampón de ADN ligasa NEB 10 X	100
T4 PNK USB	50

20

La mezcla de ligamiento de cebadores F2 se incubó a 37 °C durante 1 hora, seguido de una incubación a 65 °C durante 20 minutos. Los cebadores de PCR F2 se precipitaron en etanol y se determinó la concentración de la combinación de cebadores mediante un espectrómetro Nanodrop. La combinación de cebadores F2 se resuspendió para producir una concentración final de 1 uM cada uno/100 uM en total.

Como se muestra en la etapa 4 de la FIG. 56, los amplicones marcados (180) se amplificaron por PCR combinada. Se usaron 96 cebadores directos personalizados diferentes (F2, 145 en la FIG. 1) y un cebador de PCR universal (140) para amplificar los amplicones marcados (muestra 2 X01 del ejemplo 1) en un único volumen de reacción. La tabla 24 enumera los 96 cebadores directos personalizados diferentes.

25

La reacción de PCR combinada se preparó de la siguiente manera:

Mezcla de PCR combinada	µl
Mezcla Qiagen Multiplex 2 X	25

ES 2 711 168 T3

Mezcla de PCR combinada	µl
1 µM de cada cinasa de combinación de cebadores F2	5
PCR004 5'P 10 µM	1
Agua	18
Muestra 2 X01 de primera PCR purificada	1

5 Las condiciones de PCR combinada fueron 1 ciclo a 95 °C durante 15 minutos, seguido por 18 ciclos de 94 °C durante 30 segundos, 57 °C durante 90 segundos y 72 °C durante 1 minuto, después 1 ciclo de 68 °C durante 7 minutos y 1 ciclo de mantenimiento a 4 °C. Los amplicones combinados se purificaron por purificación Ampure y se eluyeron con 50 µ de agua. La concentración de los amplicones se determinó en 30 ng/µl mediante un espectrofotómetro Nanodrop. Se ejecutaron 5 µl de los amplicones en un gel de agarosa (FIG. 58B).

Como se muestra en la etapa 5 de la FIG. 56, se ligaron adaptadores (150, 155) a los amplicones marcados (180) para producir amplicones marcados con adaptador (190). Los amplicones marcados con adaptador se produjeron de la siguiente manera:

Mezcla de adaptadores	µl
Ligasa T4 USB 10 X	10
Agua	60
Producto de PCR con cebadores internos purificado	10
96 adaptadores ABC combinados, hibridados 50 µM	10
ADN ligasa T4 (3 µl neb hc, 7 µl usb)	10

10

La mezcla de adaptadores se incubó a 16 °C durante 4 horas. Los amplicones marcados con adaptador se purificaron por purificación Ampure y se eluyeron en 20 µl de Tris 10 mM.

Los amplicones marcados con adaptador purificados se repararon en los huecos y se amplificaron por PCR de la siguiente manera:

Mezcla de relleno y de PCR	µl
Tampón thermoPol 10 x	5
dNTP 10 mM	1.5
Agua	32
CR P1 10 µM	3
CR IDX D1 10 µM	3
Amplicones marcados con adaptador purificados	5
Vent exo- 2 u/µL	0.5

15

20 Las condiciones de PCR fueron 1 ciclo de 72 °C durante 2 minutos, seguido por 94 °C durante 1 minutos, 12 ciclos de 94 °C durante 15 segundos, 60 °C durante 15 segundos y 72 °C durante 30 segundos, 1 ciclo de 72 °C durante 4 minutos y 1 ciclo de mantenimiento a 4 °C. Los productos de PCR se purificaron por purificación Ampure y se eluyeron en 30 µl de TE. La concentración del producto de PCR purificado fue 22 ng/µl (83 nM) determinada por espectroscopia Nanodrop. Se ejecutaron 5 µl de los productos de PCR purificados en un gel de agarosa al 1 % (FIG. 58B)

Parte III. Secuenciación de la colección de amplicones marcados con adaptador

La colección de amplicones marcados con adaptador se secuenció usando un secuenciador MiSeq.

A continuación, se muestra un resumen del cartografiado de secuencias.

	Requiere coincidencia perfecta	Permite emparejamiento incorrecto de 1 pb
Pares leídos totales	7 724 955	
N.º de ARN con coincidencia de cebador universal y poliA	2 499 444 (32 %)	4 716 378 (61 %)
N.º de ARN cartografiado en las dianas (96 genes)	2 373 700	4 489 485

5 Como se muestra en el resumen de cartografiado de secuencias anterior, muchas lecturas se perdieron debido a los criterios de coincidencia rigurosos de poliA. La FIG. 59 muestra las lecturas y recuentos entre todos los genes detectados.

Las lecturas de secuenciación también se usaron para cuantificar genes específicos. La FIG. 61-62 representa un diagrama de las lecturas observadas por marcador detectado (RPLD) para diversos genes. Los valores rpkm convencionales también se muestran en los diagramas representados en la FIG. 61-62. La FIG. 59 resume una comparación de RPLD y RPKM para diversos genes.

10 La FIG. 63 representa un diagrama de las lecturas totales (marcadores) frente a rpld para diversos genes.

Los datos representados en las FIG. 4, 7 y 8 también se muestran en forma numérica en la tabla 25.

Las FIG. 64 representa un diagrama de RPKM para genes no detectados.

15 La cantidad de controles de adición en la colección de amplicones marcador con adaptador se determinó por secuenciación MiSeq. Los resultados de la secuenciación MiSeq de los controles de adición se muestran en la siguiente tabla.

20

25

30

35

Control de adición	N de entrada (mfg)	Lecturas	Marcadores (K)
Dap	6840	1 920 503	893
Phe	912	470 738	859
Thr	1824	410 664	847
Lys	456	282 174	847
Kan	7352	24	23

5 En la tabla anterior, N de entrada se refiere al número original del control de adición; Lecturas se refiere al número total de pares leídos; y Marcadores (K) se refiere al número de marcadores diferentes detectados por secuenciación. La FIG. 60A-D representa un diagrama de las lecturas observadas por marcador detectado (RPLD) para los controles de adición Lys, Phe, Thr y Dap, respectivamente. La FIG. 60E representa un diagrama de las lecturas frente a la entrada.

Tabla 23

Nombre	Secuencia
NM_144646.3F1	TTGACTTTGCCTTGGAGAGC
NR_015342.1F1	TTTTTCTTACAGTGTCTTGGCATA
NM_000193.2F1	CGTGACCCTAAGCGAGGAG
NM_001777.3F1	TTTGCAGTGATTTGAAGACCA
NM_000600.3F1	GGCATTCCCTTCTTCTGGTCA
NM_021127.2F1	CTGGGCTATATACAGTCCTCAAA
NM_004318.3F1	GGGGTGATTATGACCAGTTGA
NM_002467.4F1	TGCATGATCAAATGCAACCT
NM_001773.2F1	TCTTCCGAAAAATCCTCTTCC
NM_001770.5F1	CTGGGGTCCCAGTCCTATG
NM_001718.4F1	TGTA CTGGGAAGGCAATTTCA
NR_023920.1F1	GAGCCGCTGGGGT TACTC
NM_000267.3F1	CAGTTAGTTGCTGCACATGGA
NM_000633.2F1	TTGCATTTCTTTTGGGGAAG
NM_000314.4F1	GTCATGCATGCAGATGGAAG
NM_021151.3F1	GCTGCAGTGAGCTGTGATGT
NM_002415.1F1	GTTCTCTCCGAGCTCACC
NM_004985.3F1	TCCGAAAGTTTCCAATTCCA
NM_005375.2F1	TTGTTTGGGAGACTCTGCATT
NM_000555.3F1	GACCCCACTTGGACTGGTAG
NM_001668.3F1	GTGATCTTGATTGCGGCTTT
NM_025237.2F1	GGGGGAAAACTACAAGTGC

Tabla 23

Nombre	Secuencia
NM_021117.3F1	TGATTCCTTTTCCTGCCTGT
NM_016316.2F1	AAAAACCTCCAGGCCAGACT
NM_021975.3F1	AATCAAATAACGCCCCAGA
NM_004333.4F1	TTGCTAAAAATTGGCAGAGC
NM_001621.4F1	TTGTTAAGTGCCAAACAAAGGA
NM_005239.5F1	AAGCTGGGAAGAGCAAAGC
NM_000485.2F1	AGGACAGAGGGTGGTCGTC
NM_004048.2F1	TGAGTGCTGTCTCCATGTTTG
NM_001657.2F1	CCTCACAGCTGTTGCTGTTATT
NM_012238.4F1	AAAACACCCAGCTAGGACCA
NM_002055.4F1	AACTGAGGCACGAGCAAAGT
NM_002392.4F1	GCTTTATGGGTGGATGCTGA
NM_001625.3F1	ATAATATCGCCAGCCTCAGC
NM_002110.3F1	TCCAGAGTGTGCTGGATGAC
NM_002943.3F1	TGCAAGCCATTTATGGGAAT
NM_000059.3F1	TGGAATGAGGTCTCTTAGTACAGTT
NM_018136.4F1	TCCAGAAACACCTGTAAGGA
NM_003467.2F1	TGTCTAGGCAGGACCTGTGG
NM_004958.3F1	AGTGATGCTGCGACTCACAC
NM_006139.3F1	GGCTCAGAAAGTCTCTCTTTCC
NM_002693.2F1	CTCCCAAACCTCAGGCTTTCA
NM_001080432.2F1	AAAGCGCTGGGATTACAGG
NM_005954.2F1	CGTCCAGTTGCTTGGAGAAG
NM_024865.2F1	AATAACCTTGGCTGCCGTCT
NM_001905.2F1	GGGAATTCTCAGTGCCAACT
NM_002046.4F1	GCATCCTGGGCTACACTGAG
NM_002253.2F1	TGCTGGGAACAATGACTATAAGA
NM_002356.5F1	GCCTAAAACACTTTGGGTGGT
NM_000189.4F1	GGGTGCCACAAAAATAGAGA
NM_000546.5F1	GAGACTGGGTCTCGCTTTGT
NM_152860.1F1	TGGGAAGGCTTTCTCTAGG
NM_016231.4F1	TTCAACTTGAGTGATCTGAGCTG
NM_000518.4F1	TATGGGCAACCCTAAGGTGA

Tabla 23

Nombre	Secuencia
NM_000905.3F1	CGCTGCGACACTACATCAAC
NM_005038.2F1	TGGAGTCTTGCTCTGTCACC
NM_000041.2F1	ACGAGGTGAAGGAGCAGGT
NM_005957.4F1	CGATGCCTTTGGGTAGAGAG
NR_002785.2F1	ACTGATCGTCCAAGGACTGG
NM_000321.2F1	AAAAAGAAATCTGGTCTTGTAGAAAA
NM_152756.3F1	TTGAAAAGTGGTAAGGAATTGTGA
NM_000610.3F1	CACCAAGAATTGATTTTGTAGCC
NR_033314.1F1	AAAAATGGGGGAAAATGGTG
NM_017460.5F1	CATGGTTGAAACCCCATCTC
NR_002196.1F1	TTCAAAGCCTCCACGACTCT
NM_000591.3F1	GCTGGAACAGGTGCCTAAAG
NM_000106.5F1	CCCTAAGGGAACGACACTCA
NM_138712.3F1	ACCTGCTACAAGCCCTGGA
NM_004304.4F1	GGATCCCTAAGACCGTGGAG
NM_000754.3F1	CCACCTCAGAGGCTCCAA
NM_000492.3F1	TGCTGTATTTTAAAAGAATGATTATGA
NM_000444.4F1	GTAGCTGGGACGCTGGTTTA
NM_002463.1F1	ATTCCCTTCCCCCTACAAGA
NM_000552.3F1	CCTGAGTGCAACGACATCAC
NM_005430.3F1	GGGGGAACCAGCAGAAAT
NM_003150.3F1	GACCTAGGGCGAGGGTTC
NM_000388.3F1	AATTCCTGAAGCCAGATCCA
NM_007294.3F1	AAAATGTTTATTGTTGTAGCTCTGG
NM_005933.3F1	TTTCAAGAGCTCAACAGATGACA
NM_002343.3F1	GACTGCCCCGACAAGTTTT
NM_000376.2F1	GAGAAGGTGCCCCAAAATG
NM_002462.3F1	AGCCACTGGACTGACGACTT
NM_021005.3F1	GGAGGACTAGTGAGGGAGGTG
NM_012343.3F1	GGCAAGTGATGTGGCAATTA
NM_001741.2F1	GTTGGAGCACCTGGAAAGAA
NM_014417.4F1	ATGCCTGCCTCACCTTCAT
NM_014009.3F1	ACAGGGGCACTGTCAACAC

Tabla 23

Nombre	Secuencia
NM_006908.4F1	AAAAATCATGTGTTGCAGCTTT
NM_005228.3F1	TGCTTTCACAACATTTGCAG
NM_013994.2F1	AATGTTTCCTTGTGCCTGCT
NM_000639.1F1	ATATCCTGAGCCATCGGTGA
NM_002701.4F1	TTTTGGTACCCAGGCTATG
NM_000268.3F1	ACCCCGTGGCATTACATAAC
NM_003140.1F1	CTTCCAGGAGGCACAGAAAT
NM_000551.3F1	CTAACCTGGGCGACAGAGTG

Tabla 24

Nombre	Secuencia
NM_144646.3F2	ATATTTGGACATAACAGACTTGGAA
NR_015342.1F2	TGCTGACTTTTAAAATAAGTGATTCG
NM_000193.2F2	GCGGCAGAGTAGCCCTAAC
NM_001777.3F2	TGGGCTATTTCTATTGCTGCT
NM_000600.3F2	AATGGAAAGTGGCTATGCAG
NM_021127.2F2	GGTTGTAGTCACTTTAGATGGAAAA
NM_004318.3F2	TTTGTGACTTTGAGCACCA
NM_002467.4F2	AATGTTTCTCTGTAAATATTGCCATT
NM_001773.2F2	CACCCCATATGGTCATAGC
NM_001770.5F2	AGCACCAGGTGATCCTCAG
NM_001718.4F2	TGTTTTGCTGTAACATTGAAGGA
NR_023920.1F2	TAATGCCACAGTGGGGATG
NM_000267.3F2	GGGCCTAAACTTTGGCAGTT
NM_000633.2F2	TTTTACCTTCATGGCTCTTTT
NM_000314.4F2	GCCTTACTCTGATTCAGCCTCTT
NM_021151.3F2	CGTAACAAAATTCATTGTGGTGT
NM_002415.1F2	AGAACCGCTCCTACAGCAAG
NM_004985.3F2	GTGCTTTCTTTTGTGGGACA
NM_005375.2F2	GGGAGTTCTGCATTTGATCC
NM_000555.3F2	TGGGTCAGAGGACTTCAAGG
NM_001668.3F2	AGGGTTCTGATCACATTGCAC
NM_025237.2F2	CTGCAGGACTGGTCGTTTTT

Tabla 24

Nombre	Secuencia
NM_021117.3F2	AGGGCAGGGTAGAGAGGGTA
NM_016316.2F2	TTCTTCCATGCGGAGAAATC
NM_021975.3F2	CATGGCTGAAGGAAACCAGT
NM_004333.4F2	TTGCCAGCTATCACATGTCC
NM_001621.4F2	TCTTTTCCTGTACCAGTTTTTC
NM_005239.5F2	TGACTGGGAACATCTTGCTG
NM_000485.2F2	TGGCACCTGTACCCTTCTTC
NM_004048.2F2	TTCAATCTCTTGCACTCAAAGC
NM_001657.2F2	TGGAGTCACTGCCAAGTCAT
NM_012238.4F2	TTTGCATGATGTTTGTGTGC
NM_002055.4F2	GCACCCACTCTGCTTTGACT
NM_002392.4F2	ACCATGTAGCCAGCTTTCAA
NM_001625.3F2	GCAACTGGGCATGAGTACCT
NM_002110.3F2	CCACACCCCTTCCTACTC
NM_002943.3F2	AGTCTGCTTATTTCCAGCTGTTT
NM_000059.3F2	TCCTGTTCAAAGTCAGGATGA
NM_018136.4F2	AAATCACAAATCCCCTGCAA
NM_003467.2F2	CTGAACATTCCAGAGCGTGT
NM_004958.3F2	CAGTGGGACCACCCTCACT
NM_006139.3F2	TCTGTAGATGACCTGGCTTGC
NM_002693.2F2	TCAGAACCAAGATGCCAACA
NM_001080432.2F2	CATGACCCAGCCTATGGTTT
NM_005954.2F2	ACCTCCTGCAAGAAGAGCTG
NM_024865.2F2	TTGGGAGGCTTTGCTTATTTT
NM_001905.2F2	CTGGGAAACACTCCTTGCAT
NM_002046.4F2	CAACGAATTTGGCTACAGCA
NM_002253.2F2	CAAAGGTCATAATGCTTTCAGC
NM_002356.5F2	TTTGACGTATCTTTTCATCCAA
NM_000189.4F2	TGTTGTTGGTTTCCAAAAAGG
NM_000546.5F2	GCCAACTTTTGCATGTTTTG
NM_152860.1F2	CCCAAGCTGATCTGGTGGT
NM_016231.4F2	TGCTGTGAAAGAAACAAACATTG
NM_000518.4F2	GCACGTGGATCCTGAGAACT

Tabla 24

Nombre	Secuencia
NM_000905.3F2	CCAGCCCAGAGACTGATT
NM_005038.2F2	CACGCCAGCTAATTTTTGT
NM_000041.2F2	CCTGGTGAAGACATGCAG
NM_005957.4F2	TCACACCTGTAATCCCAGCA
NR_002785.2F2	CAGAGCTCCGCCTCATTAGT
NM_000321.2F2	TCCATTTATCATTGTTTCTGC
NM_152756.3F2	TGGTGTGGTAGGTCAGTGAACA
NM_000610.3F2	AACATGGTCCATTACCTTTATG
NR_033314.1F2	AGAGCGAGACTCCGTCTCAA
NM_017460.5F2	AGTGAGCTGAGATTGCACCA
NR_002196.1F2	AGACGGCCTTGAGTCTCAGT
NM_000591.3F2	GGGAATCCCTTCCTGGTC
NM_000106.5F2	CTTCCTGCCTTTCTCAGCAG
NM_138712.3F2	TGCAGGTGATCAAGAAGACG
NM_004304.4F2	GGTTTTGAGCATGGGTTTCAT
NM_000754.3F2	CCAGCCCCTCCTATGGAT
NM_000492.3F2	AAACTGGGACAGGGGAGAAC
NM_000444.4F2	TTTGGGTAGGTGACCTGCTT
NM_002463.1F2	TCACTGAACGAATGAGTGCTG
NM_000552.3F2	ACGATGTGCAGGACCAGTG
NM_005430.3F2	AATTTGCACTGAAACGTGGA
NM_003150.3F2	CTGTTGTGGCCATTAAAGAA
NM_000388.3F2	TTCCCTCCAGCAGTGGTATT
NM_007294.3F2	CACCAGGAAGGAAGCTGTTG
NM_005933.3F2	TTTCCTTGTGTTCTTCCAAGC
NM_002343.3F2	TCGCAGGCATTACTAATCTGAA
NM_000376.2F2	CTCTGGCTGGCTAACTGGAA
NM_002462.3F2	AGAGCCCCACCCTCAGAT
NM_021005.3F2	TGTGCAGAGTTCTCCATCTGA
NM_012343.3F2	TGCCTGTTACAAATATCAAGGAA
NM_001741.2F2	TTCCCTTCTTGCATCCTTC
NM_014417.4F2	TGTGACCACTGGCATTTCATT
NM_014009.3F2	CTCACACACACGGCCTGTTA

Tabla 24

Nombre	Secuencia
NM_006908.4F2	CACTTGACCAATACTGACCCTCT
NM_005228.3F2	GTGTGTGCCCTGTAACCTGA
NM_013994.2F2	CCACTTCCCACCTTGCAGTCT
NM_000639.1F2	TGTGTGTGTGTGTGTGTGTGT
NM_002701.4F2	TCTCCCATGCATTCAAACCTG
NM_000268.3F2	TCTAAGTGTTCTCACTGACAGG
NM_003140.1F2	TACTCTGCAGCGAAGTGCAA
NM_000551.3F2	CCAAGATCACACCATTGCAC
NM_144646.3F2	ATATTTGGACATAACAGACTTGGAA

Tabla 25

Gen	Número de lecturas	Número de marcadores	Lecturas por kb/millón (RPKM)
APOE	1585	408	0,2
APRT	11280	56	103,4
AREG	147	102	0,0
ASPM	4683	53	4,4
B2M	209362	698	3891,7
BBC3	8	1	0,0
BCL2	3627	27	33,2
BDNF	12778	116	0,3
CD19	38	5	43,1
CD44	6789	47	8,1
COMT	2828	16	10,7
CTPS1	3998	15	25,4
CXCR4	10547	54	19,2
CYP3A4	80982	267	0,1
DCX	28	24	0,0
ETS2	6	5	0,0
FASLG	3182	565	0,2
FTO	8877	58	11,3
GAPDH	227129	661	3870,8
HCK	294	2	2,4
HK2	593	2	12,7

Tabla 25

Gen	Número de lecturas	Número de marcadores	Lecturas por kb/millón (RPKM)
IGJ	119449	454	438,9
KDR	2	2	0,0
KRAS	64	31	6,8
LTF	126	90	0,0
MARCKS	1563	12	36,9
MIF	17775	89	760,4
MLL	72	9	2,6
MTHFR	4854	282	3,9
MX1	100701	285	119,0
MX2	2145	13	45,2
MYB	18361	100	2,8
MYC	6859	27	130,5
NF1	4	1	3,7
NNT	15673	78	14,1
PMAIP1	50604	244	126,9
POLG	5163	46	7,4
POU5F1	1924	12	1,0
PPID	27354	303	39,0
PTEN	20884	109	51,6
RAC1	12454	67	44,8
RB1	1420	14	46,1
RELA	3893	26	17,9
RICTOR	898	5	5,2
RORA	954	7	0,1
SOST	1	1	0,0
SP7	1	1	0,0
STAT3	706	28	16,9
TP53	900	34	14,6
VHL	11576	106	0,0

5 Aunque se han mostrado y descrito en la presente memoria realizaciones preferidas de la presente invención, será obvio para los expertos en la materia que dichas realizaciones se proporcionan a modo de ejemplo únicamente. A los expertos en la materia se les ocurrirán numerosas variaciones, cambios y sustituciones sin alejarse de la invención. Debe entenderse que pueden emplearse diversas alternativas a las realizaciones de la invención descrita en la presente memoria en la práctica de la invención.

Las publicaciones analizadas en la presente memoria se proporcionan únicamente para su divulgación anterior a la fecha de presentación de la presente solicitud. Nada en la presente memoria debe interpretarse como una admisión de que la presente invención no está facultada para antefechar dicha publicación en virtud de invención anterior. Además, las fechas de publicación proporcionadas pueden ser diferentes de las fechas de publicación reales que puede que tengan que confirmarse independientemente.

Realizaciones

En la presente memoria se divulgan métodos para analizar moléculas en dos o más muestras. El método puede comprender: a) producir una pluralidad de ácidos nucleicos con marca de una muestra mediante: i) contacto de una primera muestra que comprende una pluralidad de ácidos nucleicos con una pluralidad de primeras marcas de muestra para producir una pluralidad de primeros ácidos nucleicos con marca de muestra; y ii) contacto de una segunda muestra que comprende una pluralidad de ácidos nucleicos con una pluralidad de segundas marcas de muestra para producir una pluralidad de segundos ácidos nucleicos con marca de muestra, en el que la pluralidad de segundas marcas de muestra son diferentes de las primeras marcas de muestra; b) poner en contacto la pluralidad de ácidos nucleicos con marca de muestra con una pluralidad de marcadores identificadores moleculares para producir una pluralidad de ácidos nucleicos marcados; y c) detectar al menos uno de los ácidos nucleicos marcados, determinando de ese modo un recuento de una pluralidad de ácidos nucleicos en una pluralidad de muestras. Una o más de la pluralidad de muestras pueden comprender una célula individual o lisado celular. Una o más de la pluralidad de muestras pueden consistir en una célula individual. La marca de muestra puede comprender un marcador celular que identifica la célula de la que se originaron los ácidos nucleicos marcados. La pluralidad de muestras que consiste en una célula individual puede ser de una o más fuentes. La marca de muestra puede comprender una región de índice de muestra que identifica la fuente de la célula individual. Los marcadores identificadores moleculares pueden mencionarse como marcador molecular. Una o más de la pluralidad de muestras puede comprender menos de 1 000 000 de células. Una o más de la pluralidad de muestras puede comprender menos de 100 000 células. Una o más de la pluralidad de muestra puede comprender menos de 1000 células. Una o más de la pluralidad de muestra puede comprender menos de 100 células. Una o más de la pluralidad de muestra puede comprender un lisado celular.

Como alternativa, el método para analizar moléculas en una pluralidad de muestras puede comprender: a) producir una pluralidad de ácidos nucleicos marcados que comprende: i) poner en contacto una primera muestra con una primera pluralidad de marcas de muestra, en el que la primera pluralidad de marcas de muestras comprende secuencias de ácido nucleico idénticas; ii) poner en contacto la primera muestra con una primera pluralidad de marcadores identificadores moleculares que comprende diferentes secuencias de ácido nucleico, produciendo de ese modo una pluralidad de primeros ácidos nucleicos marcados; iii) poner en contacto una segunda muestra con una segunda pluralidad de marcas de muestras, en el que la segunda pluralidad de marcas de muestra comprende secuencias de ácido nucleico idénticas; iv) poner en contacto la segunda muestra con una segunda pluralidad de marcadores identificadores moleculares que comprende diferentes secuencias de ácido nucleico, produciendo de ese modo una pluralidad de segundos ácidos nucleicos marcados, en el que la pluralidad de ácidos nucleicos marcados comprende la pluralidad de primeros ácidos nucleicos marcados y los segundos ácidos nucleicos marcados; y b) determinar un número de diferentes ácidos nucleicos marcados, determinando de ese modo un recuento de una pluralidad de ácidos nucleicos en una pluralidad de muestras. La marca de muestra puede comprender un marcador celular que identifica la célula de la que se originaron los ácidos nucleicos marcados. La marca de muestra puede comprender una región de índice de muestra que identifica la fuente de la célula individual. Los marcadores identificadores moleculares pueden mencionarse como marcador molecular.

Como alternativa, el método para analizar moléculas en una pluralidad de muestras puede comprender: a) poner en contacto una pluralidad de muestras que comprende dos o más ácidos nucleicos diferentes con una pluralidad de marcas de muestras y una pluralidad de marcadores identificadores moleculares para producir una pluralidad de ácidos nucleicos marcados, en el que: i) la pluralidad de ácidos nucleicos marcados comprende dos o más ácidos nucleicos adheridos a dos o más marcas de muestra y dos o más marcadores identificadores moleculares; ii) las marcas de muestra adheridas a los ácidos nucleicos de una primera muestra de la pluralidad de muestras son diferentes de las marcas de muestra adheridas a las moléculas de ácido nucleico de una segunda muestra de la pluralidad de muestras; y iii) dos o más ácidos nucleicos idénticos en la misma muestra se adhieren a dos o más marcadores identificadores moleculares diferentes; y b) detectar al menos una parte de los ácidos nucleicos marcados, determinando de ese modo un recuento de dos o más ácidos nucleicos diferentes en la pluralidad de muestras. La marca de muestra puede comprender un marcador celular que identifica la célula de la que se originaron los ácidos nucleicos marcados. La marca de muestra puede comprender una región de índice de muestra que identifica la fuente de la célula individual. Los marcadores identificadores moleculares pueden mencionarse como marcador molecular.

En la presente memoria se divulgan además métodos para analizar moléculas en una pluralidad de muestras, que comprenden: a) poner en contacto una primera pluralidad de moléculas de una primera muestra de una pluralidad de muestras con un primer conjunto de códigos de barras moleculares para producir una primera pluralidad de moléculas marcadas, en los que un código de barras molecular de la primera pluralidad de códigos de barras moleculares comprende una región de marcador y una región de índice de muestra; b) poner en contacto una

segunda pluralidad de moléculas de una segunda muestra de la pluralidad de muestras con un segundo conjunto de códigos de barras moleculares para producir una segunda pluralidad de moléculas marcadas, en los que los códigos de barras moleculares de la segunda pluralidad de códigos de barras moleculares comprenden una región de marcador y una región de índice de muestra, y en los que la primera pluralidad de códigos de barras moleculares y la segunda pluralidad de códigos de barras moleculares difieren al menos en la región de índice de muestra de los códigos de barras moleculares; y c) detectar al menos una parte de dos o más moléculas de la primera pluralidad de moléculas marcadas, determinando de ese modo un recuento de las dos o más moléculas en la pluralidad de muestras. La primera pluralidad de moléculas puede comprender moléculas de ácido nucleico. La segunda pluralidad de moléculas puede comprender moléculas de ácido nucleico. La región de marcador puede mencionarse como marcador molecular. El código de barras molecular puede comprender además un marcador celular. En casos en que una muestra de la pluralidad de muestras consiste en una célula individual, la región de índice de muestra puede referirse al marcador celular.

En la presente memoria se divulga un método de selección de un cebador personalizado, que comprende: a) un primer pase, en que los cebadores elegidos comprenden: i) no más de tres guaninas secuenciales, no más de tres citosinas secuenciales, no más de cuatro adeninas secuenciales y no más de cuatro timinas secuenciales; ii) al menos 3, 4, 5 o 6 nucleótidos que son guaninas o citosinas; iii) una secuencia que no forma fácilmente una estructura de horquilla; b) un segundo pase, que comprende: i) una primera ronda de elección de una pluralidad de secuencias que tiene alta cobertura de todos los transcritos; y ii) una o más rondas posteriores, que seleccionan una secuencia que tiene la máxima cobertura de los transcritos restantes y una puntuación de complementariedad con otras secuencias elegidas de no más de 4; y c) añadir secuencias a un conjunto escogido hasta que se satura la cobertura o un número total de cebadores personalizados es menor de o igual a aproximadamente 96.

En la presente memoria se divulga además un método para producir una colección de moléculas marcadas, que comprende: a) producir una pluralidad de ácidos nucleicos con marca de muestra mediante: i) contacto de una primera muestra que comprende una pluralidad de ácidos nucleicos con una pluralidad de primeras marcas de muestra para producir una pluralidad de primeros ácidos nucleicos con marca de muestra; y ii) contacto de una segunda muestra que comprende una pluralidad de ácidos nucleicos con una pluralidad de segundas marcas de muestra para producir una pluralidad de segundos ácidos nucleicos con marca de muestra, en el que la pluralidad de primeras marcas de muestra son diferentes de las segundas marcas de muestra; y b) poner en contacto la pluralidad de ácidos nucleicos con marca de muestra con una pluralidad de marcadores identificadores moleculares para producir una pluralidad de ácidos nucleicos marcados, produciendo de ese modo una colección de ácidos nucleicos marcados.

En la presente memoria se divulgan kits para su uso en el análisis de moléculas en una pluralidad de muestras. El kit puede comprender: a) dos o más conjuntos de códigos de barras moleculares, en el que un código de barras molecular del conjunto de uno o más códigos de barras moleculares comprende una región de índice de muestra y una región de marcador, en el que i) las regiones de índice de muestra de los códigos de barras moleculares de un conjunto de códigos de barras moleculares son iguales; y ii) las regiones de índice de muestra de un primer conjunto de códigos de barras moleculares son diferentes de las regiones de índice de muestra de un segundo conjunto de códigos de barras moleculares; y b) una pluralidad de microesferas. Los dos o más conjuntos de códigos de barras moleculares pueden adherirse a la pluralidad de microesferas. Los dos o más conjuntos de códigos de barras moleculares pueden conjugarse a la microesfera. La región de marcador puede mencionarse como marcador molecular. El código de barras molecular puede comprender además un marcador celular. En casos en que una muestra de la pluralidad de muestras consiste en una célula individual, la región de índice de muestra puede referirse a un marcador celular.

El kit para analizar moléculas en una pluralidad de muestras puede comprender: a) un primer recipiente que comprende una primera pluralidad de códigos de barras moleculares, en el que: i) un código de barras molecular comprende una región de índice de muestra y una región de marcador; ii) las regiones de índice de muestra de al menos aproximadamente un 80 % del número total de códigos de barras moleculares de la primera pluralidad de códigos de barras moleculares son idénticas; y iii) las regiones de marcador de dos o más códigos de barras de la primera pluralidad de códigos de barras moleculares son diferentes; y b) un segundo recipiente que comprende una segunda pluralidad de códigos de barras moleculares, en el que: i) un código de barras molecular comprende una región de índice de muestra y una región de marcador; ii) las regiones de índice de muestra de al menos aproximadamente un 80 % del número total de códigos de barras moleculares de la primera pluralidad de códigos de barras moleculares son idénticas; y iii) las regiones de marcador de dos o más códigos de barras de la primera pluralidad de códigos de barras moleculares son diferentes; en el que las regiones de índice de muestra de la primera pluralidad de códigos de barras moleculares son diferentes de las regiones de índice de muestra de la segunda pluralidad de códigos de barras moleculares. La región de marcador puede mencionarse como marcador molecular. El código de barras molecular puede comprender además un marcador celular. En casos en que una muestra de la pluralidad de muestras consiste en una célula individual, la región de índice de muestra puede referirse a un marcador celular.

Como alternativa, el kit para analizar moléculas en una pluralidad de muestras comprende: un primer recipiente que comprende una primera pluralidad de marcas de muestra, en el que la pluralidad de marcas de muestra comprende

la misma secuencia de ácido nucleico; y b) un segundo recipiente que comprende una primera pluralidad de marcadores identificadores moleculares, en el que la pluralidad de marcadores identificadores moleculares comprende dos o más secuencias de ácido nucleico diferentes. La región de marcador puede mencionarse como marcador molecular. En casos en que una muestra de la pluralidad de muestras consiste en una célula individual, el marcador de muestra puede referirse a un marcador celular. El kit puede comprender además un tercer recipiente que comprende una primera pluralidad de marcadores celulares, en el que la pluralidad de marcadores celulares comprende dos o más secuencias de ácido nucleico diferentes.

Los kits y métodos divulgados en la presente memoria pueden comprender uno o más conjuntos de códigos de barras moleculares. Los kits y métodos divulgados en la presente memoria pueden comprender uno o más códigos de barras moleculares. Los códigos de barras moleculares pueden comprender una región de índice de muestra, región de marcador molecular, región de marcador celular o una combinación de las mismas. Al menos dos códigos de barras moleculares de un conjunto de códigos de barras moleculares pueden comprender dos o más regiones de marcador diferentes. Las regiones de marcador de dos o más códigos de barras moleculares de dos o más conjuntos de códigos de barras moleculares pueden ser idénticas. Dos o más conjuntos de códigos de barras moleculares pueden comprender códigos de barras moleculares que comprenden la misma región de marcador. En casos en que una muestra de la pluralidad de muestras consiste en una célula individual, la marca de muestra puede referirse a un marcador celular.

Los códigos de barras moleculares divulgados en la presente memoria pueden comprender una región de índice de muestra. La región de índice de muestra de códigos de barras moleculares de dos o más conjuntos de códigos de barras moleculares pueden ser diferentes. La región de índice de muestra puede comprender uno o más nucleótidos. Dos o más secuencias de regiones de índice de muestra de dos o más conjuntos diferentes de códigos de barras moleculares pueden ser menos de aproximadamente un 90 %, 85 %, 80 %, 75 %, 70 %, 65 %, 60 %, 55 %, 50 %, 45 %, 40 %, 35 %, 30 %, 25 %, 20 %, 15 %, 10 % o un 5 % homólogas. Dos o más secuencias de regiones de índice de muestra de dos o más conjuntos diferentes de códigos de barras moleculares pueden ser menos de aproximadamente un 80 % homólogas. Dos o más secuencias de regiones de índice de muestra de dos o más conjuntos diferentes de códigos de barras moleculares pueden ser menos de aproximadamente un 60 % homólogas. Dos o más secuencias de regiones de índice de muestra de dos o más conjuntos diferentes de códigos de barras moleculares pueden ser menos de aproximadamente un 40 % homólogas. Dos o más secuencias de regiones de índice de muestra de dos o más conjuntos diferentes de códigos de barras moleculares pueden ser menos de aproximadamente un 20 % homólogas.

Los códigos de barras moleculares divulgados en la presente memoria pueden comprender un marcador celular. El marcador celular de códigos de barras moleculares de dos o más conjuntos de códigos de barras moleculares pueden ser diferentes. El marcador celular puede comprender uno o más nucleótidos. Dos o más secuencias de marcadores celulares de dos o más conjuntos diferentes de códigos de barras moleculares pueden ser menos de aproximadamente un 90 %, 85 %, 80 %, 75 %, 70 %, 65 %, 60 %, 55 %, 50 %, 45 %, 40 %, 35 %, 30 %, 25 %, 20 %, 15 %, 10 % o un 5 % homólogas. Dos o más secuencias de marcadores celulares de dos o más conjuntos diferentes de códigos de barras moleculares pueden ser menos de aproximadamente un 80 % homólogas. Dos o más secuencias de marcadores celulares de dos o más conjuntos diferentes de códigos de barras moleculares pueden ser menos de aproximadamente un 60 % homólogas. Dos o más secuencias de marcadores celulares de dos o más conjuntos diferentes de códigos de barras moleculares pueden ser menos de aproximadamente un 40 % homólogas. Dos o más secuencias de marcadores celulares de dos o más conjuntos diferentes de códigos de barras moleculares pueden ser menos de aproximadamente un 20 % homólogas.

El código de barras molecular divulgado en la presente memoria puede comprender además una región de PCR universal. El código de barras molecular puede comprender además una región específica de diana. El código de barras molecular puede comprender uno o más nucleótidos. La región de marcador puede comprender uno o más nucleótidos. La región de índice de muestra puede comprender uno o más nucleótidos. La región de PCR universal puede comprender uno o más nucleótidos. La región específica de diana puede comprender uno o más nucleótidos.

Los kits y métodos divulgados en la presente memoria pueden comprender uno o más conjuntos de marcas de muestra. Los kits y métodos divulgados en la presente memoria pueden comprender una o más marcas de muestra. Las marcas de muestras pueden comprender una región de índice de muestra. La región de índice de muestra de las marcas de muestra de un primer conjunto de marcas de muestra puede ser diferente de la región de índice de muestra de las marcas de muestras de un segundo conjunto de marcas de muestra. La región de índice de muestra puede comprender uno o más nucleótidos. Dos o más secuencias de regiones de índice de muestra de dos o más conjuntos diferentes de marcas de muestra pueden ser menos de aproximadamente un 90 %, 85 %, 80 %, 75 %, 70 %, 65 %, 60 %, 55 %, 50 %, 45 %, 40 %, 35 %, 30 %, 25 %, 20 %, 15 %, 10 % o un 5 % homólogas. Dos o más secuencias de regiones de índice de muestra de dos o más conjuntos diferentes de marcas de muestra pueden ser menos de aproximadamente un 80 % homólogas. Dos o más secuencias de regiones de índice de muestra de dos o más conjuntos diferentes de marcas de muestra pueden ser menos de aproximadamente un 60 % homólogas. Dos o más secuencias de regiones de índice de muestra de dos o más conjuntos diferentes de marcas de muestra pueden ser menos de aproximadamente un 40 % homólogas. Dos o más secuencias de regiones de índice de muestra de dos o más conjuntos diferentes de marcas de muestra pueden ser menos de aproximadamente un 20 % homólogas.

Los kits y métodos divulgados en la presente memoria pueden comprender uno o más conjuntos de marcadores identificadores moleculares. Los kits y métodos divulgados en la presente memoria pueden comprender uno o más marcadores identificadores moleculares. Los marcadores identificadores moleculares pueden comprender una región de marcador. Las regiones de marcador de dos o más marcadores identificadores moleculares de un conjunto de marcadores identificadores moleculares pueden ser diferentes. La región de marcador puede comprender uno o más nucleótidos. Una secuencia de regiones de marcador de dos o más marcadores identificadores moleculares de un conjunto de marcadores identificadores moleculares puede ser menos de aproximadamente un 90 %, 85 %, 80 %, 75 %, 70 %, 65 %, 60 %, 55 %, 50 %, 45 %, 40 %, 35 %, 30 %, 25 %, 20 %, 15 %, 10 % o un 5 % homóloga. Una secuencia de regiones de marcador de dos o más marcadores identificadores moleculares de un conjunto de marcadores identificadores moleculares puede ser menos de aproximadamente un 80 % homóloga. Una secuencia de regiones de marcador de dos o más marcadores identificadores moleculares de un conjunto de marcadores identificadores moleculares puede ser menos de aproximadamente un 60 % homóloga. Una secuencia de regiones de marcador de dos o más marcadores identificadores moleculares de un conjunto de marcadores identificadores moleculares puede ser menos de aproximadamente un 40 % homóloga. Una secuencia de regiones de marcador de dos o más marcadores identificadores moleculares de un conjunto de marcadores identificadores moleculares puede ser menos de aproximadamente un 20 % homóloga. Una región de marcador puede mencionarse como región de marcador celular.

Los kits y métodos divulgados en la presente memoria pueden comprender además uno o más cebadores. El uno o más cebadores pueden comprender una secuencia que es al menos parcialmente complementaria a la región de PCR universal. El uno o más cebadores pueden comprender una secuencia que es al menos aproximadamente un 50 % complementaria a la región de PCR universal. El uno o más cebadores pueden comprender una secuencia que es al menos aproximadamente un 80 % complementaria a la región de PCR universal.

Los kits y métodos divulgados en la presente memoria pueden comprender además uno o más agentes de amplificación. Los agentes de amplificación pueden comprender un panel fijo de cebadores. Los agentes de amplificación pueden comprender uno o más cebadores personalizados. Los agentes de amplificación pueden comprender uno o más cebadores de control. Los agentes de amplificación pueden comprender uno o más cebadores de genes constitutivos. Los agentes de amplificación pueden comprender uno o más reactivos de PCR. El uno o más reactivos de PCR pueden comprender polimerasas, trifosfatos de desoxirribonucleótido (dNTP), tampones o una combinación de los mismos.

Los kits y métodos divulgados en la presente memoria pueden comprender además una o más microesferas. Los códigos de barras moleculares pueden adherirse a la una o más microesferas. Las marcas de muestra pueden adherirse a la una o más microesferas. Los marcadores identificadores moleculares pueden adherirse a la una o más microesferas.

En la presente memoria se divulgan además métodos para generar uno o más conjuntos de microesferas. El método puede comprender: a) depositar una pluralidad de primeros ácidos nucleicos en una pluralidad de pocillos, en el que dos o más pocillos diferentes de la pluralidad de pocillos pueden comprender dos o más ácidos nucleicos diferentes de la pluralidad de ácidos nucleicos; b) poner en contacto uno o más pocillos de la pluralidad de pocillos con una o más microesferas para producir una pluralidad de microesferas con marcador individuales, en el que una microesfera con marcador individual de la pluralidad de primeras microesferas marcadas comprende una microesfera adherida a un ácido nucleico de la pluralidad de primeros ácidos nucleicos; c) combinar la pluralidad de primeras microesferas marcadas de la pluralidad de pocillos para producir una combinación de primeras microesferas marcadas; d) distribuir la combinación de primeras microesferas marcadas a una pluralidad posterior de pocillos, en el que dos o más pocillos de la pluralidad posterior de pocillos comprenden dos o más ácidos nucleicos diferentes de una pluralidad de posteriores ácidos nucleicos; y e) adherir uno o más ácidos nucleicos de la pluralidad de ácidos nucleicos posteriores a una o más primeras microesferas marcadas para producir una pluralidad de microesferas marcadas de forma unívoca.

Los métodos y kits divulgados en la presente memoria pueden usarse para analizar una pluralidad de ácidos nucleicos. Los métodos y kits divulgados en la presente memoria pueden usarse para analizar menos de aproximadamente 100 000 000 de ácidos nucleicos. Los métodos y kits divulgados en la presente memoria pueden usarse para analizar menos de aproximadamente 10 000 000 de ácidos nucleicos. Los métodos y kits divulgados en la presente memoria pueden usarse para analizar menos de aproximadamente 1 000 000 de ácidos nucleicos. En la presente memoria se divulgan además métodos de análisis de una pluralidad de proteínas. El método puede comprender: a) producir una pluralidad de polipéptidos con marca de muestra mediante: i) contacto de una primer muestra que comprende una pluralidad de polipéptidos con una pluralidad de primeras marcas de muestra para producir una pluralidad de primeros polipéptidos con marca de muestra; y ii) contacto de una segunda muestra que comprende una pluralidad de polipéptidos con una pluralidad de segundas marcas de muestra para producir una pluralidad de segundos polipéptidos con marca de muestra, en el que la pluralidad de primeras marcas de muestras son diferentes de la pluralidad de segundas marcas de muestra; b) poner en contacto la pluralidad de polipéptidos con marca de muestra con una pluralidad de marcadores identificadores moleculares para producir una pluralidad de polipéptidos marcados; y c) detectar al menos una parte de los polipéptidos marcados, determinando de ese modo un recuento de la pluralidad de polipéptidos en la pluralidad de muestras.

Los métodos de análisis de polipéptidos en una pluralidad de muestras pueden comprender además determinar la identidad de uno o más polipéptidos marcados. La determinación de la identidad del uno o más polipéptidos marcados puede comprender espectrometría de masas. El método puede comprender además combinar los polipéptidos marcados de la primera muestra con los polipéptidos marcados de la segunda muestra. Los polipéptidos marcados pueden combinarse antes de determinar el número de polipéptidos marcados diferentes. El método puede comprender además combinar los primeros polipéptidos con marca de muestra y los segundos polipéptidos con marca de muestra. Los primeros polipéptidos con marca de muestra y los segundos polipéptidos con marca de muestra pueden combinarse antes del contacto con la pluralidad de marcadores identificadores moleculares. La determinación del número de polipéptidos marcados diferentes puede comprender detectar al menos una parte del polipéptido marcado con marca. La detección de al menos una parte del polipéptido marcado con marca puede comprender la detección de al menos una parte de la marca de muestra, marca específica de molécula, polipéptido o una combinación de los mismos.

Los métodos divulgados en la presente memoria pueden comprender el contacto de una pluralidad de muestras con una pluralidad de marcas de muestra y una pluralidad de marcadores identificadores moleculares. El contacto de la pluralidad de muestras con la pluralidad de marcas de muestra y la pluralidad de marcadores identificadores moleculares puede producirse simultáneamente. El contacto de la pluralidad de muestras con la pluralidad de marcas de muestra y la pluralidad de marcadores identificadores moleculares puede producirse de forma concurrente. El contacto de la pluralidad de muestras con la pluralidad de marcas de muestra y la pluralidad de marcadores identificadores moleculares puede producirse secuencialmente. El contacto de la pluralidad de muestras con la pluralidad de marcas de muestra puede producirse antes del contacto de la pluralidad de muestras con la pluralidad de marcadores identificadores moleculares. El contacto de la pluralidad de muestras con la pluralidad de marcas de muestra puede producirse después del contacto de la pluralidad de muestras con la pluralidad de marcadores identificadores moleculares.

Los métodos divulgados en la presente memoria pueden comprender el contacto de una primera muestra con una primera pluralidad de marcas de muestras y una primera pluralidad de marcadores identificadores moleculares. El contacto de la primera muestra con la primera pluralidad de marcas de muestras y la primera pluralidad de marcadores identificadores moleculares puede producirse simultáneamente. El contacto de la primera muestra con la primera pluralidad de marcas de muestras y la primera pluralidad de marcadores identificadores moleculares puede producirse de forma concurrente. El contacto de la primera muestra con la primera pluralidad de marcas de muestras y la primera pluralidad de marcadores identificadores moleculares puede producirse secuencialmente. El contacto de la primera muestra con la primera pluralidad de marcas de muestra puede producirse antes del contacto de la primera muestra con la primera pluralidad de marcadores identificadores moleculares. El contacto de la primera muestra con la primera pluralidad de marcas de muestra puede producirse después del contacto de la primera muestra con la primera pluralidad de marcadores identificadores moleculares.

Los métodos divulgados en la presente memoria pueden comprender el contacto de una segunda muestra con una segunda pluralidad de marcas de muestra y una segunda pluralidad de marcadores identificadores moleculares. El contacto de la segunda muestra con la segunda pluralidad de marcas de muestras y la segunda pluralidad de marcadores identificadores moleculares puede producirse simultáneamente. El contacto de la segunda muestra con la segunda pluralidad de marcas de muestras y la segunda pluralidad de marcadores identificadores moleculares puede producirse de forma concurrente. El contacto de la segunda muestra con la segunda pluralidad de marcas de muestras y la segunda pluralidad de marcadores identificadores moleculares puede producirse secuencialmente. El contacto de la segunda muestra con la segunda pluralidad de marcas de muestra puede producirse antes del contacto de la segunda muestra con la segunda pluralidad de marcadores identificadores moleculares. El contacto de la segunda muestra con la segunda pluralidad de marcas de muestra puede producirse después del contacto de la segunda muestra con la segunda pluralidad de marcadores identificadores moleculares.

Los métodos y kits divulgados en la presente memoria pueden comprender además combinar dos o más muestras. Los métodos y kits divulgados en la presente memoria pueden comprender además combinar la primera muestra y la segunda muestra. La primera y segunda muestras pueden combinarse antes del contacto con la pluralidad de marcadores identificadores moleculares. La primera y segunda muestras pueden combinarse antes de detectar los ácidos nucleicos marcados. Las dos o más muestras pueden combinarse antes del marcaje estocástico de dos o más moléculas en las dos o más muestras. Las dos o más muestras pueden combinarse después del marcaje estocástico de dos o más moléculas en las dos o más muestras. Las dos o más muestras pueden combinarse antes de detectar dos o más moléculas en las dos o más muestras. Las dos o más muestras pueden combinarse después de detectar dos o más moléculas en las dos o más muestras. Las dos o más muestras pueden combinarse antes de analizar dos o más moléculas en las dos o más muestras. Las dos o más muestras pueden combinarse después de analizar dos o más moléculas en las dos o más muestras. Las dos o más muestras pueden combinarse antes de realizar uno o más ensayos en dos o más moléculas en las dos o más muestras. Las dos o más muestras pueden combinarse después de realizar uno o más ensayos en dos o más moléculas en las dos o más muestras.

Los métodos y kits divulgados en la presente memoria pueden combinar la realización de uno o más ensayos en dos o más moléculas en una muestra. El uno o más ensayos pueden comprender una o más reacciones de amplificación. Los métodos y kits divulgados en la presente memoria pueden comprender además la realización de una o más reacciones de amplificación para producir amplicones de ácido nucleico marcados. Los ácidos nucleicos

marcados pueden amplificarse antes de detectar los ácidos nucleicos marcados. El método puede comprender además la combinación de la primera y segunda muestras antes de realizar la una o más reacciones de amplificación.

- Las reacciones de amplificación pueden comprender la amplificación de al menos una parte de la marca de muestra.
- 5 Las reacciones de amplificación pueden comprender la amplificación de al menos una parte del marcador. Las reacciones de amplificación pueden comprender la amplificación de al menos una parte de la marca de muestra, marcador, ácido nucleico o una combinación de los mismos. Las reacciones de amplificación pueden comprender la amplificación de al menos aproximadamente un 1 %, 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 % o 100 % del número total de ácidos nucleicos de la pluralidad de ácidos nucleicos. Las reacciones de amplificación pueden comprender la amplificación de al menos aproximadamente un 1 % del número total de ácidos nucleicos de la pluralidad de ácidos nucleicos. Las reacciones de amplificación pueden comprender la amplificación de al menos aproximadamente un 5 % del número total de ácidos nucleicos de la pluralidad de ácidos nucleicos. Las reacciones de amplificación pueden comprender la amplificación de al menos aproximadamente un 1 %, 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 % o 100 % del número total de ácidos nucleicos marcados de la pluralidad de ácidos nucleicos marcados. Las reacciones de amplificación pueden comprender la amplificación de al menos aproximadamente un 1 % del número total de ácidos nucleicos marcados de la pluralidad de ácidos nucleicos marcados. Las reacciones de amplificación pueden comprender la amplificación de al menos aproximadamente un 5 % del número total de ácidos nucleicos marcados de la pluralidad de ácidos nucleicos marcados. Las reacciones de amplificación pueden comprender la amplificación de al menos aproximadamente un 10 % del número total de ácidos nucleicos marcados de la pluralidad de ácidos nucleicos marcados. Las reacciones de amplificación pueden comprender la amplificación de menos de aproximadamente un 95 %, 90 %, 80 %, 70 %, 60 % o 50 % del número total de ácidos nucleicos de la pluralidad de ácidos nucleicos. Las reacciones de amplificación pueden comprender la amplificación de menos de aproximadamente un 50 % del número total de ácidos nucleicos de la pluralidad de ácidos nucleicos. Las reacciones de amplificación pueden comprender la amplificación de menos de aproximadamente un 20 % del número total de ácidos nucleicos de la pluralidad de ácidos nucleicos. Las reacciones de amplificación pueden comprender la amplificación de menos de aproximadamente un 10 % del número total de ácidos nucleicos de la pluralidad de ácidos nucleicos. Las reacciones de amplificación pueden comprender la amplificación de menos de aproximadamente un 95 %, 90 %, 80 %, 70 %, 60 % o 50 % del número total de ácidos nucleicos marcados de la pluralidad de ácidos nucleicos marcados. Las reacciones de amplificación pueden comprender la amplificación de menos de aproximadamente un 40 % del número total de ácidos nucleicos marcados de la pluralidad de ácidos nucleicos marcados. Las reacciones de amplificación pueden comprender la amplificación de menos de aproximadamente un 25 % del número total de ácidos nucleicos marcados de la pluralidad de ácidos nucleicos marcados. Las reacciones de amplificación pueden comprender la amplificación de menos de aproximadamente un 10 % del número total de ácidos nucleicos marcados de la pluralidad de ácidos nucleicos marcados.

- La una o más reacciones de amplificación pueden provocar la amplificación de aproximadamente 1, 2, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 200, 300, 400, 500, 600, 700, 800, 900 o 1000 ácidos nucleicos diana en una muestra. La una o más reacciones de amplificación pueden provocar la amplificación de aproximadamente 2000 ácidos nucleicos diana en una muestra. La una o más reacciones de amplificación pueden provocar la amplificación de aproximadamente 1000 ácidos nucleicos diana en una muestra. La una o más reacciones de amplificación pueden provocar la amplificación de aproximadamente 2000 moléculas diana. La una o más reacciones de amplificación pueden provocar la amplificación de aproximadamente 100 ácidos nucleicos diana en una muestra.

- 45 Las reacciones de amplificación pueden comprender una o más reacciones en cadena de la polimerasa (PCR). La una o más reacciones en cadena de la polimerasa pueden comprender PCR combinada, PCR con cebadores internos, PCR absoluta, HD-PCR, PCR de la siguiente generación, RTA digital o cualquier combinación de las mismas. La una o más reacciones en cadena de la polimerasa pueden comprender PCR combinada. La una o más reacciones en cadena de la polimerasa pueden comprender PCR con cebadores internos.

- 50 La realización de la una o más reacciones de amplificación puede comprender el uso de uno o más cebadores. El uno o más cebadores pueden comprender uno o más oligonucleótidos. El uno o más oligonucleótidos pueden comprender al menos aproximadamente 7-9 nucleótidos. El uno o más oligonucleótidos pueden comprender menos de 12-15 nucleótidos. El uno o más cebadores pueden hibridar con al menos una parte de la pluralidad de ácidos nucleicos marcados. El uno o más cebadores pueden hibridar con el extremo 3' y/o extremo 5' de la pluralidad de ácidos nucleicos marcados. El uno o más cebadores pueden hibridar con una región interna de la pluralidad de ácidos nucleicos marcados. La región interna puede estar al menos aproximadamente 50, 100, 150, 200, 220, 230, 240, 250, 260, 270, 280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 380, 390, 400, 410, 420, 430, 440, 450, 460, 470, 480, 490, 500, 510, 520, 530, 540, 550, 560, 570, 580, 590, 600, 650, 700, 750, 800, 850, 900 o 1000 nucleótidos desde los extremos 3' de la pluralidad de ácidos nucleicos marcados. La región interna puede estar al menos aproximadamente 2000 nucleótidos desde los extremos 3' de la pluralidad de ácidos nucleicos marcados. 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 oligonucleótidos pueden comprender una secuencia seleccionada de un grupo que consiste en las secuencias de la tabla 1. El uno o más cebadores pueden comprender un cebador universal. El cebador universal puede hibridar con un sitio de unión a cebador universal. El cebador universal puede hibridar con una región de PCR universal. El uno o más cebadores personalizados pueden hibridar con al menos una parte de una marca de muestra. El uno o más cebadores personalizados pueden hibridar con al menos una parte de un marcador identificador molecular. El uno o más cebadores personalizados pueden hibridar con al menos una parte de un código de barras molecular. El uno o más cebadores personalizados pueden hibridar con la primera marca de muestra, la segunda marca de muestra, el marcador identificador molecular, el ácido nucleico o un producto del mismo. El uno o más cebadores pueden comprender un cebador universal y un cebador personalizado. El uno o más cebadores pueden comprender al menos aproximadamente 96 o más cebadores personalizados. El uno o más cebadores pueden comprender al menos aproximadamente 960 o más cebadores personalizados. El uno o más cebadores pueden comprender al menos aproximadamente 9600 o más cebadores personalizados. El uno o más cebadores personalizados pueden hibridar con dos o más ácidos nucleicos marcados diferentes. Los dos o más ácidos nucleicos marcados diferentes pueden corresponder a uno o más genes.
- 5 Las reacciones de PCR combinada pueden comprender una reacción de PCR con cebadores internos. La reacción de PCR con cebadores internos puede comprender un par de cebadores que comprenden un primer cebador y un segundo cebador. El primer cebador puede hibridar con una región de uno o más ácidos nucleicos de la pluralidad de ácidos nucleicos. La región del uno o más ácidos nucleicos puede estar al menos aproximadamente de 300 a 400 nucleótidos desde el extremo 3' del uno o más ácidos nucleicos. El segundo cebador puede hibridar con una región de uno o más ácidos nucleicos de la pluralidad de ácidos nucleicos. La región del uno o más ácidos nucleicos puede estar al menos de 200 a 300 nucleótidos desde el extremo 3' del uno o más ácidos nucleicos.
- 10 Los métodos y kits divulgados en la presente memoria pueden comprender además la realización de una o más reacciones de síntesis de ADNc para producir una o más copias de ADNc de las moléculas o derivados de las mismas (por ejemplo, moléculas marcadas). La una o más reacciones de síntesis de ADNc pueden comprender una o más reacciones de transcripción inversa.
- 15 Los métodos y kits divulgados en la presente memoria pueden comprender una o más muestras. Los métodos y kits divulgados en la presente memoria pueden comprender una pluralidad de muestras. La pluralidad de muestras puede comprender al menos aproximadamente 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 1000 o más muestras. La pluralidad de muestras puede comprender al menos aproximadamente 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000 muestras, 9000 o 10 000 muestras o 100 000 muestras o 1 000 000 muestras o más muestras. La pluralidad de muestras puede comprender al menos aproximadamente 10 000 muestras. La pluralidad de muestras puede comprender al menos aproximadamente 2 muestras. La pluralidad de muestras puede comprender al menos aproximadamente 5 muestras. La pluralidad de muestras puede comprender al menos aproximadamente 10 muestras. La pluralidad de muestras puede comprender al menos aproximadamente 50 muestras. La pluralidad de muestras puede comprender al menos aproximadamente 100 muestras.
- 20 Los métodos y kits divulgados en la presente memoria pueden comprender una o más muestras que comprenden uno o más células. Los métodos y kits divulgados en la presente memoria pueden comprender dos o más muestras que comprenden una o más células. Una primera muestra puede comprender una o más células. Una segunda muestra puede comprender una o más células. La una o más células de la primera muestra pueden ser del mismo tipo celular como la una o más muestras de la segunda muestra.
- 25 Los métodos y kits divulgados en la presente memoria pueden comprender una pluralidad de muestras. La pluralidad de muestras puede ser de uno o más sujetos. La pluralidad de muestras puede ser de dos o más sujetos. La pluralidad de muestras puede ser del mismo sujeto. Los dos o más sujetos pueden ser de la misma especie. Los dos o más sujetos pueden ser de diferentes especies. La pluralidad de muestras puede ser de una o más fuentes. La pluralidad de muestras puede ser de dos o más fuentes. La pluralidad de muestras puede ser del mismo sujeto. Las dos o más fuentes puede ser de la misma especie. La dos o más fuentes pueden ser de diferentes especies.
- 30 La pluralidad de muestras puede obtenerse de forma concurrente. La pluralidad de muestras puede obtenerse secuencialmente. La pluralidad de muestras puede obtenerse durante dos o más periodos de tiempo. Los dos o más periodos de tiempo pueden estar una o más horas separados. Los dos o más periodos de tiempo pueden estar uno o más días separados. Los dos o más periodos de tiempo pueden estar una o más semanas separados. Los dos o más periodos de tiempo pueden estar uno o más meses separados. Los dos o más periodos de tiempo pueden estar uno o más años separados.
- 35 La pluralidad de muestras puede ser de uno o más líquidos corporales, tejidos, células, órganos o músculos. La pluralidad de muestras puede comprender una o más muestras de sangre.
- 40 Los métodos y kits divulgados en la presente memoria pueden comprender una muestra que comprenden uno o más ácidos nucleicos. Dos o más muestras pueden comprender uno o más ácidos nucleicos. Dos o más muestras pueden comprender dos o más ácidos nucleicos. El uno o más ácidos nucleicos de una primera muestra pueden ser diferentes de uno o más ácidos nucleicos de una segunda muestra. Los ácidos nucleicos en una primera muestra
- 45
- 50
- 55

pueden ser al menos aproximadamente un 50 % idénticos a los ácidos nucleicos en una segunda muestra. Los ácidos nucleicos en una primera muestra pueden ser al menos aproximadamente un 70 % idénticos a los ácidos nucleicos en una segunda muestra. Los ácidos nucleicos en una primera muestra pueden ser al menos aproximadamente un 80 % idénticos a los ácidos nucleicos en una segunda muestra.

5 La pluralidad de ácidos nucleicos en la una o más muestras puede comprender dos o secuencias idénticas. Al menos aproximadamente un 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 100 % de los ácidos nucleicos totales en la una o más muestras puede comprender la misma secuencia. La pluralidad de ácidos nucleicos en una o más muestras puede comprender al menos dos secuencias diferentes. Al menos aproximadamente un 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 100 % de los ácidos nucleicos totales en la una o más muestras puede comprender diferentes secuencias.

10 La pluralidad de ácidos nucleicos puede comprender ARN, ADN, ADNc, ARNm, ADN genómico, ARN pequeño, ARN no codificante u otros contenidos de ácido nucleico de una célula. La pluralidad de ácidos nucleicos puede comprender ARNm. La pluralidad de ácidos nucleicos puede comprender ARN. La pluralidad de ácidos nucleicos puede comprender ARNm. La pluralidad de ácidos nucleicos puede comprender ADN.

15 Los métodos y kits divulgados en la presente memoria pueden comprender una o más marcas de muestra. Los métodos y kits divulgados en la presente memoria pueden comprender una o más pluralidades de marcas de muestra. Las marcas de muestra pueden comprender una región de índice de muestra. La región de índice de muestra de una primera pluralidad de marcas de muestra puede ser diferente de la región de índice de muestra de una segunda pluralidad de marcas de muestra. Las marcas de muestra pueden comprender uno o más nucleótidos.

20 Las marcas de muestra pueden comprender al menos aproximadamente 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95 o 100 nucleótidos. Las marcas de muestra pueden comprender al menos aproximadamente 5 o más nucleótidos. Las marcas de muestra pueden comprender al menos aproximadamente 10 o más nucleótidos. Las marcas de muestra pueden comprender menos de aproximadamente 200 nucleótidos. Las marcas de muestra pueden comprender menos de aproximadamente 100 nucleótidos. Las marcas de muestra pueden comprender menos de aproximadamente 60 nucleótidos.

25 Las marcas de muestra pueden comprender además un sitio de unión a cebador universal. Las marcas de muestra pueden comprender además una región de PCR universal. Las marcas de muestra pueden comprender además una o más regiones adaptadoras. Las marcas de muestra pueden comprender además una o más regiones específicas de diana.

30 Los métodos y kits divulgados en la presente memoria pueden comprender uno o más marcadores identificadores moleculares. Los métodos y kits divulgados en la presente memoria pueden comprender una o más pluralidades de marcadores identificadores moleculares. La una o más pluralidades de marcadores identificadores moleculares puede comprender dos o más marcadores identificadores moleculares diferentes. La una o más pluralidades de marcadores identificadores moleculares puede comprender 50 o más marcadores identificadores moleculares diferentes. La una o más pluralidades de marcadores identificadores moleculares puede comprender 90 o más marcadores identificadores moleculares diferentes. La una o más pluralidades de marcadores identificadores moleculares puede comprender 100 o más marcadores identificadores moleculares diferentes. La una o más pluralidades de marcadores identificadores moleculares puede comprender 300 o más marcadores identificadores moleculares diferentes. La una o más pluralidades de marcadores identificadores moleculares puede comprender 500 o más marcadores identificadores moleculares diferentes. La una o más pluralidades de marcadores identificadores moleculares puede comprender 960 o más marcadores identificadores moleculares diferentes. La una o más pluralidades de marcadores identificadores moleculares puede comprender múltiples copias de uno o más marcadores identificadores moleculares. Dos o más pluralidades de marcadores identificadores moleculares pueden comprender uno o más marcadores identificadores moleculares idénticos. Dos o más pluralidades de marcadores identificadores moleculares pueden comprender 10 o más marcadores identificadores moleculares idénticos. Los marcadores identificadores moleculares de una primera pluralidad de marcadores identificadores moleculares pueden ser al menos aproximadamente un 30 % idénticos a los marcadores identificadores moleculares de una segunda pluralidad de marcadores identificadores moleculares. Los marcadores identificadores moleculares de una primera pluralidad de marcadores identificadores moleculares pueden ser al menos aproximadamente un 50 % idénticos a los marcadores identificadores moleculares de una segunda pluralidad de marcadores identificadores moleculares. Los marcadores identificadores moleculares de una primera pluralidad de marcadores identificadores moleculares pueden ser al menos aproximadamente un 80 % idénticos a los marcadores identificadores moleculares de una segunda pluralidad de marcadores identificadores moleculares.

35 40 45 50 55 60 Los marcadores identificadores moleculares pueden comprender una región de marcador (por ejemplo, región de marcador molecular, región de índice molecular). La región de marcador de dos o más marcadores identificadores moleculares de una primera pluralidad de marcadores identificadores moleculares puede ser diferente. Una o más pluralidades de marcadores identificadores moleculares pueden comprender al menos aproximadamente 20 regiones de marcador diferentes. Una o más pluralidades de marcadores identificadores moleculares pueden comprender al menos aproximadamente 50 regiones de marcador diferentes. Una o más pluralidades de

- 5 marcadores identificadores moleculares pueden comprender al menos aproximadamente 96 regiones de marcador diferentes. Una o más pluralidades de marcadores identificadores moleculares pueden comprender al menos aproximadamente 200 regiones de marcador diferentes. Una o más pluralidades de marcadores identificadores moleculares pueden comprender al menos aproximadamente 500 regiones de marcador diferentes. Una o más pluralidades de marcadores identificadores moleculares pueden comprender al menos aproximadamente 960 regiones de marcador diferentes.
- 10 Los marcadores identificadores moleculares pueden comprender al menos aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más nucleótidos. Los marcadores identificadores moleculares pueden comprender al menos aproximadamente 20, 30, 40, 50 o más nucleótidos. Los marcadores identificadores moleculares pueden comprender al menos aproximadamente 21 nucleótidos.
- Los marcadores identificadores moleculares pueden comprender además una región específica de diana. La región específica de diana puede comprender una secuencia de oligodT.
- 15 Los marcadores identificadores moleculares pueden comprender además uno o más marcadores de tinte. Los marcadores identificadores moleculares pueden comprender además un tinte Cy3. Los marcadores identificadores moleculares pueden comprender además un tinte Tye563.
- 20 Los métodos y kits divulgados en la presente memoria pueden comprender una o más moléculas marcadas. La una o más moléculas marcadas pueden producirse poniendo en contacto una pluralidad de moléculas con una pluralidad de marcas de muestra. La una o más moléculas marcadas pueden producirse poniendo en contacto una pluralidad de ácidos nucleicos con una pluralidad de marcas de muestra. El contacto de la pluralidad de ácidos nucleicos con la pluralidad de marcas de muestra puede comprender el ligamiento de una o más marcas de muestra con uno o más ácidos nucleicos. El contacto de la pluralidad de ácidos nucleicos con la pluralidad de marcas de muestra puede comprender la hibridación de una o más marcas de muestra con uno o más ácidos nucleicos. El contacto de la pluralidad de ácidos nucleicos con la pluralidad de marcas de muestra puede comprender la realización de una o más reacciones de prolongación de ácido nucleico. La una o más reacciones de prolongación de ácido nucleico pueden comprender transcripción inversa.
- 25 Los métodos y kits divulgados en la presente memoria pueden comprender además la adhesión de uno o más conectores oligonucleotídicos a la pluralidad de ácidos nucleicos. El método y kits pueden comprender además la adhesión de uno o más conectores oligonucleotídicos a los ácidos nucleicos con marca de muestra. Los métodos y kits pueden comprender además la adhesión de uno o más conectores oligonucleotídicos a los ácidos nucleicos marcados. El uno o más conectores pueden comprender al menos aproximadamente 10 nucleótidos.
- 30 Los métodos y kits divulgados en la presente memoria pueden comprender además la adhesión de uno o más ácidos nucleicos marcados a un soporte. El soporte puede comprender un soporte sólido. El soporte puede comprender una microesfera. El soporte puede comprender una matriz. El soporte puede comprender un portaobjetos de vidrio.
- 35 La adhesión de los ácidos nucleicos marcados al soporte puede comprender la reticulación con amina-tiol, reticulación con maleimida, *N*-hidroxisuccinimida o *N*-hidroxisulfosuccinimida, Zenon, SiteClick o una combinación de los mismos. La adhesión de los ácidos nucleicos marcados al soporte puede comprender la adhesión de biotina al uno o más ácidos nucleicos marcados.
- 40 El soporte puede comprender una o más microesferas. La una o más microesferas pueden ser una microesfera recubierta. La microesfera recubierta puede estar recubierta con estreptavidina.
- El soporte puede comprender una matriz. La matriz puede comprender una o más sondas. Los ácidos nucleicos marcados pueden adherirse a la una o más sondas. La una o más sondas pueden comprender uno o más oligonucleótidos. La una o más sondas pueden adherirse a al menos una parte de los ácidos nucleicos marcados. La parte de los ácidos nucleicos marcados adheridos a la una o más sondas pueden comprender al menos una parte de la marca muestra, marcador identificador molecular, código de barras molecular, ácido nucleico o una combinación de los mismos.
- 45 El soporte puede comprender un portaobjetos de vidrio. El portaobjetos de vidrio puede comprender uno o más pocillos. El uno o más pocillos pueden estar grabados en el portaobjetos de vidrio. El uno o más pocillos pueden comprender al menos 960 pocillos. El portaobjetos de vidrio puede comprender una o más sondas. La una o más sondas pueden estar impresas en el portaobjetos de vidrio. El uno o más pocillos pueden comprender además una o más sondas. La una o más sondas pueden estar impresas dentro del uno o más pocillos. La una o más sondas pueden comprender 960 ácidos nucleicos. Los ácidos nucleicos pueden ser diferentes. Los ácidos nucleicos pueden ser iguales.
- 50 Los métodos y kits divulgados en la presente memoria pueden usarse para determinar un recuento de una o más moléculas en una o más muestras. La determinación del recuento de una o más moléculas puede comprender la determinación del número de ácidos nucleicos marcados diferentes. La determinación del número de ácidos nucleicos marcados diferentes puede comprender la detección de al menos una parte del ácido nucleico marcado.
- 55

- La detección de al menos una parte del ácido nucleico marcado puede comprender la detección de al menos una parte de la marca de muestra, marcador identificador molecular, código de barras molecular, ácido nucleico o una combinación de los mismos.
- 5 La determinación del número de diferentes ácidos nucleicos marcados puede comprender secuenciación. La secuenciación puede comprender secuenciación MiSeq. La secuenciación puede comprender secuenciación HiSeq. La determinación del número de diferentes ácidos nucleicos marcados puede comprender una matriz. La determinación del número de diferentes ácidos nucleicos marcados puede comprender el contacto de los ácidos nucleicos marcados con la una o más sondas.
- 10 La determinación del número de diferentes ácidos nucleicos marcados puede comprender el contacto de los ácidos nucleicos marcados con una matriz. La matriz puede comprender una pluralidad de sondas. La determinación del número de diferentes ácidos nucleicos marcados puede comprender el contacto de los ácidos nucleicos marcados con un portaobjetos de vidrio de una pluralidad de sondas.
- 15 La determinación del número de diferentes ácidos nucleicos marcados puede comprender hibridación de sonda marcada, amplificación específica de diana, secuenciación específica de diana, secuenciación con nucleótidos marcados específicos para un polimorfismo nucleotídico pequeño diana, secuenciación con nucleótidos marcados específicos para patrones de digestión con enzima de restricción, secuenciación con nucleótidos marcados específicos para mutaciones o una combinación de los mismos.
- 20 La determinación del número de diferentes ácidos nucleicos marcados puede comprender clasificación por citometría de flujo de un marcador específico de secuencia. La determinación del número de diferentes ácidos nucleicos marcados puede comprender la detección de los ácidos nucleicos marcados adheridos a las microesferas. La detección de los ácidos nucleicos marcados adheridos a las microesferas puede comprender detección de fluorescencia.
- 25 La determinación del número de diferentes ácidos nucleicos marcados puede comprender el recuento de la pluralidad de ácidos nucleicos marcados por transferencia de energía de resonancia de fluorescencia (FRET), entre una sonda específica de diana y un ácido nucleico marcado o una sonda marcada específica de diana. La determinación del número de diferentes ácidos nucleicos marcados puede comprender la adhesión del ácido nucleico marcado al soporte.
- Los métodos y kits divulgados en la presente memoria pueden comprender además inmunoprecipitación de una secuencia diana con una proteína de unión a ácido nucleico.
- 30 Los métodos y kits divulgados en la presente memoria pueden comprender además la distribución de la pluralidad de muestras en una pluralidad de pocillos de una placa de micropocillos. Una o más de la pluralidad de muestras pueden comprender una pluralidad de células. Una o más de la pluralidad de muestras pueden comprender una pluralidad de ácidos nucleicos. Los métodos y kits divulgados en la presente memoria pueden comprender además la distribución de una o menos células en la pluralidad de pocillos. La pluralidad de células puede lisarse en la placa de micropocillos. Los métodos y kits divulgados en la presente memoria pueden comprender además la síntesis de ADNc en la placa de micropocillos. La síntesis de ADNc puede comprender transcripción inversa de ARNm.
- 35 Los métodos y kits divulgados en la presente memoria pueden comprender además la distribución de la pluralidad de primeras marcas de muestra, la pluralidad de segundas de muestras, la pluralidad de marcadores identificadores moleculares o cualquier combinación de los mismos en una placa de micropocillos.
- 40 Los métodos y kits divulgados en la presente memoria pueden comprender además la distribución de una o más microesferas en la placa de micropocillos. La placa de micropocillos puede comprender una placa de micropocillos fabricada de PDMS por litografía suave, gravada en una oblea de silicio, gravada en un portaobjetos de vidrio, estampado fotorresistente en un portaobjetos de vidrio o una combinación de los mismos. El micropocillo puede comprender un orificio en una placa microcapilar. La placa de micropocillos puede comprender una emulsión de agua en aceite. La placa de micropocillos puede comprender al menos uno o más pocillos. La placa de micropocillos puede comprender al menos aproximadamente 6 pocillos, 12 pocillos, 48 pocillos, 96 pocillos, 384 pocillos, 960 pocillos o 1000 pocillos.
- 45 Los métodos y kits divulgados en la presente memoria pueden comprender además un chip. La placa de micropocillos puede adherirse al chip. El chip puede comprender al menos aproximadamente 6 pocillos, 12 pocillos, 48 pocillos, 96 pocillos, 384 pocillos, 960 pocillos, 1000 pocillos, 2000 pocillos, 3000 pocillos, 4000 pocillos, 5000 pocillos, 6000 pocillos, 7000 pocillos, 8000 pocillos, 9000 pocillos, 10 000 pocillos, 20 000 pocillos, 30 000 pocillos, 40 000 pocillos, 50 000 pocillos, 60 000 pocillos, 70 000 pocillos, 80 000 pocillos, 90 000 pocillos, 100 000 pocillos, 200 000 pocillos, 500 000 pocillos o un millón de pocillos. Los pocillos pueden comprender un área de al menos aproximadamente 300 micrómetros², 400 micrómetros², 500 micrómetros², 600 micrómetros², 700 micrómetros², 800 micrómetros², 900 micrómetros², 1000 micrómetros², 1100 micrómetros², 1200 micrómetros², 1300 micrómetros², 1400 micrómetros², 1500 micrómetros². Los métodos y kits divulgados en la presente memoria pueden comprender además la distribución entre aproximadamente 10 000 y 30 000 muestras en el chip.
- 50
- 55

Los métodos y kits divulgados en la presente memoria pueden comprender además el diagnóstico de una afección, enfermedad o trastorno en un sujeto que lo necesita.

5 Los métodos y kits divulgados en la presente memoria pueden comprender además el pronóstico de una afección, enfermedad o trastorno en un sujeto que lo necesita. Los métodos y kits divulgados en la presente memoria pueden comprender además la determinación de un tratamiento para una afección, enfermedad o trastorno en un sujeto que lo necesita.

La pluralidad de muestras puede comprender una o más muestras de un sujeto que padece una enfermedad o afección. La pluralidad de muestras puede comprender una o más muestras de un sujeto sano.

10 En la presente memoria se divulga además un método de análisis forense que comprende: a) marcaje estocástico de dos o más moléculas en dos o más muestras para producir dos o más moléculas marcadas; y b) detección de las dos o más moléculas marcadas.

15 El método de selección del cebador personalizado puede comprender además seleccionar el cebador personalizado basándose en uno o más ácidos nucleicos. El uno o más ácidos nucleicos pueden comprender transcritos de ARNm, transcritos no codificantes que incluyen ARN estructurales, pseudogenes transcritos, ARN de modelo proporcionado por un proceso de anotación de genoma, secuencias correspondientes a un cóntigo genómico o cualquier combinación de los mismos. El uno o más ácidos nucleicos pueden ser ARN. El uno o más ácidos nucleicos pueden ser ARNm. El uno o más ácidos nucleicos pueden comprender uno o más exones. El método de selección del cebador personalizado puede comprender además enriquecer uno o más subconjuntos de ácidos nucleicos. El uno o más subconjuntos comprenden baja abundancia de ARNm. El método de selección del cebador personalizado puede comprender además un algoritmo informático.

20 Los métodos y kits divulgados en la presente memoria pueden comprender el uso de uno o más controles. El uno o más controles pueden ser controles de adición. El uno o más controles pueden comprender ácidos nucleicos. La una o más muestras que comprenden una pluralidad de ácidos nucleicos pueden tener uno o más ácidos nucleicos de control añadidos. El uno o más ácidos nucleicos de control pueden usarse para medir una eficacia de producción de la colección de ácidos nucleicos marcados.

25 Los métodos y kits divulgados en la presente memoria pueden comprender usarse en la producción de una o más colecciones de ácidos nucleicos. La una o más colecciones de ácidos nucleicos pueden comprender una pluralidad de ácidos nucleicos marcados o derivados de los mismos (por ejemplo, amplicones marcados). El método de producción de la colección de ácidos nucleicos marcados puede comprender el marcaje estocástico de dos o más ácidos nucleicos en dos o más muestras con dos o más conjuntos de códigos de barras moleculares para producir una pluralidad de ácidos nucleicos marcados. El método de producción de una colección de ácidos nucleicos marcados puede comprender el contacto de dos o más muestras con una pluralidad de marcas de muestra y una pluralidad de marcadores específicos de molécula para producir una pluralidad de ácidos nucleicos marcados. Los ácidos nucleicos marcados pueden comprender una región de índice de muestra, una región de marcador y una región de ácido nucleico. La región de índice de muestra puede usarse para conferir una identidad de muestra o submuestra al ácido nucleico. La región de índice de muestra puede usarse para determinar la fuente del ácido nucleico. La región de marcador puede usarse para conferir una identidad única al ácido nucleico, posibilitando de ese modo la diferenciación de dos o más ácidos nucleicos idénticos en la misma muestra o submuestra.

30 El método de producción de una colección de ácidos nucleicos puede comprender además amplificar uno o más ácidos nucleicos marcados para producir uno o más ácidos nucleicos marcados enriquecidos. El método puede comprender además realizar uno o más ensayos de arrastre del uno o más ácidos nucleicos enriquecidos. El método puede comprender además purificar el uno o más ácidos nucleicos marcados enriquecidos.

35 Los kits divulgados en la presente memoria pueden comprender una pluralidad de microesferas, un cebador y/o agentes de amplificación. Puede usarse uno o más kits en el análisis de al menos aproximadamente 10, 20, 30, 40, 50, 60, 70, 80, 90 o 100 o más muestras o submuestras. Puede usarse uno o más kits en el análisis de al menos aproximadamente 96 muestras. Puede usarse uno o más kits en el análisis de al menos aproximadamente 384 muestras. El kit puede comprender además instrucciones para el diseño y optimización de cebadores.

40 El kit puede comprender además una o más placas de micropocillos. La una o más placas de micropocillos pueden usarse para la distribución de una o más microesferas. La una o más placas de micropocillos pueden usarse para la distribución de una o más moléculas o derivados de las mismas (por ejemplo, moléculas marcadas, amplicones marcados) de una o más muestras.

45 El kit puede comprender además uno o más recipientes adicionales. El uno o más recipientes adicionales pueden comprender una o más pluralidades adicionales de marcas de muestra. La una o más pluralidades adicionales de marcas de muestra en el uno o más recipientes adicionales pueden ser diferentes de la primera pluralidad de marcas de muestra del primer recipiente. El uno o más recipientes adicionales pueden comprender una o más pluralidades adicionales de marcadores identificadores moleculares. La una o más pluralidades adicionales de marcadores identificadores moleculares del uno o más recipientes adicionales pueden ser al menos aproximadamente un 50 % idénticas al uno o más marcadores identificadores moleculares adicionales del segundo

recipiente. La una o más pluralidades adicionales de marcadores identificadores moleculares del uno o más recipientes adicionales pueden ser al menos aproximadamente un 80 % idénticas al uno o más marcadores identificadores moleculares adicionales del segundo recipiente. La una o más pluralidades adicionales de marcadores identificadores moleculares del uno o más recipientes adicionales pueden ser al menos aproximadamente un 90 % idénticas al uno o más marcadores identificadores moleculares adicionales del segundo recipiente.

En la presente memoria se divulgan además métodos de producción de uno o más conjuntos de microesferas marcadas. El método de producción del uno o más conjuntos de microesferas marcadas puede comprender la adhesión de uno o más ácidos nucleicos a una o más microesferas, produciendo de ese modo uno o más conjuntos de microesferas marcadas. El uno o más ácidos nucleicos pueden comprender uno o más códigos de barras moleculares. El uno o más ácidos nucleicos pueden comprender una o más marcas de muestra. El uno o más ácidos nucleicos pueden comprender uno o más marcadores identificadores moleculares. El uno o más ácidos nucleicos pueden comprender a) una región de cebador; b) una región de muestra; y c) un conector o región adaptadora. El uno o más ácidos nucleicos pueden comprender a) una región de cebador; b) una región de marcador; y c) un conector o región adaptadora. El uno o más ácidos nucleicos pueden comprender a) una región de índice de muestra; y b) una región de marcador. El uno o más ácidos nucleicos pueden comprender además una región de cebador. El uno o más ácidos nucleicos pueden comprender además una región específica de diana. El uno o más ácidos nucleicos pueden comprender además una región de conector. El uno o más ácidos nucleicos pueden comprender además una región adaptadora. El uno o más ácidos nucleicos pueden comprender además una región de índice de muestra. El uno o más ácidos nucleicos pueden comprender además una región de marcador.

La región de cebador de los ácidos nucleicos para un conjunto de microesferas marcadas puede ser al menos aproximadamente un 70 % idéntica. La región de cebador de los ácidos nucleicos para un conjunto de microesferas marcadas puede ser al menos aproximadamente un 90 % idéntica. La región de cebador de los ácidos nucleicos para un conjunto de microesferas marcadas puede ser igual.

La región de índice de muestra de los ácidos nucleicos para un conjunto de microesferas marcadas puede ser al menos aproximadamente un 70 % idéntica. La región de índice de muestra de los ácidos nucleicos para un conjunto de microesferas marcadas puede ser al menos aproximadamente un 90 % idéntica. La región de índice de muestra de los ácidos nucleicos para un conjunto de microesferas marcadas puede ser igual. La región de índice de muestra de los ácidos nucleicos de dos o más conjuntos de microesferas con índice de muestra puede ser menos de aproximadamente un 40 % idéntica. La región de índice de muestra de los ácidos nucleicos de dos o más conjuntos de microesferas con índice de muestra puede ser menos de aproximadamente un 50 % idéntica. La región de índice de muestra de los ácidos nucleicos de dos o más conjuntos de microesferas con índice de muestra puede ser menos de aproximadamente un 60 % idéntica. La región de índice de muestra de los ácidos nucleicos de dos o más conjuntos de microesferas con índice de muestra puede ser diferente.

La región de marcador de los ácidos nucleicos para dos o más conjuntos de microesferas marcadas puede ser al menos aproximadamente un 70 % idéntica. La región de marcador de los ácidos nucleicos para dos o más conjuntos de microesferas marcadas puede ser al menos aproximadamente un 90 % idéntica. La región de marcador de los ácidos nucleicos para dos o más conjuntos de microesferas marcadas puede ser igual. La región de marcador de los ácidos nucleicos para un conjunto de microesferas marcadas puede ser menos de aproximadamente un 40 % idéntica. La región de marcador de los ácidos nucleicos para un conjunto de microesferas marcadas puede ser menos de aproximadamente un 50 % idéntica. La región de marcador de los ácidos nucleicos para un conjunto de microesferas marcadas puede ser menos de aproximadamente un 60 % idéntica. La región de marcador de dos o más ácidos nucleicos para un conjunto de microesferas marcadas puede ser diferente.

El conector o región adaptadora de los ácidos nucleicos para un conjunto de microesferas marcadas puede ser al menos aproximadamente un 70 % idéntica. El conector o región adaptadora de los ácidos nucleicos para un conjunto de microesferas marcadas puede ser al menos aproximadamente un 90 % idéntica. El conector o región adaptadora de los ácidos nucleicos para un conjunto de microesferas marcadas puede ser igual.

La región específica de diana de los ácidos nucleicos para dos o más conjuntos de microesferas específicas de diana puede ser al menos aproximadamente un 70 % idéntica. La región específica de diana de los ácidos nucleicos para dos o más conjuntos de microesferas específicas de diana puede ser al menos aproximadamente un 90 % idéntica. La región específica de diana de los ácidos nucleicos para dos o más conjuntos de microesferas específicas de diana puede ser igual. La región específica de diana de los ácidos nucleicos para un conjunto de microesferas específicas de diana puede ser menos de aproximadamente un 40 % idéntica. La región específica de diana de los ácidos nucleicos para un conjunto de microesferas específicas de diana puede ser menos de aproximadamente un 50 % idéntica. La región específica de diana de los ácidos nucleicos para un conjunto de microesferas específicas de diana puede ser menos de aproximadamente un 60 % idéntica. La región específica de diana de dos o más ácidos nucleicos para un conjunto de microesferas específicas de diana puede ser diferente.

El uno o más conjuntos de microesferas marcadas pueden comprender un millón o más microesferas marcadas. El uno o más conjuntos de microesferas marcadas pueden comprender diez millones o más microesferas marcadas.

La adhesión del uno o más ácidos nucleicos a las microesferas puede comprender adhesión covalente. La adhesión del uno o más ácidos nucleicos a las microesferas puede comprender conjugación. La adhesión del uno o más ácidos nucleicos a las microesferas puede comprender interacciones iónicas.

5 Las microesferas pueden ser microesferas recubiertas. Los ácidos nucleicos pueden adherirse a una o más marcas. Las microesferas pueden recubrirse con estreptavidina. Los ácidos nucleicos pueden adherirse a biotina. Las microesferas también pueden recubrirse con anticuerpos o ácidos nucleicos, y los ácidos nucleicos pueden adherirse a las microesferas indirectamente mediante dichos materiales recubiertos en superficie.

10 En un aspecto, la divulgación proporciona una composición que comprende: un soporte sólido, en el que dicho soporte sólido comprende una pluralidad de oligonucleótidos, en el que al menos dos de dicha pluralidad de oligonucleótidos comprenden un marcador celular y un marcador molecular, en el que dichos marcadores celulares de dichos al menos dos de dicha pluralidad de oligonucleótidos son iguales, y en el que dichos marcadores moleculares de dichos al menos dos de dicha pluralidad de oligonucleótidos son diferentes. En algunas realizaciones, la pluralidad de oligonucleótidos comprende además un marcador de muestra. En algunas realizaciones, la pluralidad de oligonucleótidos comprende además una región de unión a diana. En algunas realizaciones, la región de unión a diana comprende una secuencia que está adaptada para hibridar con un ácido nucleico diana. En algunas realizaciones, la región de unión a diana comprende una secuencia seleccionada del grupo que consiste en: un multímero aleatorio, por ejemplo, un dímero, trímero, tetrámero, pentámero, hexámero, heptámero octámero, nonámero, decámero aleatorio o secuencia sólida multimérica superior de cualquier longitud; un cebador específico de gen; y oligodT; o cualquier combinación de los mismos. En algunas realizaciones, la pluralidad de oligonucleótidos comprende un marcador universal. En algunas realizaciones, el marcador universal comprende un sitio de unión para un cebador de secuenciación. En algunas realizaciones, la pluralidad de oligonucleótidos comprende un conector. En algunas realizaciones, el conector comprende un grupo funcional. En algunas realizaciones, el conector está ubicado 5' a dicho oligonucleótido. En algunas realizaciones, el conector se selecciona del grupo que consiste en: C6, biotina, estreptavidina, aminos primarias, aldehídos y cetonas, o cualquier combinación de los mismos. En algunas realizaciones, el soporte sólido está compuesto de poliestireno. En algunas realizaciones, el soporte sólido es magnético. En algunas realizaciones, el soporte sólido se selecciona del grupo que consiste en: un soporte sólido de PDMS, un soporte sólido de vidrio, un soporte sólido de polipropileno, un soporte sólido de agarosa, un soporte sólido de gelatina, un soporte sólido magnético y un soporte sólido de Pluronic, o cualquier combinación de los mismos. En algunas realizaciones, el soporte sólido comprende una microesfera. En algunas realizaciones, el soporte sólido comprende un diámetro de aproximadamente 20 micrómetros. En algunas realizaciones, el soporte sólido comprende un diámetro de aproximadamente 5 micrómetros a aproximadamente 40 micrómetros. En algunas realizaciones, el soporte sólido comprende un grupo funcional. En algunas realizaciones, el grupo funcional se selecciona del grupo que consiste en: C6, biotina, estreptavidina, aminos primarias, aldehídos y cetonas, o cualquier combinación de los mismos. En algunas realizaciones, el marcador celular comprende una pluralidad de marcadores celulares. En algunas realizaciones, la pluralidad de marcadores celulares está intercalada con una pluralidad de secuencias de marcador de conector. En algunas realizaciones, la pluralidad de oligonucleótidos comprende de 10 000 a 1000 millones de oligonucleótidos.

40 En un aspecto, la divulgación proporciona un soporte sólido que comprende: un primer oligonucleótido que comprende: un primer marcador celular que comprende una primera secuencia aleatoria, una segunda secuencia aleatoria y una primera secuencia de marcador de conector, en el que dicha primera secuencia de marcador de conector conecta dicha primera secuencia aleatoria y dicha segunda secuencia aleatoria; y un primer marcador molecular que comprende una secuencia aleatoria; y un segundo oligonucleótido que comprende: un segundo marcador celular que comprende una tercera secuencia aleatoria, una cuarta secuencia aleatoria y una segunda secuencia de marcador de conector, en el que dicha segunda secuencia de marcador de conector conecta dicha tercera secuencia aleatoria y dicha cuarta secuencia aleatoria; y un segundo marcador molecular que comprende una secuencia aleatoria, en el que dicho primer marcador celular y dicho segundo marcador celular son iguales y dicho primer marcador molecular y dicho segundo marcador molecular son diferentes. En algunas realizaciones, el primer y segundo oligonucleótido comprenden además regiones de índice de muestra idénticas. En algunas realizaciones, la región de índice de muestra comprende una secuencia aleatoria. En algunas realizaciones, la región de índice de muestra es de 4-12 nucleótidos de longitud. En algunas realizaciones, el marcador celular se adhiere directamente a dicho marcador molecular. En algunas realizaciones, el marcador celular y dicho marcador molecular se adhieren mediante una secuencia de marcador de conector. En algunas realizaciones, la secuencia aleatoria de dicho marcador celular es de 4-12 nucleótidos de longitud. En algunas realizaciones, la secuencia constante de dicho marcador celular es de al menos 4 nucleótidos de longitud. En algunas realizaciones, el marcador celular tiene una longitud total de al menos 12 nucleótidos. En algunas realizaciones, el marcador celular comprende además una o más secuencias aleatorias adicionales. En algunas realizaciones, el marcador celular comprende además una o más secuencias de marcador conector adicionales. En algunas realizaciones, la una o más secuencias de marcador conector adicionales conectan la una o más secuencias aleatorias adicionales. En algunas realizaciones, la secuencia aleatoria del marcador molecular es de 4-12 nucleótidos de longitud.

60 En un aspecto, la divulgación proporciona una composición que comprende: un soporte sólido, en el que dicho soporte sólido comprende una pluralidad de oligonucleótidos, en el que al menos dos de dicha pluralidad de oligonucleótidos comprenden: un marcador celular, un marcador molecular; y una región de unión a diana; y una pluralidad de ácidos nucleicos diana, en el que dichos marcadores celulares de dichos al menos dos de dicha

pluralidad de oligonucleótidos son iguales, y en el que dichos marcadores moleculares de dichos al menos dos de dicha pluralidad de oligonucleótidos son diferentes. En algunas realizaciones, la región de unión a diana comprende una secuencia que está adaptada para hibridar con al menos uno de dicha pluralidad de ácidos nucleicos diana. En algunas realizaciones, la región de unión a diana comprende una secuencia seleccionada del grupo que consiste en: un multímero aleatorio, por ejemplo, un dímero, trímero, tetrámero, pentámero, hexámero, heptámero octámero, nonámero, decámero aleatorio o secuencia multimérica superior de cualquier longitud; un cebador específico de gen; un oligodT; o cualquier combinación de los mismos. En algunas realizaciones, la pluralidad de oligonucleótidos comprende 10 000 a 1000 millones de oligonucleótidos. En algunas realizaciones, la pluralidad de oligonucleótidos comprende un número de oligonucleótidos mayor que el número de ácidos nucleicos diana de dicha pluralidad de ácidos nucleicos diana. En algunas realizaciones, la pluralidad de ácidos nucleicos diana comprende múltiples copias de un mismo ácido nucleico diana. En algunas realizaciones, la pluralidad de ácidos nucleicos diana comprende múltiples copias de diferentes ácidos nucleicos diana. En algunas realizaciones, la pluralidad de ácidos nucleicos diana se une a dicha pluralidad de oligonucleótidos. En algunas realizaciones, el oligonucleótido comprende además un marcador de muestra. En algunas realizaciones, la pluralidad de oligonucleótidos comprende un marcador universal. En algunas realizaciones, el marcador universal comprende un sitio de unión para un cebador de secuenciación. En algunas realizaciones, la pluralidad de oligonucleótidos comprende un conector. En algunas realizaciones, el conector comprende un grupo funcional. En algunas realizaciones, el conector está ubicado 5' a dicho oligonucleótido. En algunas realizaciones, el grupo funcional comprende un grupo amino. En algunas realizaciones, el conector se selecciona del grupo que consiste en: C6, biotina, estreptavidina, aminas primarias, aldehídos y cetonas, o cualquier combinación de los mismos. En algunas realizaciones, el soporte sólido está compuesto de poliestireno. En algunas realizaciones, el soporte sólido es magnético. En algunas realizaciones, el soporte sólido se selecciona del grupo que consiste en: un soporte sólido de PDMS, un soporte sólido de vidrio, un soporte sólido de polipropileno, un soporte sólido de agarosa, un soporte sólido de gelatina, un soporte sólido magnético y un soporte sólido de Pluronic, o cualquier combinación de los mismos. En algunas realizaciones, el soporte sólido comprende una microesfera. En algunas realizaciones, el soporte sólido comprende un diámetro de aproximadamente 20 micrómetros. En algunas realizaciones, el soporte sólido comprende un diámetro de aproximadamente 5 micrómetros a aproximadamente 40 micrómetros. En algunas realizaciones, el soporte sólido comprende un grupo funcional. En algunas realizaciones, el grupo funcional comprende un grupo carboxi. En algunas realizaciones, el grupo funcional se selecciona del grupo que consiste en: C6, biotina, estreptavidina, aminas primarias, aldehídos y cetonas, o cualquier combinación de los mismos. En algunas realizaciones, el marcador celular comprende una pluralidad de marcadores celulares. En algunas realizaciones, la pluralidad de marcadores celulares está intercalada con una pluralidad de secuencias de marcador de conector.

En un aspecto, la divulgación proporciona un kit que comprende: un primer soporte sólido, en el que dicho primer soporte sólido comprende una primera pluralidad de oligonucleótidos, en el que dicha primera pluralidad de oligonucleótidos comprende el mismo primer marcador celular, un segundo soporte sólido, en el que dicho segundo soporte sólido comprende una segunda pluralidad de oligonucleótidos, en el que dicha segunda pluralidad de oligonucleótidos comprende el mismo segundo marcador celular, instrucciones para su uso, en el que dicho primer marcador celular y dicho segundo marcador celular son diferentes. En algunas realizaciones, los oligonucleótidos forman dicha primera pluralidad de oligonucleótidos y dicha segunda pluralidad de oligonucleótidos comprende un marcador molecular. En algunas realizaciones, los marcadores moleculares de dichos oligonucleótidos son diferentes. En algunas realizaciones, los marcadores moleculares de dichos oligonucleótidos son iguales. En algunas realizaciones, el marcador molecular de algunos de dichos oligonucleótidos es diferente y alguno es igual. En algunas realizaciones, los oligonucleótidos de dicha primera pluralidad de oligonucleótidos y dicha segunda pluralidad de oligonucleótidos comprenden una región de unión a diana. En algunas realizaciones, el kit comprende además: una matriz de micropocillos. En algunas realizaciones, el kit comprende además: un tampón. En algunas realizaciones, el tampón se selecciona del grupo que consiste en: un tampón de reconstitución, un tampón de dilución y un tampón de estabilización, o cualquier combinación de los mismos.

En un aspecto, la divulgación proporciona un método para determinar una cantidad de un ácido nucleico diana, que comprende: poner en contacto una muestra con un soporte sólido, en el que dicho soporte sólido comprende una pluralidad de oligonucleótidos, en el que al menos dos de dicha pluralidad de oligonucleótidos comprenden un marcador celular y un marcador molecular, en el que dichos marcadores celulares de dichos al menos dos de dicha pluralidad de oligonucleótidos son iguales y en el que dichos marcadores moleculares de dichos al menos dos de dicha pluralidad de oligonucleótidos son diferentes; e hibridar dicho ácido nucleico diana de dicha muestra con un oligonucleótido de dicha pluralidad de oligonucleótidos. En algunas realizaciones, la muestra comprende células. En algunas realizaciones, la muestra se lisa antes de dicha hibridación. En algunas realizaciones, la hibridación comprende hibridar múltiples copias de un mismo ácido nucleico diana con dicha pluralidad de oligonucleótidos. En algunas realizaciones, el método comprende además: amplificar dicho ácido nucleico diana. En algunas realizaciones, la amplificación comprende transcribir de forma inversa dicho ácido nucleico diana. En algunas realizaciones, la amplificación comprende la amplificación usando un método seleccionado del grupo que consiste en: PCR, PCR cuantitativa, PCR en tiempo real y PCR digital, o cualquier combinación de los mismos. En algunas realizaciones, la amplificación se realiza directamente en dicho soporte sólido. En algunas realizaciones, la amplificación se realiza en un molde transcrito a partir de dicho soporte sólido. En algunas realizaciones, el método comprende además: secuenciar dicho ácido nucleico diana. En algunas realizaciones, la secuenciación comprende secuenciar dicho ácido nucleico diana y dicho marcador molecular. En algunas realizaciones, el método comprende

además: determinar una cantidad de dicho ácido nucleico diana. En algunas realizaciones, la determinación comprende cuantificar niveles de dicho ácido nucleico diana. En algunas realizaciones, la determinación comprende contar el número de marcadores moleculares secuenciados para dicho ácido nucleico diana. En algunas realizaciones, el contacto se produce en un micropocillo. En algunas realizaciones, el micropocillo está hecho de un material seleccionado del grupo que consiste en: plástico hidrófilo, plástico, elastómero e hidrogel, o cualquier combinación de los mismos. En algunas realizaciones, el micropocillo comprende agarosa. En algunas realizaciones, el micropocillo es un micropocillo de una matriz de micropocillos. En algunas realizaciones, la matriz de micropocillos comprende al menos 90 micropocillos. En algunas realizaciones, la matriz de micropocillos comprende al menos 150 000 micropocillos. En algunas realizaciones, el micropocillo comprende al menos un soporte sólido por pocillo. En algunas realizaciones, el micropocillo comprende como mucho dos soportes sólidos por pocillo. En algunas realizaciones, el micropocillo es de un tamaño que acomoda como mucho dos de dichos sopores sólidos. En algunas realizaciones, el micropocillo es de un tamaño que acomoda como mucho un soporte sólido. En algunas realizaciones, el micropocillo es de al menos 25 micrómetros de profundidad. En algunas realizaciones, el micropocillo es de al menos 25 micrómetros de diámetro.

En un aspecto, la divulgación proporciona un método para reducir el sesgo de amplificación de un ácido nucleico diana, que comprende: poner en contacto una muestra con un soporte sólido, en el que dicho soporte sólido comprende una pluralidad de oligonucleótidos, en el que al menos dos de dicha pluralidad de oligonucleótidos comprenden un marcador celular y un marcador molecular, en el que dichos marcadores celulares de dichos al menos dos de dicha pluralidad de oligonucleótidos son iguales y en el que dichos marcadores moleculares de dichos al menos dos de dicha pluralidad de oligonucleótidos son diferentes; e hibridar un ácido nucleico diana de dicha muestra con dicha pluralidad de oligonucleótidos; amplificar dicho ácido nucleico diana; secuenciar dicho ácido nucleico diana, en el que dicha secuenciación secuencía dicho ácido nucleico diana y dicho marcador molecular de dicho oligonucleótido al que se une dicho ácido nucleico diana; y determinar una cantidad de dicho ácido nucleico diana. En algunas realizaciones, la hibridación comprende hibridar múltiples copias de un mismo ácido nucleico diana con dicha pluralidad de oligonucleótidos. En algunas realizaciones, la determinación comprende contar un número de marcadores moleculares secuenciados para un mismo ácido nucleico diana. En algunas realizaciones, el recuento cuenta el número de copias de dicho mismo ácido nucleico diana. En algunas realizaciones, la muestra comprende células. En algunas realizaciones, la amplificación comprende transcribir de forma inversa dicho ácido nucleico diana. En algunas realizaciones, la amplificación comprende la amplificación usando un método seleccionado del grupo que consiste en: PCR, PCR cuantitativa, PCR en tiempo real y PCR digital, o cualquier combinación de los mismos. En algunas realizaciones, la amplificación se realiza directamente en dicho soporte sólido. En algunas realizaciones, la amplificación se realiza en un molde transcrito a partir de dicho soporte sólido.

En un aspecto, la divulgación proporciona una composición que comprende: un micropocillo; una célula; y un soporte sólido, en la que dicho soporte sólido comprende una pluralidad de oligonucleótidos, en la que al menos dos de dicha pluralidad de oligonucleótidos comprenden un marcador celular y un marcador molecular, en la que dichos marcadores celulares de dichos al menos dos de dicha pluralidad de oligonucleótidos son iguales y en la que dichos marcadores moleculares de dichos al menos dos de dicha pluralidad de oligonucleótidos son diferentes. En algunas realizaciones, los al menos dos de dicha pluralidad de oligonucleótidos comprenden además un marcador de muestra. En algunas realizaciones, los al menos dos de dicha pluralidad de oligonucleótidos comprenden además una región de unión a diana. En algunas realizaciones, la región de unión a diana comprende una secuencia seleccionada del grupo que consiste en: un multímero aleatorio, por ejemplo, un dímero, trímero, tetrámero, pentámero, hexámero, heptámero octámero, nonámero, decámero aleatorio o secuencia multimérica superior de cualquier longitud; un cebador específico de gen; un oligodT; o cualquier combinación de los mismos. En algunas realizaciones, la pluralidad de oligonucleótidos comprende un marcador universal. En algunas realizaciones, el marcador universal comprende un sitio de unión para un cebador de secuenciación. En algunas realizaciones, el soporte sólido está compuesto de poliestireno. En algunas realizaciones, el soporte sólido es magnético. En algunas realizaciones, el soporte sólido se selecciona del grupo que consiste en: un soporte sólido de PDMS, un soporte sólido de vidrio, un soporte sólido de polipropileno, un soporte sólido de agarosa, un soporte sólido de gelatina, un soporte sólido magnético y un soporte sólido de Pluronic, o cualquier combinación de los mismos. En algunas realizaciones, el soporte sólido comprende una microesfera. En algunas realizaciones, el soporte sólido comprende un diámetro de aproximadamente 20 micrómetros. En algunas realizaciones, el soporte sólido comprende un diámetro de aproximadamente 5 micrómetros a aproximadamente 40 micrómetros. En algunas realizaciones, el marcador celular comprende una pluralidad de marcadores celulares. En algunas realizaciones, la pluralidad de marcadores celulares está intercalada con una pluralidad de secuencias de marcador de conector. En algunas realizaciones, el micropocillo está hecho de un material seleccionado del grupo que consiste en: plástico hidrófilo, plástico, elastómero e hidrogel, o cualquier combinación de los mismos. En algunas realizaciones, el micropocillo comprende agarosa. En algunas realizaciones, el micropocillo es un micropocillo de una matriz de micropocillos. En algunas realizaciones, la matriz de micropocillos comprende al menos 90 micropocillos. En algunas realizaciones, el micropocillo comprende al menos un soporte sólido por pocillo. En algunas realizaciones, el micropocillo comprende como mucho dos soportes sólidos por pocillo. En algunas realizaciones, el micropocillo es de un tamaño que acomoda al menos uno de dichos sopores sólidos y al menos una de dichas células. En algunas realizaciones, el micropocillo es de un tamaño que acomoda como mucho uno de dichos soportes sólidos y al menos una de dichas células. En algunas realizaciones, el micropocillo es de al menos 25 micrómetros de profundidad. En algunas realizaciones, el micropocillo es de al menos 25 micrómetros de diámetro. En algunas realizaciones, el micropocillo

es plano.

5 En un aspecto, la divulgación proporciona un dispositivo que comprende: un primer sustrato que comprende una primera matriz de micropocillos; en el que dicha primera matriz de micropocillos comprende una pluralidad de primeros micropocillos en una primera disposición espacial predeterminada configurada para realizar marcaje estocástico combinado de células individuales y ensayos de indización molecular.

10 En algunas realizaciones, el dispositivo comprende un primer sustrato que comprende al menos una segunda matriz de micropocillos, en la que dicha al menos segunda matriz de micropocillos comprende una pluralidad de al menos segundos micropocillos en al menos una segunda disposición espacial predeterminada. En algunas realizaciones, los primeros micropocillos y los al menos segundos micropocillos son iguales. En algunas realizaciones, los primeros micropocillos y los al menos segundos micropocillos son diferentes. En algunas realizaciones, la primera disposición espacial predeterminada y la al menos segunda disposición espacial predeterminada son iguales. En algunas realizaciones, la primera disposición espacial predeterminada y la al menos segunda disposición espacial predeterminada son diferentes. En algunas realizaciones, una disposición espacial predeterminada comprende un patrón de matriz unidimensional o bidimensional. En algunas realizaciones, el patrón de matriz bidimensional comprende una rejilla cuadrada, una rejilla rectangular o una rejilla hexagonal. En algunas realizaciones, los micropocillos comprenden una geometría cilíndrica, una geometría cónica, una geometría hemisférica, una geometría rectangular, una geometría poliédrica o una combinación de las mismas. En algunas realizaciones, un diámetro de los micropocillos es entre aproximadamente 5 micrómetros y aproximadamente 50 micrómetros. En algunas realizaciones, una profundidad de los micropocillos es entre aproximadamente 10 micrómetros y aproximadamente 60 micrómetros. En algunas realizaciones, un espaciado de centro a centro entre dos micropocillos adyacentes es entre aproximadamente 15 micrómetros y aproximadamente 75 micrómetros. En algunas realizaciones, un número total de micropocillos en una primera o al menos segunda matriz de micropocillos es entre aproximadamente 96 y aproximadamente 5 000 000. En algunas realizaciones, el primer sustrato comprende silicio, sílice fundida, vidrio, un polímero, un metal o una combinación de los mismos. En algunas realizaciones, el primer sustrato comprende además agarosa o un hidrogel. En algunas realizaciones, una matriz de micropocillos comprende además al menos una característica de superficie, en la que dicha característica de superficie rodea uno más micropocillos individuales o traspasa una superficie entre micropocillos individuales, y en la que dicha característica de superficie está abovedada, estriada o puntiaguda.

30 En un aspecto, la divulgación proporciona un dispositivo que comprende: un primer sustrato que comprende al menos una primera matriz de micropocillos; y una fijación mecánica que comprende una placa superior, una placa inferior y una junta; en el que cuando el primer sustrato y la fijación mecánica están en forma ensamblada, el primer sustrato está ubicado entre la junta y la placa inferior, la junta forma un sellamiento a prueba de filtraciones con el primer sustrato y la placa superior y la junta forman al menos una primera cámara que abarca dicha al menos primera matriz de micropocillos de modo que puede distribuirse una muestra celular y un marcador oligonucleotídico basado en microesfera en dicha al menos primera cámara para realizar marcaje estocástico combinado de células individuales y ensayos de indización molecular.

40 En algunas realizaciones, la al menos primera matriz de micropocillos es cualquiera descrita en la presente memoria. En algunas realizaciones, la junta se fabrica de polidimetilsiloxano (PDMS) o un material elastomérico similar. En algunas realizaciones, las placas superior e inferior se fabrican de aluminio, aluminio anodizado, acero inoxidable, teflón, polimetilmetacrilato, policarbonato o un material polimérico rígido similar.

45 En un aspecto, la divulgación proporciona un dispositivo que comprende: al menos un sustrato que comprende además al menos una matriz de micropocillos; y una celda de flujo; en el que la celda de flujo encierra o está adherida a dicho al menos un sustrato, e incluye al menos un acceso de entrada y al menos un acceso de salida con el fin de suministrar los fluidos a dichas matrices de micropocillos; y en el que dispositivo está configurado para realizar marcaje estocástico combinado de células individuales y ensayos de indización molecular.

50 En algunas realizaciones, dicho al menos un sustrato comprende al menos una matriz de micropocillos como se describe en la presente memoria. En algunas realizaciones, la celda de flujo comprende además una pluralidad de cámaras de micromatriz que entran en contacto con una pluralidad de matrices de micropocillos de modo que puede procesarse una o más muestras en paralelo. En algunas realizaciones, la celda de flujo comprende además una barrera porosa o difusor de flujo para proporcionar suministro más uniforme de las células y las microesferas a la al menos una matriz de micropocillos. En algunas realizaciones, la celda de flujo comprende además divisores que dividen cada cámara que contiene una matriz de micropocillos en subsecciones que cubren colectivamente la misma zona de matriz total y proporciona suministro más uniforme de células y microesferas a la al menos una matriz de micropocillos. En algunas realizaciones, la anchura de los canales de fluido incorporados en el dispositivo está entre aproximadamente 50 micrómetros y 20 mm. En algunas realizaciones, la profundidad de los canales de fluido incorporados en el dispositivo es entre aproximadamente 50 micrómetros y aproximadamente 2 mm. En algunas realizaciones, la celda de flujo se fabrica de un material seleccionado del grupo que consiste en silicio, sílice fundida, vidrio, polidimetilsiloxano (PDMS; elastómero), polimetilmetacrilato (PMMA), policarbonato (PC), polipropileno (PP), polietileno (PE), polietileno de alta densidad (HDPE), poliimida, polímeros de olefina cíclicos (COP), copolímeros de olefina cíclicos (COC), tereftalato de polietileno (PET), resina epoxi, metal o una combinación de estos materiales. En algunas realizaciones, el dispositivo comprende un componente fijo de un sistema instrumental configurado para

realizar marcaje estocástico combinado de células individuales automatizado y ensayos de indización molecular. En algunas realizaciones, el dispositivo comprende un componente desmontable de un sistema instrumental configurado para realizar marcaje estocástico combinado de células individuales automatizado y ensayos de indización molecular.

- 5 En un aspecto, la divulgación proporciona un cartucho que comprende: al menos un primer sustrato que comprende además al menos una primera matriz de micropocillos; al menos una primera celda de flujo o cámara de matriz de micropocillos; uno o más depósitos de muestra o reactivo; y en el que el cartucho comprende además al menos un acceso de entrada y al menos un acceso de salida con el fin de suministrar fluidos a dicha al menos primera matriz de micropocillos; y en el que el cartucho está configurado para realizar marcaje estocástico combinado de células individuales y ensayos de indización molecular.

10 En algunas realizaciones, dicho al menos primer sustrato comprende al menos una primera matriz de micropocillos como se describe en la presente memoria. En algunas realizaciones, el cartucho comprende una pluralidad de matrices de micropocillos y está configurado para procesar una o más muestras en paralelo. En algunas realizaciones, la al menos primera celda de flujo o cámara de matriz de micropocillos comprende además una barrera porosa o difusor de flujo para proporcionar suministro más uniforme de células y microesferas a las al menos primeras matrices de micropocillos. En algunas realizaciones, la al menos primera celda de flujo o cámara de matriz de micropocillos comprende además divisores que dividen la al menos primera celda de flujo o cámara de matriz de micropocillos en subsecciones que cubren colectivamente la misma zona de matriz total y proporcionan suministro más uniforme de células y microesferas a las matrices de micropocillos. En algunas realizaciones, la anchura de los canales de fluido incorporados en el cartucho está entre aproximadamente 50 micrómetros y 200 micrómetros. En algunas realizaciones, la anchura de los canales de fluido incorporados en el cartucho está entre aproximadamente 200 micrómetros y 2 mm. En algunas realizaciones, la anchura de los canales de fluido incorporados en el cartucho está entre aproximadamente 2 mm y 10 mm. En algunas realizaciones, la anchura de los canales de fluido incorporados en el cartucho está entre aproximadamente 10 mm y 20 mm. En algunas realizaciones, la profundidad de los canales de fluido incorporados en el cartucho está entre aproximadamente 50 micrómetros y aproximadamente 2 mm. En algunas realizaciones, la profundidad de los canales de fluido incorporados en el cartucho está entre aproximadamente 500 micrómetros y 1 mm. En algunas realizaciones, la profundidad de los canales de fluido incorporados en el cartucho está entre aproximadamente 1 mm y aproximadamente 2 mm. En algunas realizaciones, la una o más celdas de flujo o cámaras de matriz de micropocillos se fabrican de un material seleccionado del grupo que consiste en silicio, sílice fundida, vidrio, polidimetilsiloxano (PDMS; elastómero), polimetilmetacrilato (PMMA), policarbonato (PC), polipropileno (PP), polietileno (PE), polietileno de alta densidad (HDPE), poliimida, polímeros de olefina cíclicos (COP), copolímeros de olefina cíclicos (COC), tereftalato de polietileno (PET), resina epoxi, metal o una combinación de estos materiales. En algunas realizaciones, el dispositivo comprende un componente consumible, desmontable de un sistema instrumental configurado para realizar marcaje estocástico combinado de células individuales automatizado y ensayos de indización molecular. En algunas realizaciones, el cartucho comprende además canales de desviación u otras características de diseño para proporcionar automedición de las muestras de células o suspensiones de microesferas dispensadas o inyectadas en el cartucho. En algunas realizaciones, el cartucho comprende además bombas en miniatura integradas para controlar el flujo de fluido a través del dispositivo. En algunas realizaciones, el cartucho comprende además válvulas en miniatura integradas para compartimentar los reactivos precargados y para controlar el flujo de fluido a través del dispositivo. En algunas realizaciones, el cartucho comprende además ventilaciones para proporcionar una trayectoria de escape para el aire atrapado. En algunas realizaciones, el cartucho comprende además elementos de diseño para crear barreras físicas o químicas que aumentan de forma eficaz la longitud de trayectoria y evitan o minimizar la difusión de moléculas entre micropocillos, en el que los elementos de diseño se seleccionan del grupo que consiste en: un patrón de canales en serpentina para suministrar células y microesferas a la al menos primera matriz de micropocillos, una platina retráctil o membrana deformable que se presiona en contacto con la superficie de la al menos primera matriz de micropocillos, o la liberación de un fluido hidrófobo inmiscible desde un depósito dentro del cartucho. En algunas realizaciones, el cartucho comprende además componentes integrados de control de la temperatura o una conexión térmica integrada para proporcionar un buen contacto térmico con un sistema instrumental externo. En algunas realizaciones, el cartucho comprende además una interfaz óptica o ventana para tomar imágenes ópticas de la al menos primera matriz de micropocillos. En algunas realizaciones, el cartucho comprende además una o más cámaras desmontables de recogida de muestras que están configuradas para conectar con termocicladores de PCR y/o instrumentos de secuenciación independientes. En algunas realizaciones, el propio cartucho está configurado para conectar directamente con termocicladores de PCR y/o instrumentos de secuenciación independientes.

55 En un aspecto, la divulgación proporciona un sistema instrumental que comprende: al menos una primera celda de flujo o cartucho que comprende además al menos una primera matriz de micropocillos; y un controlador de flujo; en el que el controlador de flujo controla el suministro de muestras de células, reactivos de marcaje oligonucleotídicos basados en microesfera y otros reactivos de ensayo a la al menos primera matriz de micropocillos, y el sistema instrumental está configurado para realizar marcaje estocástico combinado de células individuales y ensayos de indización molecular.

60 En algunas realizaciones, la al menos primera matriz de micropocillos como se describe en la presente memoria. En algunas realizaciones, la al menos primera celda de flujo es un componente fijo del sistema. En algunas

realizaciones, la al menos primera celda de flujo es un componente consumible desmontable del sistema. En algunas realizaciones, el al menos primera cartucho es un componente consumible desmontable del sistema. En algunas realizaciones, las muestras de células y reactivos oligonucleotídicos basados en microesfera se dispensan e inyectan directamente en el cartucho por el usuario. En algunas realizaciones, se precargan reactivos de ensayo diferentes de las muestras de células en el cartucho. En algunas realizaciones, el sistema instrumental comprende además un sistema de imágenes para tomar imágenes de la al menos primera matriz de micropocillos. En algunas realizaciones, el sistema instrumental comprende además un sistema de distribución de células o microesferas para facilitar la distribución uniforme de células y microesferas a través de la al menos primera matriz de micropocillos, en el que el mecanismo subyacente de dicho sistema de distribución se selecciona del grupo que consiste en balanceo, agitación, removido, flujo de recirculación, agitación de alta frecuencia o agitación de alta frecuencia. En algunas realizaciones, el sistema instrumental comprende además un sistema de lisis celular en el que el sistema usa un transductor piezoeléctrico de alta frecuencia para sonicar las células. En algunas realizaciones, el sistema instrumental comprende además un controlador de la temperatura para mantener una temperatura especificada por el usuario o para incrementar gradualmente la temperatura entre dos o más temperaturas especificadas durante dos o más intervalos de tiempo especificados. En algunas realizaciones, el sistema instrumental comprende además un controlador de campo magnético para su uso en la elución de microesferas de los micropocillos. En algunas realizaciones, el sistema instrumental comprende además un ordenador o procesador programado para proporcionar una interfaz de usuario y control de las funciones del sistema. En algunas realizaciones, el sistema instrumental comprende además código de programa para proporcionar capacidad de análisis de imágenes en tiempo real. En algunas realizaciones, el análisis de imágenes en tiempo real y las funciones de control del instrumento están acoplados, de modo que las etapas de carga de muestra de células y microesferas pueden prolongarse o repetirse hasta que se consiguen distribuciones óptimas de células/microesferas. En algunas realizaciones, el sistema instrumental comprende además un termociclador de PCR integrado para la amplificación de marcadores oligonucleotídicos. En algunas realizaciones, el sistema instrumental comprende además un secuenciador integrado para secuenciar colecciones de oligonucleótidos, proporcionando de ese modo capacidad de respuesta a la muestra. En algunas realizaciones, las muestras de células comprenden muestras de pacientes y los resultados del marcaje estocástico combinado de células individuales y ensayo de indización molecular se usan para aplicaciones de diagnóstico clínico. En algunas realizaciones, las muestras de células comprenden muestras de pacientes y los resultados del marcaje estocástico combinado de células individuales y ensayo de indización molecular se usan por un profesional sanitario para tomar decisiones fundadas sobre el tratamiento asistencial.

En un aspecto, la divulgación proporciona un programa informático que reside en un medio legible por ordenador programado para realizar una o más de las siguientes funciones de análisis de datos de secuencia: determinar el número de lecturas por gen por célula, y el número de moléculas de transcrito únicas por gen por célula; análisis de componentes principales u otro análisis estadístico para predecir los intervalos de confianza para las determinaciones del número de moléculas de transcrito por gen por célula; alineación de datos de secuencias génicas con secuencias de referencia conocidas; descodificación/separación de códigos de barras de muestra, códigos de barras celulares y códigos de barras moleculares; y agrupación automatizada de marcadores moleculares para compensar los errores de amplificación o secuenciación; en el que los datos de secuencia se generan realizando marcaje estocástico combinado de células individuales y ensayos de indización molecular.

En un aspecto, la divulgación proporciona una composición que comprende: un soporte sólido, en la que el soporte sólido comprende una pluralidad de oligonucleótidos, en la que al menos dos de la pluralidad de oligonucleótidos comprenden un marcador celular y un marcador molecular, en la que los marcadores celulares de los al menos dos de la pluralidad de oligonucleótidos son iguales, y en la que los marcadores moleculares de los al menos dos de la pluralidad de oligonucleótidos son diferentes.

En algunas realizaciones, la pluralidad de oligonucleótidos comprende además un marcador de muestra. En algunas realizaciones, la pluralidad de oligonucleótidos comprende además una región de unión a diana. En algunas realizaciones, la región de unión a diana comprende una secuencia adaptada para hibridar con un ácido nucleico diana. En algunas realizaciones, el ácido nucleico diana comprende una pluralidad de ácidos nucleicos diana que comprende al menos un 0,01 %, 0,02 %, 0,05 %, 0,1 %, 0,2 %, 0,3 %, 0,4 %, 0,5 %, 1 %, 2 %, 3 %, 4 %, 5 %, 6 %, 7 %, 8 %, 9 %, 10 %, 11 %, 12 %, 13 %, 14 %, 15 %, 16 %, 17 %, 18 %, 19 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 60 %, 70 %, 80 %, 90 % o 100 % de los transcritos de un transcriptoma de un organismo. En algunas realizaciones, el ácido nucleico diana es ADN. En algunas realizaciones, el ácido nucleico diana es ARN. En algunas realizaciones, el ácido nucleico diana es ARNm. En algunas realizaciones, el ADN es ADN genómico. En algunas realizaciones, el ADN genómico está cortado. En algunas realizaciones, el ADN genómico cortado comprende al menos un 0,01 %, 0,02 %, 0,05 %, 0,1 %, 0,2 %, 0,3 %, 0,4 %, 0,5 %, 1 %, 2 %, 3 %, 4 %, 5 %, 6 %, 7 %, 8 %, 9 %, 10 %, 11 %, 12 %, 13 %, 14 %, 15 %, 16 %, 17 %, 18 %, 19 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 60 %, 70 %, 80 %, 90 % o 100 % de los genes de un genoma de un organismo. En algunas realizaciones, la región de unión a diana comprende una secuencia seleccionada del grupo que consiste en: un multímero aleatorio, por ejemplo, un dímero, trímero, tetrámero, pentámero, hexámero, heptámero, octámero, nonámero, decámero aleatorio, o secuencia multimérica superior de cualquier longitud; un cebador específico de gen; y oligodT; o cualquier combinación de los mismos. En algunas realizaciones, la pluralidad de oligonucleótidos comprende un marcador universal. En algunas realizaciones, el marcador universal comprende un sitio de unión para un cebador de secuenciación. En algunas realizaciones, la pluralidad de oligonucleótidos comprende un conector. En algunas

realizaciones, el conector comprende un grupo funcional. En algunas realizaciones, el conector está ubicado 5' del oligonucleótido. En algunas realizaciones, el conector se selecciona del grupo que consiste en: C6, biotina, estreptavidina, aminas primarias, aldehídos y cetonas, o cualquier combinación de los mismos. En algunas realizaciones, el soporte sólido está compuesto de poliestireno. En algunas realizaciones, el soporte sólido es magnético. En algunas realizaciones, el soporte sólido se selecciona del grupo que consiste en: un soporte sólido de PDMS, un soporte sólido de vidrio, un soporte sólido de polipropileno, un soporte sólido de agarosa, un soporte sólido de gelatina, un soporte sólido magnético y un soporte sólido de pluronic, o cualquier combinación de los mismos. En algunas realizaciones, el soporte sólido comprende una microesfera. En algunas realizaciones, el soporte sólido comprende un diámetro de aproximadamente 20 micrómetros. En algunas realizaciones, el soporte sólido comprende un diámetro de aproximadamente 5 micrómetros a aproximadamente 40 micrómetros. En algunas realizaciones, el soporte sólido comprende un grupo funcional. En algunas realizaciones, el grupo funcional se selecciona del grupo que consiste en: C6, biotina, estreptavidina, aminas primarias, aldehídos y cetonas, o cualquier combinación de los mismos. En algunas realizaciones, el marcador celular comprende una pluralidad de marcadores celulares. En algunas realizaciones, la pluralidad de marcadores celulares está intercalada con una pluralidad de secuencias de marcador de conector. En algunas realizaciones, la pluralidad de oligonucleótidos comprende de 10 000 a 1000 millones de oligonucleótidos. En algunas realizaciones, la pluralidad de oligonucleótidos comprende de 10 000 a 1000 millones de regiones de unión a diana. En algunas realizaciones, la pluralidad de oligonucleótidos comprende de 10 000 a 1000 millones de regiones de unión a diana diferentes. En algunas realizaciones, la pluralidad de oligonucleótidos comprende de 10 000 a 1000 millones de regiones de unión a diana iguales. En algunas realizaciones, las diferentes regiones de unión a diana pueden hibridar con al menos un 0,01 %, 0,02 %, 0,05 %, 0,1 %, 0,2 %, 0,3 %, 0,4 %, 0,5 %, 1 %, 2 %, 3 %, 4 %, 5 %, 6 %, 7 %, 8 %, 9 %, 10 %, 11 %, 12 %, 13 %, 14 %, 15 %, 16 %, 17 %, 18 %, 19 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 60 %, 70 %, 80 %, 90 % o 100 % de los transcritos de un transcriptoma de un organismo. En algunas realizaciones, las diferentes regiones de unión a diana pueden hibridar con al menos un 0,01 %, 0,02 %, 0,05 %, 0,1 %, 0,2 %, 0,3 %, 0,4 %, 0,5 %, 1 %, 2 %, 3 %, 4 %, 5 %, 6 %, 7 %, 8 %, 9 %, 10 %, 11 %, 12 %, 13 %, 14 %, 15 %, 16 %, 17 %, 18 %, 19 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 60 %, 70 %, 80 %, 90 % o 100 % de los transcritos de un transcriptoma de un organismo.

En un aspecto, la divulgación proporciona una composición que comprende: un soporte sólido, en la que el soporte sólido comprende una pluralidad de oligonucleótidos, en la que al menos dos de la pluralidad de oligonucleótidos comprende un marcador celular y un marcador molecular, en la que los marcadores celulares de los al menos dos de la pluralidad de oligonucleótidos son iguales, y en la que los marcadores moleculares de los al menos dos de la pluralidad de oligonucleótidos son diferentes.

En algunas realizaciones, la pluralidad de oligonucleótidos comprende además un marcador de muestra. En algunas realizaciones, la pluralidad de oligonucleótidos comprende además una región de unión a diana. En algunas realizaciones, la región de unión a diana comprende una secuencia adaptada para hibridar con un ácido nucleico diana. En algunas realizaciones, el ácido nucleico diana comprende una pluralidad de ácidos nucleicos diana que comprende al menos un 0,01 %, 0,02 %, 0,05 %, 0,1 %, 0,2 %, 0,3 %, 0,4 %, 0,5 %, 1 %, 2 %, 3 %, 4 %, 5 %, 6 %, 7 %, 8 %, 9 %, 10 %, 11 %, 12 %, 13 %, 14 %, 15 %, 16 %, 17 %, 18 %, 19 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 60 %, 70 %, 80 %, 90 % o 100 % de los transcritos de un transcriptoma de un organismo. En algunas realizaciones, el ácido nucleico diana comprende ADN genómico cortado, en el que el ADN genómico cortado comprende al menos un 0,01 %, 0,02 %, 0,05 %, 0,1 %, 0,2 %, 0,3 %, 0,4 %, 0,5 %, 1 %, 2 %, 3 %, 4 %, 5 %, 6 %, 7 %, 8 %, 9 %, 10 %, 11 %, 12 %, 13 %, 14 %, 15 %, 16 %, 17 %, 18 %, 19 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 60 %, 70 %, 80 %, 90 % o 100 % de los genes de un genoma de un organismo. En algunas realizaciones, la región de unión a diana comprende un oligodT. En algunas realizaciones, los al menos dos de la pluralidad de oligonucleótidos comprende un primer oligonucleótido y un segundo oligonucleótido, en el que el primer oligonucleótido comprende un primer marcador celular y un primer marcador molecular, en el que el primer marcador celular comprende una primera secuencia aleatoria, una segunda secuencia aleatoria, y una primera secuencia de marcador de conector, en el que la primera secuencia de marcador de conector conecta la primera secuencia aleatoria y la segunda secuencia aleatoria; y el primer marcador molecular comprende una secuencia aleatoria; y el segundo oligonucleótido comprende un segundo marcador celular y un segundo marcador molecular, en el que el segundo marcador celular comprende una tercera secuencia aleatoria, una cuarta secuencia aleatoria, y una segunda secuencia de marcador de conector, en el que la segunda secuencia de marcador de conector conecta la tercera secuencia aleatoria y la cuarta secuencia aleatoria; y el segundo marcador molecular comprende una secuencia aleatoria, y en el que el primer marcador celular y el segundo marcador celular son iguales y el primer marcador molecular y el segundo marcador molecular son diferentes.

En un aspecto, la divulgación proporciona un kit que comprende cualquier composición descrita en la presente memoria e instrucciones para su uso.

En un aspecto, la divulgación proporciona un método, que comprende: poner en contacto una muestra con un soporte sólido, en el que el soporte sólido comprende una pluralidad de oligonucleótidos, en el que al menos dos de la pluralidad de oligonucleótidos comprenden un marcador celular y un marcador molecular, en el que los marcadores celulares de los al menos dos de la pluralidad de oligonucleótidos son iguales, y en el que los marcadores moleculares de los al menos dos de la pluralidad de oligonucleótidos son diferentes; e hibridar el ácido nucleico diana de la muestra con un oligonucleótido de la pluralidad de oligonucleótidos.

En algunas realizaciones, la muestra comprende células. En algunas realizaciones, la muestra se lisa antes de la hibridación. En algunas realizaciones, la hibridación comprende hibridar múltiples copias de un mismo ácido nucleico diana con la pluralidad de oligonucleótidos. En algunas realizaciones, el método comprende además transcribir de forma inversa el ácido nucleico diana. En algunas realizaciones, el método comprende además realizar una
 5 amplificación de oligonucleótidos. En algunas realizaciones, la amplificación comprende la amplificación usando un método seleccionado del grupo que consiste en: PCR, PCR cuantitativa, PCR en tiempo real y PCR digital, o cualquier combinación de las mismas.

En un aspecto, la divulgación proporciona un soporte sólido que comprende: un primer oligonucleótido que comprende: un primer marcador celular que comprende una primera secuencia aleatoria, una segunda secuencia aleatoria, y una primera secuencia de marcador de conector, en el que la primera secuencia de marcador de conector conecta la primera secuencia aleatoria y la segunda secuencia aleatoria; y un primer marcador molecular que comprende una secuencia aleatoria; y un segundo oligonucleótido que comprende: un segundo marcador celular que comprende una tercera secuencia aleatoria, una cuarta secuencia aleatoria, y una segunda secuencia de marcador de conector, en el que la segunda secuencia de marcador de conector conecta la tercera secuencia aleatoria y la cuarta secuencia aleatoria; y un segundo marcador molecular que comprende una secuencia aleatoria, en el que el primer marcador celular y el segundo marcador celular son iguales y el primer marcador molecular y el segundo marcador molecular son diferentes. En algunas realizaciones, el primer y segundo oligonucleótidos comprenden además regiones de índice de muestra idénticas. En algunas realizaciones, la región de índice de muestra comprende una secuencia aleatoria. En algunas realizaciones, la región de índice de muestra es de 4-12 nucleótidos de longitud. En algunas realizaciones, el marcador celular se adhiere directamente al marcador molecular. En algunas realizaciones, el marcador celular y el marcador molecular se adhieren mediante una secuencia de marcador de conector. En algunas realizaciones, la secuencia aleatoria del marcador celular es de 4-12 nucleótidos de longitud. En algunas realizaciones, la secuencia constante del marcador celular es de al menos 4 nucleótidos de longitud. En algunas realizaciones, el marcador celular tiene una longitud total de al menos 12 nucleótidos. En algunas realizaciones, el marcador celular comprende además una o más secuencias aleatorias adicionales. En algunas realizaciones, el marcador celular comprende además una o más secuencias de marcador de conector adicionales. En algunas realizaciones, la una o más secuencias de marcador de conector adicionales conectan la una o más secuencias aleatorias adicionales. En algunas realizaciones, la secuencia aleatoria del marcador molecular es de 4-12 nucleótidos de longitud.

En un aspecto, la divulgación proporciona una composición que comprende: un soporte sólido, en la que el soporte sólido comprende una pluralidad de oligonucleótidos, en la que al menos dos de la pluralidad de oligonucleótidos comprende: un marcador celular, un marcador molecular; y una región de unión a diana; y una pluralidad de ácidos nucleicos diana, en la que los marcadores celulares de los al menos dos de la pluralidad de oligonucleótidos son iguales, y en la que los marcadores moleculares de los al menos dos de la pluralidad de oligonucleótidos son
 35 diferentes.

En algunas realizaciones, la región de unión a diana comprende una secuencia que está adaptada para hibridar con al menos uno de la pluralidad de ácidos nucleicos diana. En algunas realizaciones, la región de unión a diana comprende una secuencia seleccionada del grupo que consiste en: un multímero aleatorio, por ejemplo, un dímero, trímero, tetrámero, pentámero, hexámero, heptámero, octámero, nonámero, decámero aleatorio, o secuencia multimérica superior de cualquier longitud; un cebador específico de gen; y oligodT; o cualquier combinación de los mismos. En algunas realizaciones, la pluralidad de oligonucleótidos comprende de 10 000 a 1000 millones de oligonucleótidos. En algunas realizaciones, la pluralidad de oligonucleótidos comprende un número de oligonucleótidos mayor que el número de ácidos nucleicos diana de la pluralidad de ácidos nucleicos diana. En algunas realizaciones, la pluralidad de ácidos nucleicos diana comprende múltiples copias de un mismo ácido nucleico diana. En algunas realizaciones, la pluralidad de ácidos nucleicos diana comprende múltiples copias de diferentes ácidos nucleicos diana. En algunas realizaciones, la pluralidad de ácidos nucleicos diana se une a la pluralidad de oligonucleótidos. En algunas realizaciones, el oligonucleótido comprende además un marcador de muestra. En algunas realizaciones, la pluralidad de oligonucleótidos comprende un marcador universal. En algunas realizaciones, el marcador universal comprende un sitio de unión para un cebador de secuenciación. En algunas realizaciones, la pluralidad de oligonucleótidos comprende un conector. En algunas realizaciones, el conector comprende un grupo funcional. En algunas realizaciones, el conector está ubicado 5' del oligonucleótido. En algunas realizaciones, el grupo funcional comprende un grupo amino. En algunas realizaciones, el conector se selecciona del grupo que consiste en: C6, biotina, estreptavidina, aminos primarias, aldehídos y cetonas, o cualquier combinación de los mismos. En algunas realizaciones, el soporte sólido está compuesto de poliestireno. En algunas realizaciones, el soporte sólido es magnético. En algunas realizaciones, el soporte sólido se selecciona del grupo que consiste en: un soporte sólido de PDMS, un soporte sólido de vidrio, un soporte sólido de polipropileno, un soporte sólido de agarosa, un soporte sólido de gelatina, un soporte sólido magnético y un soporte sólido de pluronic, o cualquier combinación de los mismos. En algunas realizaciones, el soporte sólido comprende una microesfera. En algunas realizaciones, el soporte sólido comprende un diámetro de aproximadamente 20 micrómetros. En algunas realizaciones, el soporte sólido comprende un diámetro de aproximadamente 5 micrómetros a aproximadamente 40 micrómetros. En algunas realizaciones, el soporte sólido comprende un grupo funcional. En algunas realizaciones, el grupo funcional comprende un grupo carboxi. En algunas realizaciones, el grupo funcional se selecciona del grupo que consiste en: C6, biotina, estreptavidina, aminos primarias, aldehídos y

cetonas, o cualquier combinación de los mismos. En algunas realizaciones, el marcador celular comprende una pluralidad de marcadores celulares. En algunas realizaciones, la pluralidad de marcadores celulares está intercalada con una pluralidad de secuencias de marcador de conector.

5 En un aspecto, la divulgación proporciona un kit que comprende: un primer soporte sólido, en el que el primer soporte sólido comprende una primera pluralidad de oligonucleótidos, en el que la primera pluralidad de oligonucleótidos comprende un mismo primer marcador celular, un segundo soporte sólido, en el que el segundo soporte sólido comprende una segunda pluralidad de oligonucleótidos, en el que la segunda pluralidad de oligonucleótidos comprende un mismo segundo marcador celular, e instrucciones para su uso, en el que el primer marcador celular y el segundo marcador celular son diferentes.

10 En algunas realizaciones, los oligonucleótidos de la primera pluralidad de oligonucleótidos y la segunda pluralidad de oligonucleótidos comprenden un marcador molecular. En algunas realizaciones, el marcador molecular de los oligonucleótidos es diferente. En algunas realizaciones, el marcador molecular de los oligonucleótidos son iguales. En algunas realizaciones, el marcador molecular de algunos de los oligonucleótidos son diferentes y algunos son iguales. En algunas realizaciones, los oligonucleótidos de la primera pluralidad de oligonucleótidos y la segunda pluralidad de oligonucleótidos comprenden una región de unión a diana. En algunas realizaciones, el kit comprende además una matriz de micropocillos. En algunas realizaciones, el kit comprende además un tampón. En algunas realizaciones, el tampón se selecciona del grupo que consiste en: un tampón de reconstitución, un tampón de dilución y un tampón de estabilización, o cualquier combinación de los mismos.

20 En un aspecto, la divulgación proporciona un método para determinar una cantidad de un ácido nucleico diana, que comprende: poner en contacto una muestra con un soporte sólido, en el que el soporte sólido comprende una pluralidad de oligonucleótidos, en el que al menos dos de la pluralidad de oligonucleótidos comprenden un marcador celular y un marcador molecular, en el que los marcadores celulares de los al menos dos de la pluralidad de oligonucleótidos son iguales, y en el que los marcadores moleculares de los al menos dos de la pluralidad de oligonucleótidos son diferentes; e hibridar el ácido nucleico diana de la muestra con un oligonucleótido de la pluralidad de oligonucleótidos.

25 En algunas realizaciones, la muestra comprende células. En algunas realizaciones, la muestra se lisa antes de la hibridación. En algunas realizaciones, la hibridación comprende hibridar múltiples copias de un mismo ácido nucleico diana con la pluralidad de oligonucleótidos. En algunas realizaciones, el método comprende además amplificar el ácido nucleico diana. En algunas realizaciones, la amplificación comprende transcribir de forma inversa el ácido nucleico diana. En algunas realizaciones, la amplificación comprende amplificación usando un método seleccionado del grupo que consiste en: PCR, PCR cuantitativa, PCR en tiempo real y PCR digital, o cualquier combinación de las mismas. En algunas realizaciones, la amplificación se realiza directamente sobre el soporte sólido. En algunas realizaciones, la amplificación se realiza sobre un molde transcrito desde el soporte sólido. En algunas realizaciones, el método comprende además secuenciar el ácido nucleico diana. En algunas realizaciones, la secuenciación comprende secuenciar el ácido nucleico diana y el marcador molecular. En algunas realizaciones, el método comprende además determinar una cantidad del ácido nucleico diana. En algunas realizaciones, la determinación comprende cuantificar niveles del ácido nucleico diana. En algunas realizaciones, la determinación comprende contar el número de marcadores moleculares secuenciados para el ácido nucleico diana. En algunas realizaciones, el contacto se produce en un micropocillo. En algunas realizaciones, el micropocillo está hecho de un material seleccionado del grupo que consiste en: plástico hidrófilo, plástico, elastómero e hidrogel, o cualquier combinación de los mismos. En algunas realizaciones, el micropocillo comprende agarosa. En algunas realizaciones, el micropocillo es un micropocillo de una matriz de micropocillos. En algunas realizaciones, la matriz de micropocillos comprende al menos 90 micropocillos. En algunas realizaciones, la matriz de micropocillos comprende al menos 150 000 micropocillos. En algunas realizaciones, el micropocillo comprende al menos un soporte sólido por pocillo.

30 En algunas realizaciones, el micropocillo comprende como mucho dos soportes sólidos por pocillo. En algunas realizaciones, el micropocillo es de un tamaño que acomoda como mucho dos soportes sólidos. En algunas realizaciones, el micropocillo es de un tamaño que acomoda como mucho un soporte sólido. En algunas realizaciones, el micropocillo es de al menos 25 micrómetros de profundidad. En algunas realizaciones, el micropocillo es al menos de 25 micrómetros de diámetro.

35 En un aspecto, la divulgación proporciona un método para reducir el sesgo de amplificación de un ácido nucleico diana, que comprende: poner en contacto una muestra con un soporte sólido, en el que el soporte sólido comprende una pluralidad de oligonucleótidos, en el que al menos dos de la pluralidad de oligonucleótidos comprenden un marcador celular y un marcador molecular, en el que los marcadores celulares de los al menos dos de la pluralidad de oligonucleótidos son iguales, y en el que los marcadores moleculares de los al menos dos de la pluralidad de oligonucleótidos son diferentes; e hibridar un ácido nucleico diana de la muestra con la pluralidad de oligonucleótidos; amplificar el ácido nucleico diana o complemento del mismo, secuenciar el ácido nucleico diana o complemento del mismo, en el que la secuenciación secuenciar el ácido nucleico diana o complemento del mismo y el marcador molecular del oligonucleótido al que se une el ácido nucleico diana o complemento del mismo, determinar una cantidad del ácido nucleico diana.

40 En algunas realizaciones, la hibridación comprende hibridar múltiples copias de un mismo ácido nucleico diana con la pluralidad de oligonucleótidos. En algunas realizaciones, la determinación comprende contar un número de

5 marcadores moleculares secuenciados para un mismo ácido nucleico diana. En algunas realizaciones, el recuento cuenta el número de copias del mismo ácido nucleico diana. En algunas realizaciones, la muestra comprende células. En algunas realizaciones, la amplificación comprende transcribir de forma inversa el ácido nucleico diana. En algunas realizaciones, la amplificación comprende amplificación usando un método seleccionado del grupo que consiste en: PCR, PCR cuantitativa, PCR en tiempo real y PCR digital, o cualquier combinación de las mismas. En algunas realizaciones, la amplificación se realiza directamente sobre el soporte sólido. En algunas realizaciones, la amplificación se realiza sobre un molde transcrito del soporte sólido.

10 En un aspecto, la divulgación proporciona una composición que comprende un micropocillo; una célula; y un soporte sólido, en la que el soporte sólido comprende una pluralidad de oligonucleótidos, en la que al menos dos de la pluralidad de oligonucleótidos comprenden un marcador celular y un marcador molecular, en la que los marcadores celulares de los al menos dos de la pluralidad de oligonucleótidos son iguales, y en la que los marcadores moleculares de los al menos dos de la pluralidad de oligonucleótidos son diferentes.

15 En algunas realizaciones, los al menos dos de la pluralidad de oligonucleótidos comprenden además un marcador de muestra. En algunas realizaciones, los al menos dos de la pluralidad de oligonucleótidos comprenden además una región de unión a diana. En algunas realizaciones, la región de unión a diana comprende una secuencia seleccionada del grupo que consiste en: un multímero aleatorio, por ejemplo, un dímero, trímero, tetrámero, pentámero, hexámero, heptámero, octámero, nonámero, decámero aleatorio, o secuencia multimérica superior de cualquier longitud; un cebador específico de gen; y oligodT; o cualquier combinación de los mismos. En algunas realizaciones, la pluralidad de oligonucleótidos comprende un marcador universal. En algunas realizaciones, el marcador universal comprende un sitio de unión para un cebador de secuenciación. En algunas realizaciones, el soporte sólido está compuesto de poliestireno. En algunas realizaciones, el soporte sólido es magnético. En algunas realizaciones, el soporte sólido se selecciona del grupo que consiste en: un soporte sólido de PDMS, un soporte sólido de vidrio, un soporte sólido de polipropileno, un soporte sólido de agarosa, un soporte sólido de gelatina, un soporte sólido magnético y un soporte sólido de pluronic, o cualquier combinación de los mismos. En algunas realizaciones, el soporte sólido comprende una microesfera. En algunas realizaciones, el soporte sólido tiene un diámetro de aproximadamente 20 micrómetros. En algunas realizaciones, el soporte sólido tiene un diámetro de aproximadamente 5 micrómetros a aproximadamente 40 micrómetros. En algunas realizaciones, el marcador celular comprende una pluralidad de marcadores celulares. En algunas realizaciones, la pluralidad de marcadores celulares está intercalada con una pluralidad de secuencias de conector. En algunas realizaciones, el micropocillo está hecho de un material seleccionado del grupo que consiste en: plástico hidrófilo, plástico, elastómero e hidrogel, o cualquier combinación de los mismos. En algunas realizaciones, el micropocillo comprende agarosa. En algunas realizaciones, el micropocillo es un micropocillo de una matriz de micropocillos. En algunas realizaciones, el micropocillo comprende al menos un soporte sólido por pocillo. En algunas realizaciones, el micropocillo comprende como mucho dos soportes sólidos por pocillo. En algunas realizaciones, el micropocillo es de un tamaño que acomoda al menos uno de los soportes sólidos y al menos una de las células. En algunas realizaciones, el micropocillo es de un tamaño que acomoda como mucho uno de los soportes sólidos y al menos una de las células. En algunas realizaciones, el micropocillo es de al menos 25 micrómetros de profundidad. En algunas realizaciones, el micropocillo es de al menos 25 micrómetros de diámetro. En algunas realizaciones, el micropocillo es plano.

40 En un aspecto, la divulgación proporciona un dispositivo, que comprende una pluralidad de micropocillos, en el que la pluralidad de micropocillos comprende al menos dos micropocillos; y en el que cada micropocillo de la pluralidad de micropocillos tiene un volumen que varía de aproximadamente $1000 \mu\text{m}^3$ a aproximadamente $120\,000 \mu\text{m}^3$. En algunas realizaciones, cada micropocillo de la pluralidad de micropocillos tiene un volumen de aproximadamente $20\,000 \mu\text{m}^3$. En algunas realizaciones, la pluralidad de micropocillos comprende de aproximadamente 1000 a aproximadamente 5 000 000 de micropocillos. En algunas realizaciones, la pluralidad de micropocillos comprende de aproximadamente 100 000 a aproximadamente 200 000 micropocillos. En algunas realizaciones, los micropocillos están comprendidos en una única capa de un material. En algunas realizaciones, al menos aproximadamente un 10 % de los micropocillos comprenden además una célula. En algunas realizaciones, al menos aproximadamente un 10 % de los micropocillos comprenden además un soporte sólido que comprende una pluralidad de oligonucleótidos, en el que al menos dos de la pluralidad de oligonucleótidos comprenden un marcador celular y un marcador molecular, en el que los marcadores celulares de los al menos dos de la pluralidad de oligonucleótidos son iguales, y en el que los marcadores moleculares de los al menos dos de la pluralidad de oligonucleótidos son diferentes. En algunas realizaciones, los soportes sólidos están magnetizados.

En un aspecto, la divulgación proporciona un aparato que comprende cualquier dispositivo descrito en la presente memoria, y un manipulador de líquidos.

55 En algunas realizaciones, el manipulador de líquidos suministra líquido a la pluralidad de micropocillos en aproximadamente 1 segundo. En algunas realizaciones, el aparato suministra líquido a la pluralidad de micropocillos desde un único acceso de entrada. En algunas realizaciones, el aparato comprende además un imán. En algunas realizaciones, el aparato comprende además al menos uno de: un acceso de entrada, un acceso de salida, una bomba, una válvula, una ventilación, un depósito, una cámara de recogida de muestra, un aparato de control de la temperatura o cualquier combinación de los mismos. En algunas realizaciones, el aparato comprende la cámara de recogida de muestra, en el que la cámara de recogida de muestra se desmontable del aparato. En algunas realizaciones, el aparato comprende además un generador de imágenes óptico. En algunas realizaciones, el

generador de imágenes óptico produce una señal de salida que se usa para control del manipulador de líquidos. En algunas realizaciones, el aparato comprende además un mecanismo termociclador configurado para realizar amplificación por reacción en cadena de la polimerasa (PCR) de oligonucleótidos.

5 En un aspecto, la divulgación proporciona un método para producir un resultado de ensayo de diagnóstico clínico, que comprende producir el resultado de ensayo de diagnóstico clínico con cualquier dispositivo o aparato descrito en la presente memoria. En algunas realizaciones, el resultado de ensayo de diagnóstico clínico se transmite mediante un medio de comunicación.

10 En un aspecto, la divulgación proporciona un dispositivo que comprende: uno o más sustratos que comprenden además una o más matrices de micropocillos; en el que las matrices de micropocillos se usan para realizar marcaje estocástico combinado de células individuales y ensayos de indización molecular.

En algunas realizaciones, las matrices de micropocillos de los sustratos comprenden micropocillos dispuestos en un patrón de matriz unidimensional o bidimensional. En algunas realizaciones, el patrón de matriz bidimensional de micropocillos se selecciona del grupo que incluye una rejilla cuadrada, una rejilla rectangular o una rejilla hexagonal.

15 En algunas realizaciones, los micropocillos de las matrices de micropocillos se fabrican usando una geometría de pocillo seleccionada del grupo que incluye cilíndrica, cónica, hemisférica, rectangular o poliédrica. En algunas realizaciones, los micropocillos de las matrices de micropocillos se fabrican usando una geometría global que comprende dos o más componentes geométricos seleccionados del grupo que incluye cilíndrico, cónico, hemisférico, rectangular o poliédrico. En algunas realizaciones, el diámetro de los micropocillos en las matrices de micropocillos está entre aproximadamente 5 micrómetros y aproximadamente 50 micrómetros. En algunas realizaciones, la profundidad de los micropocillos en las matrices de micropocillos está entre aproximadamente 10 micrómetros y aproximadamente 60 micrómetros. En algunas realizaciones, el espaciado de un centro a otro entre micropocillos en las matrices de micropocillos está entre aproximadamente 15 micrómetros y aproximadamente 75 micrómetros. En algunas realizaciones, el número total de micropocillos en cada una de las matrices de micropocillos está entre aproximadamente 96 y aproximadamente 5 000 000. En algunas realizaciones, el uno o más sustratos se fabrican de un material seleccionado del grupo que incluye silicio, sílice fundida, vidrio, un polímero o un metal. En algunas realizaciones, el uno o más sustratos se fabrican de agarosa o un hidrogel. En algunas realizaciones, las matrices de micropocillos comprenden además características de superficie entre micropocillos que rodean los micropocillos o traspasan la superficie entre micropocillos, y se seleccionan del grupo que incluye características de superficie abovedadas, estriadas o puntiagudas.

30 En un aspecto, la divulgación proporciona un dispositivo que comprende: un sustrato que comprende además una o más matrices de micropocillos; y una fijación mecánica que comprende una placa superior, una placa inferior y una junta; en el que cuando se ensamblan, el sustrato se ubica entre la junta y la placa inferior, la junta forma un sellamiento a prueba de filtraciones con el sustrato y la placa superior y la junta forman una o más cámaras que abarcan las matrices de micropocillos de modo que una o más muestras de células y marcadores oligonucleotídicos basados en microesfera pueden dispensarse en las cámaras con el fin de realizar marcaje estocástico combinado de células individuales y ensayos de indización molecular.

40 En algunas realizaciones, el sustrato comprende una cualquiera o más matrices de micropocillos como se describe en la presente memoria. En algunas realizaciones, la junta se fabrica de polidimetilsiloxano (PDMS) o un material elastomérico similar. En algunas realizaciones, las placas superior e inferior se fabrican de aluminio, aluminio anodizado, acero inoxidable, teflón, polimetilmetacrilato, policarbonato o un material polimérico rígido similar.

45 En un aspecto, la divulgación proporciona un dispositivo que comprende: uno o más sustratos que comprenden además una o más matrices de micropocillos; y una o más celdas de flujo; en el que la una o más celdas de flujo encierran o se adhieren al uno o más sustratos, e incluyen al menos un acceso de entrada y al menos un acceso de salida con el fin de suministrar fluidos a las matrices de micropocillos; y en el que el dispositivo se usa para realizar marcaje estocástico combinado de células individuales y ensayos de indización molecular.

50 En algunas realizaciones, el uno o más sustratos comprenden una cualquiera o más matrices de micropocillos como se describe en la presente memoria. En algunas realizaciones, cada una de la una o más celdas de flujo comprende además una pluralidad de cámaras de micromatriz que entra en contacto con una pluralidad de matrices de micropocillos de modo que puede procesarse una o más muestras en paralelo. En algunas realizaciones, la una o más celdas de flujo comprenden además una barrera porosa o difusor de flujo para proporcionar suministro más uniforme de células y microesferas a las matrices de micropocillos. En algunas realizaciones, la una o más celdas de flujo comprenden además divisores que dividen las cámaras que contienen matrices de micropocillos en subsecciones que cubren colectivamente la misma zona de matriz total y proporcionan suministro más uniforme de células y microesferas a las matrices de micropocillos. En algunas realizaciones, la anchura de los canales de fluido incorporados en el dispositivo está entre aproximadamente 50 micrómetros y 20 mm. En algunas realizaciones, la profundidad de los canales de fluido incorporados en el dispositivo está entre aproximadamente 50 micrómetros y aproximadamente 2 mm. En algunas realizaciones, la una o más celdas de flujo se fabrican de un material seleccionado del grupo que consiste en silicio, sílice fundida, vidrio, polidimetilsiloxano (PDMS; elastómero), polimetilmetacrilato (PMMA), policarbonato (PC), polipropileno (PP), polietileno (PE), polietileno de alta densidad

(HDPE), poliimida, polímeros de olefina cíclicos (COP), copolímeros de olefina cíclicos (COC), tereftalato de polietileno (PET), resina epoxi, metal o una combinación de estos materiales. En algunas realizaciones, el dispositivo comprende un componente fijo de un sistema instrumental para realizar marcaje estocástico combinado de células individuales automatizado y ensayos de indización molecular. En algunas realizaciones, el dispositivo comprende un componente desmontable de un sistema instrumental para realizar marcaje estocástico combinado de células individuales automatizado y ensayos de indización molecular.

En un aspecto, la divulgación proporciona un cartucho que comprende: uno o más sustratos que comprenden además una o más matrices de micropocillos; una o más celdas de flujo o cámaras de matriz de micropocillos; uno o más depósitos de muestra o reactivo; y en el que el cartucho comprende además al menos un acceso de entrada y al menos un acceso de salida con el fin de suministrar fluidos a las matrices de micropocillos; y en el que el cartucho se usa para realizar marcaje estocástico combinado de células individuales y ensayos de indización molecular.

En algunas realizaciones, el uno o más sustratos comprenden una cualquiera o más matrices de micropocillos como se describe en la presente memoria. En algunas realizaciones, la una o más celdas de flujo o cámaras de matriz de micropocillos entra en contacto con una pluralidad de matrices de micropocillos de modo que puede procesarse una o más muestras en paralelo. En algunas realizaciones, la una o más celdas de flujo o cámaras de matriz de micropocillos comprenden además una barrera porosa o difusor de flujo para proporcionar suministro más uniforme de células y microesferas a las matrices de micropocillos. En algunas realizaciones, la una o más celdas de flujo o cámaras de matriz de micropocillos comprenden además divisores que dividen las celdas de flujo o cámaras en subsecciones que cubren colectivamente la misma zona de matriz total y proporcionan suministro más uniforme de células y microesferas a las matrices de micropocillos. En algunas realizaciones, la anchura de los canales de fluido incorporados en el cartucho está entre aproximadamente 50 micrómetros y 200 micrómetros. En algunas realizaciones, la anchura de los canales de fluido incorporados en el cartucho está entre aproximadamente 200 micrómetros y 2 mm. En algunas realizaciones, la anchura de los canales de fluido incorporados en el cartucho está entre aproximadamente 2 mm y 10 mm. En algunas realizaciones, la anchura de los canales de fluido incorporados en el cartucho está entre aproximadamente 10 mm y 20 mm. En algunas realizaciones, la profundidad de los canales de fluido incorporados en el cartucho está entre aproximadamente 50 micrómetros y aproximadamente 10 mm. En algunas realizaciones, la profundidad de los canales de fluido incorporados en el cartucho está entre aproximadamente 500 micrómetros y 1 mm. En algunas realizaciones, la profundidad de los canales de fluido incorporados en el cartucho está entre aproximadamente 1 mm y aproximadamente 2 mm. En algunas realizaciones, la una o más celdas de flujo o cámaras de matriz de micropocillos se fabrican de un material seleccionado del grupo que consiste en silicio, sílice fundida, vidrio, polidimetilsiloxano (PDMS; elastómero), polimetilmetacrilato (PMMA), policarbonato (PC), polipropileno (PP), polietileno (PE), polietileno de alta densidad (HDPE), poliimida, polímeros de olefina cíclicos (COP), copolímeros de olefina cíclicos (COC), tereftalato de polietileno (PET), resina epoxi, metal o una combinación de estos materiales. En algunas realizaciones, el dispositivo comprende un componente consumible desmontable de un sistema instrumental para realizar marcaje estocástico combinado de células individuales automatizado y ensayos de indización molecular. En algunas realizaciones, el cartucho comprende además canales de desviación u otras características de diseño para proporcionar automedición de muestras de células o suspensiones de microesferas dispensadas o inyectadas en el cartucho. En algunas realizaciones, el cartucho comprende además bombas en miniatura integradas para controlar el flujo de fluido a través del dispositivo. En algunas realizaciones, el cartucho comprende además válvulas en miniatura integradas para compartimentar reactivos precargados y para controlar el flujo de fluido a través del dispositivo. En algunas realizaciones, el cartucho comprende además ventilaciones para proporcionar una trayectoria de escape para el aire atrapado. En algunas realizaciones, el cartucho comprende además elementos de diseño para crear barreras físicas o químicas que aumentan de forma eficaz la longitud de trayectoria y evitar o minimizar la difusión de moléculas entre micropocillos, en el que los elementos de diseño se seleccionan del grupo que consiste en: un patrón de canales en serpentín para el suministro de células y microesferas a la matriz de micropocillos, una platina retráctil o membrana deformable que se presiona en contacto con la superficie de la matriz de micropocillos, o la liberación de un fluido hidrófobo inmiscible desde un depósito dentro del cartucho. En algunas realizaciones, el cartucho comprende además componentes integrados de control de la temperatura o una conexión térmica integrada para proporcionar buen contacto térmico con un sistema instrumental externo. En algunas realizaciones, el cartucho comprende además una interfaz óptica o ventana para tomar imágenes ópticas de la una o más matrices de micropocillos. En algunas realizaciones, el cartucho comprende además una o más cámaras de recogida de muestras desmontables que están configuradas para conectar con termocicladores de PCR o instrumentos de secuenciación independientes. En algunas realizaciones, el propio cartucho está configurado para conectar directamente con termocicladores de PCR y/o instrumentos de secuenciación independientes.

En un aspecto, la divulgación proporciona un sistema instrumental que comprende: una o más celdas de flujo o cartuchos que comprenden además una o más matrices de micropocillos; y un controlador de flujo; en el que el controlador de flujo controla el suministro de muestras de células, reactivos de marcaje oligonucleotídicos basados en microesfera y otros reactivos de ensayo a las matrices de micropocillos, y el sistema instrumental se usa para realizar marcaje estocástico combinado de células individuales y ensayos de indización molecular.

En algunas realizaciones, la una o más matrices de micropocillos son cualquiera descrita en la presente memoria. En algunas realizaciones, la una o más celdas de flujo son un componente fijo del sistema. En algunas realizaciones, la una o más celdas de flujo son un componente consumible desmontable del sistema. En algunas

realizaciones, el uno o más cartuchos son componentes consumibles desmontables del sistema. En algunas realizaciones, las muestras de células y reactivos oligonucleotídicos basados en microesfera se dispensan o inyectan directamente en el cartucho por el usuario. En algunas realizaciones, se precargan reactivos de ensayo diferentes de las muestras de células en el cartucho. En algunas realizaciones, el sistema instrumental comprende además un sistema de imágenes para tomar imágenes de las matrices de micropocillos. En algunas realizaciones, el sistema instrumental comprende además un sistema de distribución de células o microesferas para facilitar la distribución uniforme de células y microesferas entre las matrices de micropocillos, en el que el mecanismo subyacente del sistema de distribución se selecciona del grupo que consiste en balanceo, agitación, removido, flujo de recirculación, agitación de baja frecuencia o agitación de alta frecuencia. En algunas realizaciones, el sistema instrumental comprende además un sistema de lisis celular en el que el sistema usa un transductor piezoeléctrico de alta frecuencia para sonicar las células. En algunas realizaciones, el sistema instrumental comprende además un controlador de la temperatura para mantener una temperatura especificada por el usuario, o para incrementar gradualmente la temperatura entre dos o más temperaturas especificadas durante dos o más intervalos de tiempo especificados. En algunas realizaciones, el sistema instrumental comprende además un controlador de campo magnético para su uso en eluir las microesferas de los micropocillos. En algunas realizaciones, el sistema instrumental comprende además un ordenador o procesador programado para proporcionar una interfaz de usuario y control de las funciones del sistema. En algunas realizaciones, el sistema instrumental comprende además código de programa para proporcionar capacidad de análisis de imágenes en tiempo real. En algunas realizaciones, el análisis de imágenes en tiempo real y las funciones de control del instrumento están acoplados, de modo que las etapas de carga de muestra de células y microesferas pueden prolongarse o repetirse hasta que se consiguen distribuciones óptimas de células/microesferas. En algunas realizaciones, el sistema instrumental comprende además un termociclador de PCR integrado para la amplificación de marcadores oligonucleotídicos. En algunas realizaciones, el sistema instrumental comprende además un secuenciador integrado para secuenciar colecciones de oligonucleótidos, proporcionando de ese modo capacidad de respuesta a muestra. En algunas realizaciones, las muestras de células comprenden muestras de pacientes y los resultados del marcaje estocástico combinado de células individuales y ensayo de indización molecular se usan para aplicaciones de diagnóstico clínico. En algunas realizaciones, las muestras de células comprenden muestras de pacientes y los resultados del marcaje estocástico combinado de células individuales y ensayo de indización molecular se usan por un profesional sanitario para tomar decisiones fundadas sobre el tratamiento asistencial.

En un aspecto, la divulgación proporciona programa informático que reside en un medio legible por ordenador programado para realizar uno o más de los siguientes análisis de datos de secuencia: determinar el número de lecturas por gen por célula, y el número de moléculas de transcrito únicas por gen por célula; análisis de componentes principales u otro análisis estadístico para predecir intervalos de confianza para determinaciones del número de moléculas de transcrito por gen por célula; alineación de datos de secuencias génicas con secuencias de referencia conocidas; descodificación/separación de códigos de barras de muestra, códigos de barras celulares y códigos de barras moleculares; y agrupación automatizada de marcadores moleculares para compensar los errores de amplificación o secuenciación; en el que los datos de secuencia se generan realizando marcaje estocástico combinado de células individuales y ensayos de indización molecular.

40

45

50

REIVINDICACIONES

1. Un método que comprende:
 - (a) distribuir una muestra que comprende una pluralidad de células en cualquier matriz de micropocillos;
 - 5 (b) poner en contacto una célula individual en un único micropocillo con un único soporte sólido, en el que el soporte sólido es una microesfera que comprende una pluralidad de oligonucleótidos, comprendiendo cada oligonucleótido un marcador celular, un marcador molecular y una región de unión a diana;
 - (c) hibridar o ligar un ácido nucleico diana de la célula individual a la región de unión a diana de un oligonucleótido de la pluralidad de oligonucleótidos para generar un ácido nucleico diana marcado;
 - (d) opcionalmente amplificar el ácido nucleico diana marcado o complemento del mismo; y
 - 10 (e) determinar una cantidad del ácido nucleico diana o complemento del mismo detectando y/o secuenciando el ácido nucleico diana marcado, amplicón o complemento del mismo, y contando el número de marcadores moleculares para el ácido nucleico diana,

en el que el marcador celular comprende una secuencia de ácido nucleico que proporciona información para la célula que están en contacto con el oligonucleótido, y en el que el marcador celular de cada uno de la pluralidad de

 - 15 oligonucleótidos en una microesfera es igual, y en el que los marcadores celulares de los oligonucleótidos en una primera microesfera son diferentes de los marcadores celulares de los oligonucleótidos en una segunda microesfera,

en el que el marcador molecular comprende una secuencia de ácido nucleico que proporciona información de identificación para el ácido nucleico diana marcado, y en el que cada marcador molecular de cada uno de la pluralidad de oligonucleótidos en una microesfera es diferente, y

 - 20 en el que la región de unión a diana comprende una secuencia de ácido nucleico que puede unirse al ácido nucleico diana.
2. El método de la reivindicación 1, en el que dicha matriz de micropocillos comprende al menos 1000 micropocillos, opcionalmente en el que cada uno de al menos aproximadamente un 10 % de dichos al menos 1000 micropocillos contiene una célula individual.
 - 25 3. El método de la reivindicación 1 o reivindicación 2, en el que la distribución comprende poner en contacto uno o más micropocillos con una disolución diluida de células diluidas de modo que se distribuye como mucho una célula al uno o más micropocillos.
 4. El método de una cualquiera de las reivindicaciones 1 a 3, que comprende, además:

distribuir como mucho una microesfera por micropocillo; y/o

 - 30 medir el transcriptoma de la célula individual en la muestra, opcionalmente en el que, después de la hibridación o ligamiento, la microesfera comprende aproximadamente un 0,501 % a aproximadamente un 100 % de los transcritos de un transcriptoma de un organismo o complementos de los mismos.
5. El método de una cualquiera de las reivindicaciones 1 a 4, en el que dicha muestra de células

comprende una célula seleccionada del grupo que consiste en: una célula infrecuente, una célula de un tumor, una célula de un ser humano, una célula de un tejido multicelular y una célula que comprende un virus, y cualquier combinación de las mismas, y/o

 - 35 comprende una pluralidad de diferentes tipos de células.
6. El método de una cualquiera de las reivindicaciones 1 a 5, en el que las células en los micropocillos se lisan, opcionalmente en el que las células se lisan antes de dicha hibridación o ligamiento.
 - 40 7. El método de una cualquiera de las reivindicaciones 1 a 6, que comprende además recuperar la microesfera del micropocillo, en el que opcionalmente la recuperación se realiza usando un imán.
 8. El método de una cualquiera de las reivindicaciones 1 a 7, en el que la región de unión a diana comprende una secuencia seleccionada del grupo que consiste en: una secuencia de oligodT, una secuencia específica de gen y una secuencia multimérica aleatoria.
 - 45 9. El método de una cualquiera de las reivindicaciones 1 a 8, que comprende además amplificar el ácido nucleico diana o complemento del mismo, opcionalmente en el que la amplificación se realiza directamente sobre la microesfera, sobre un molde transcrito de la microesfera o una combinación de los mismos.
 10. El método de la reivindicación 9, en el que la amplificación comprende transcribir de forma inversa el ácido

nucleico diana, opcionalmente para producir moléculas de ADNc marcadas.

11. El método de la reivindicación 10, que comprende además realizar síntesis de segunda hebra sobre dichos ADNc diana marcados, generando de ese modo polinucleótidos diana marcados bicatenarios.

5 12. El método de una cualquiera de las reivindicaciones 9 a 11, en el que la amplificación emplea un método seleccionado del grupo que consiste en: PCR, PCR con cebadores internos, PCR cuantitativa, PCR en tiempo real, PCR digital y cualquier combinación de las mismas.

13. El método de una cualquiera de las reivindicaciones 1 a 12, en el que cada uno de la pluralidad de oligonucleótidos comprende además un marcador universal, opcionalmente en el que el marcador universal comprende un primer sitio de unión.

10 14. El método de una cualquiera de las reivindicaciones 1 a 13, en el que dicha pluralidad de oligonucleótidos comprende de 10 000 a 1000 millones de oligonucleótidos.

15. El método de una cualquiera de las reivindicaciones 1 a 14, en el que el ácido nucleico diana es un ARNm.

15

20

25

30

35

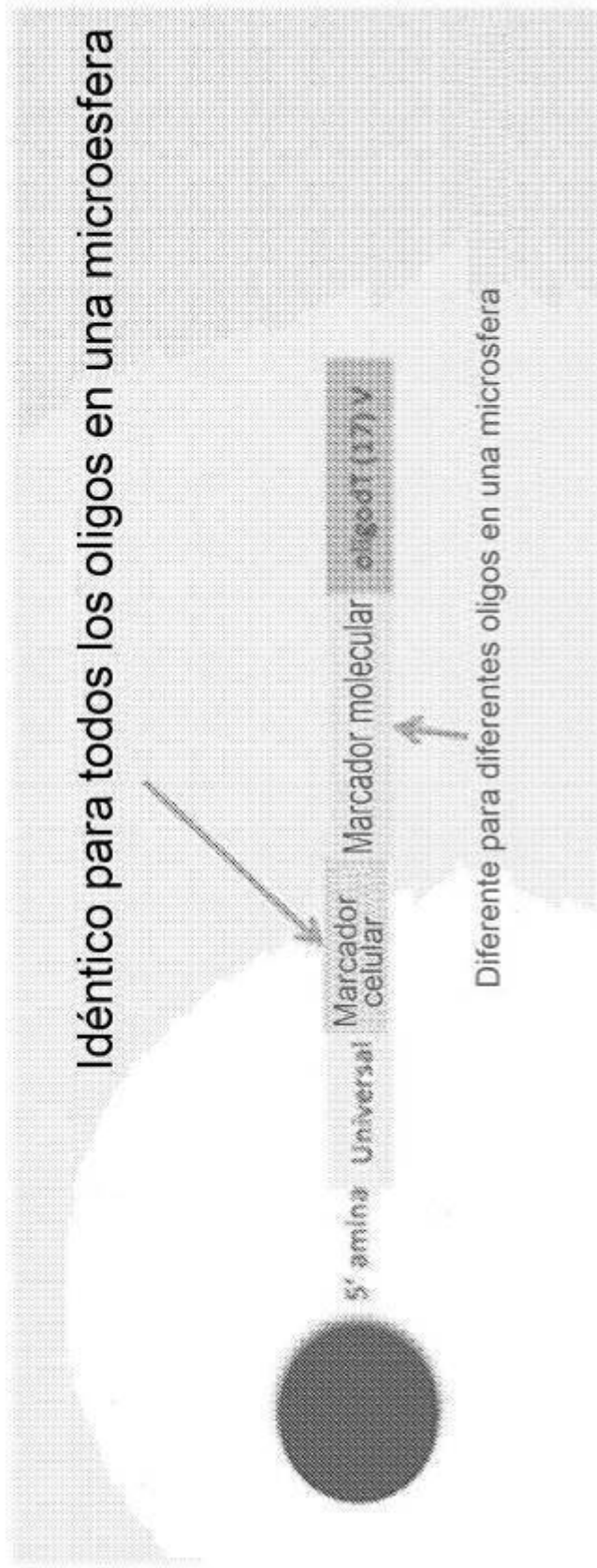


FIG. 1

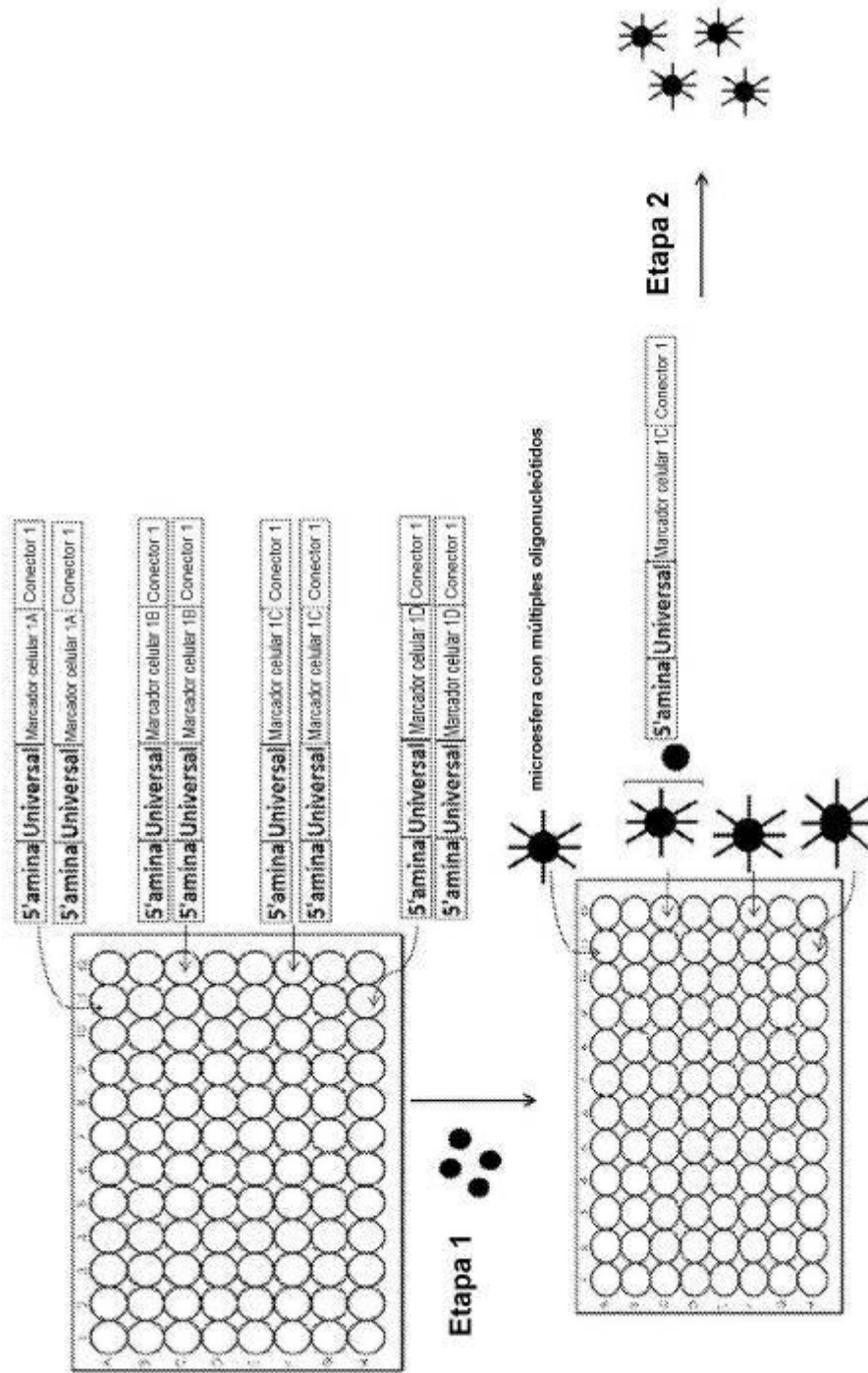


FIG. 2A

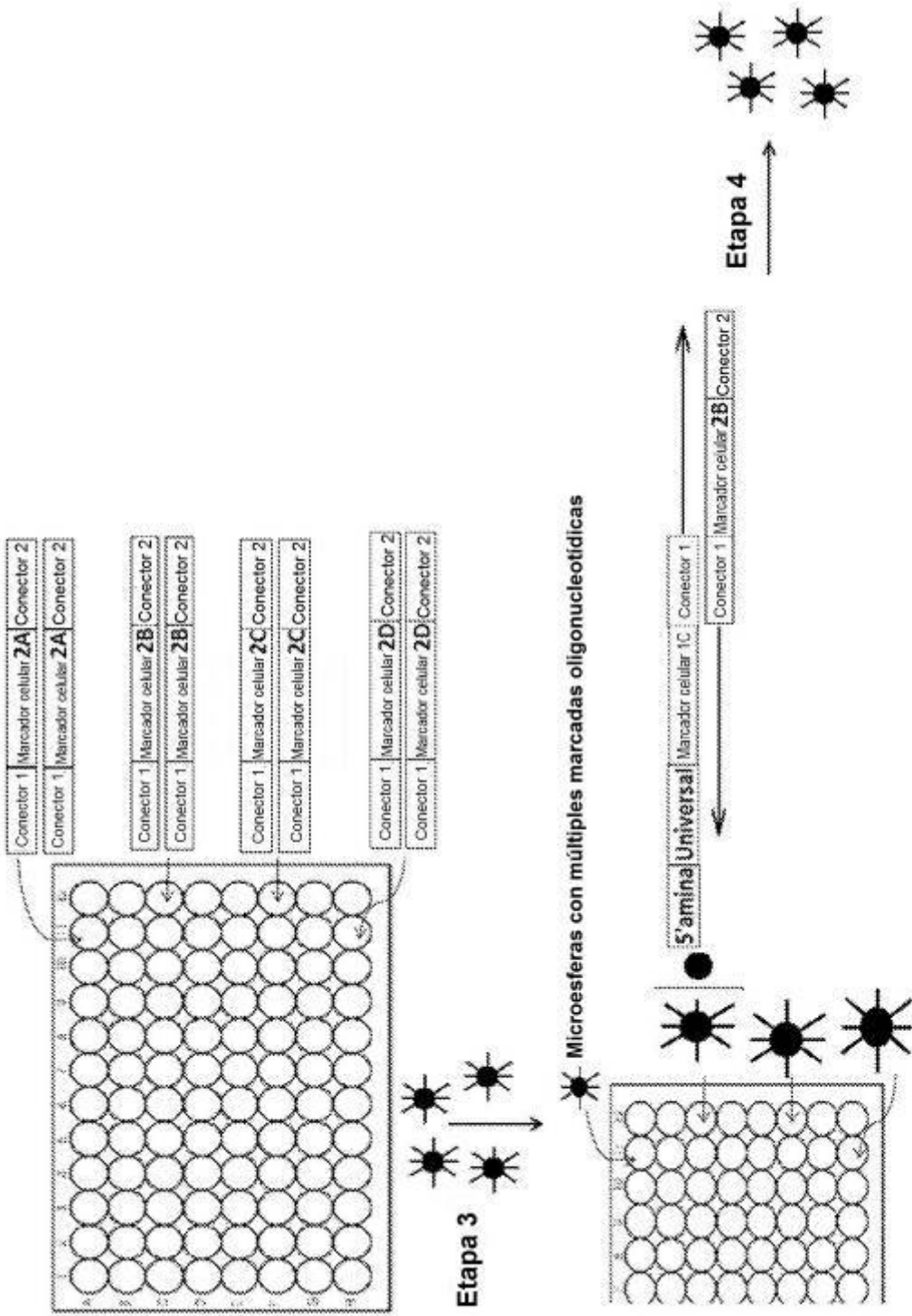


FIG. 2B

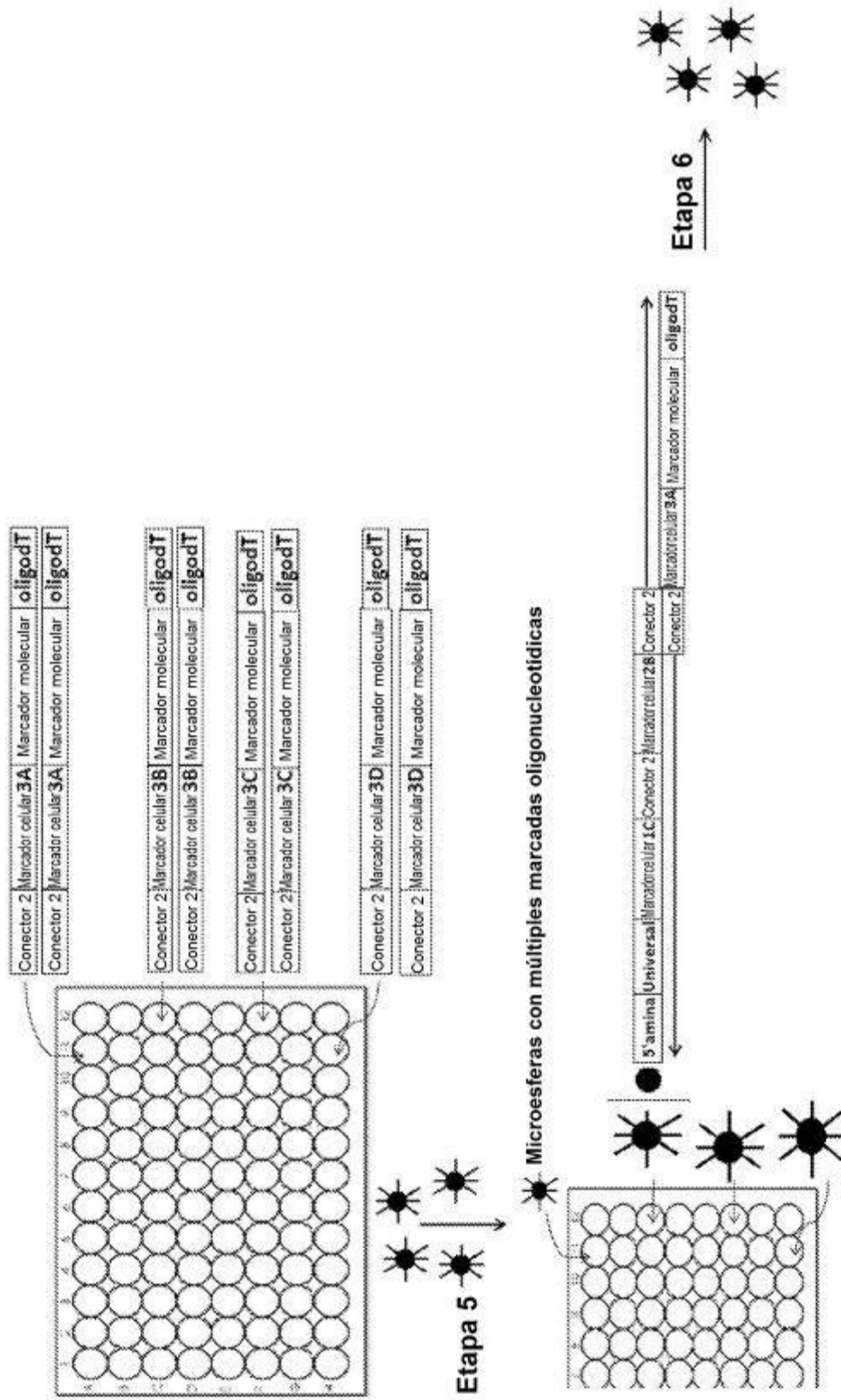


FIG. 2C

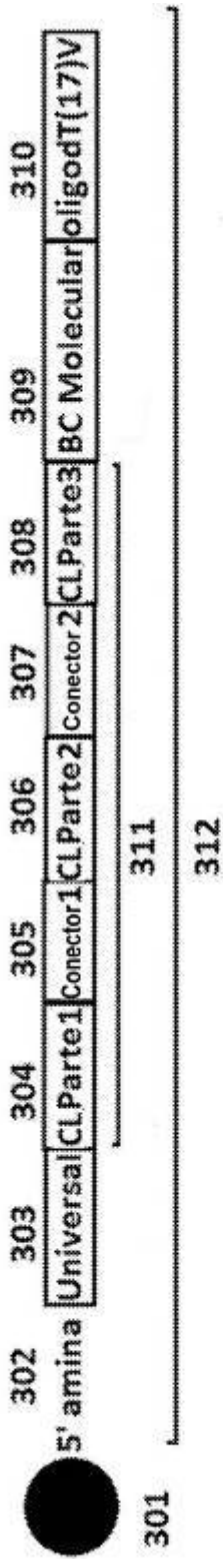


FIG. 3

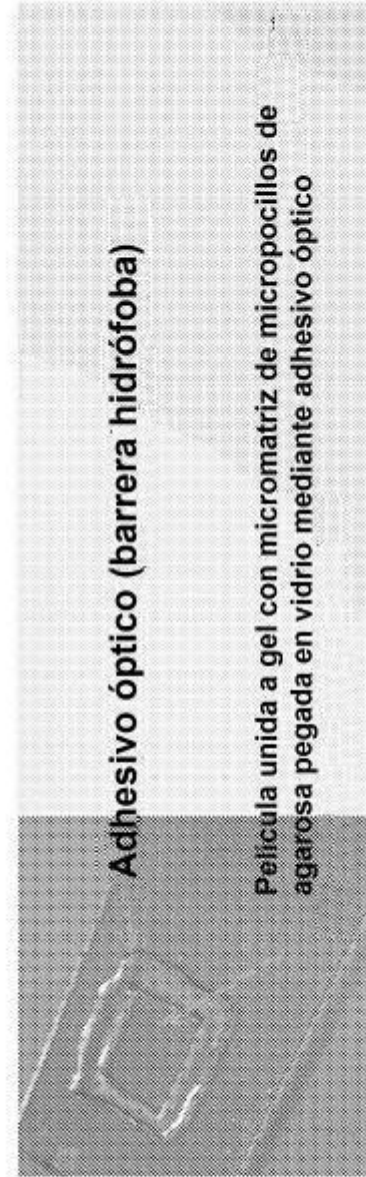


FIG. 4

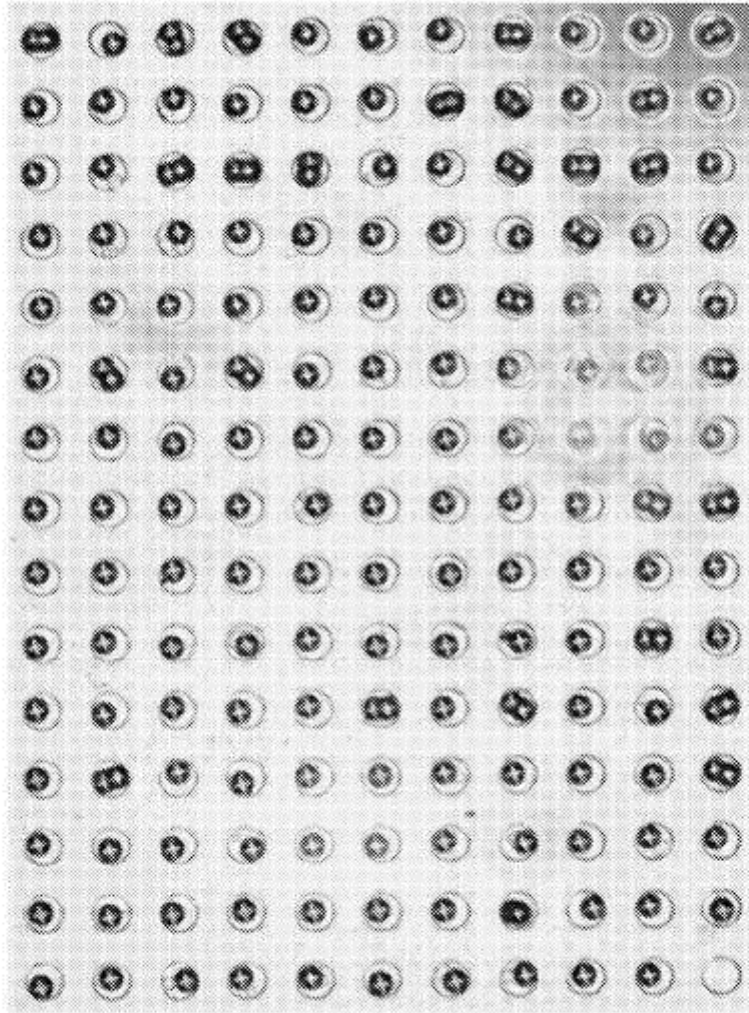


FIG. 5

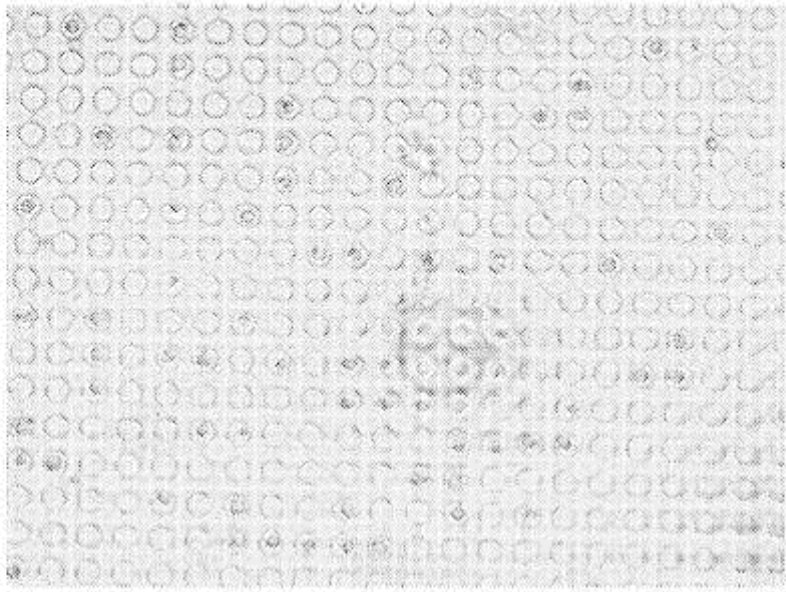


FIG. 6A

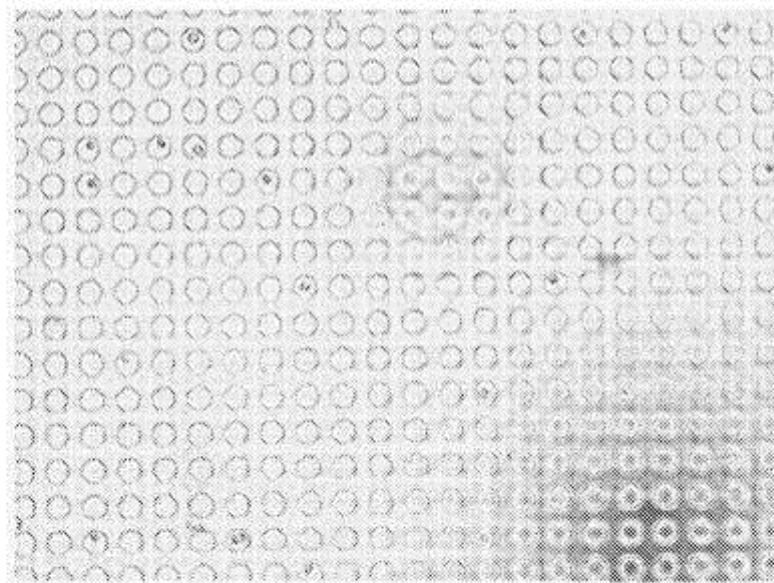


FIG. 6B

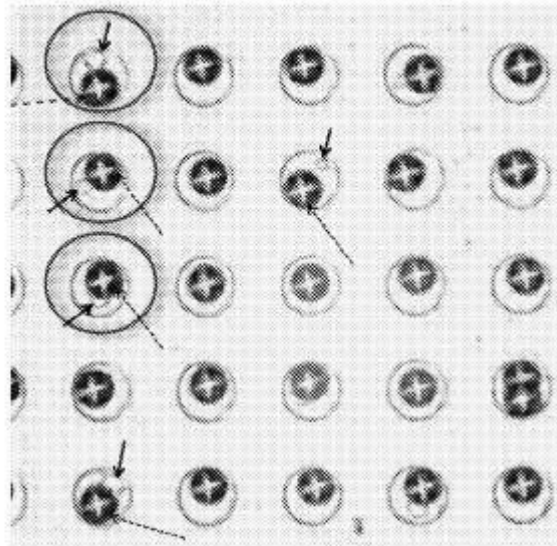


FIG. 6C

volumen del pocillo	2.12E-14	m ³
volumen de la microesfera	4.19E-15	m ³
volumen del pocillo - volumen de microesfera	1.70E-14	m ³
volumen del pocillo - volumen de microesfera	1.70E-11	L
n.º de ARNm por célula	2.50E+05	moléculas
concentración de ARNm	1.47E+16	moléculas por L
concentración de ARNm	2.44E-08	moles por L
concentración de ARNm	24.40	nM

FIG. 7

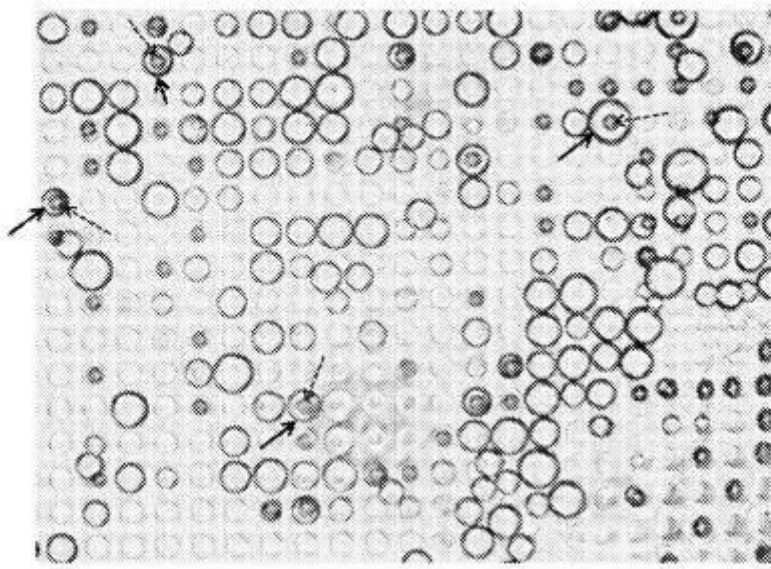


FIG. 8A

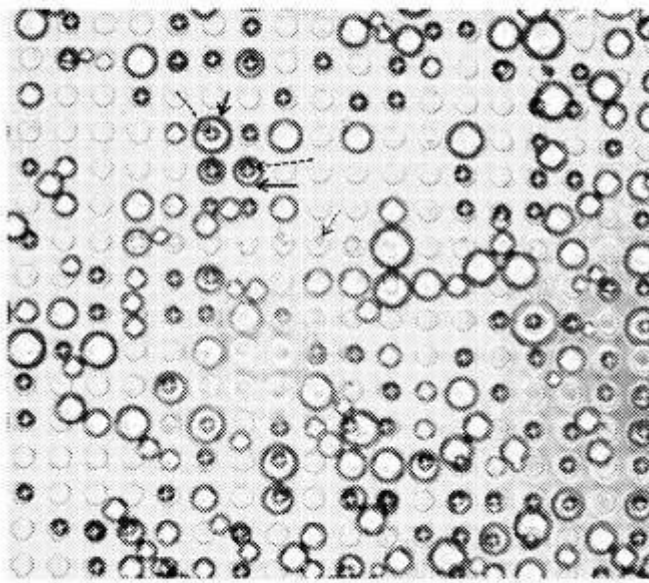


FIG. 8B



FIG. 8C

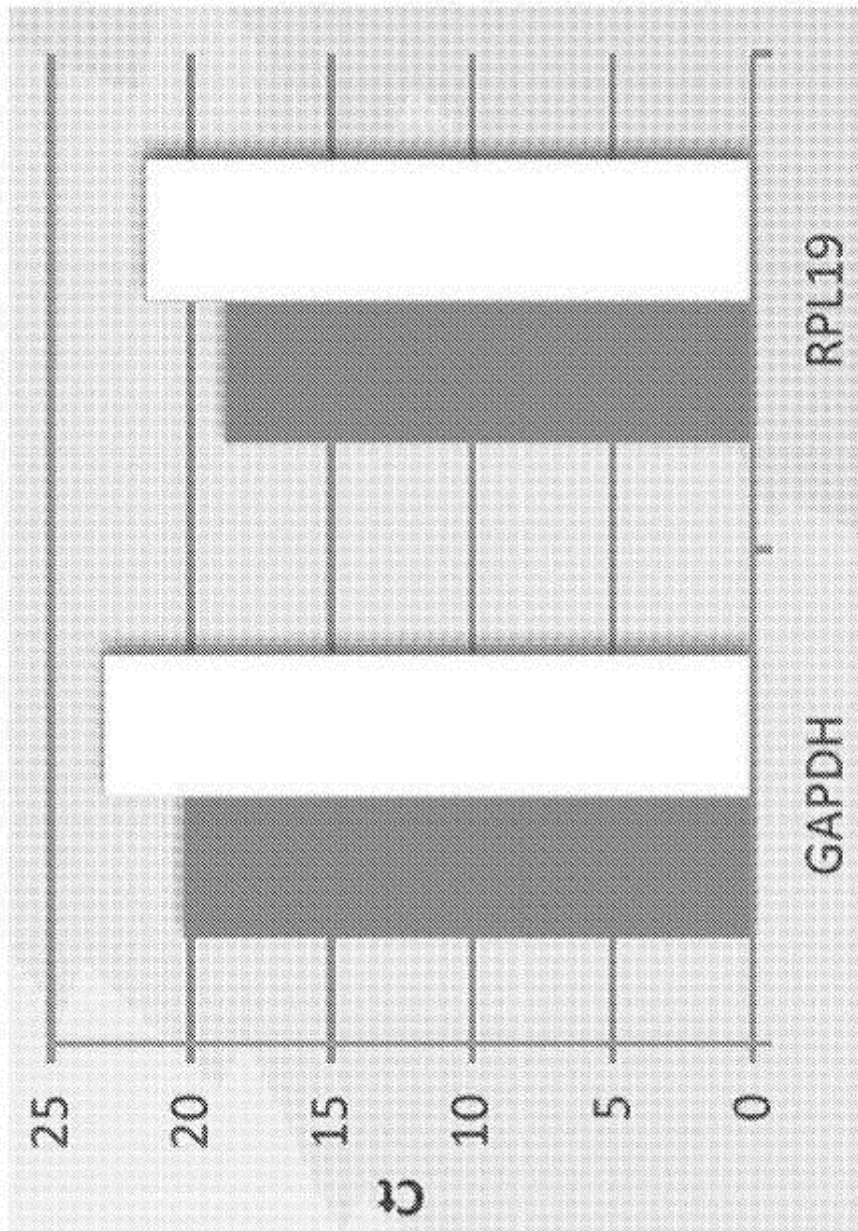


FIG. 9



FIG. 10

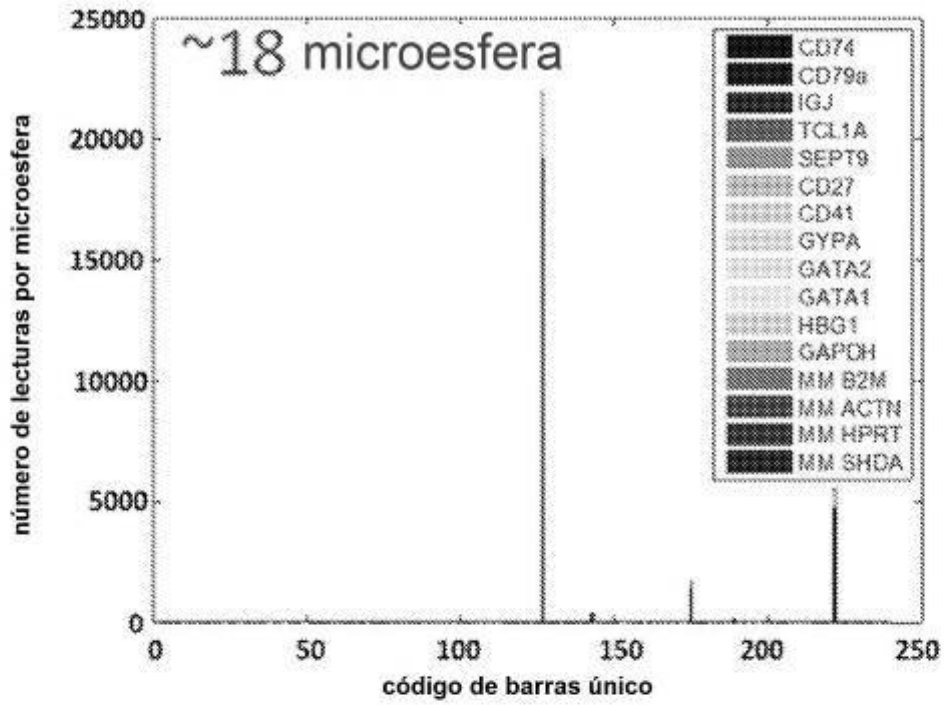


FIG. 11A

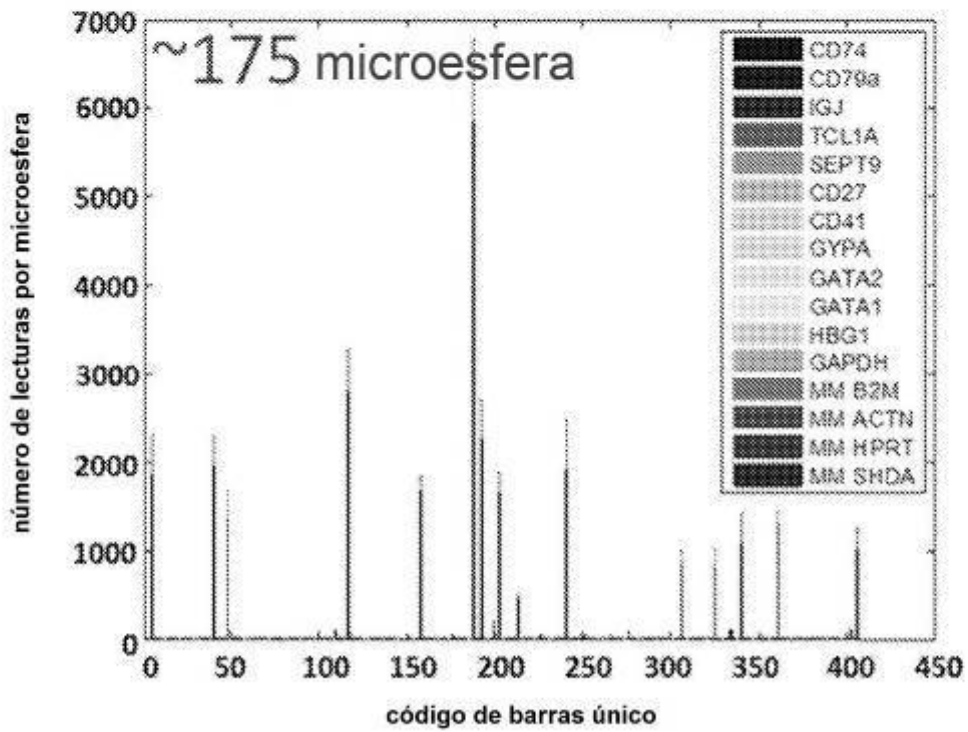


FIG. 11B

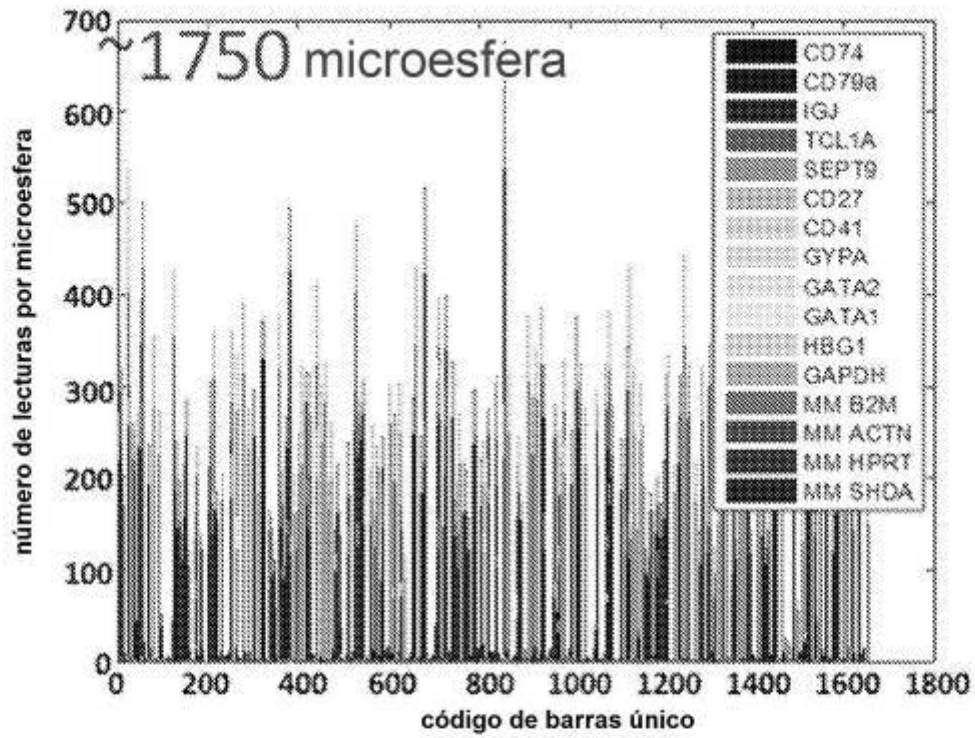


FIG. 11C

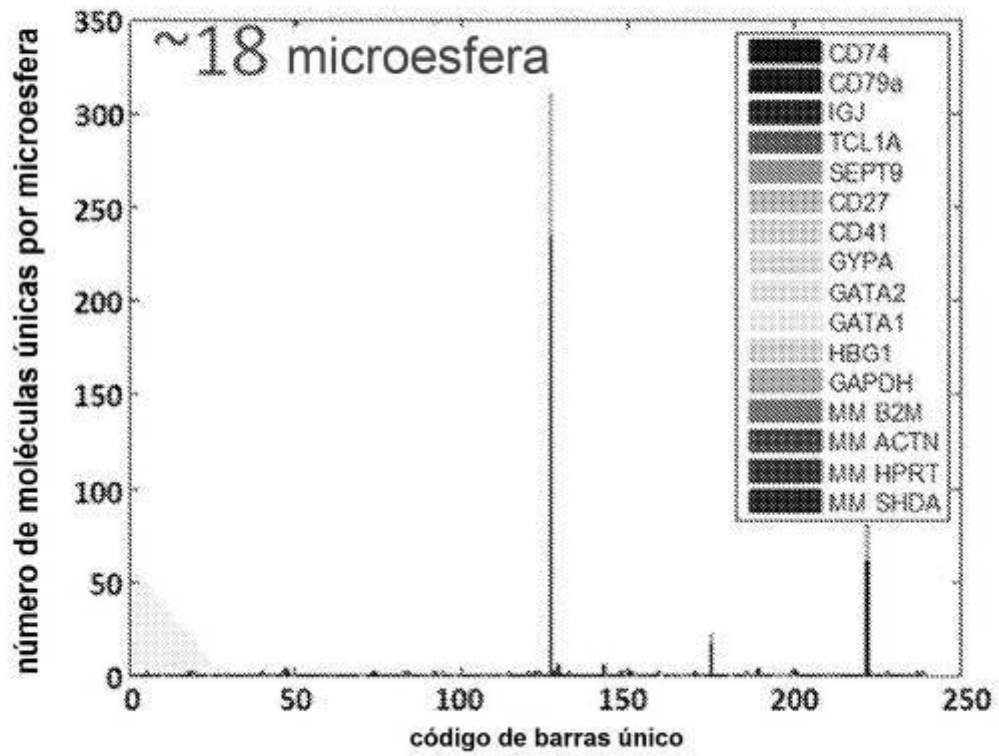


FIG. 11D

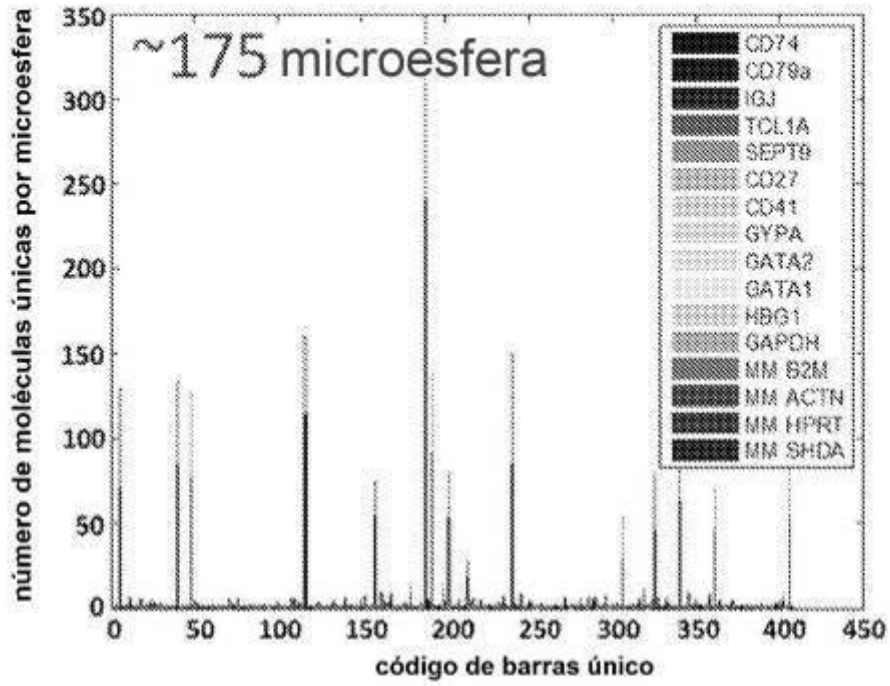


FIG. 11E

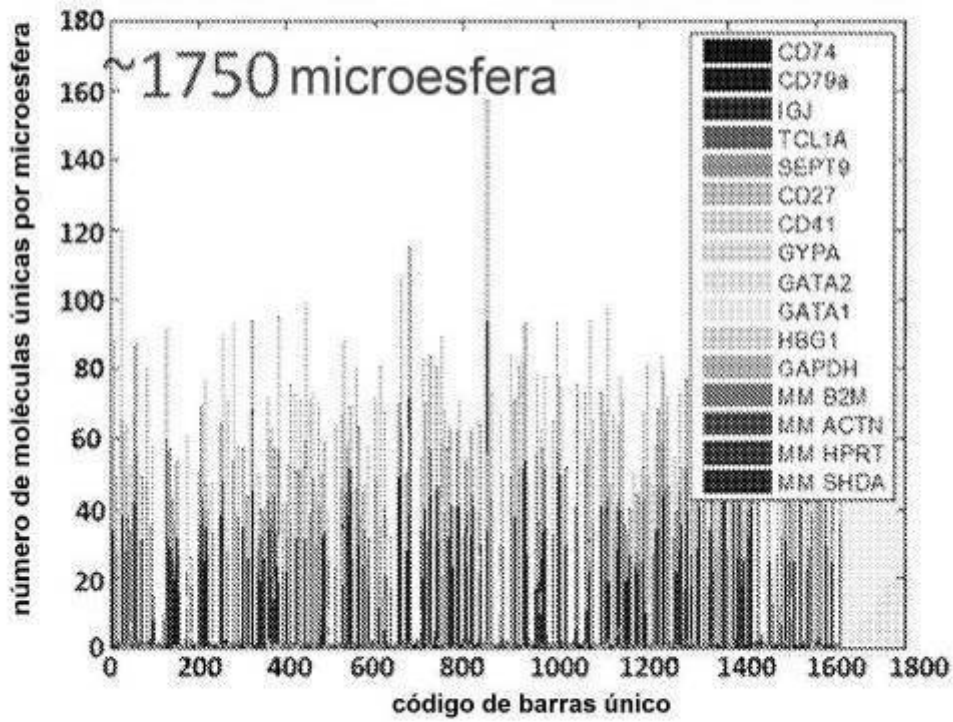


FIG. 11F

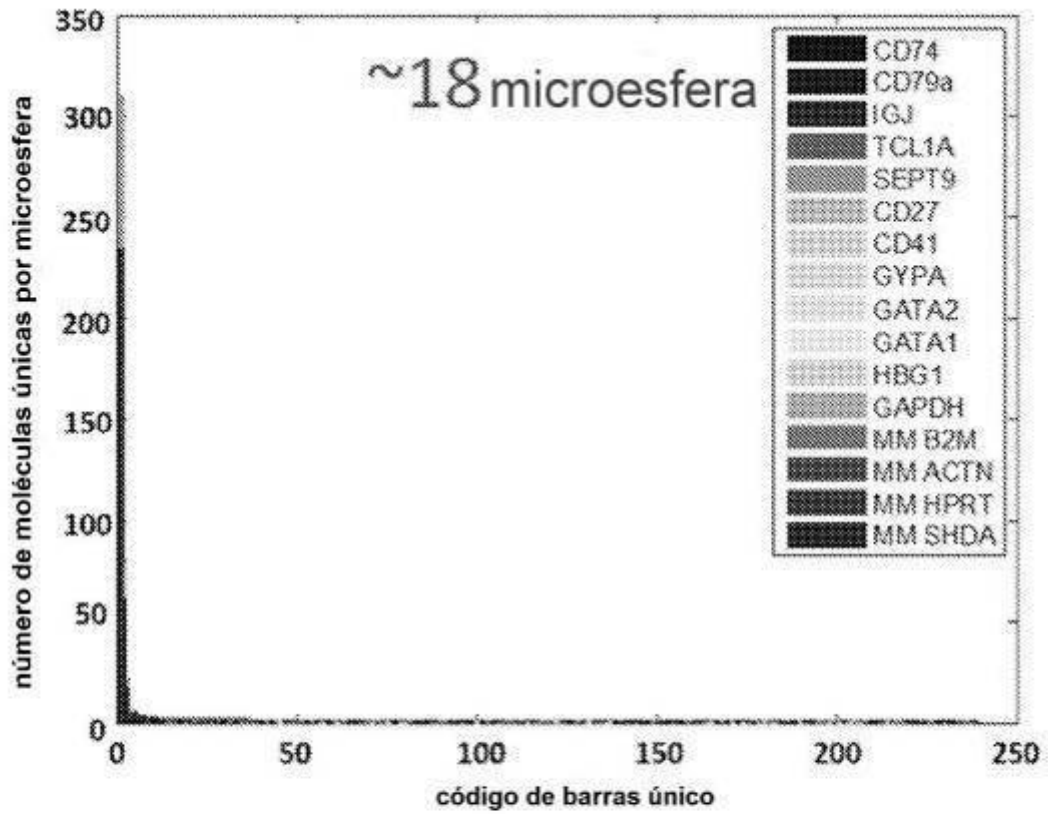


FIG. 11G

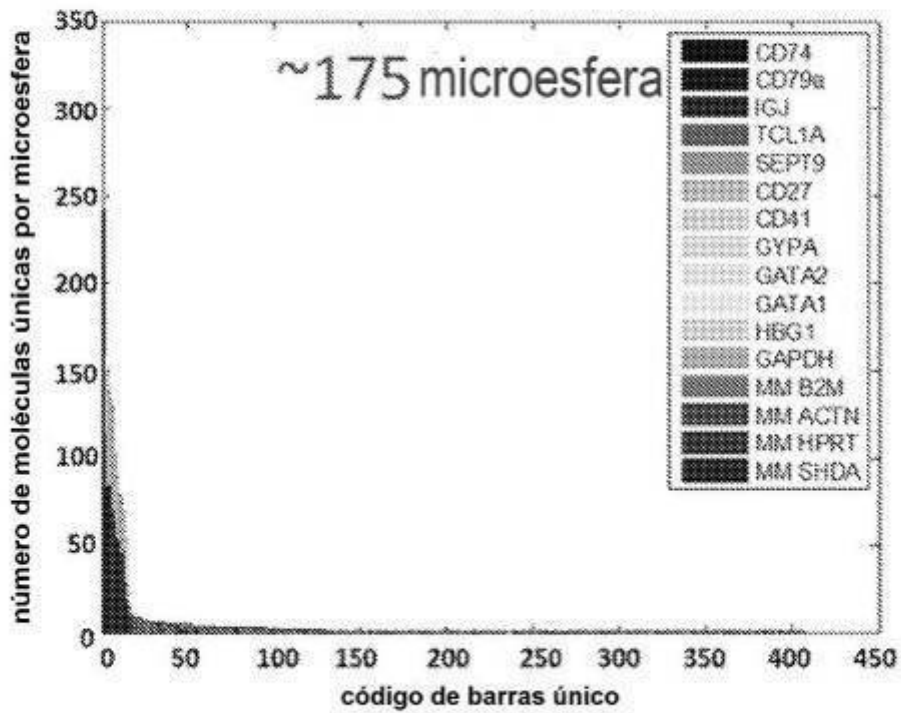


FIG. 11H

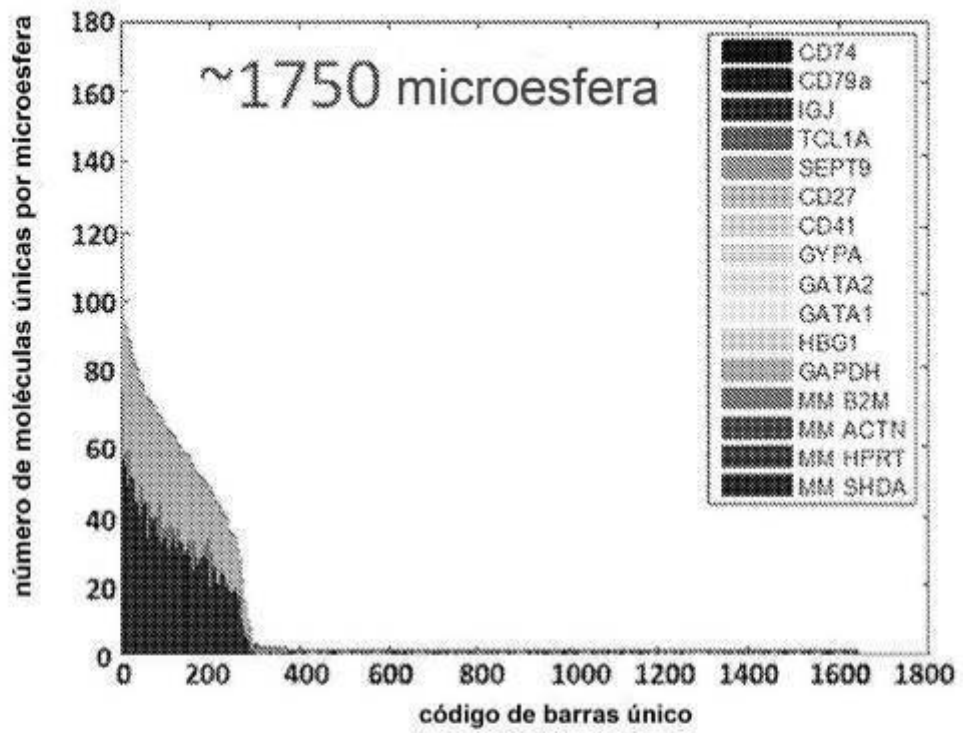


FIG. 11I

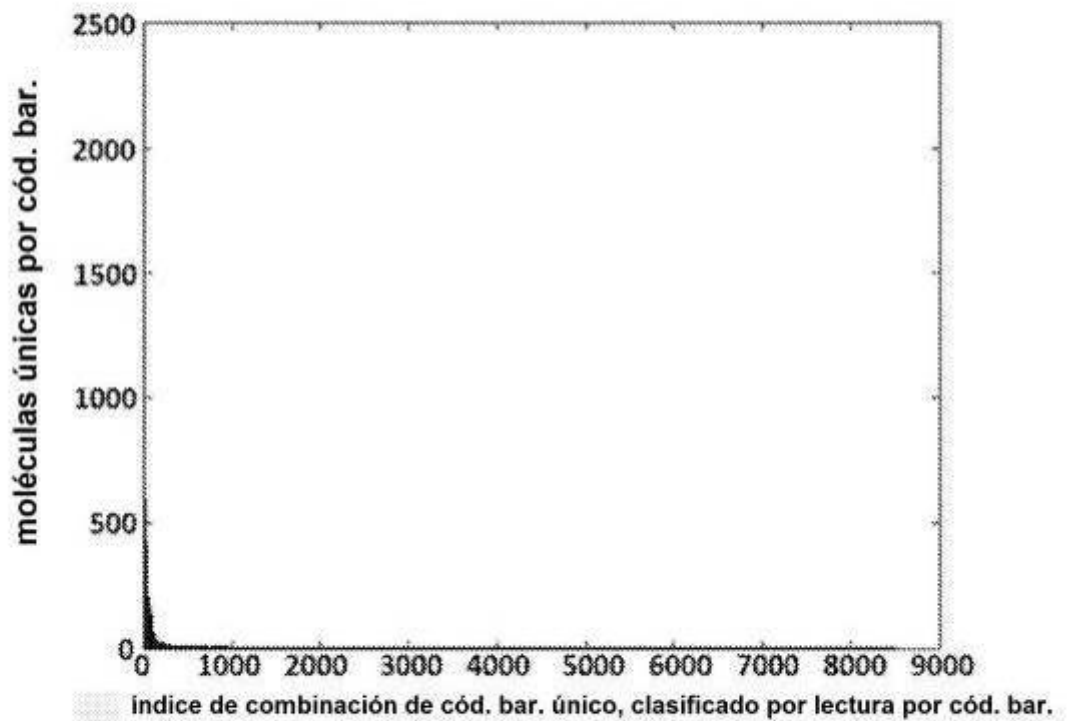


FIG. 12A

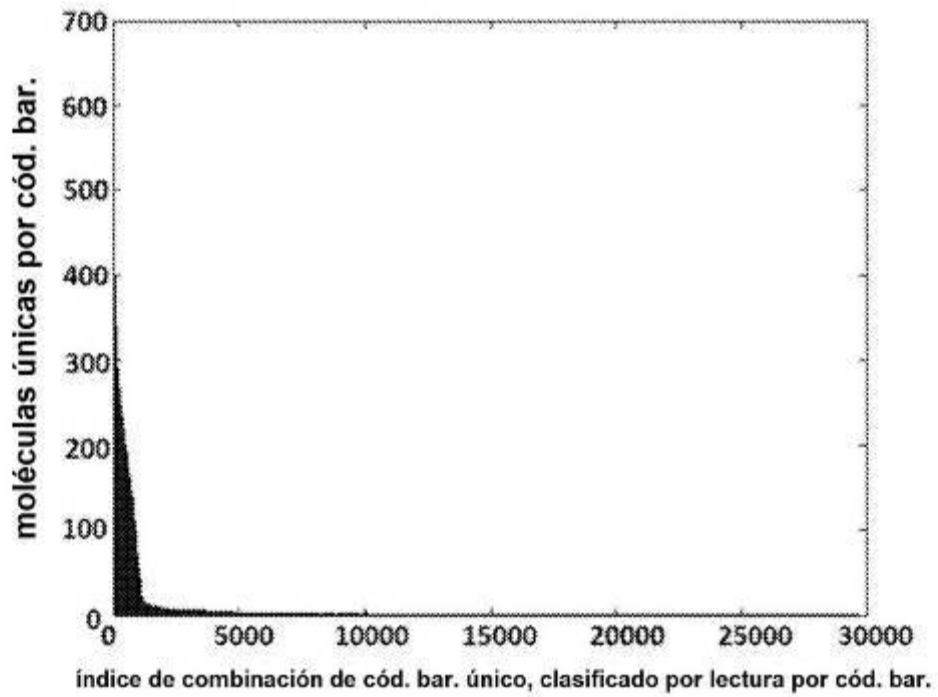


FIG. 12B

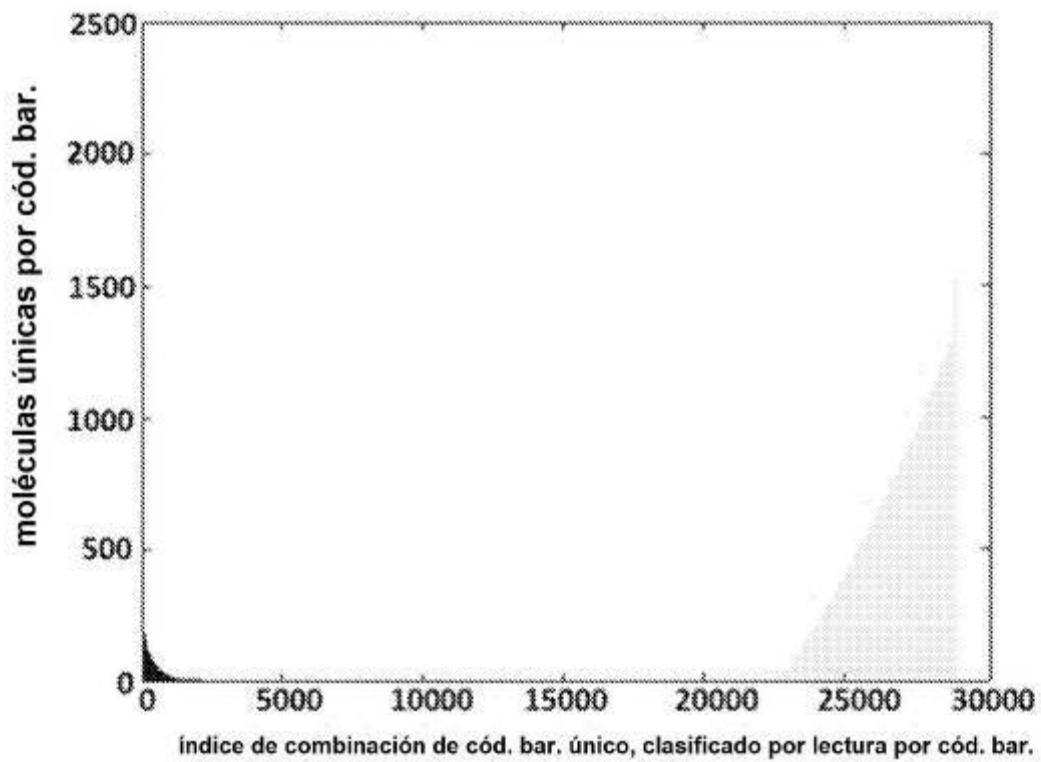


FIG. 12C

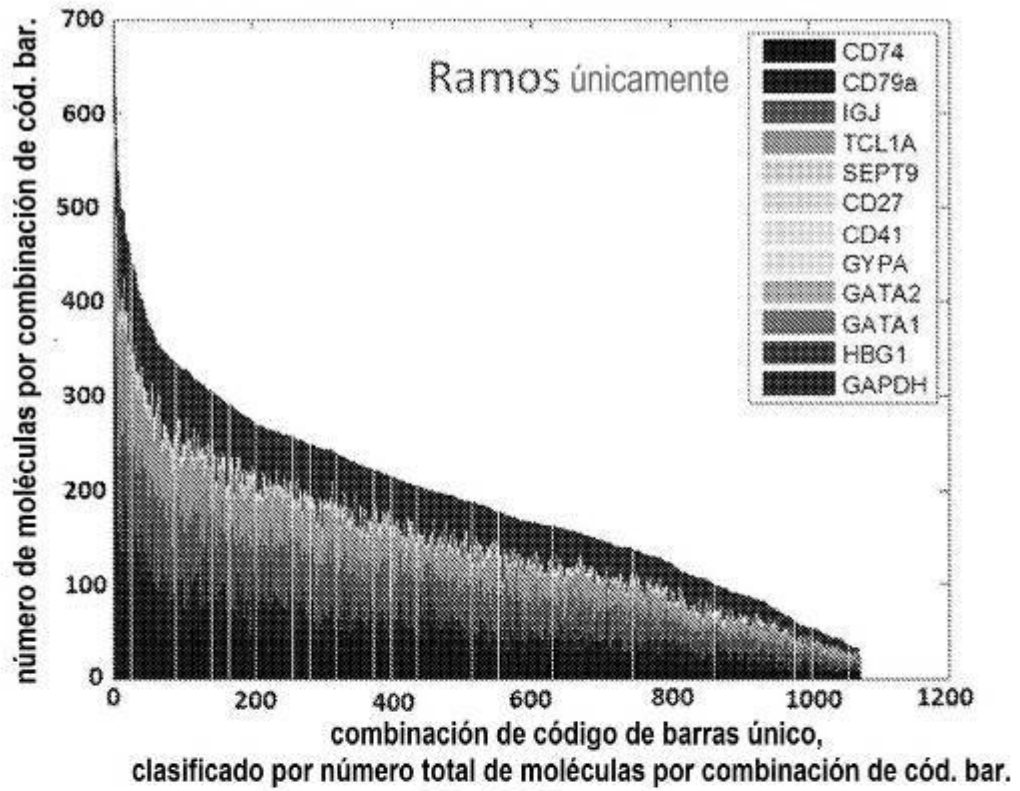


FIG. 12D

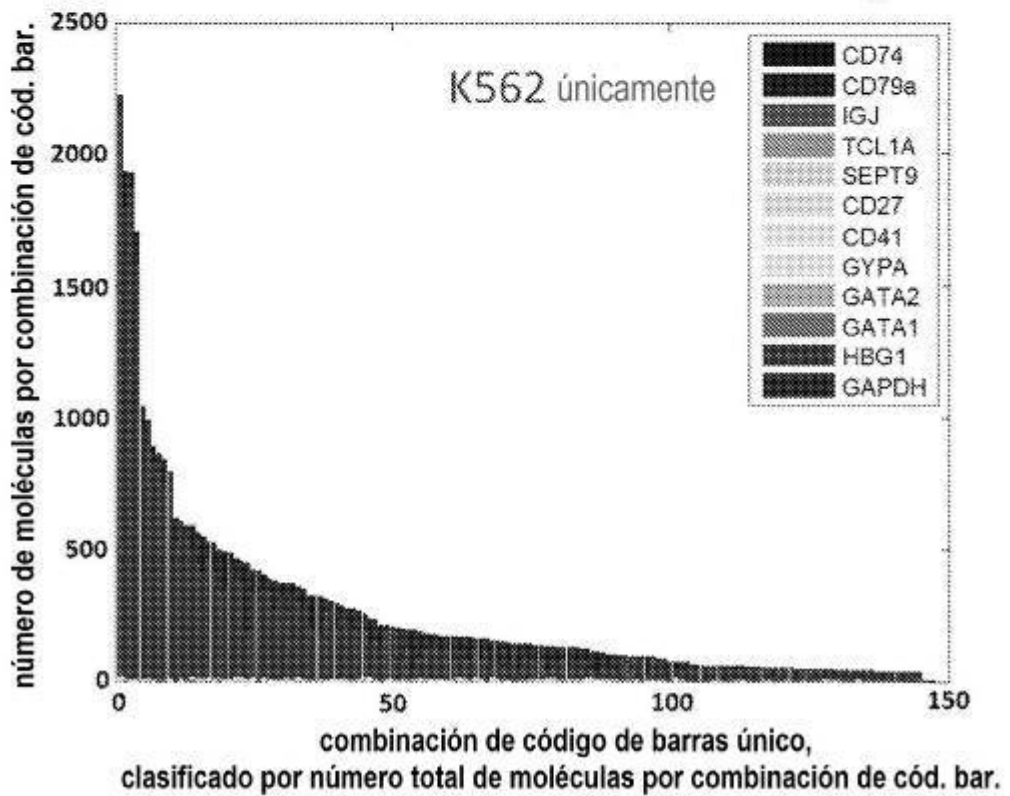


FIG. 12E

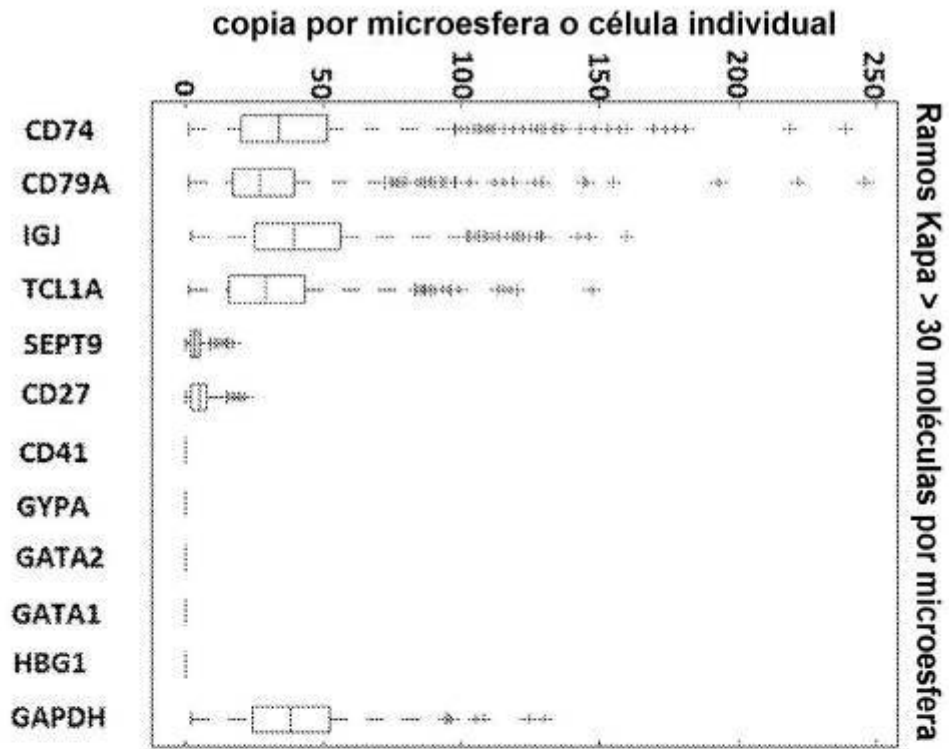


FIG. 12F

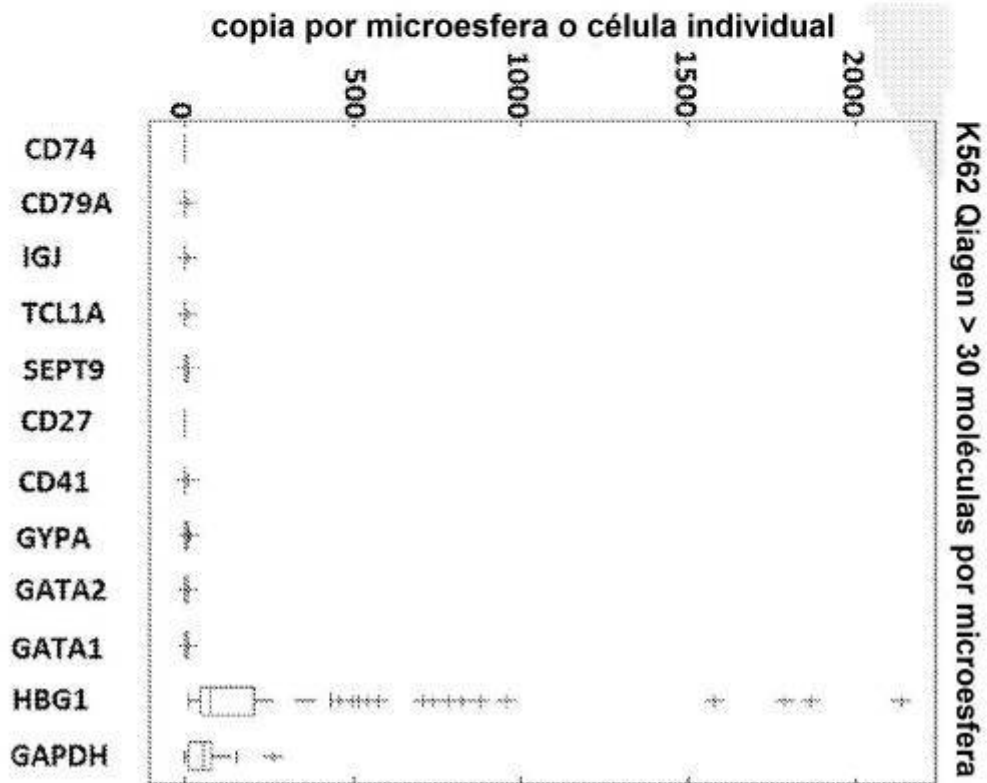


FIG. 12G

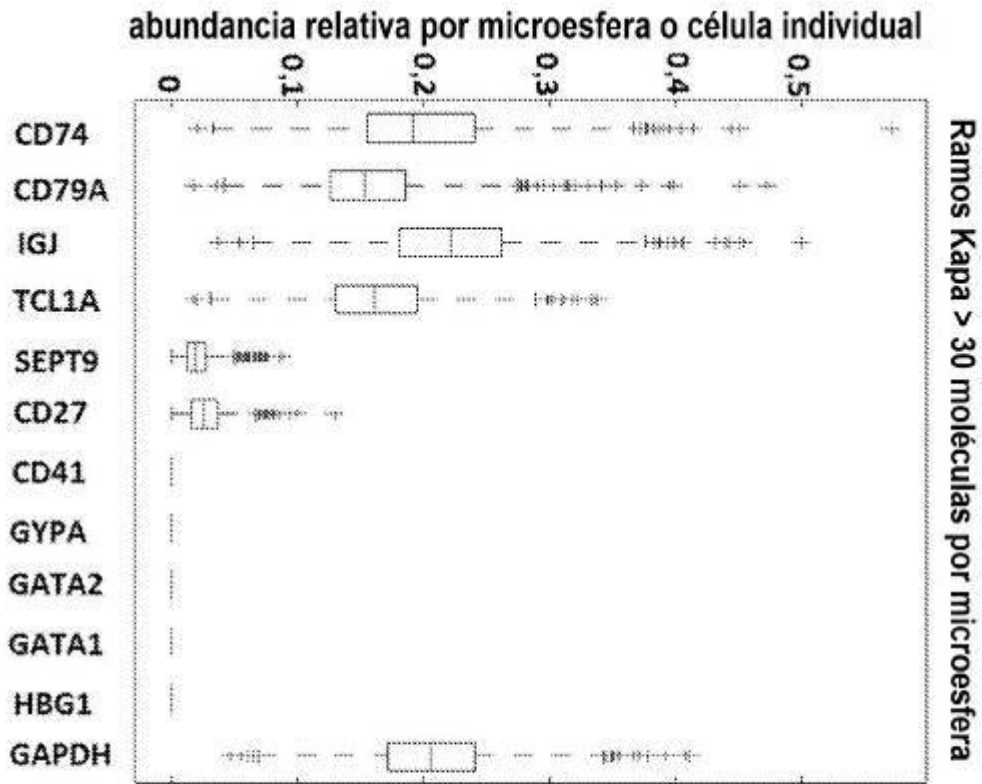


FIG. 12H

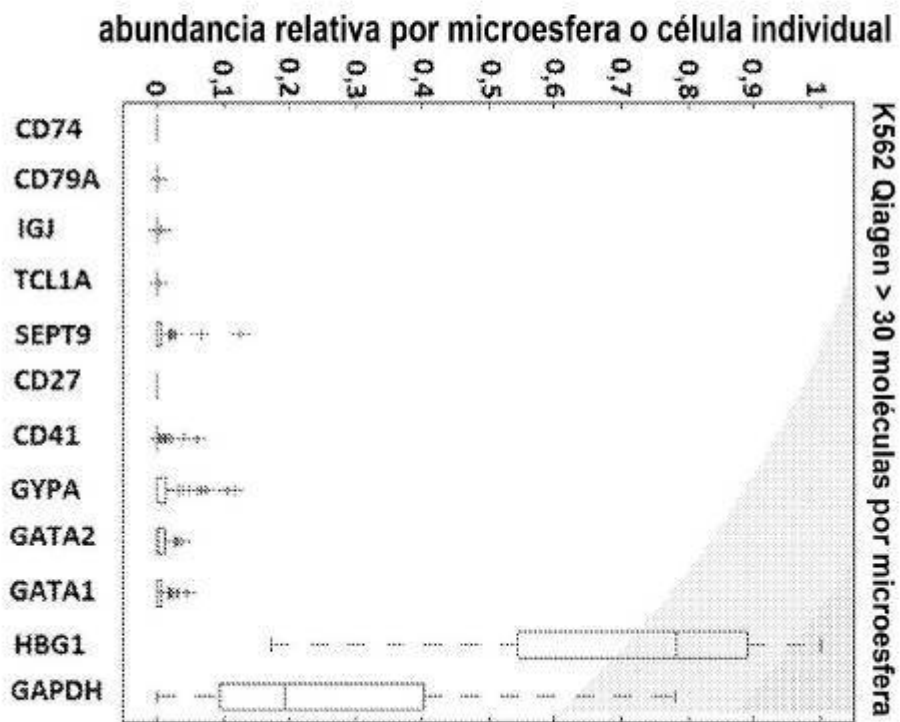


FIG. 12I

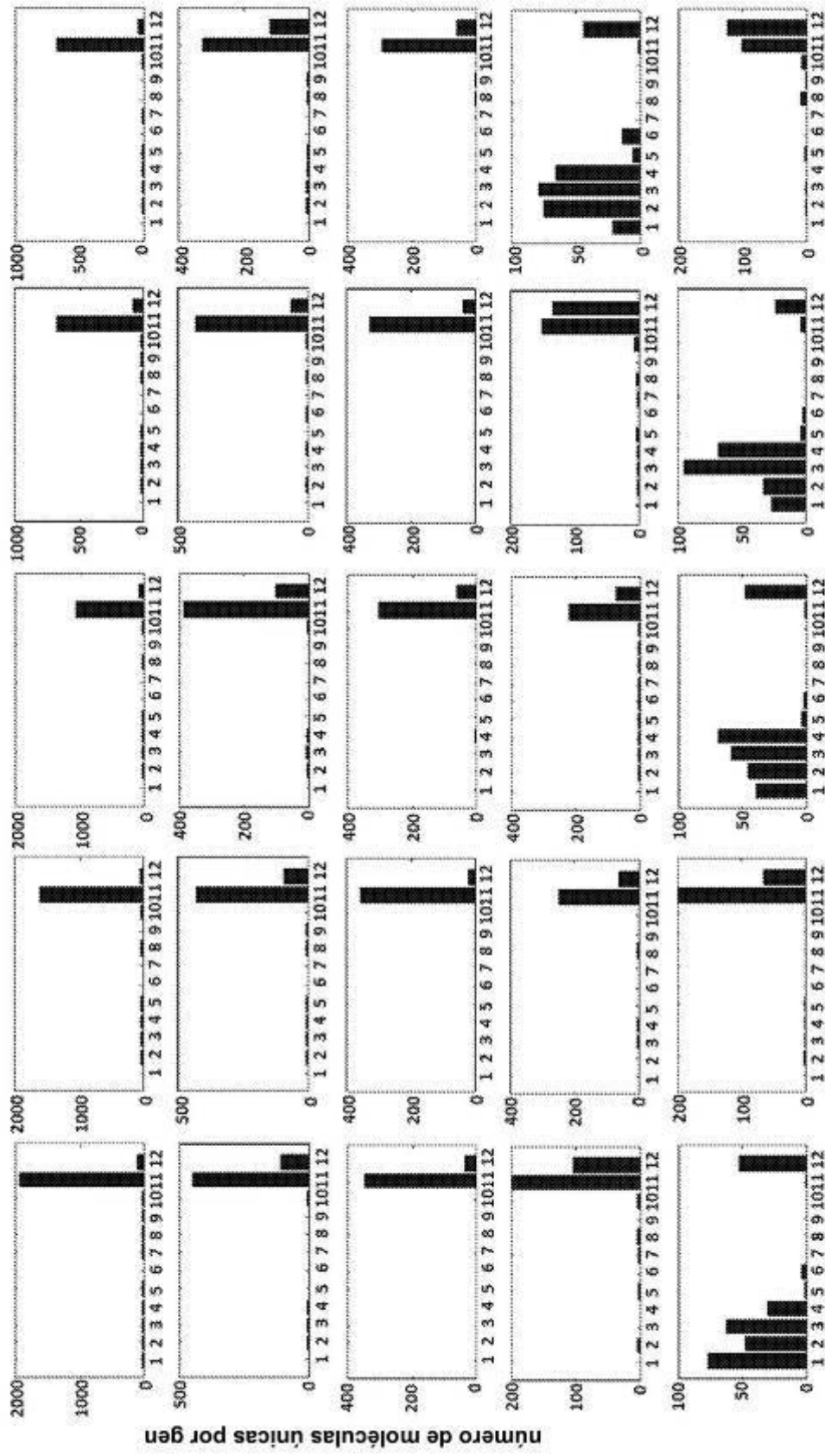


FIG. 12J

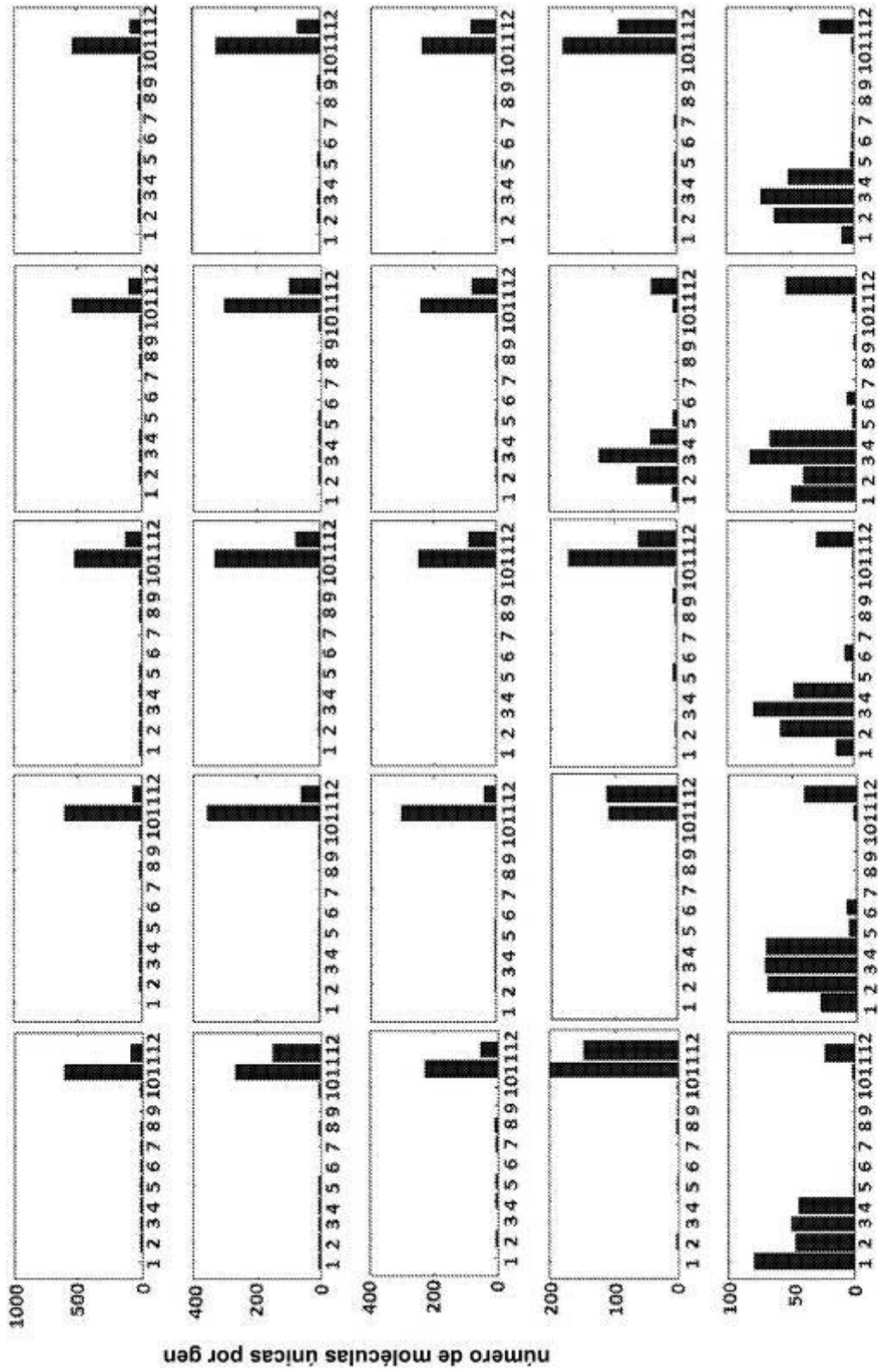


FIG. 12K

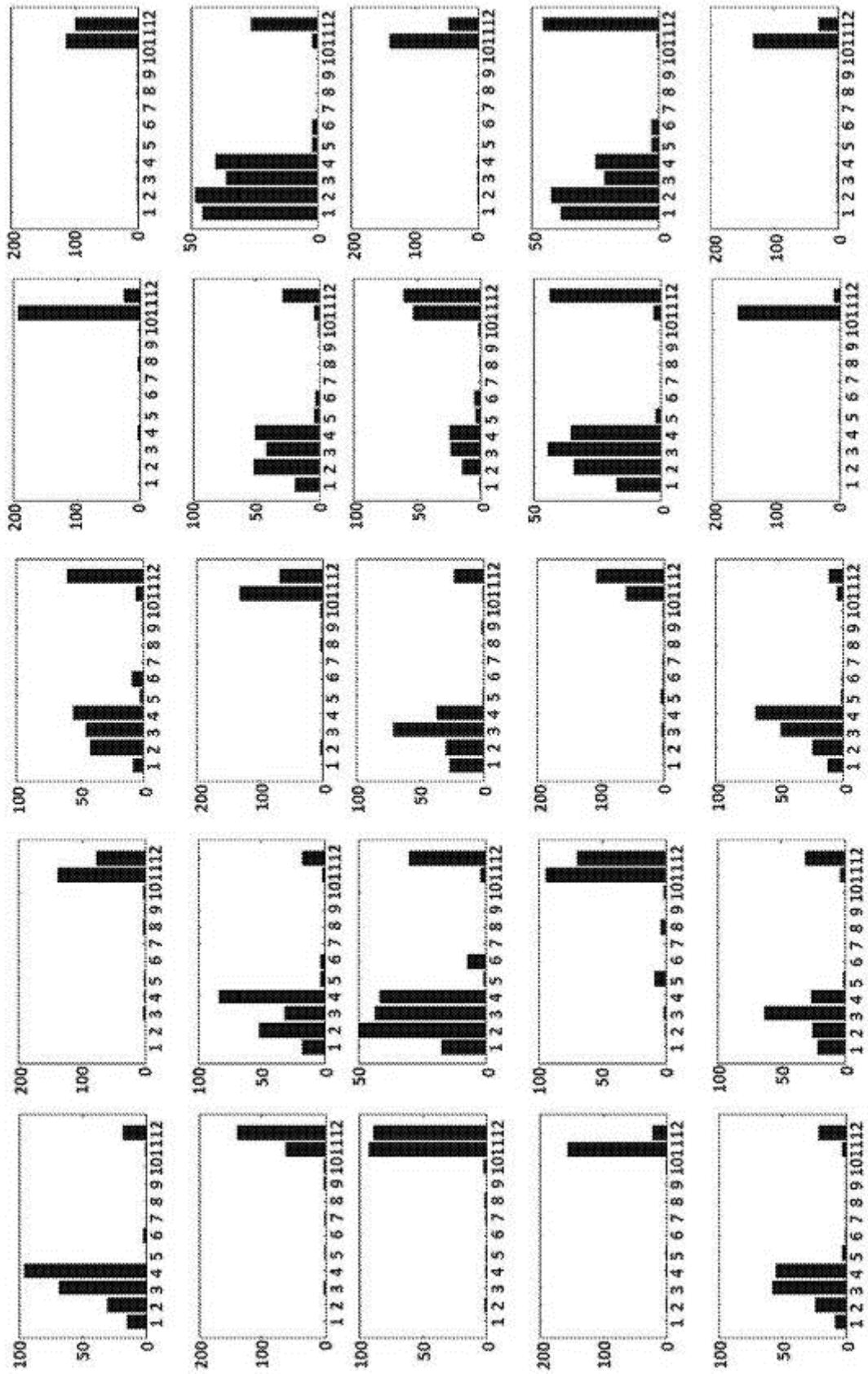


FIG. 12L

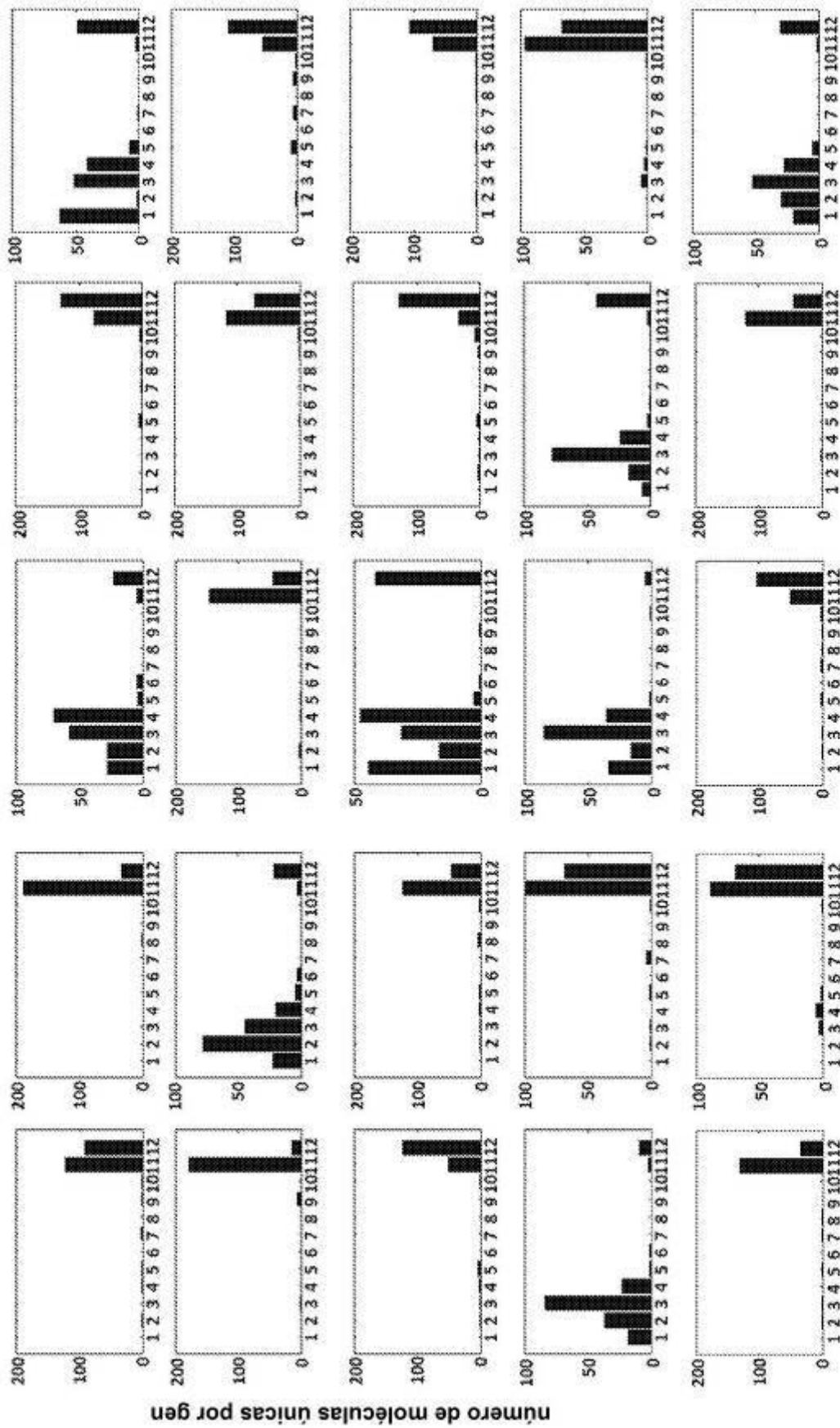


FIG. 12M

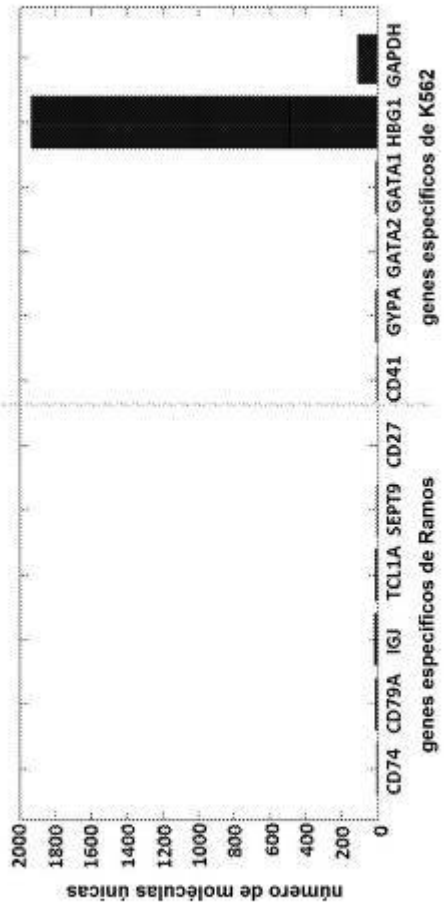


FIG. 12N

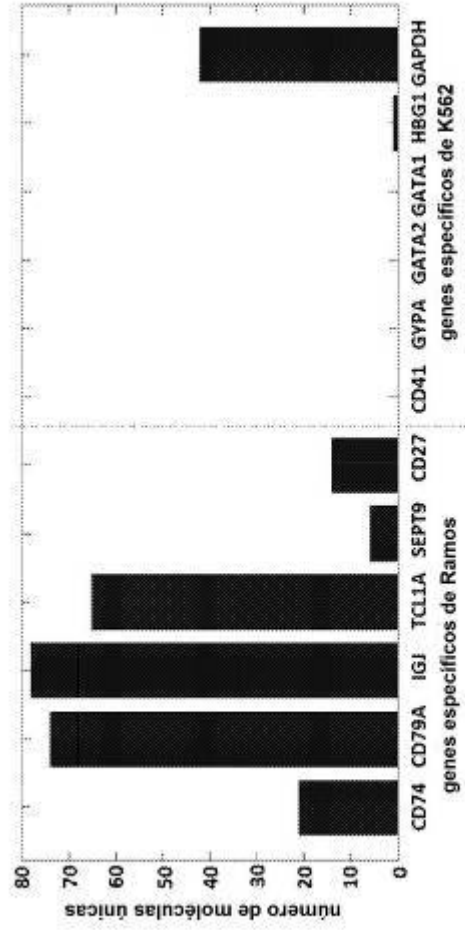


FIG. 12O

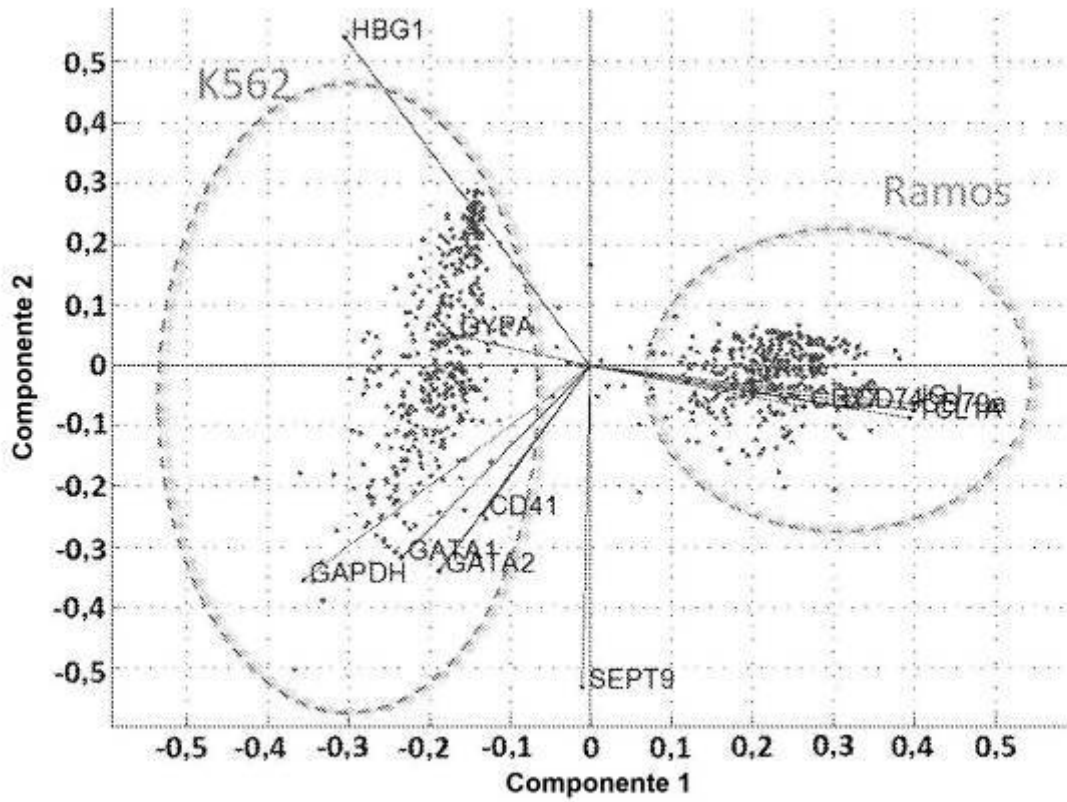


FIG. 12P

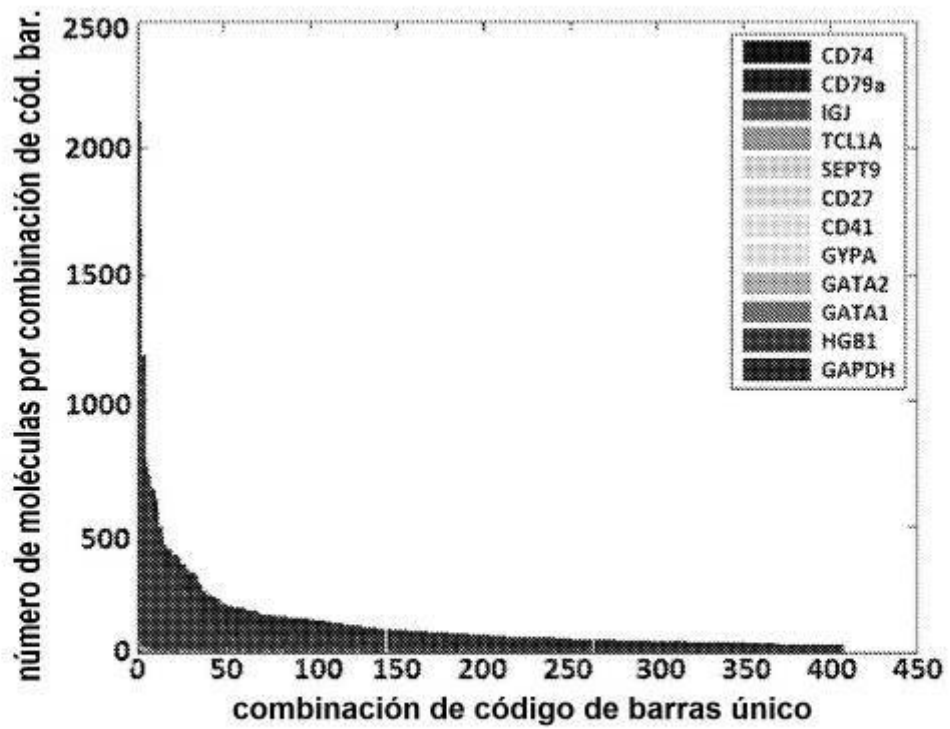


FIG. 12Q

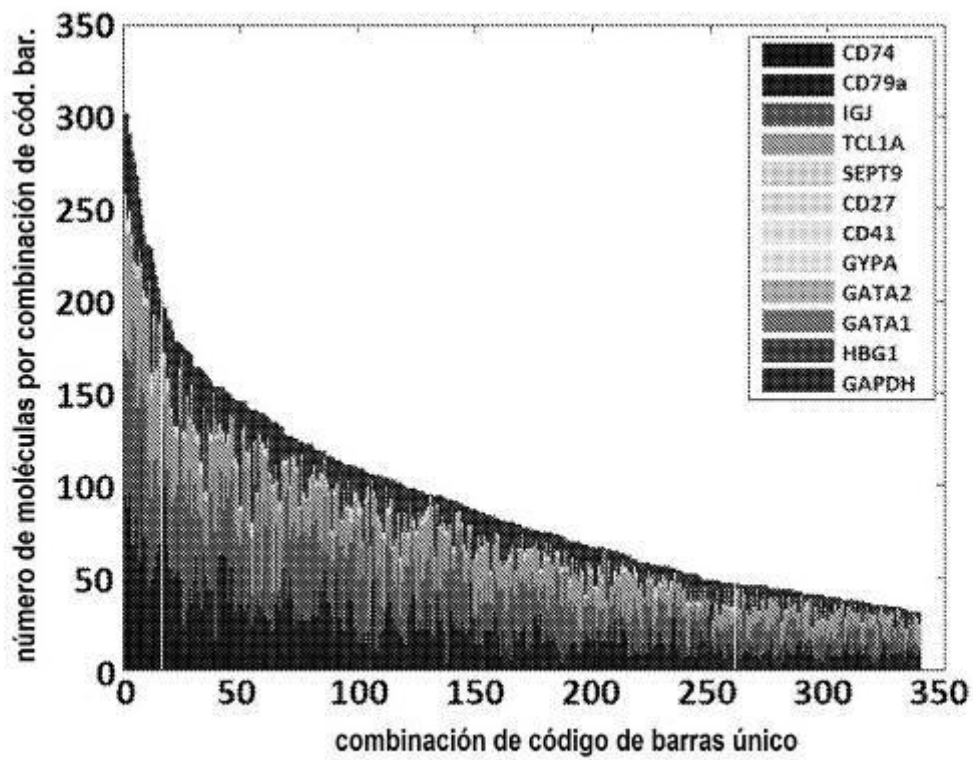


FIG. 12R

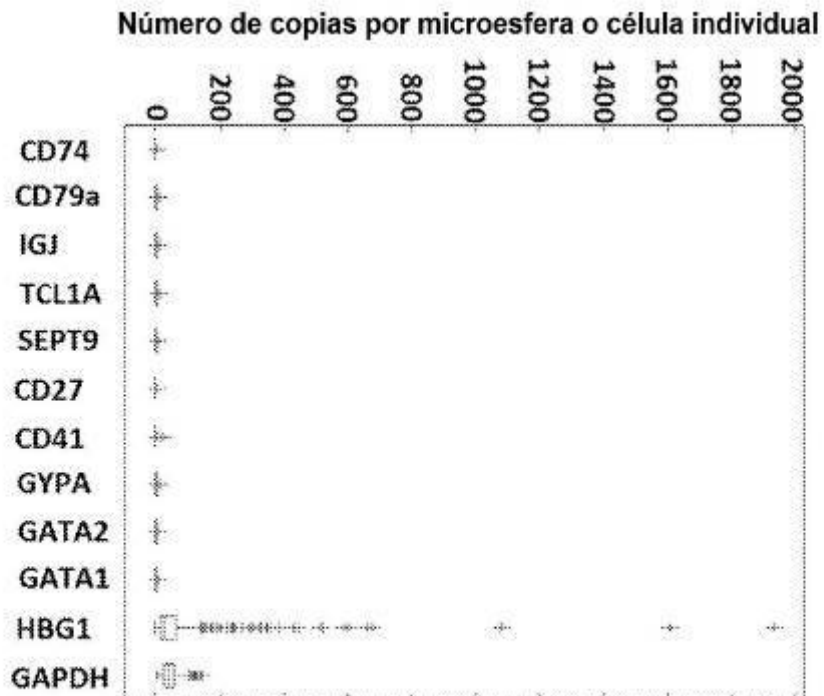


FIG. 12S

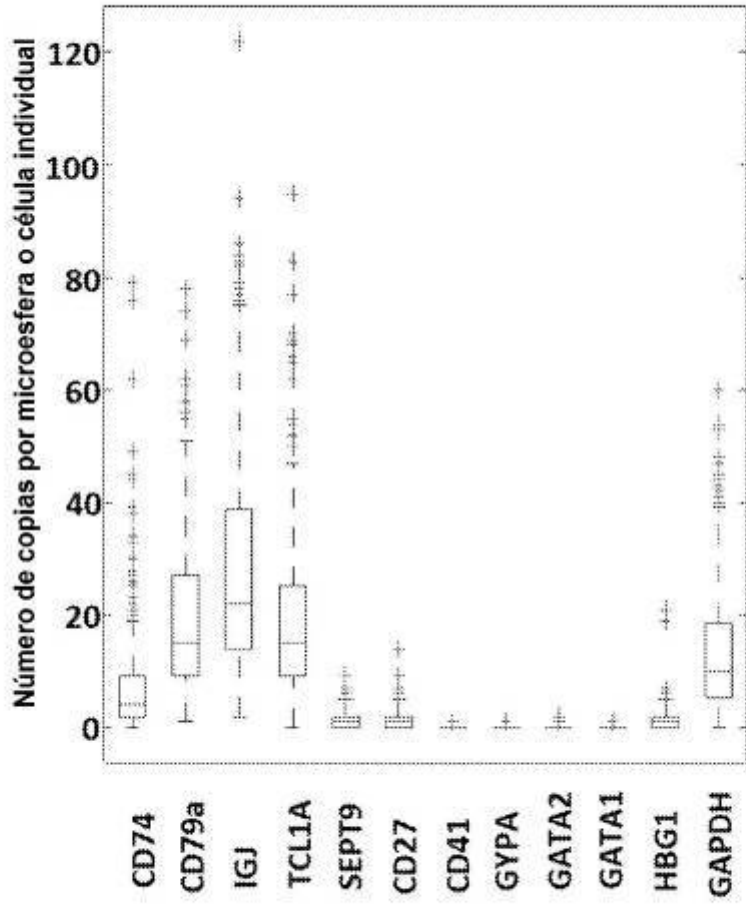


FIG. 12T

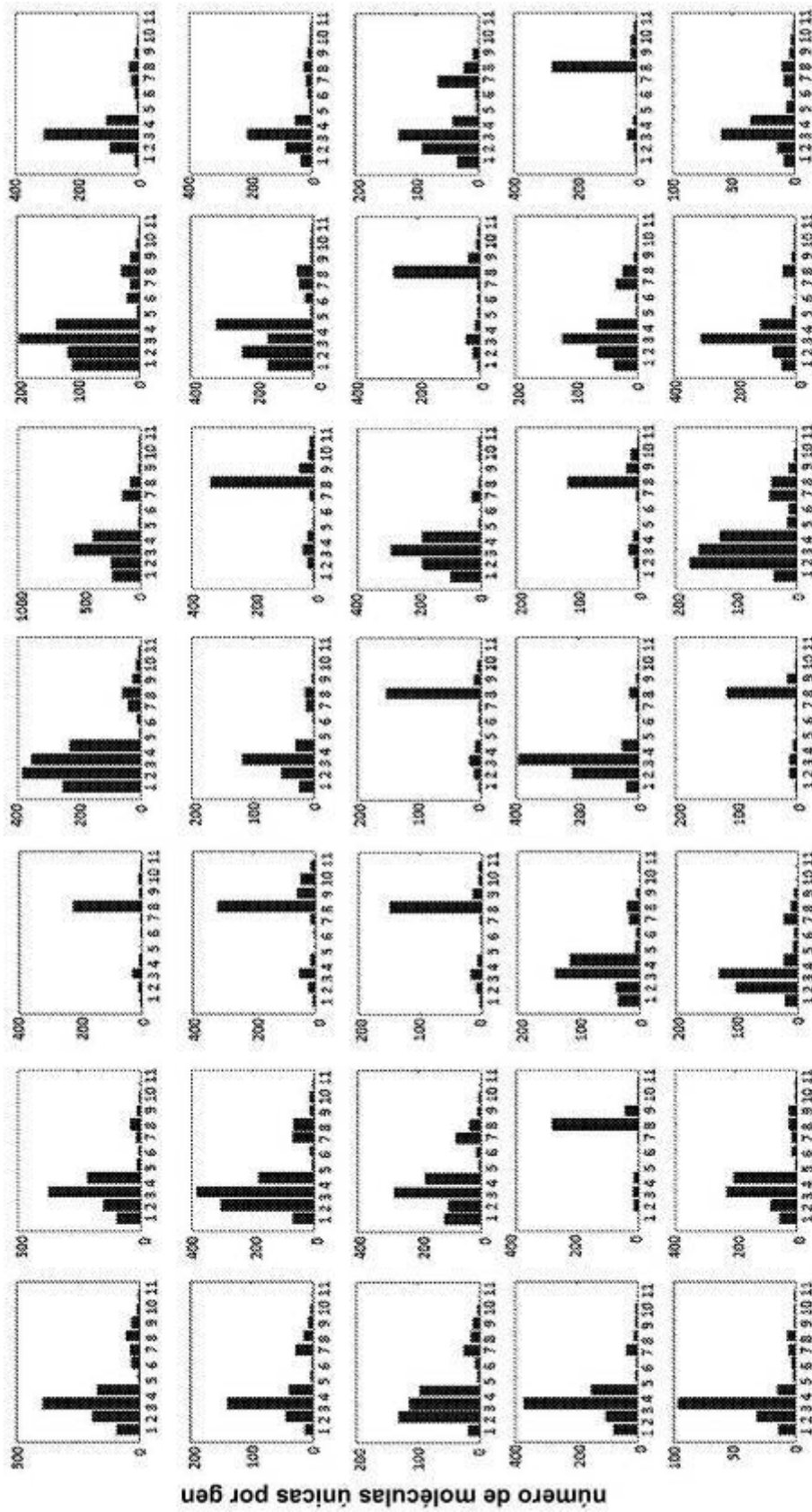


FIG. 13A

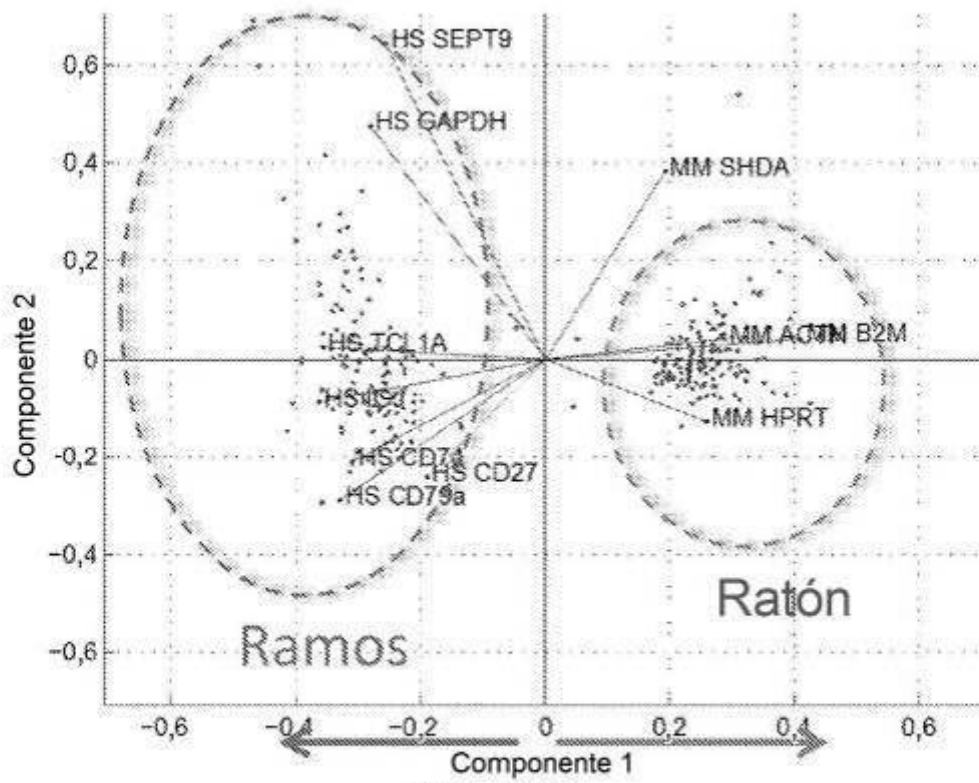


FIG. 13B

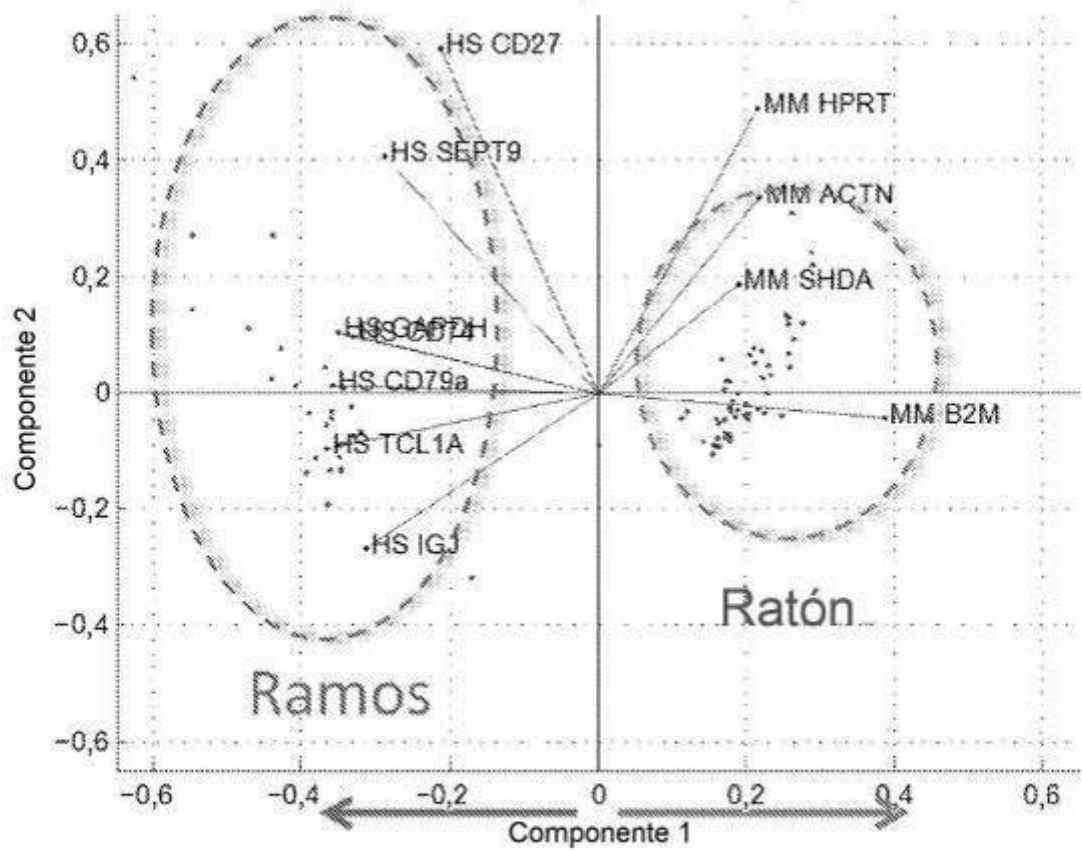


FIG. 13C

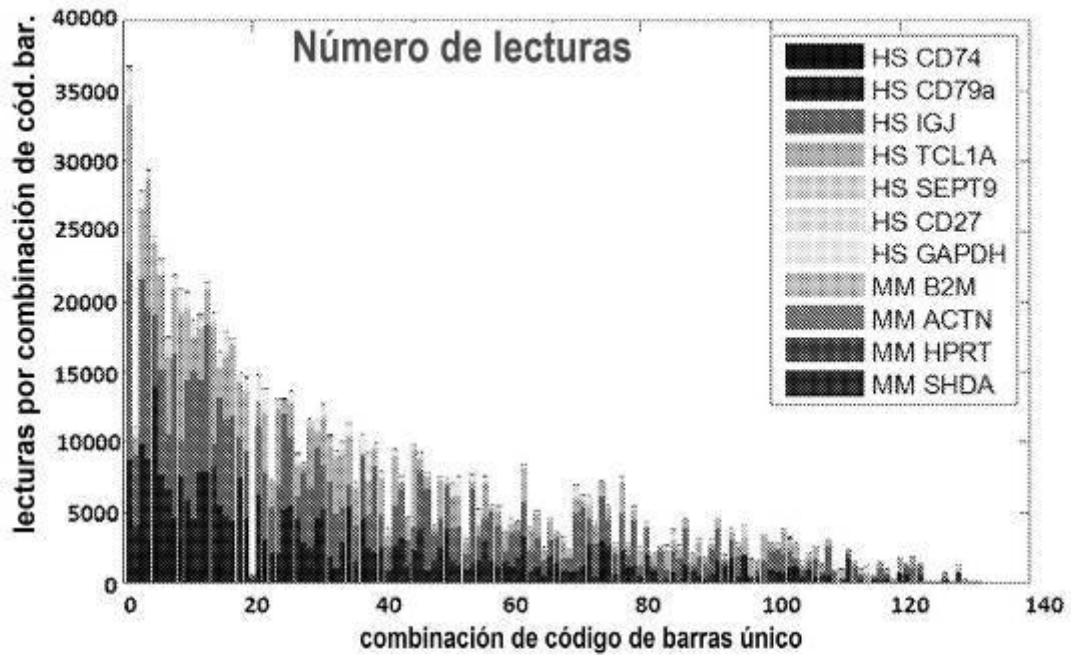


FIG. 13D

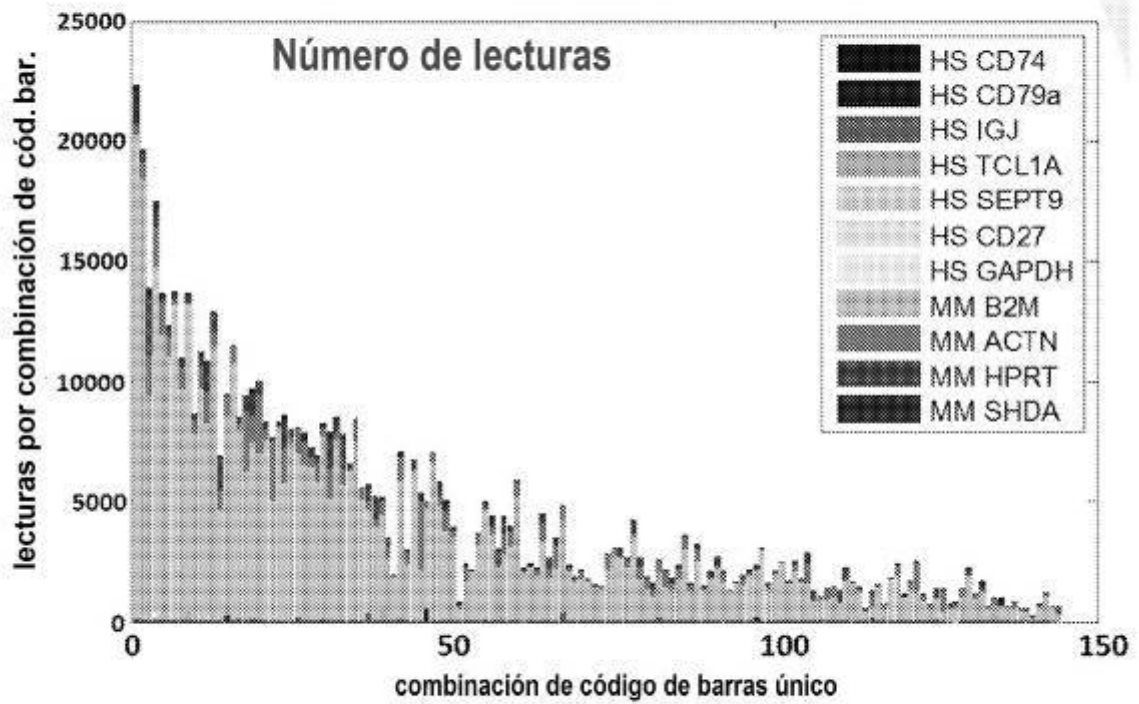


FIG. 13E

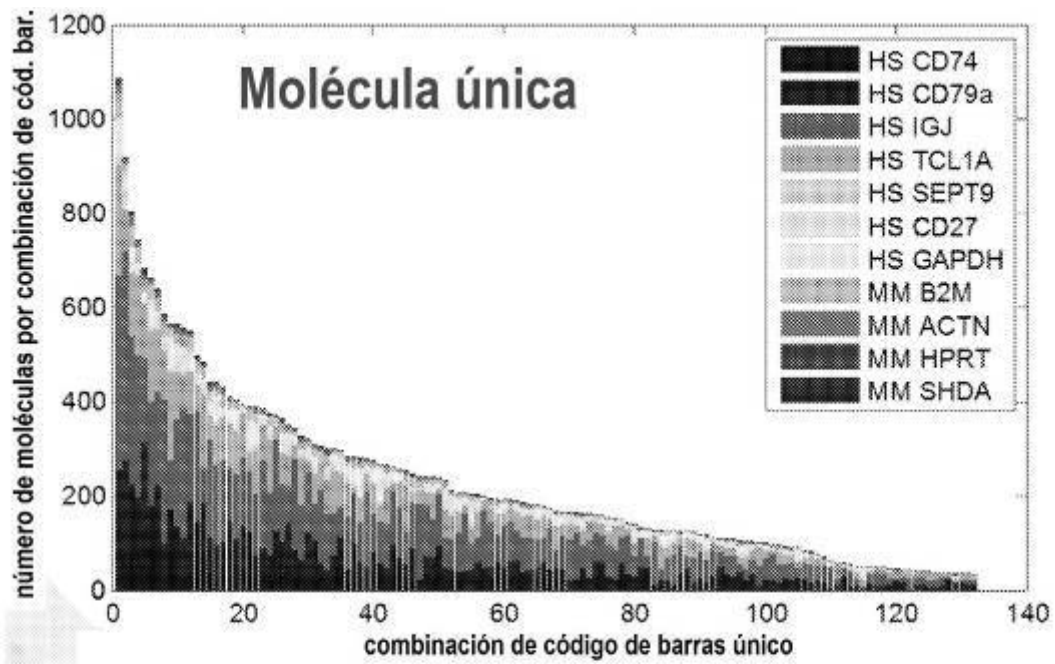


FIG. 13F

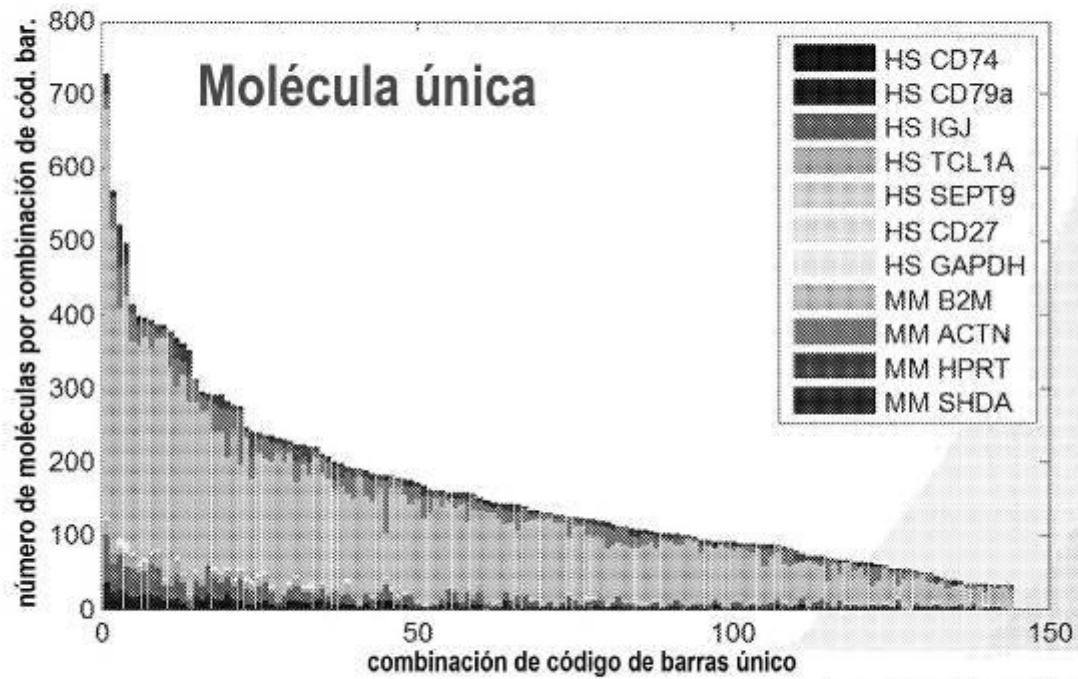


FIG. 13G

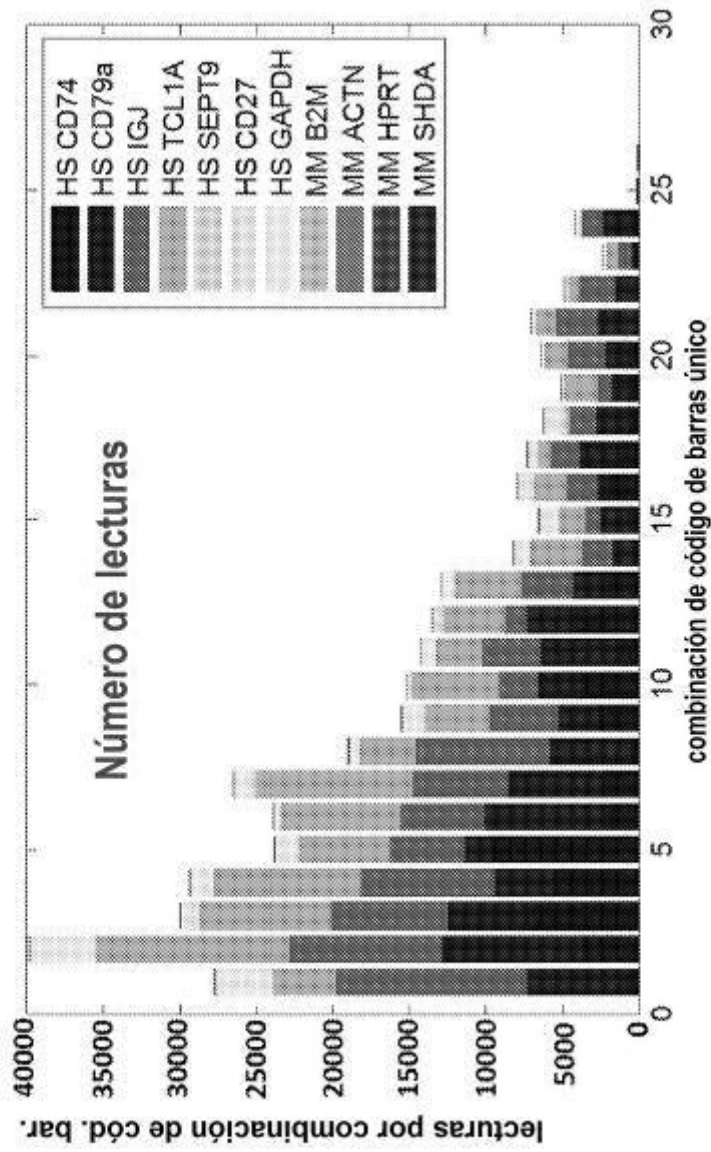


FIG. 13H

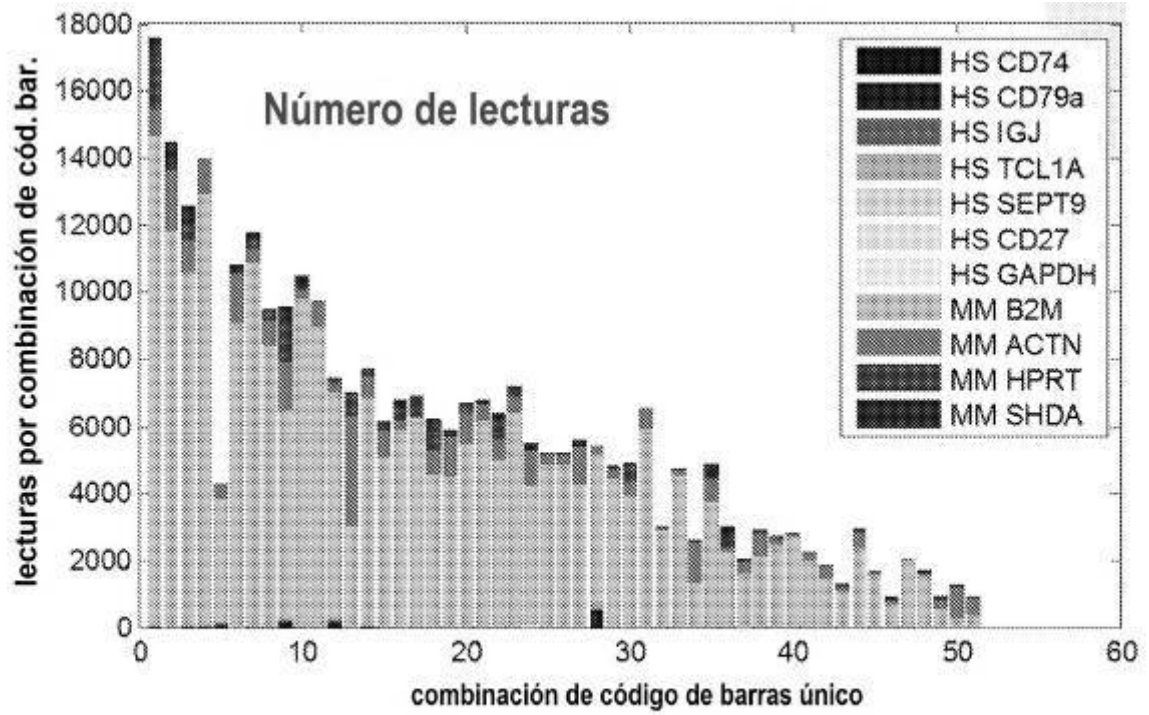


FIG. 13I

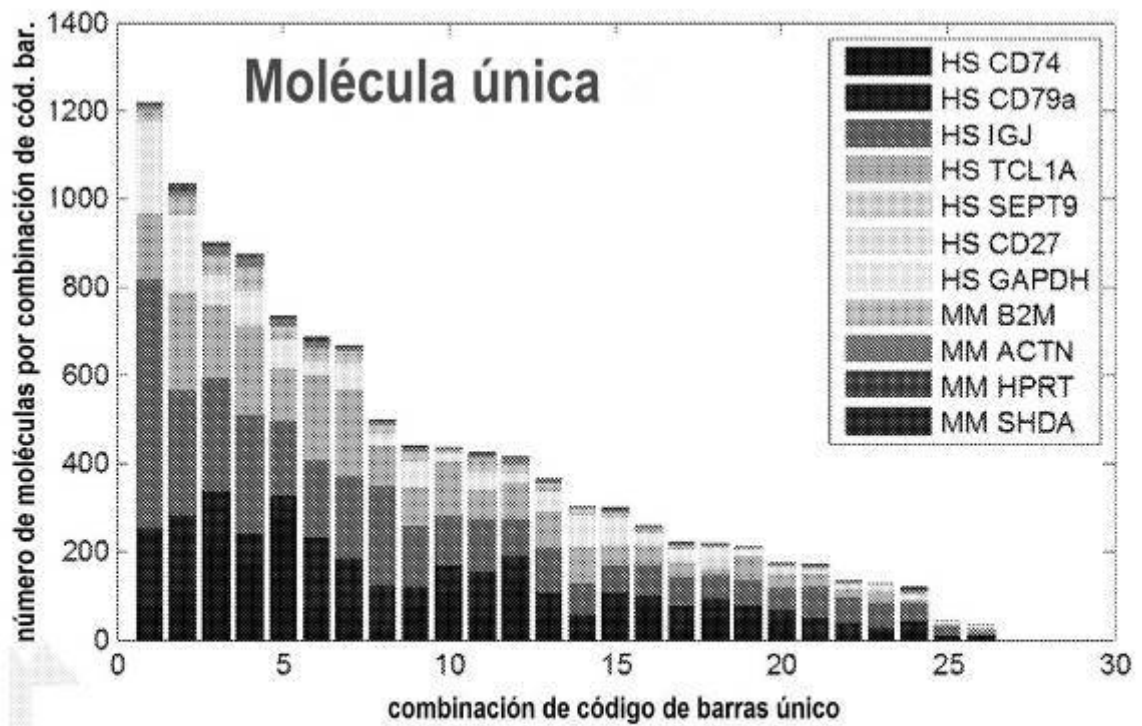


FIG. 13J

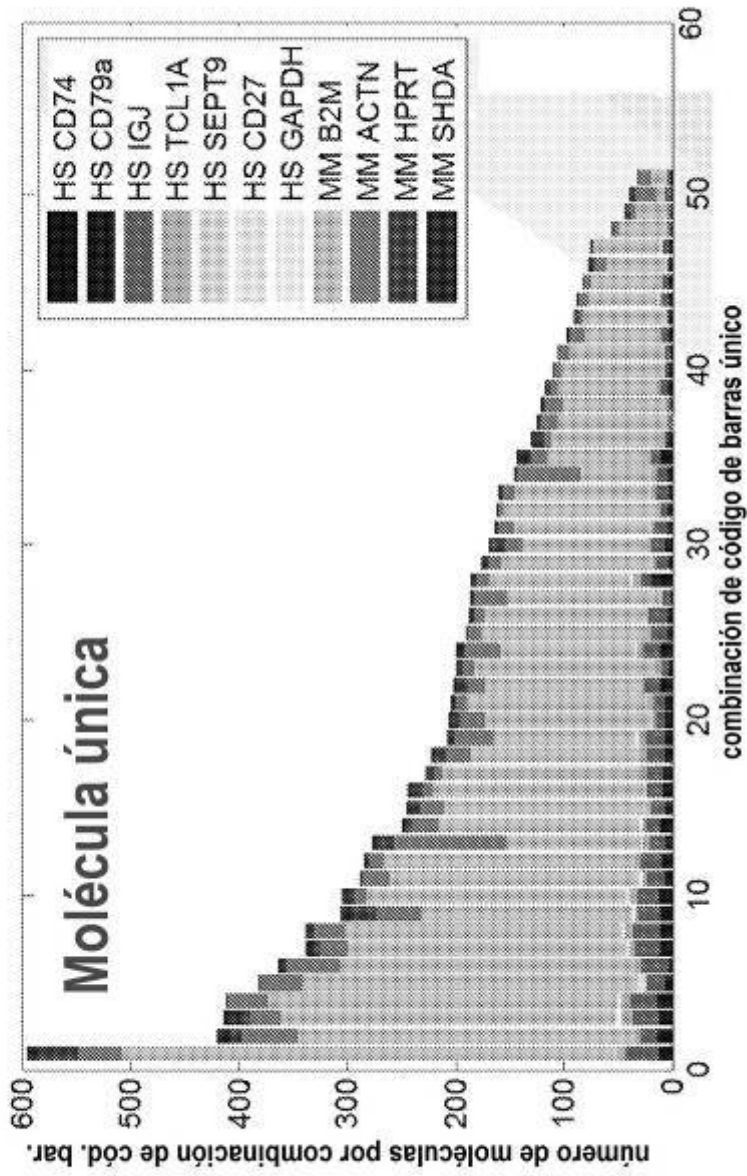


FIG. 13K

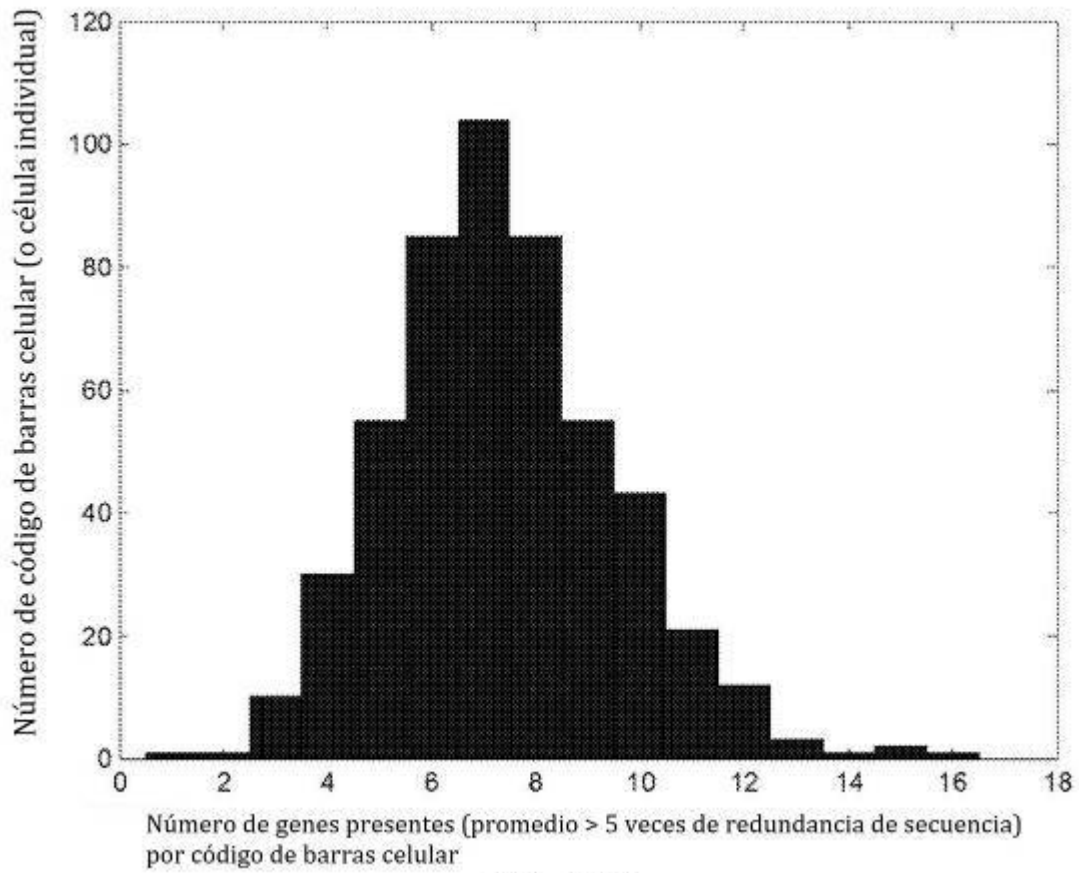


FIG. 15A

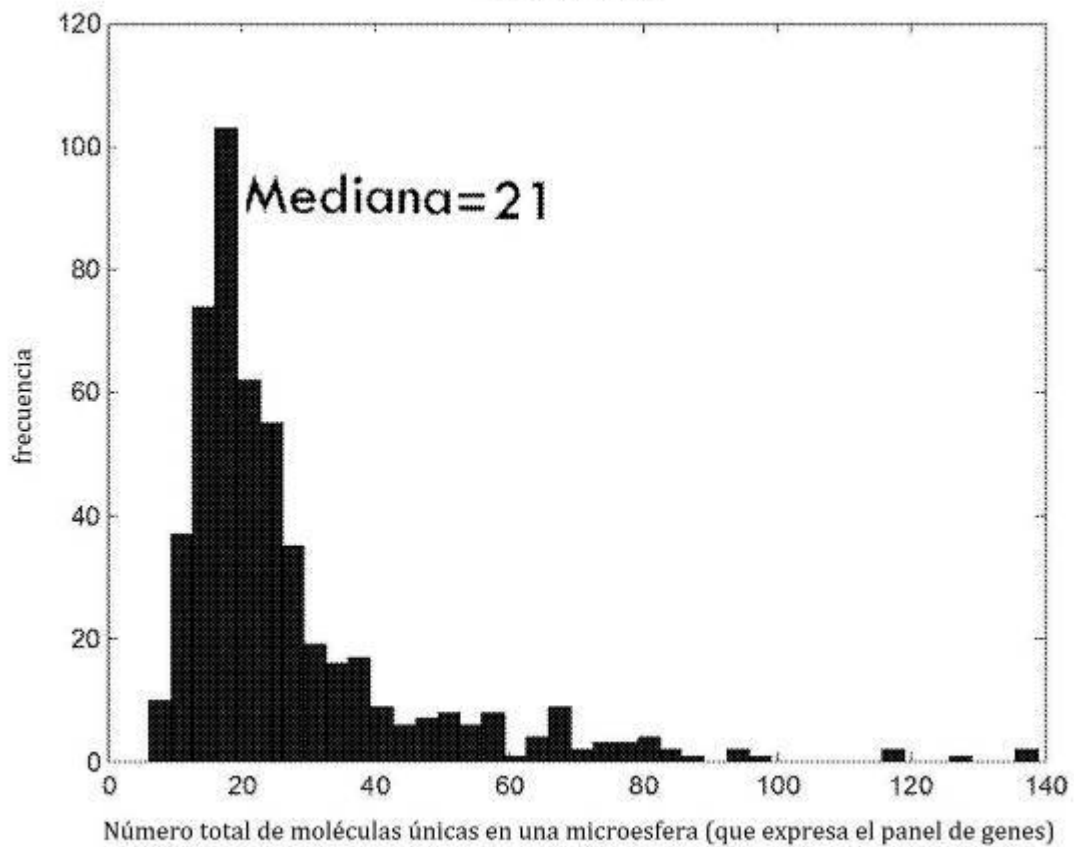


FIG. 15B

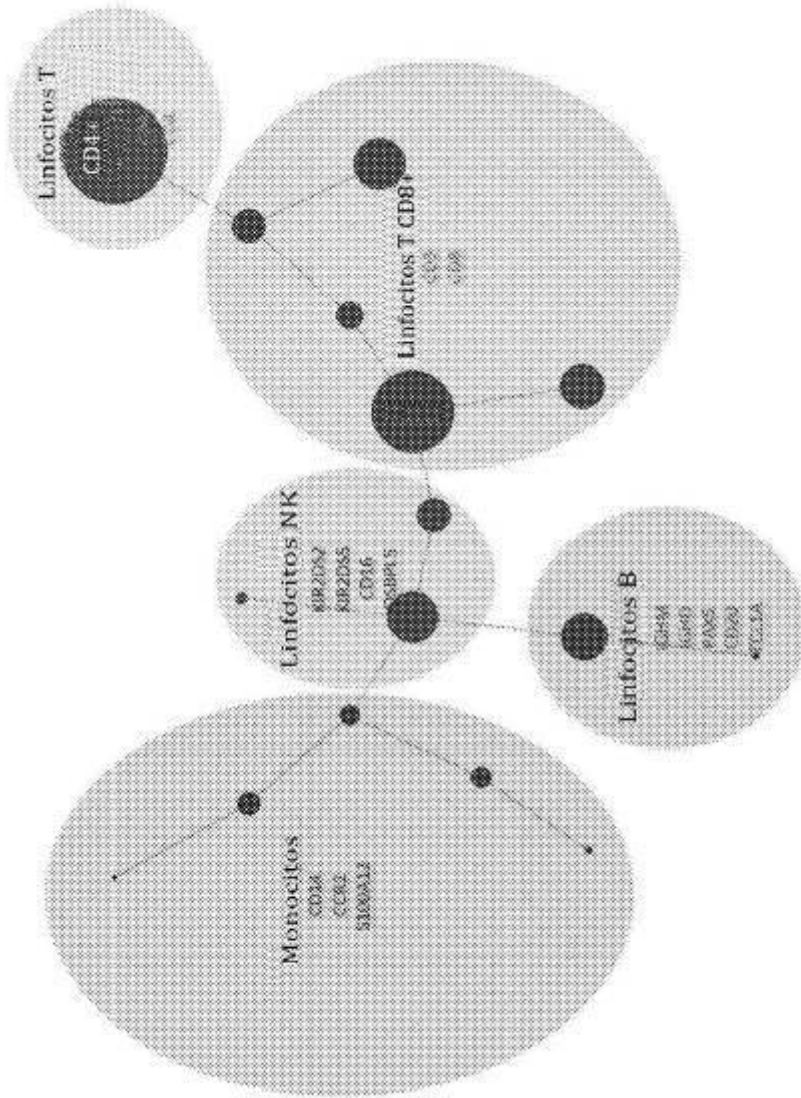


FIG. 16

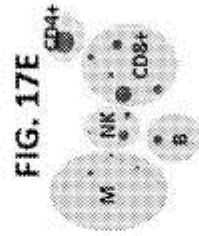
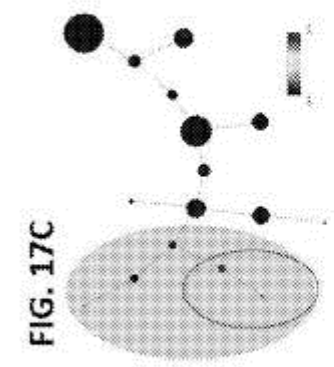
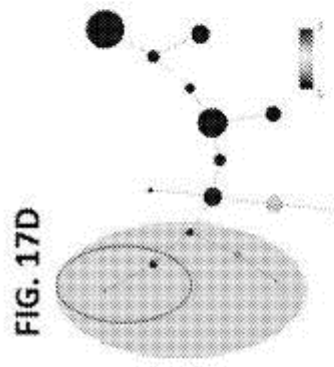
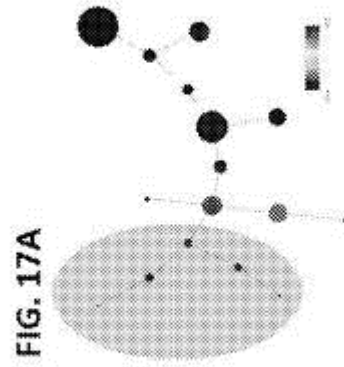
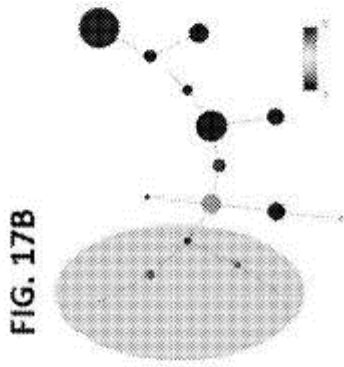


FIG. 17A-E

FIG. 18A

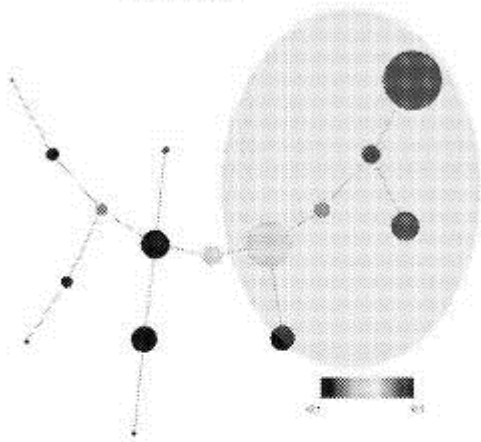


FIG. 18B

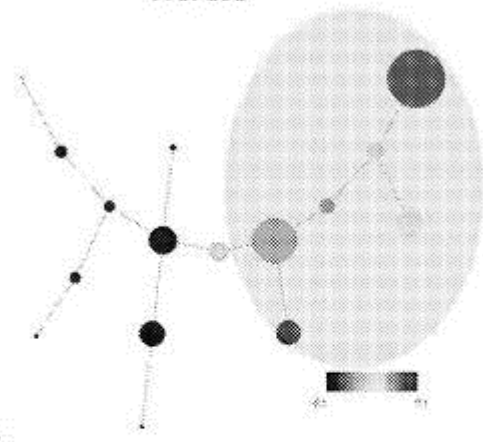


FIG. 18C

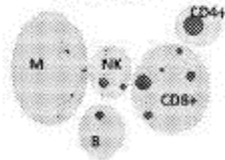


FIG. 18A-C

FIG. 19A

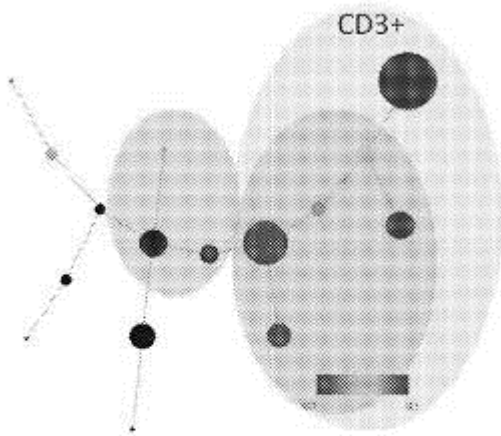


FIG. 19B

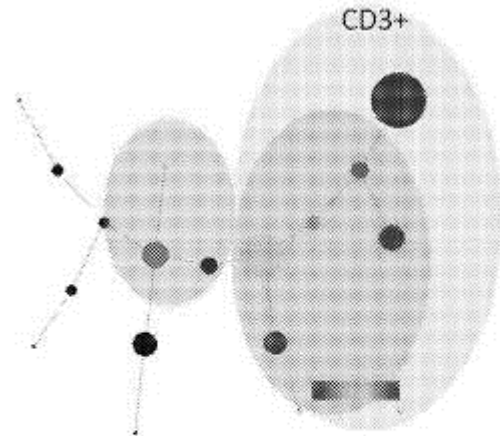


FIG. 19C

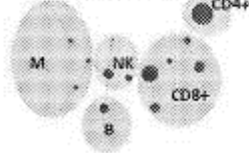


FIG. 19A-C

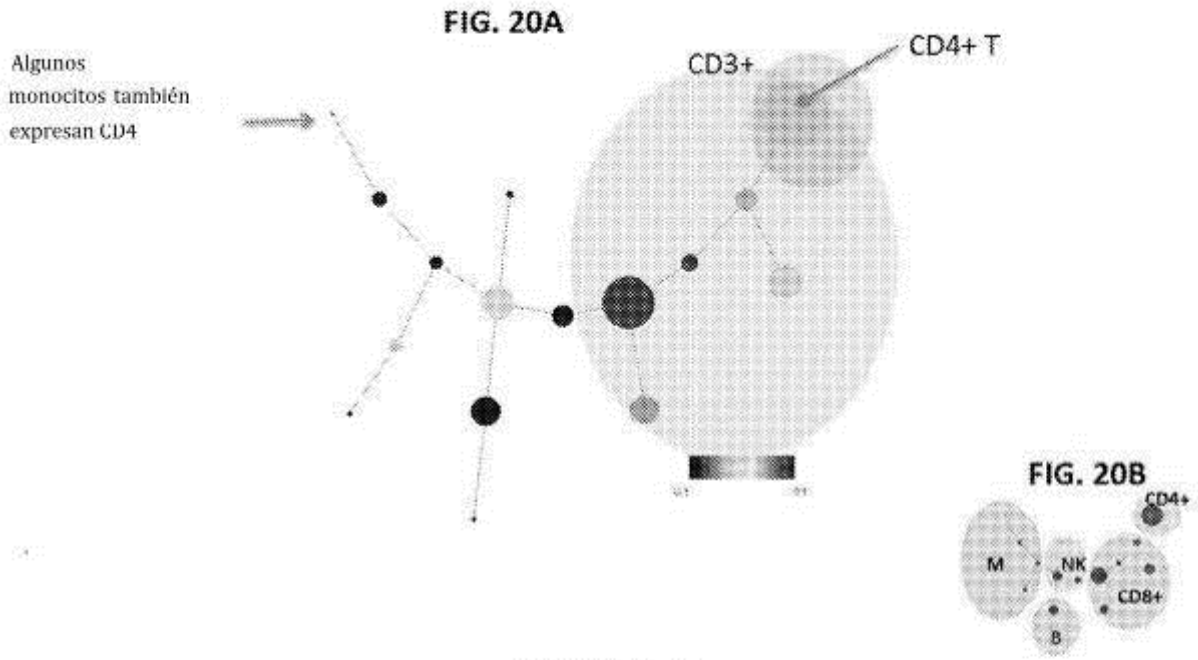


FIG. 20A-B

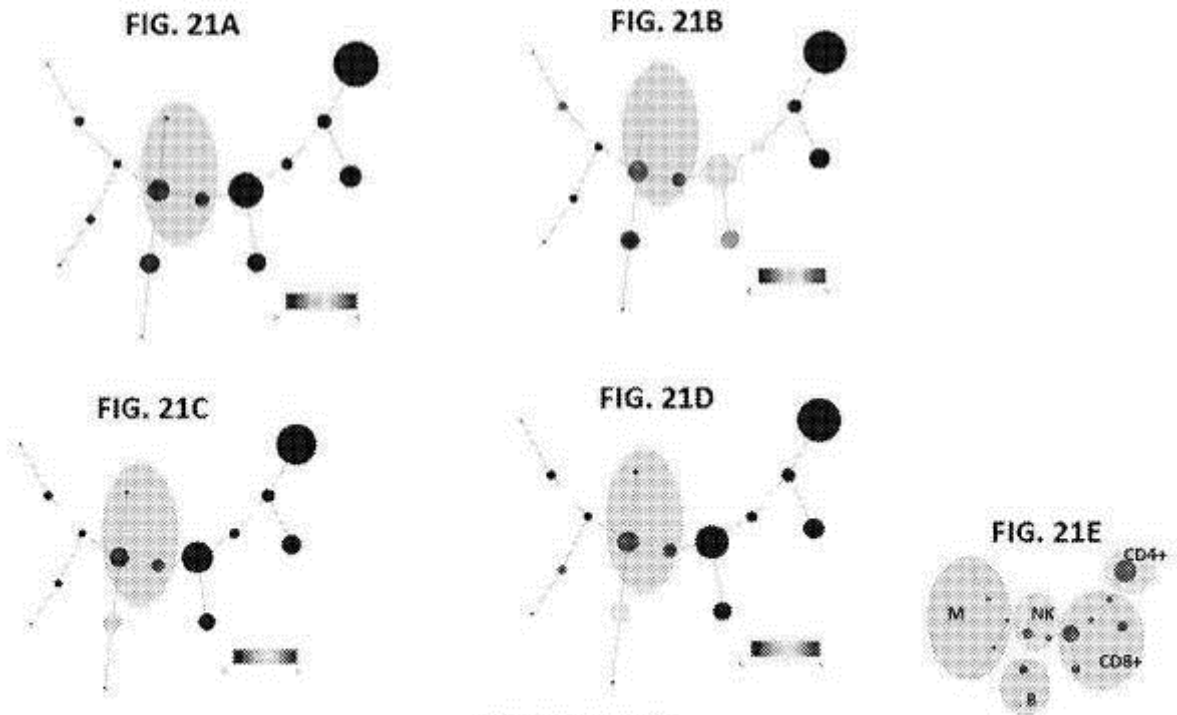


FIG. 21A-E

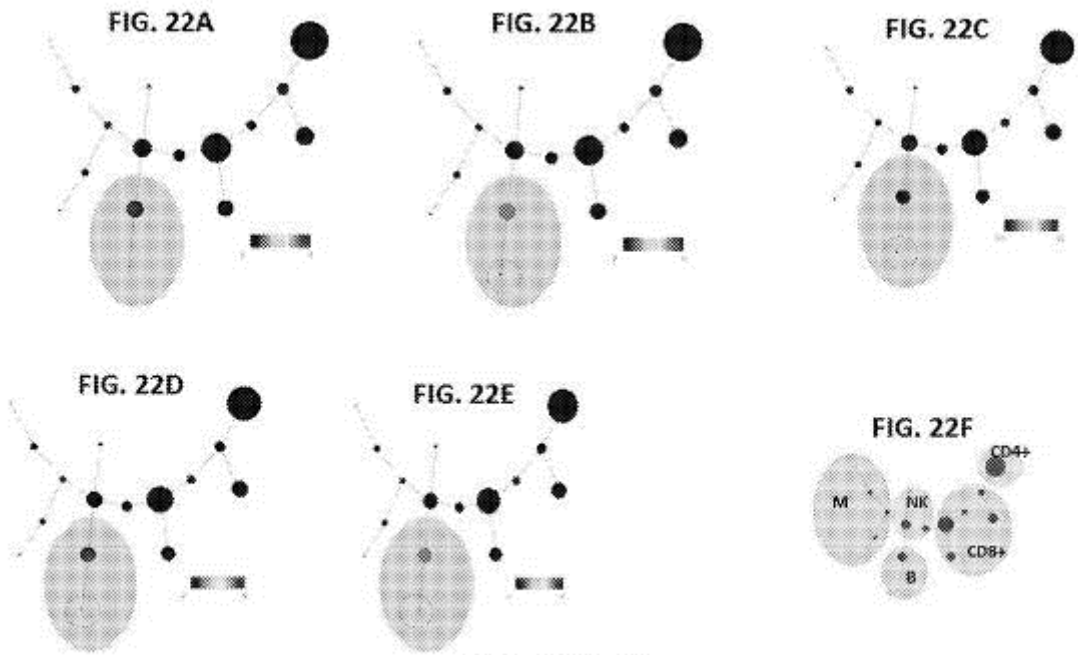


FIG. 22A-F

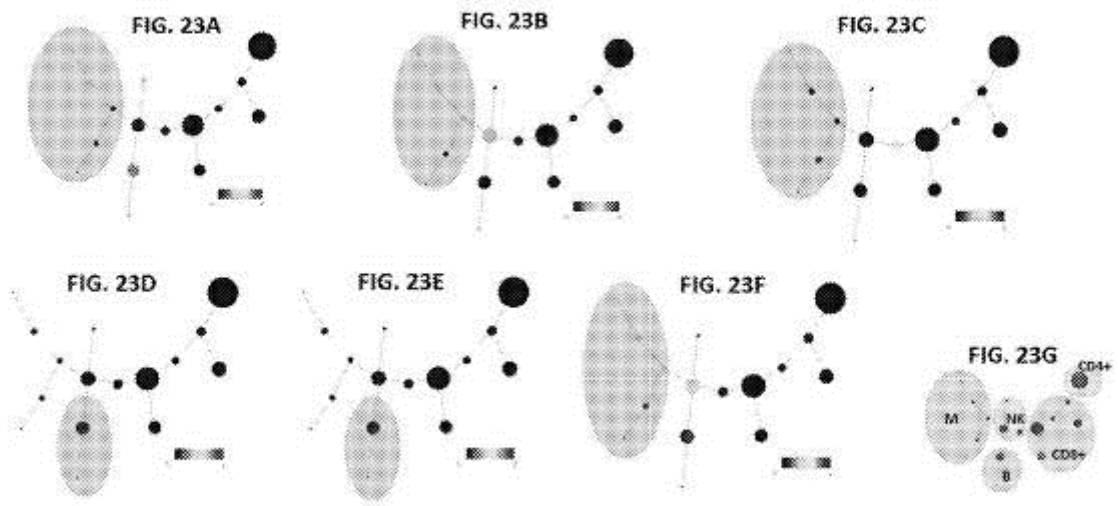


FIG. 23A-G

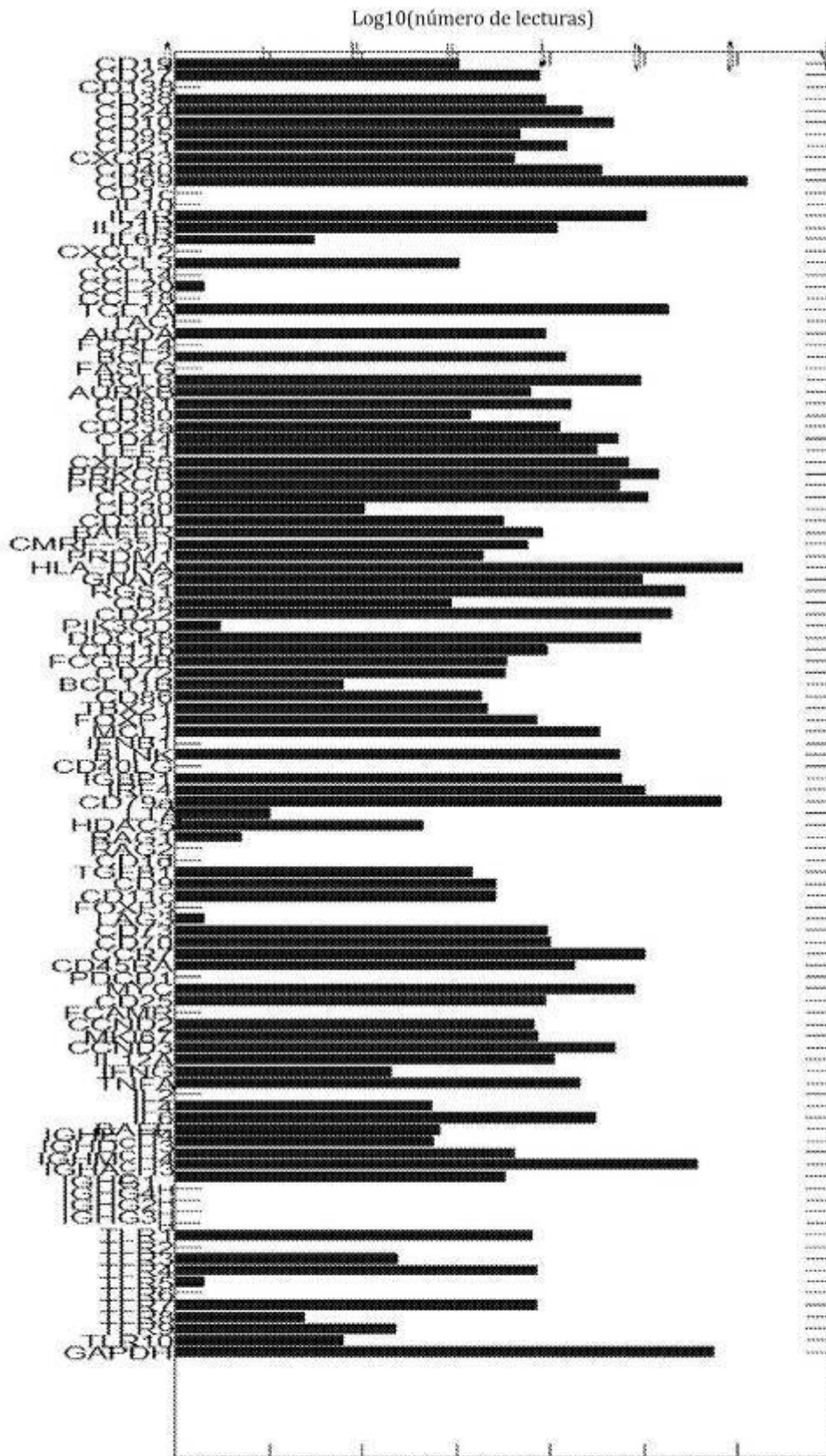


FIG. 24

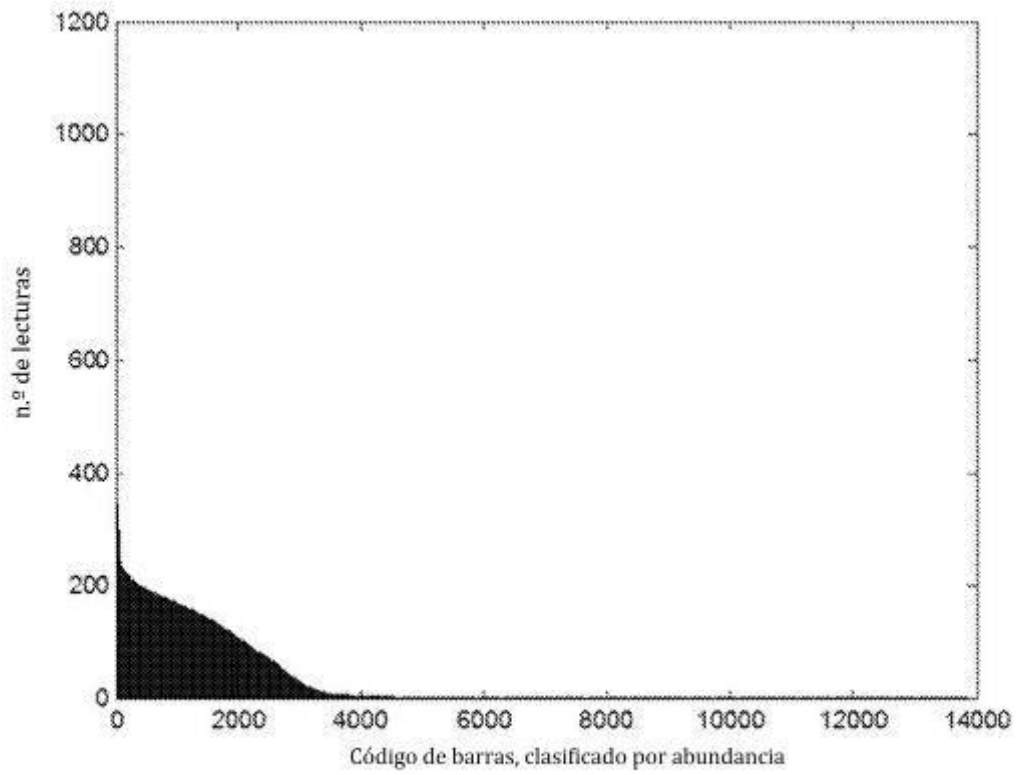


FIG. 25A

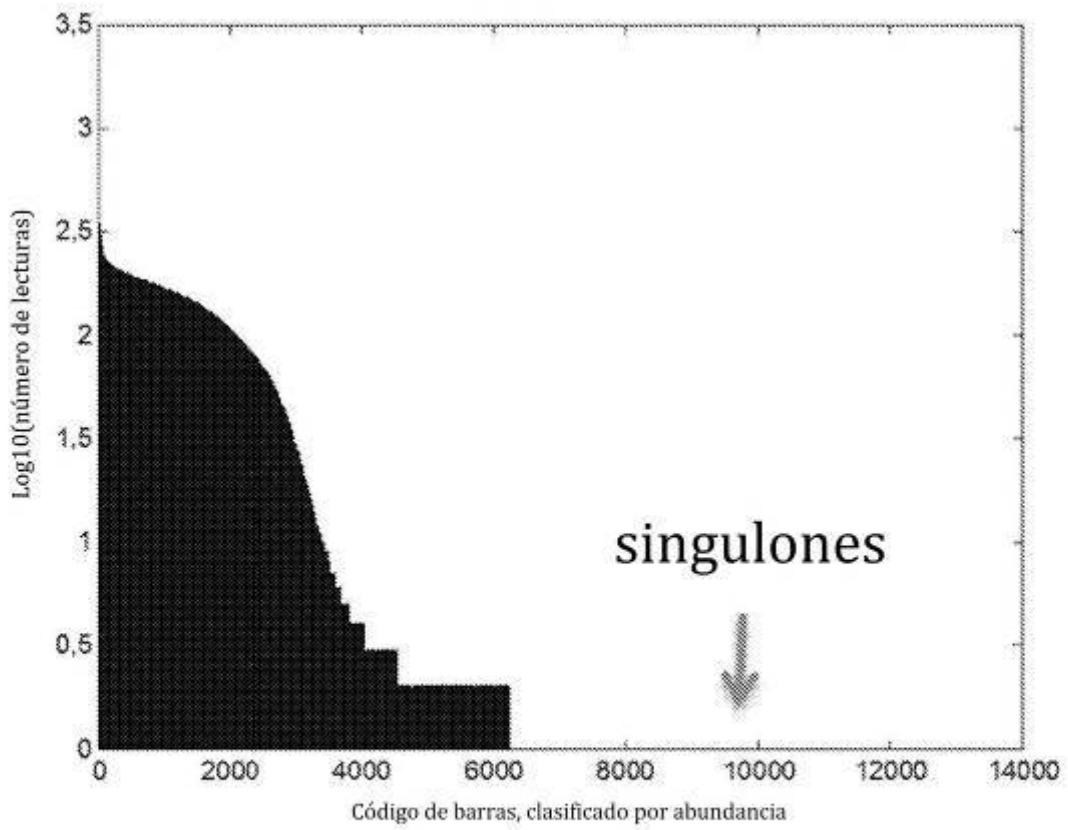


FIG. 25B

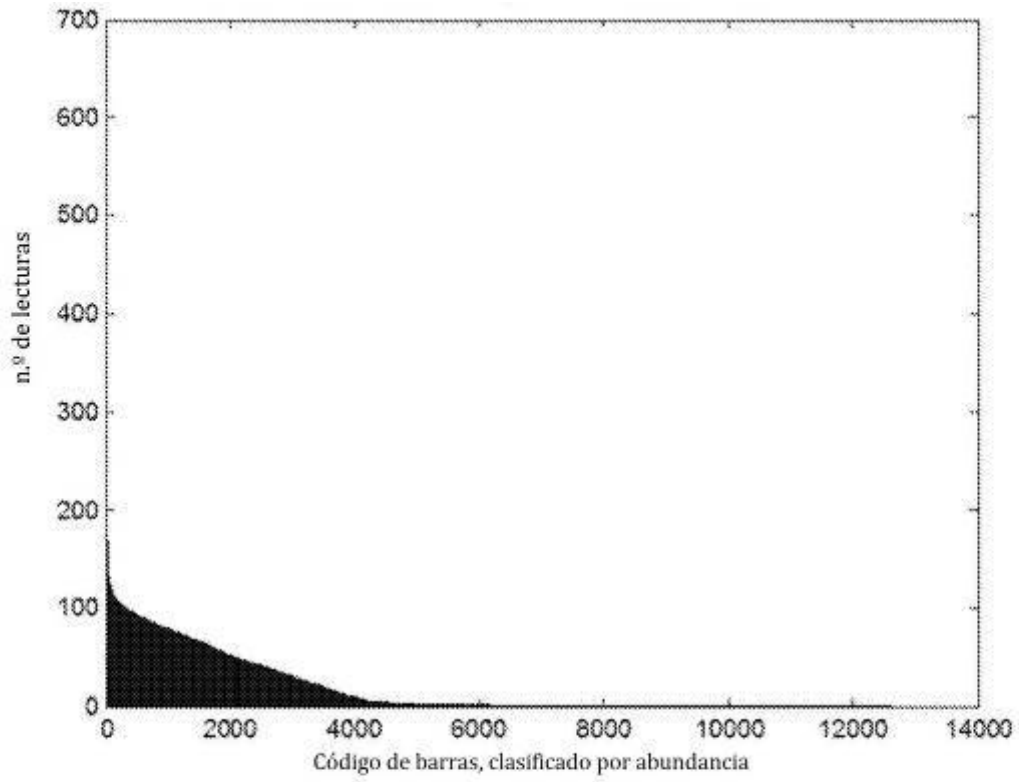


FIG. 25C

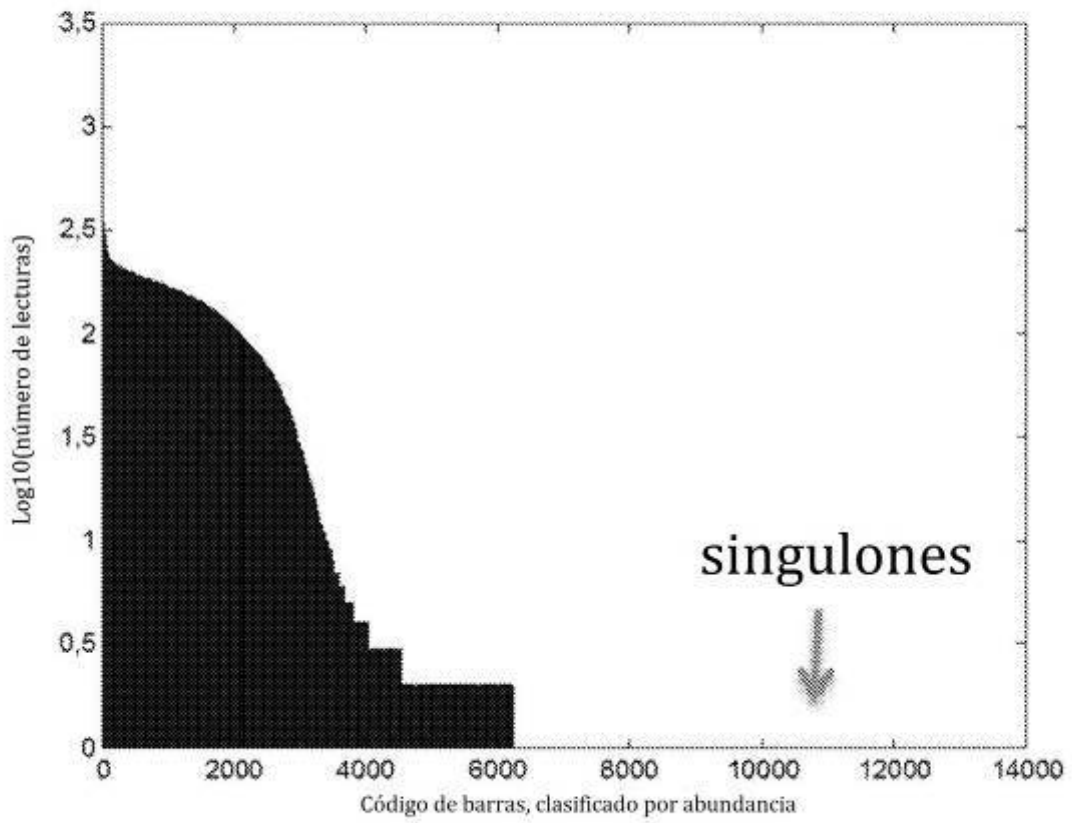


FIG. 25D

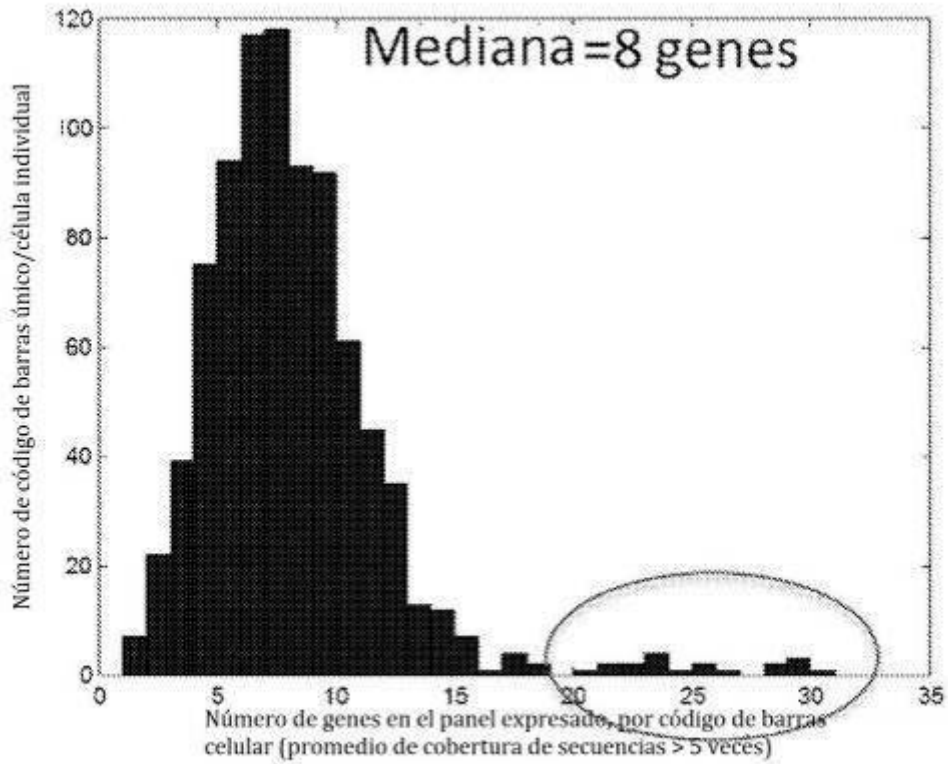


FIG. 26A

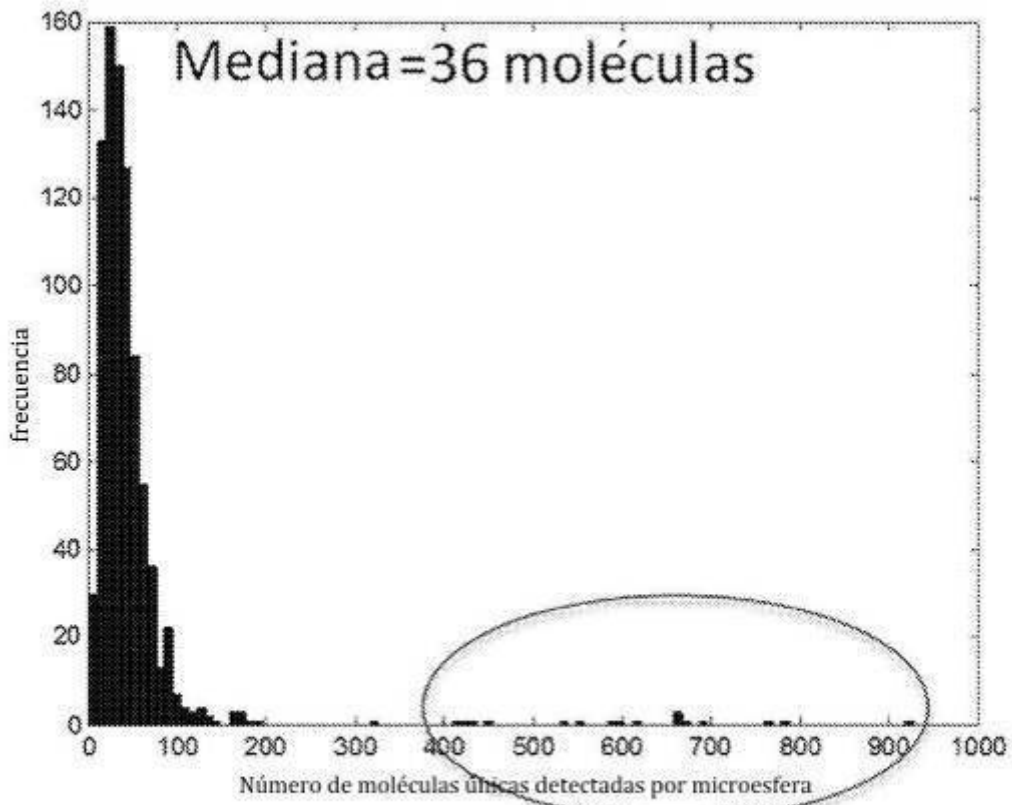


FIG. 26B

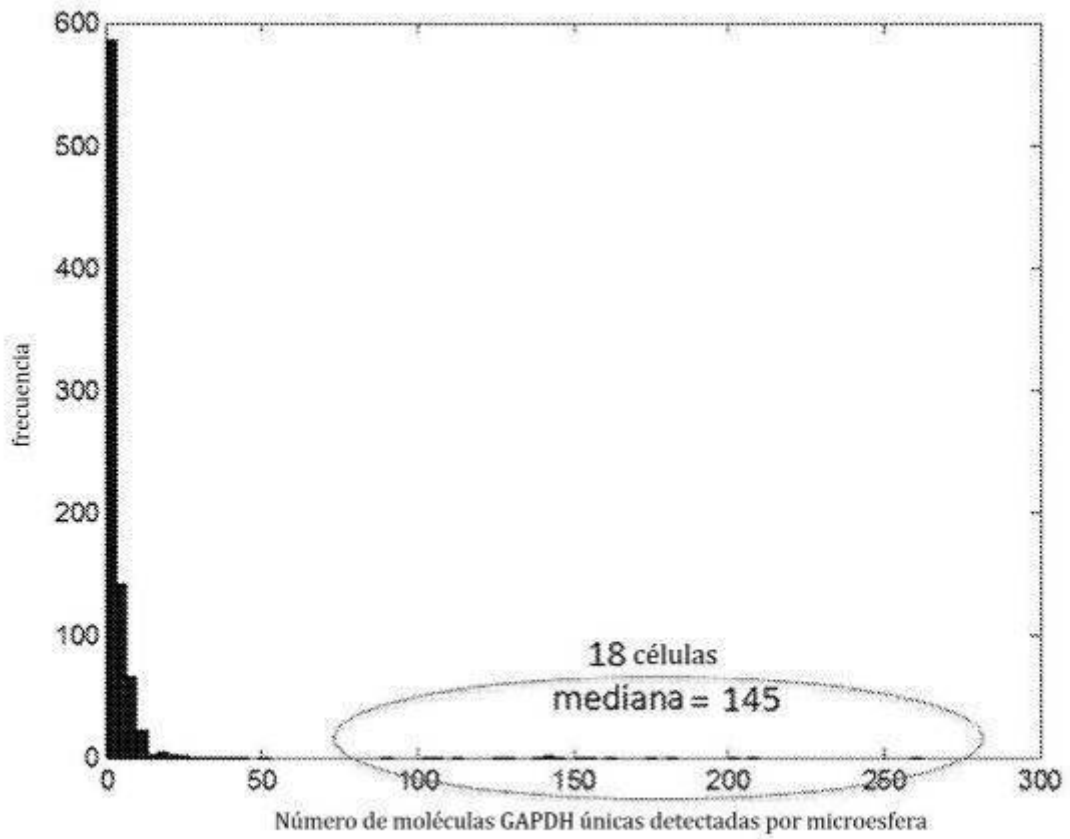


FIG. 26C

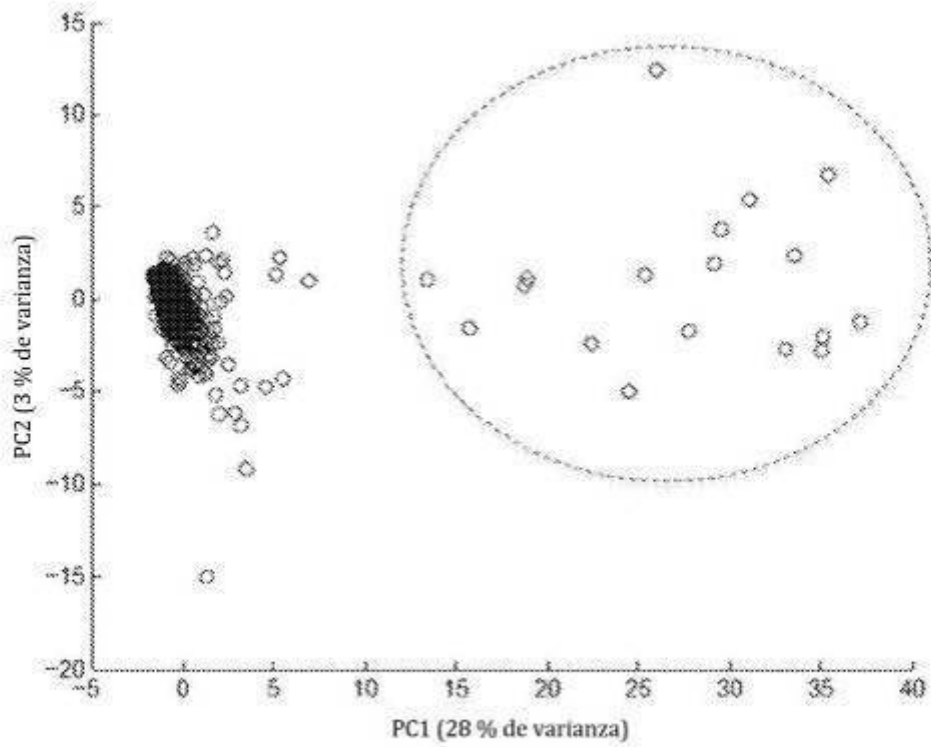


FIG. 27

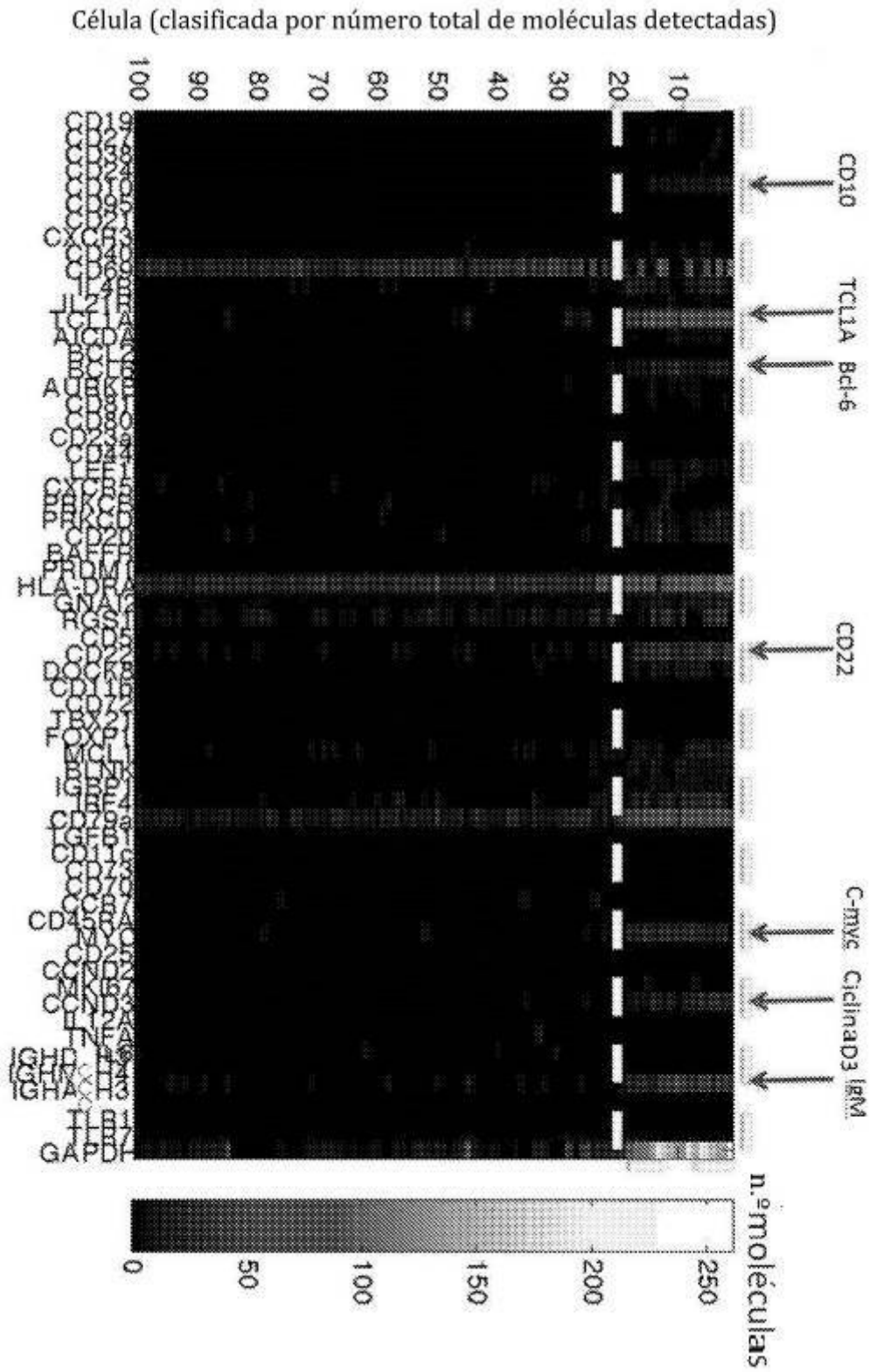


FIG. 28

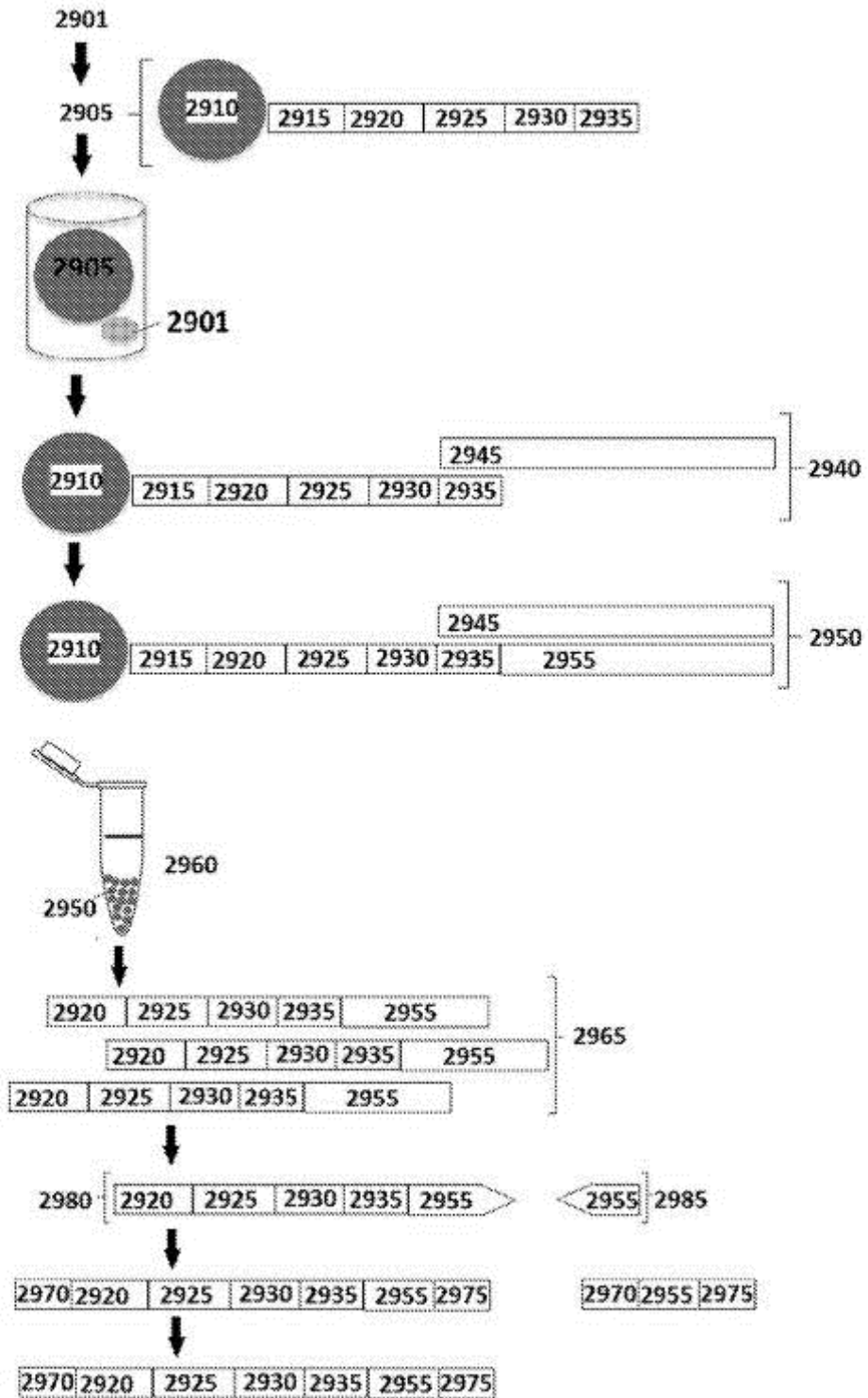


FIG. 29

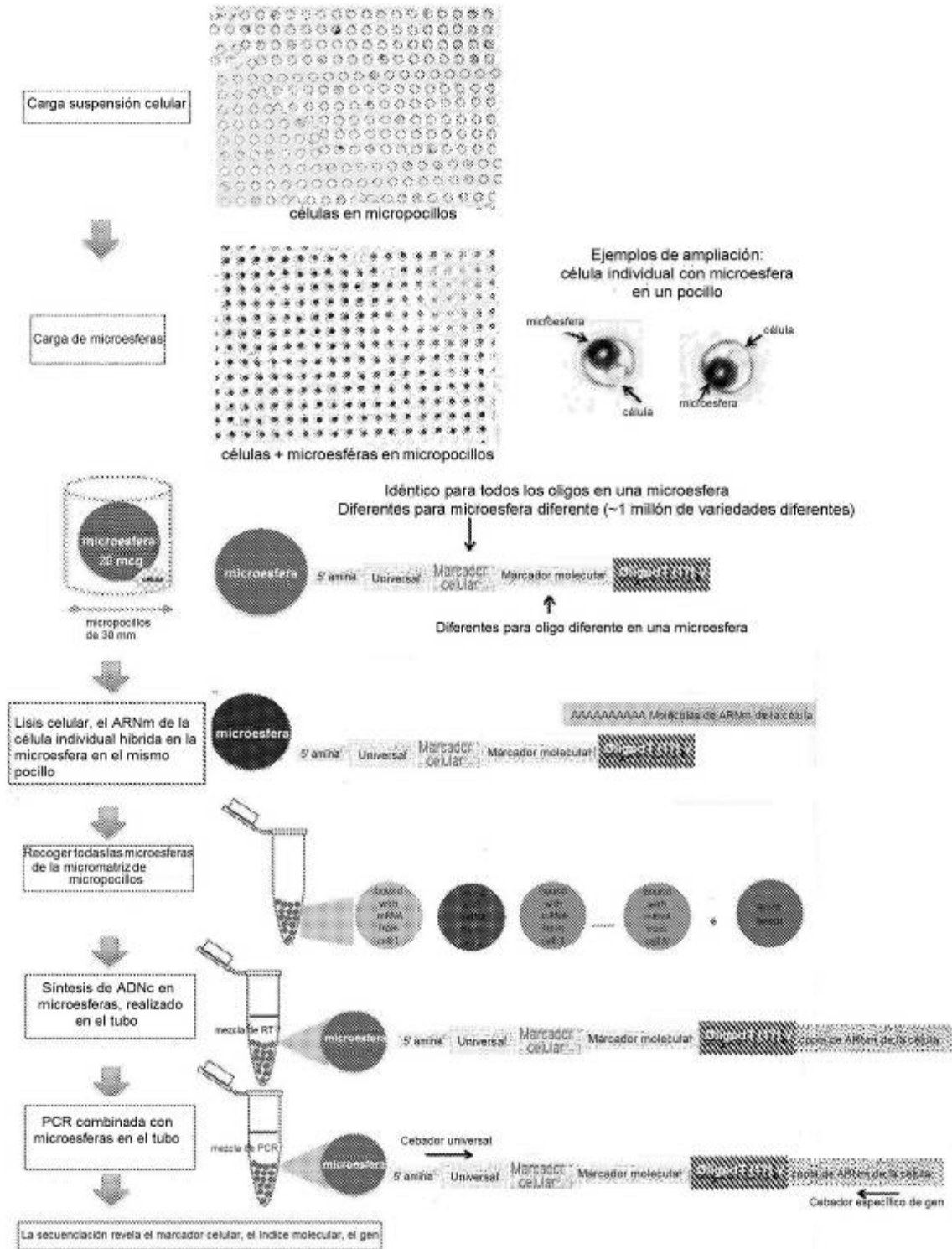


FIG. 30

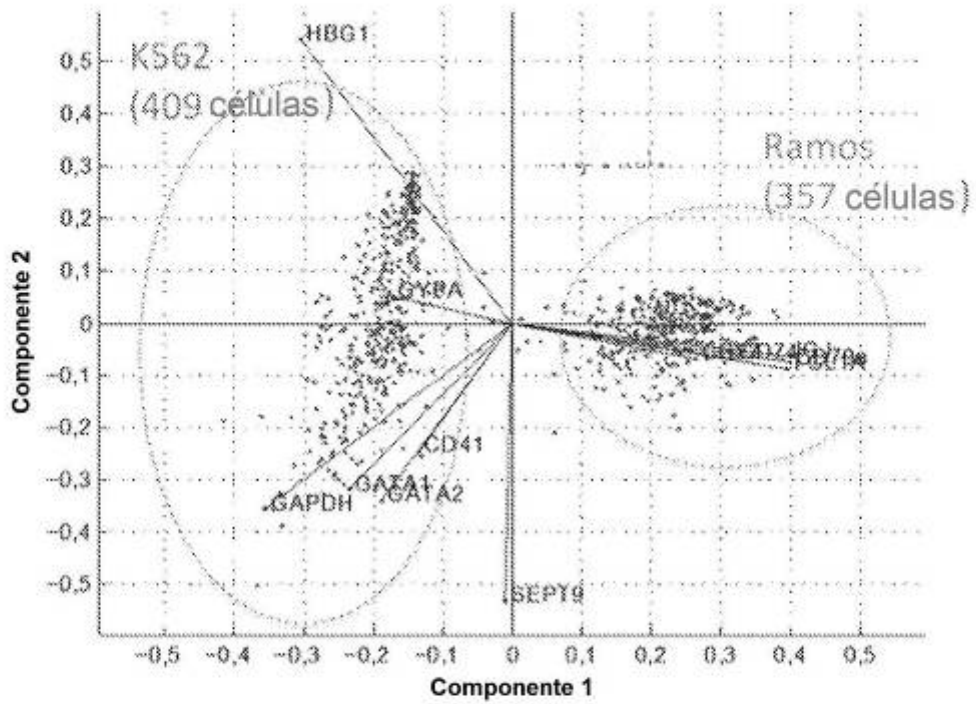


FIG. 31A

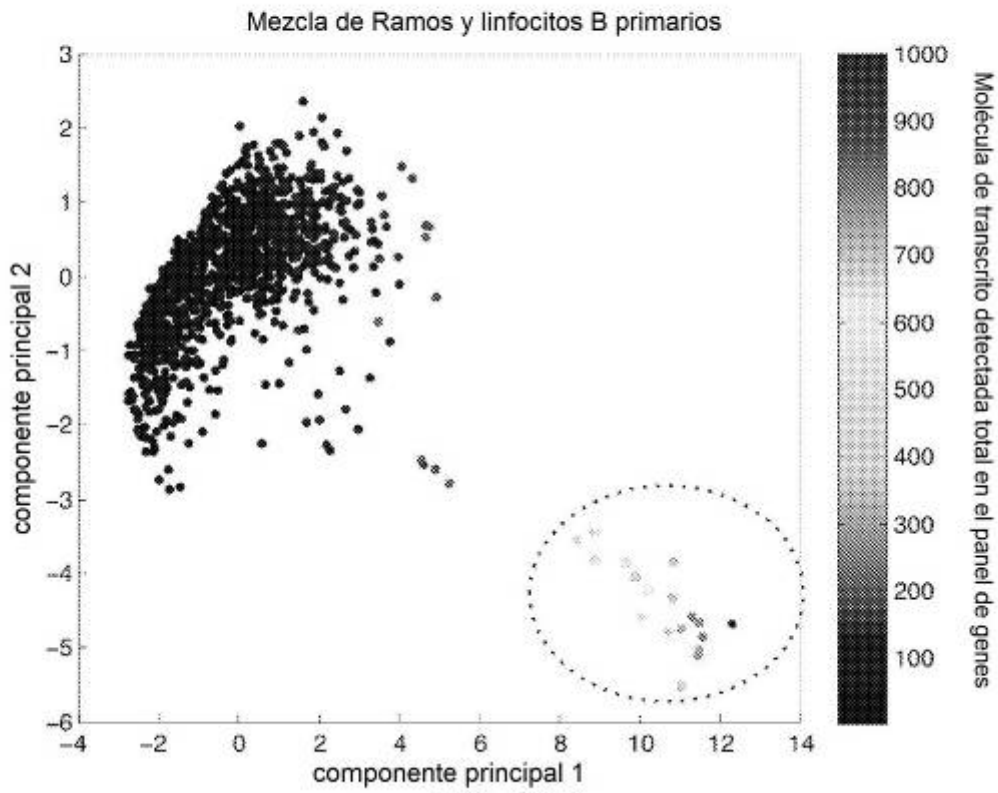


FIG. 31B

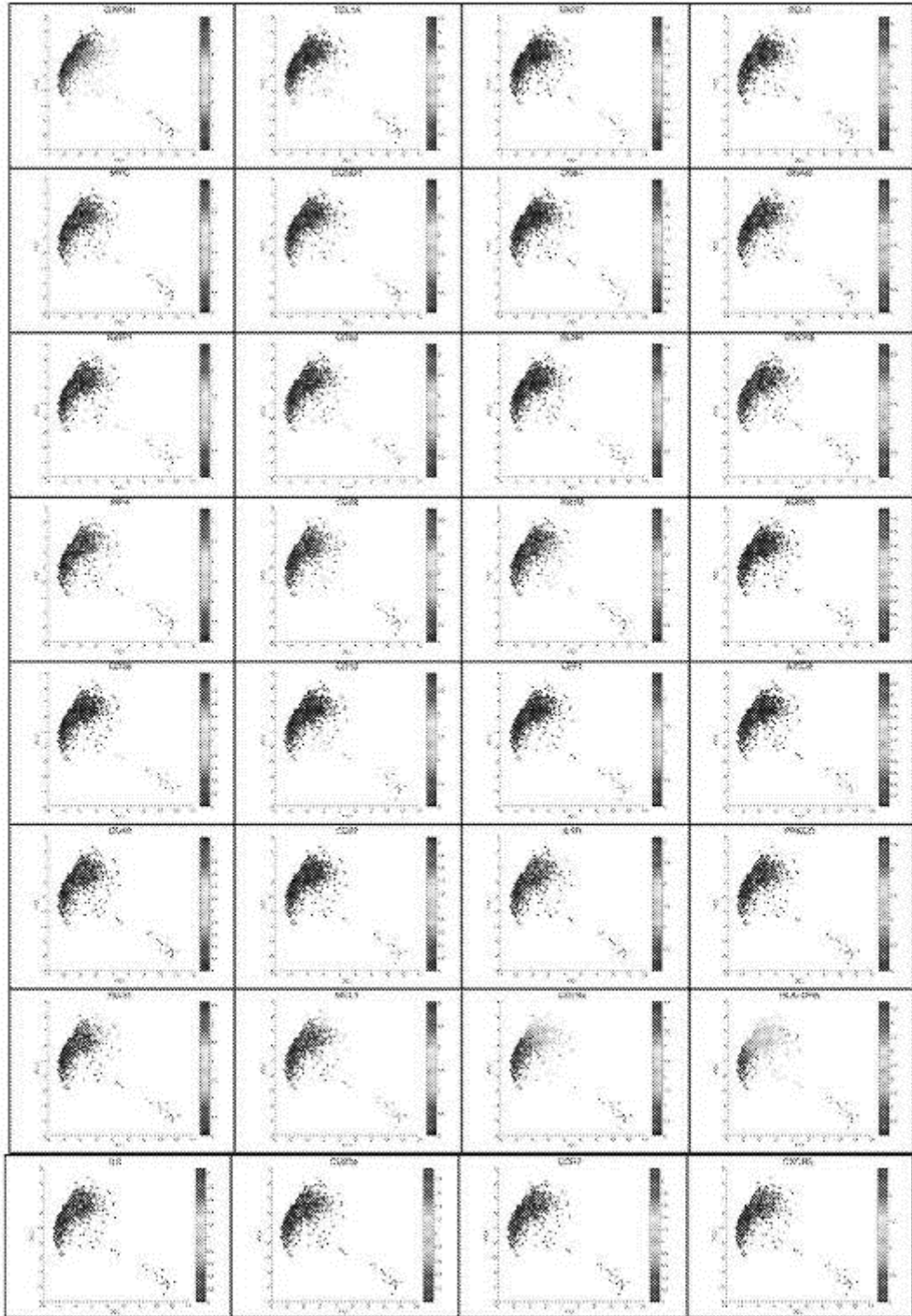


FIG. 31D

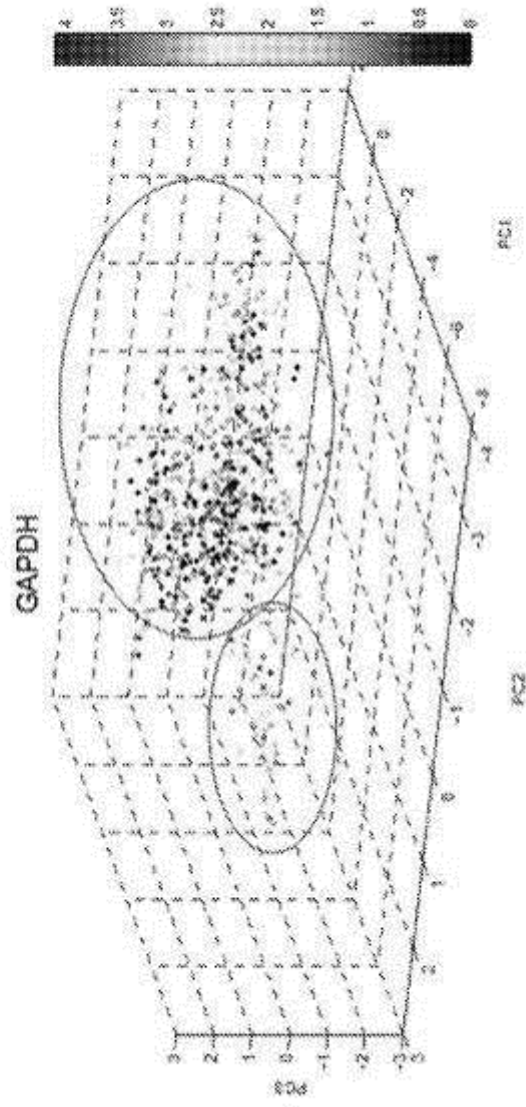


FIG. 32

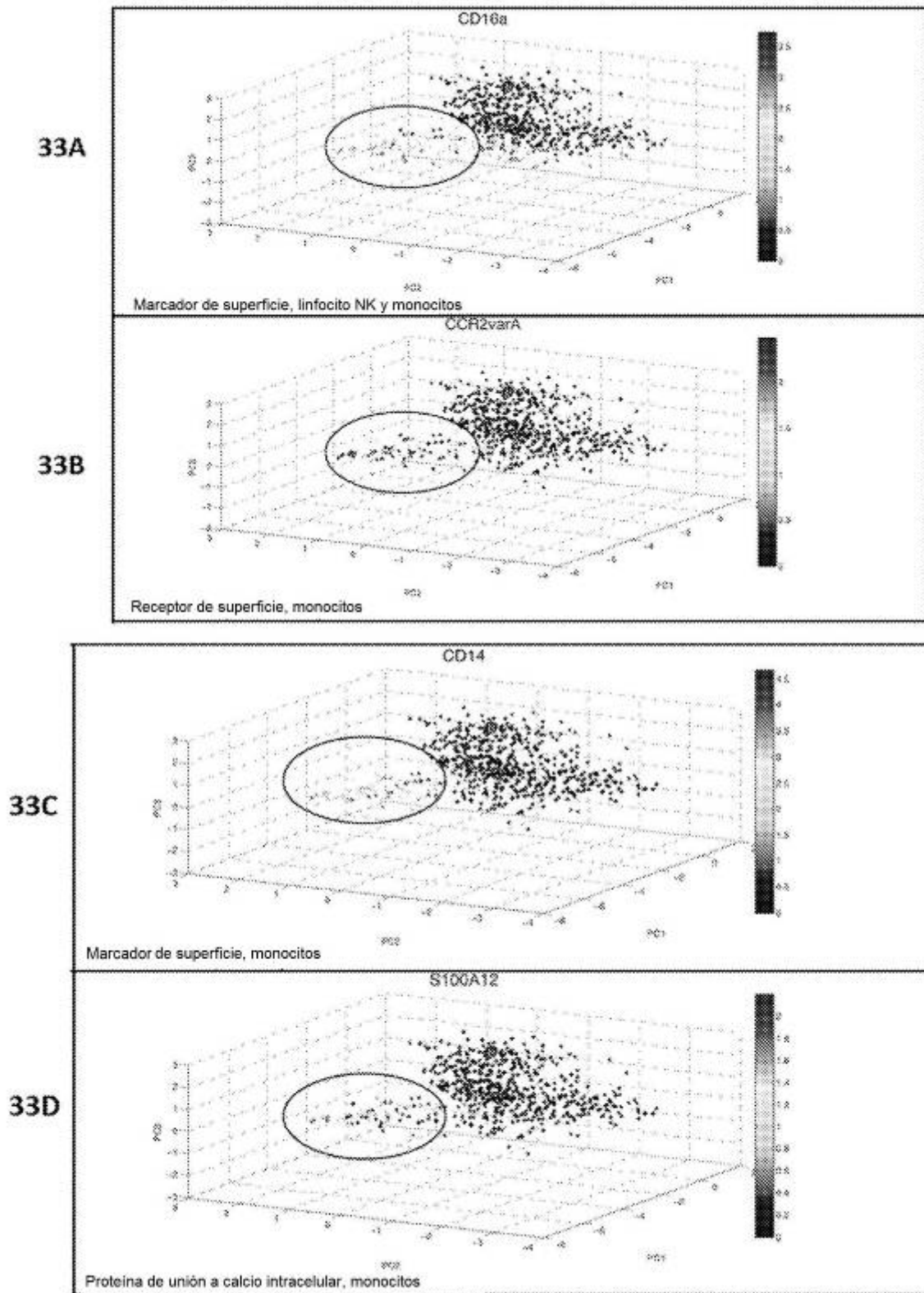


FIG. 33A-D

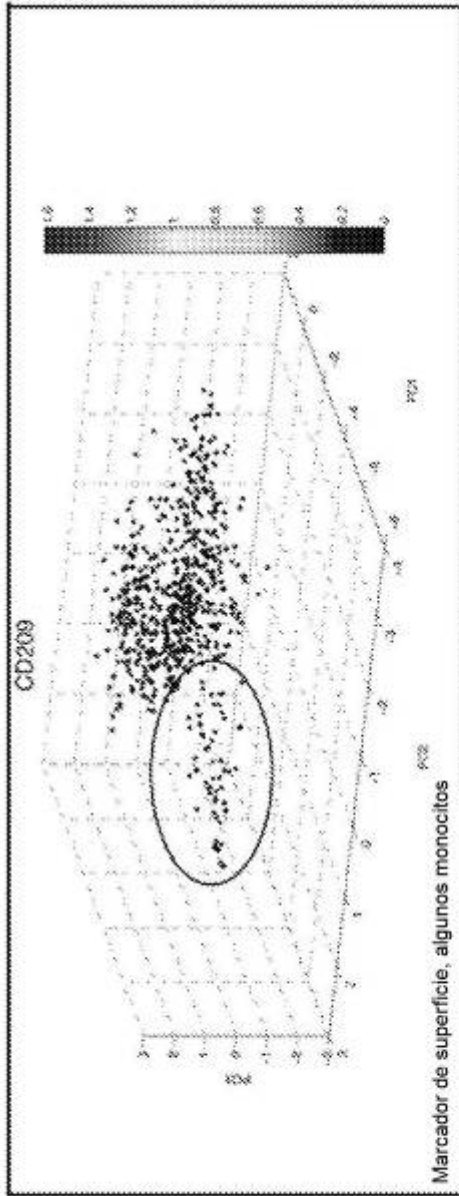


FIG. 33E

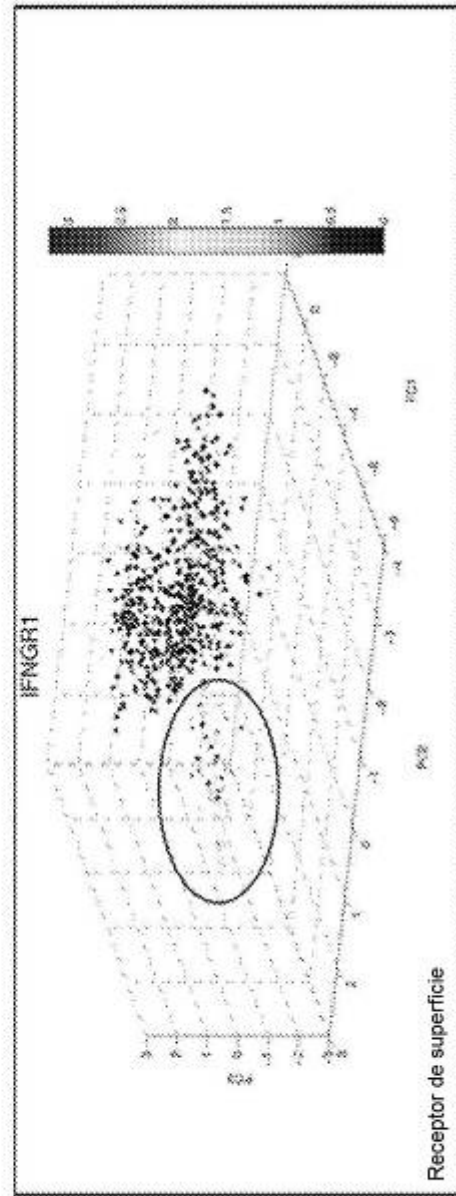


FIG. 33F

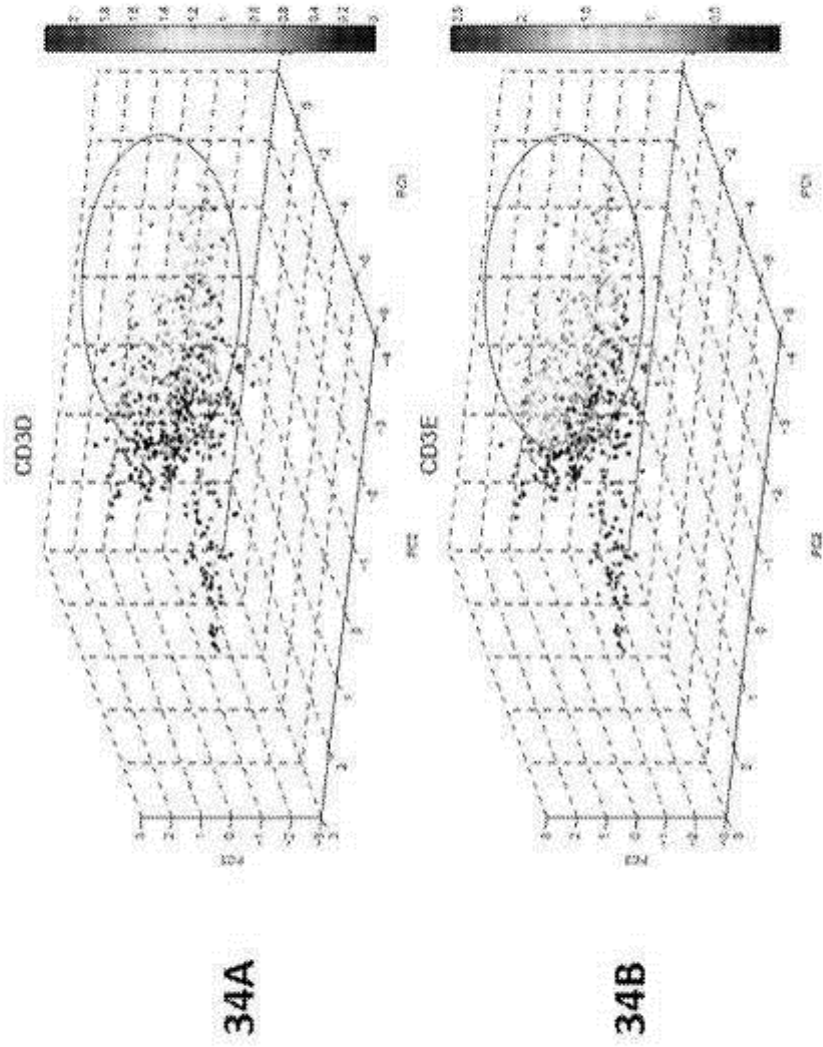


FIG. 34A-B

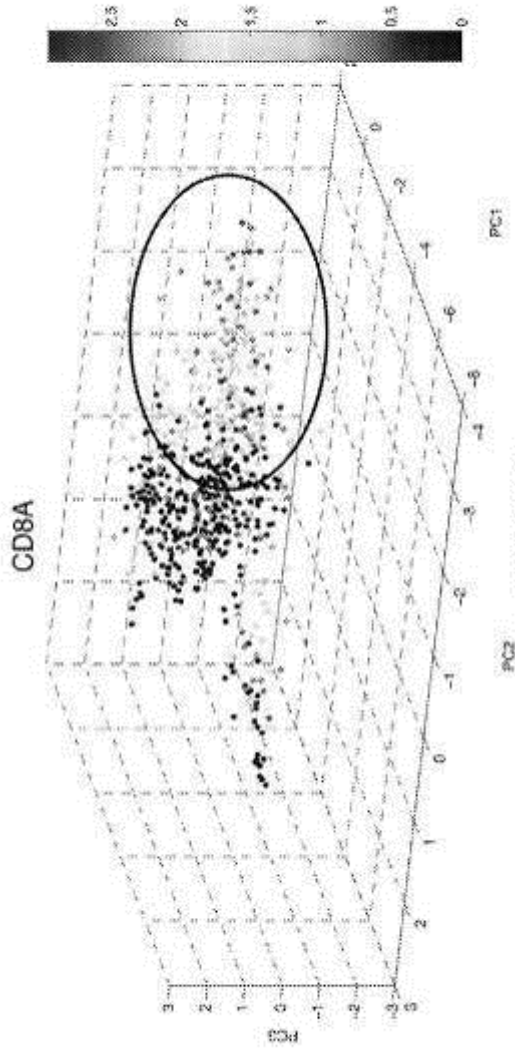


FIG. 35A

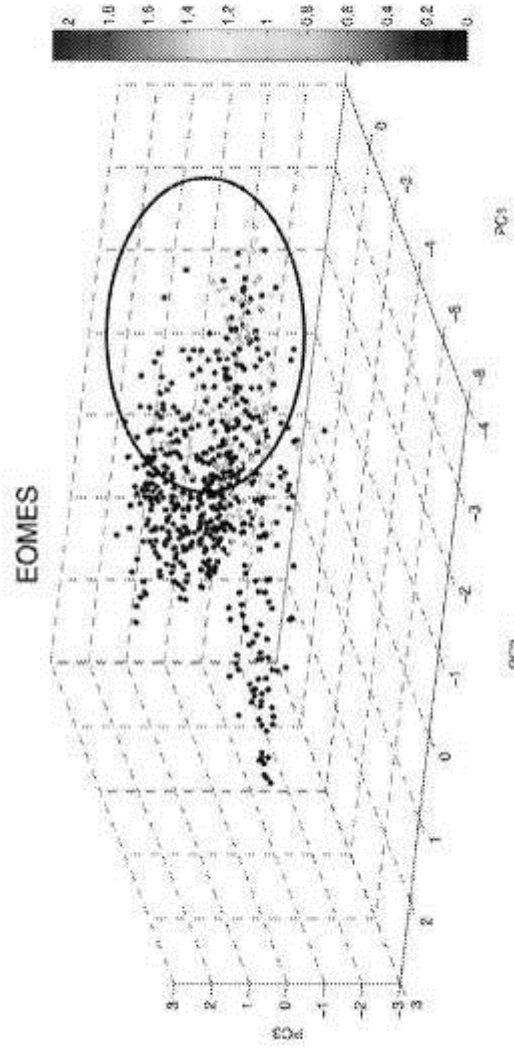


FIG. 35B

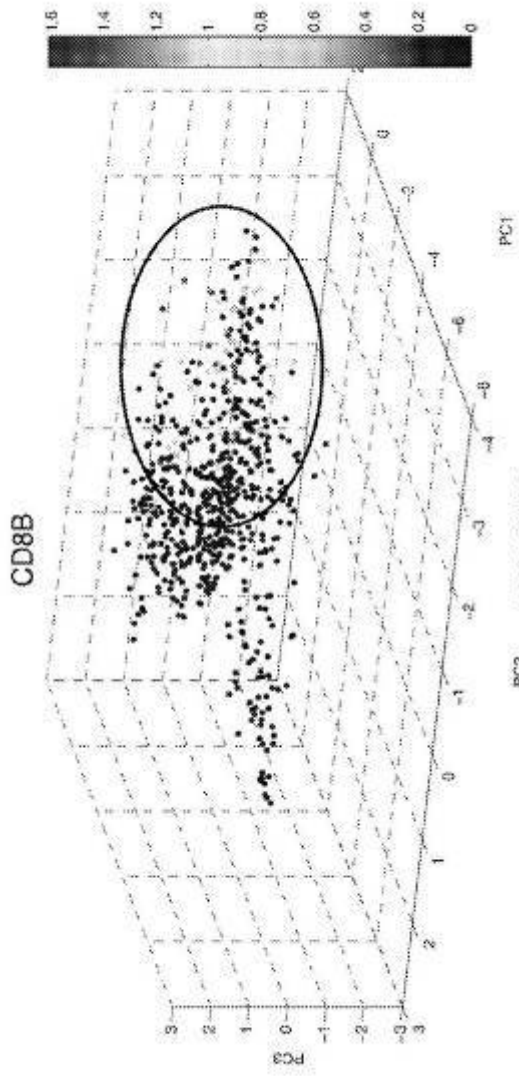


FIG. 35C

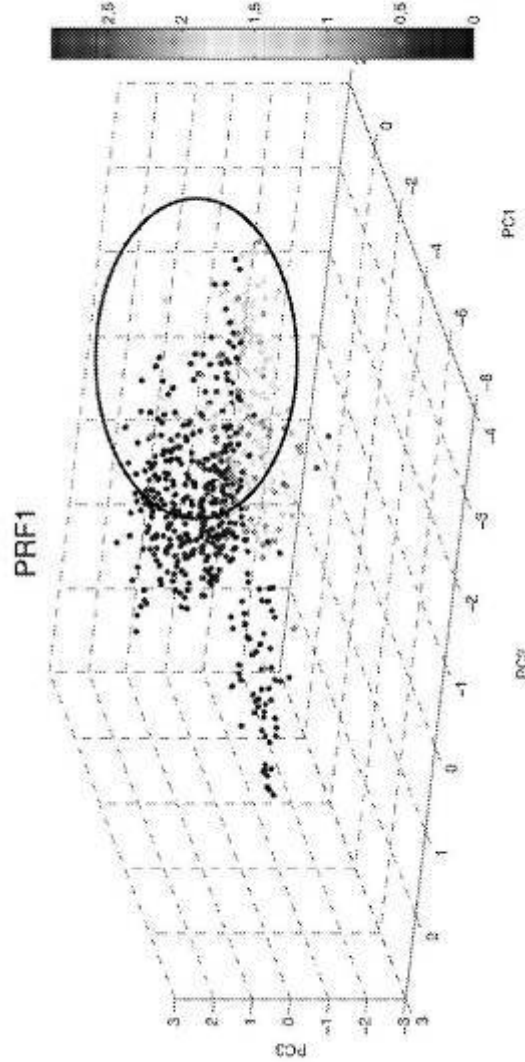


FIG. 35D

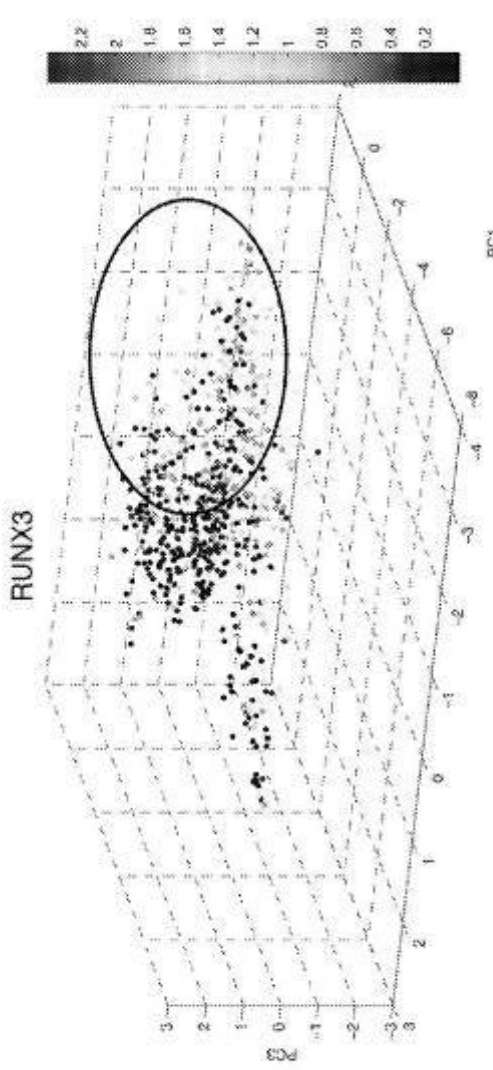


FIG. 35E

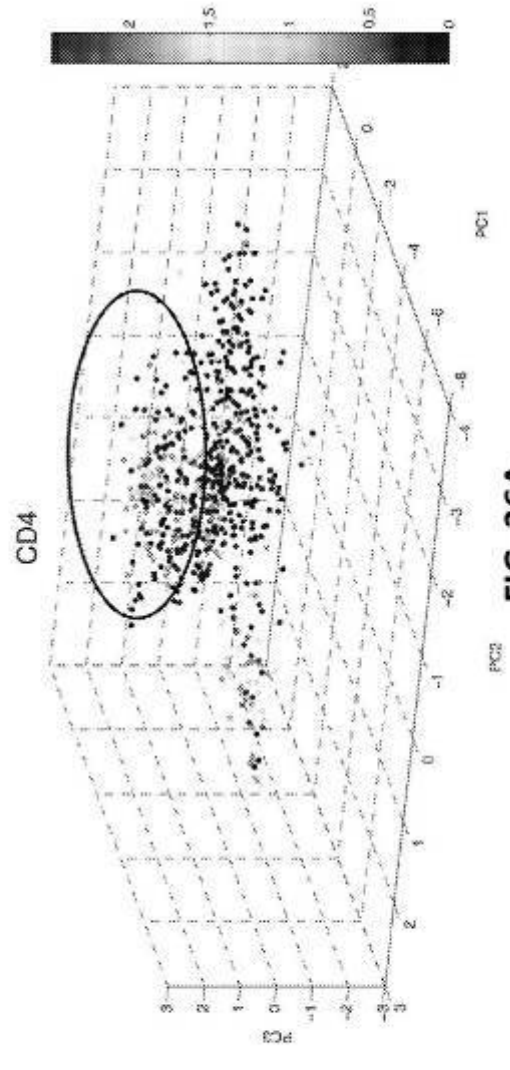


FIG. 36A

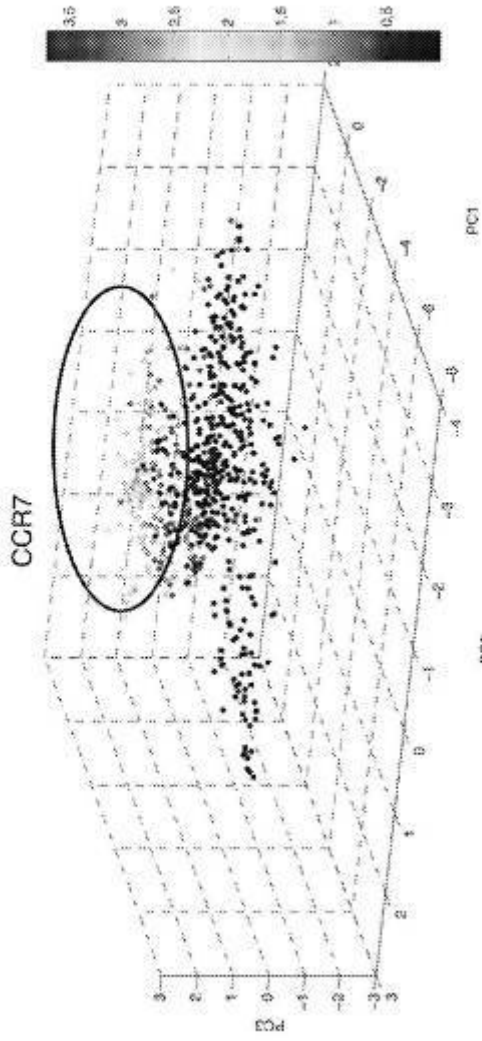


FIG. 36B

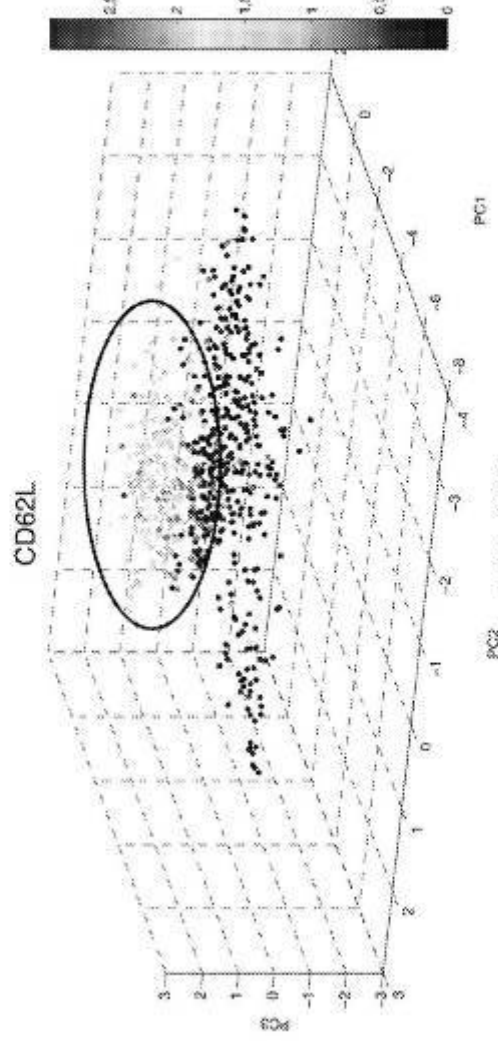
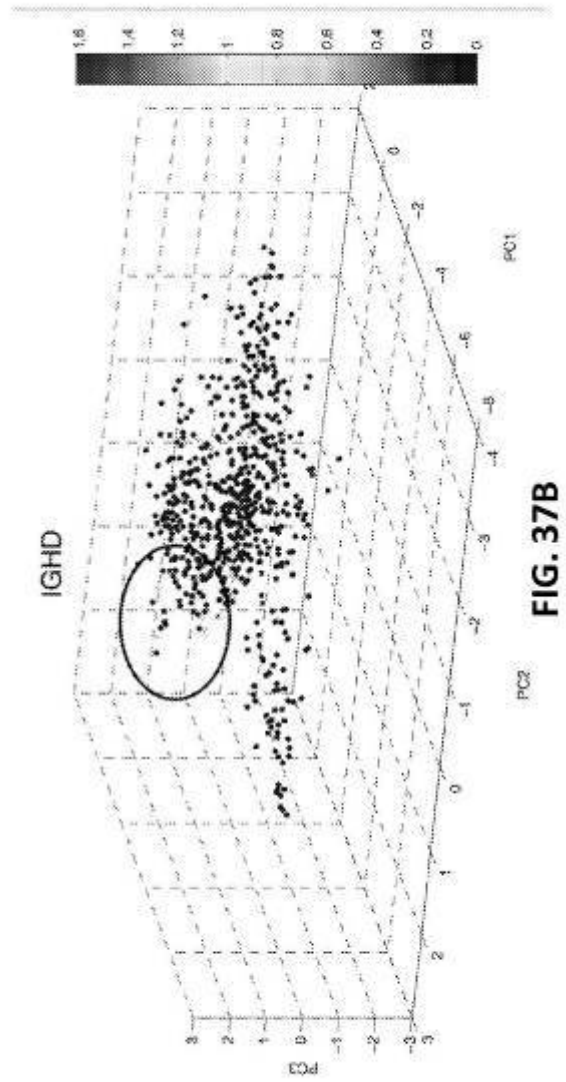
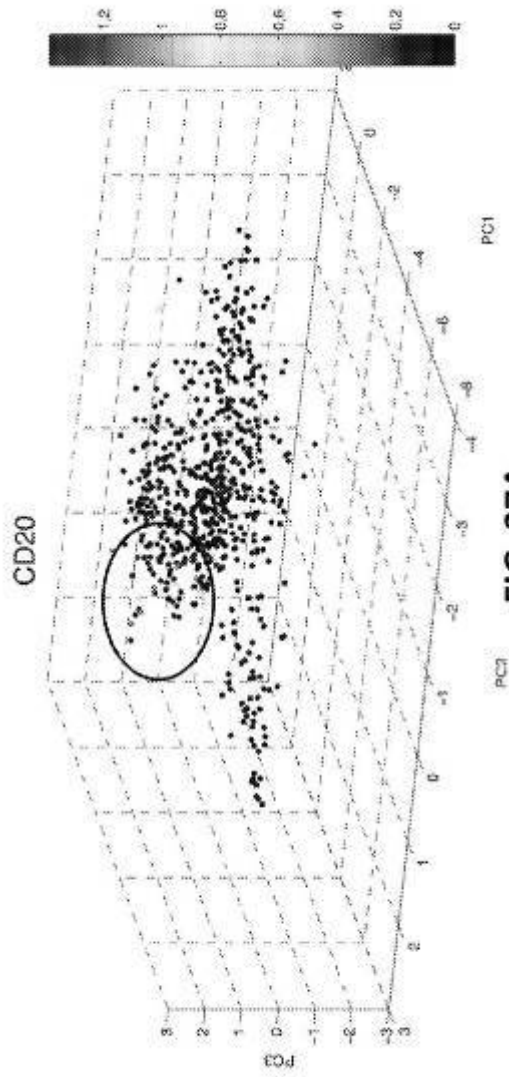
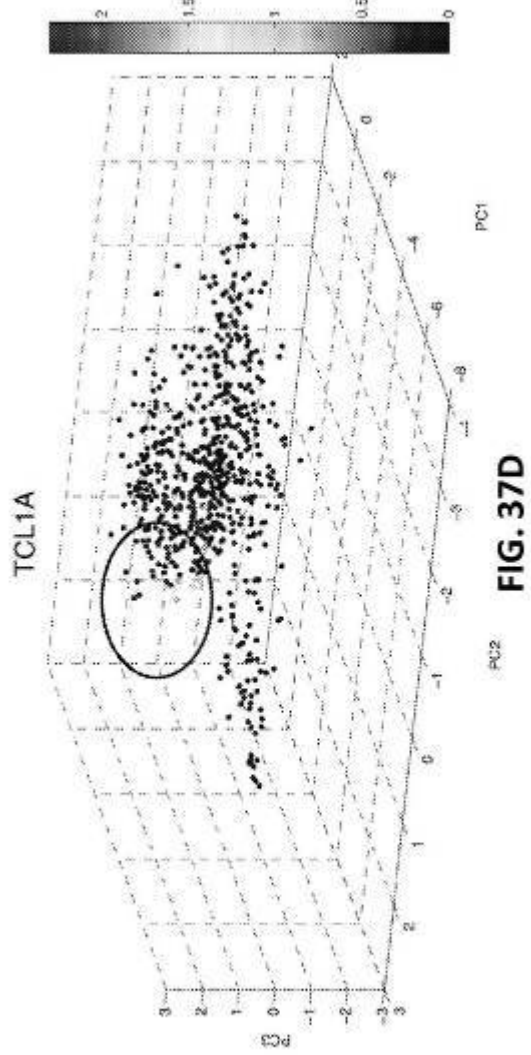
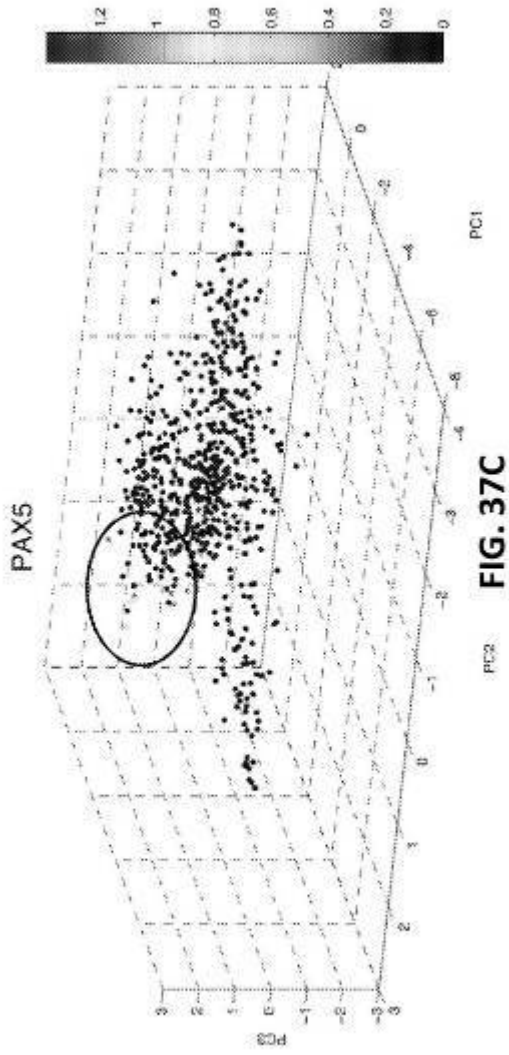
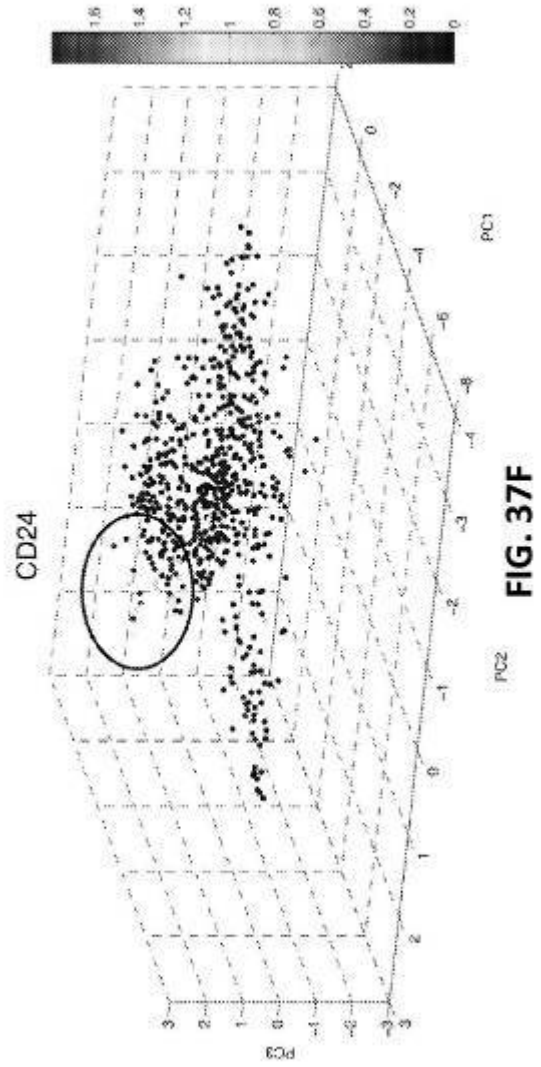
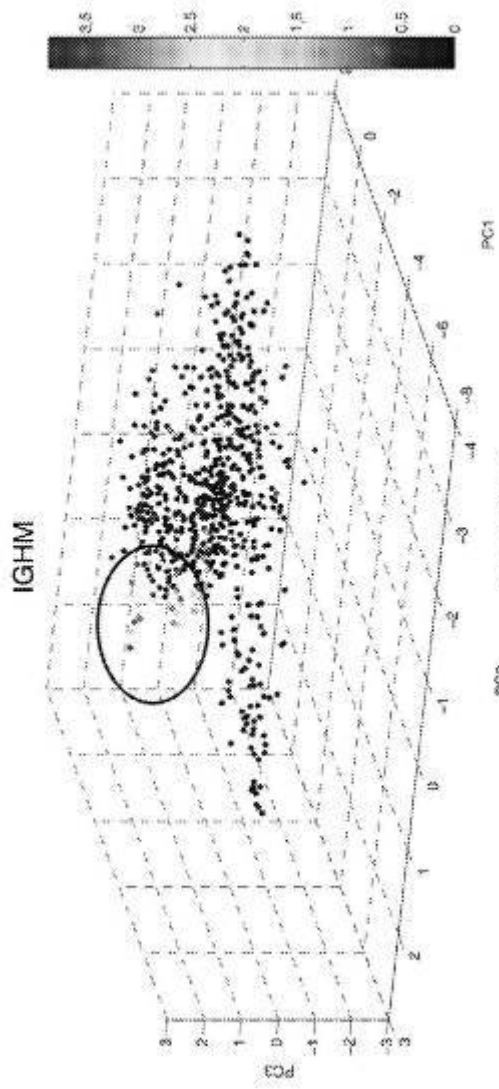


FIG. 36C







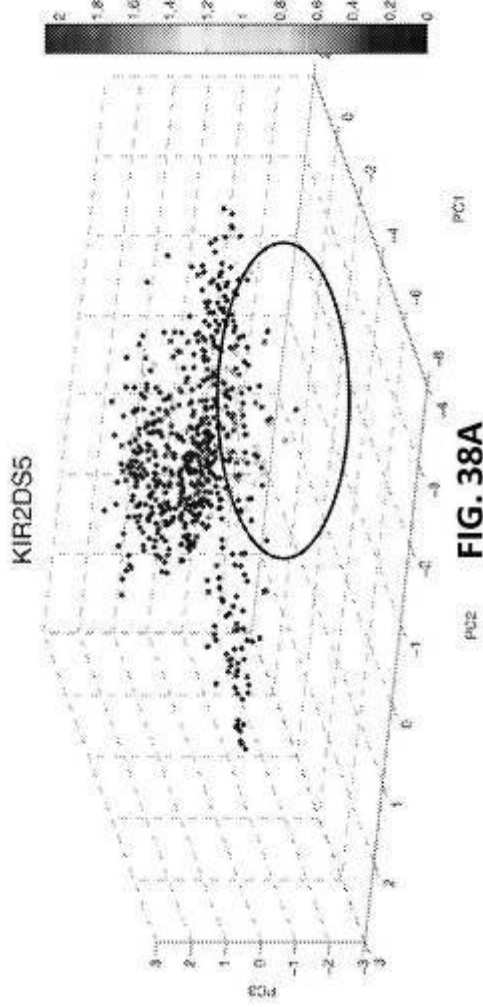


FIG. 38A

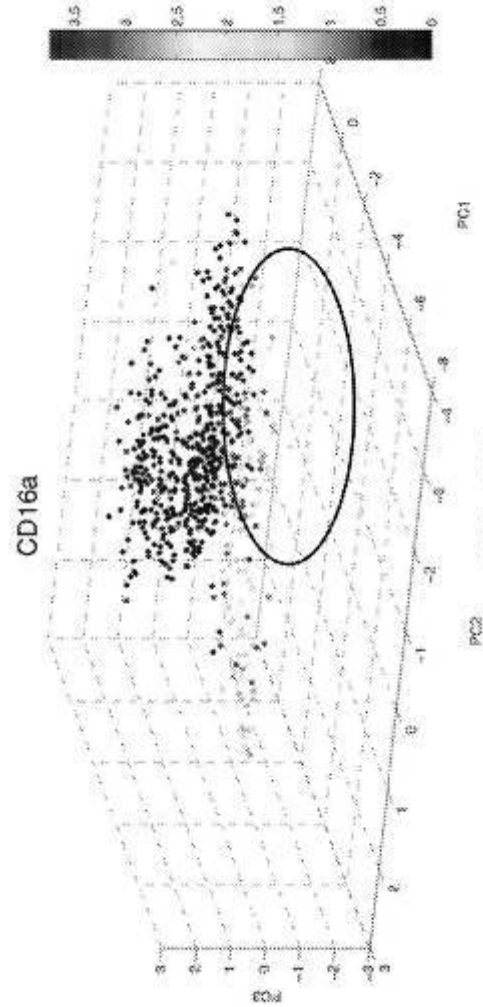


FIG. 38B

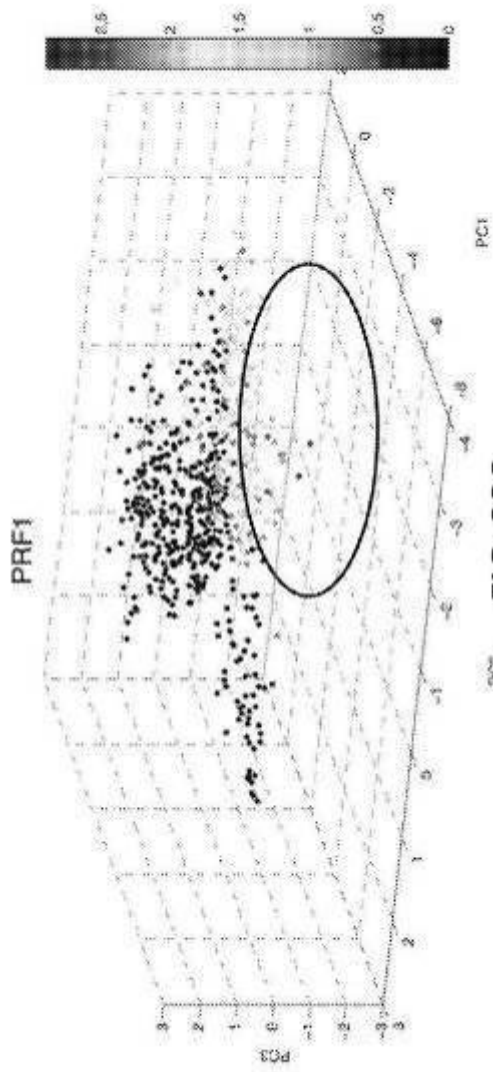


FIG. 38C

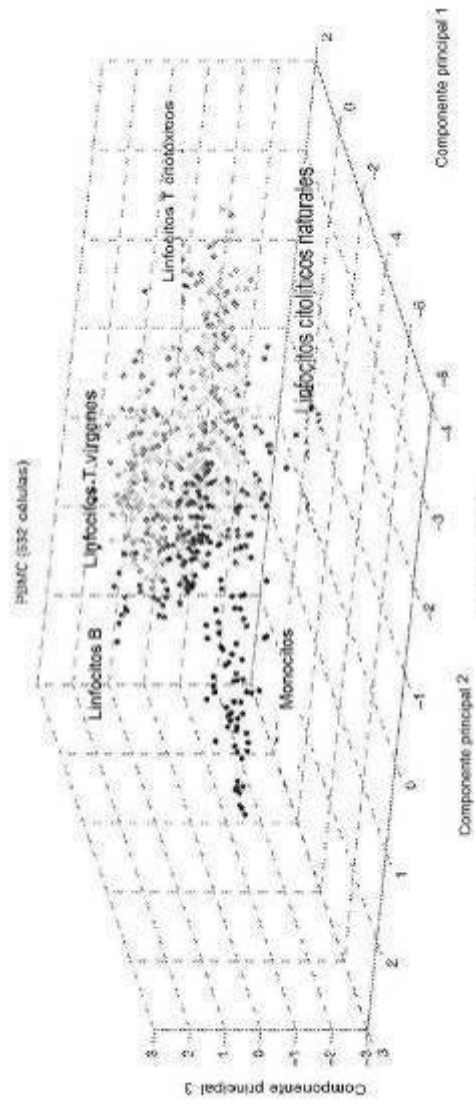


FIG. 39

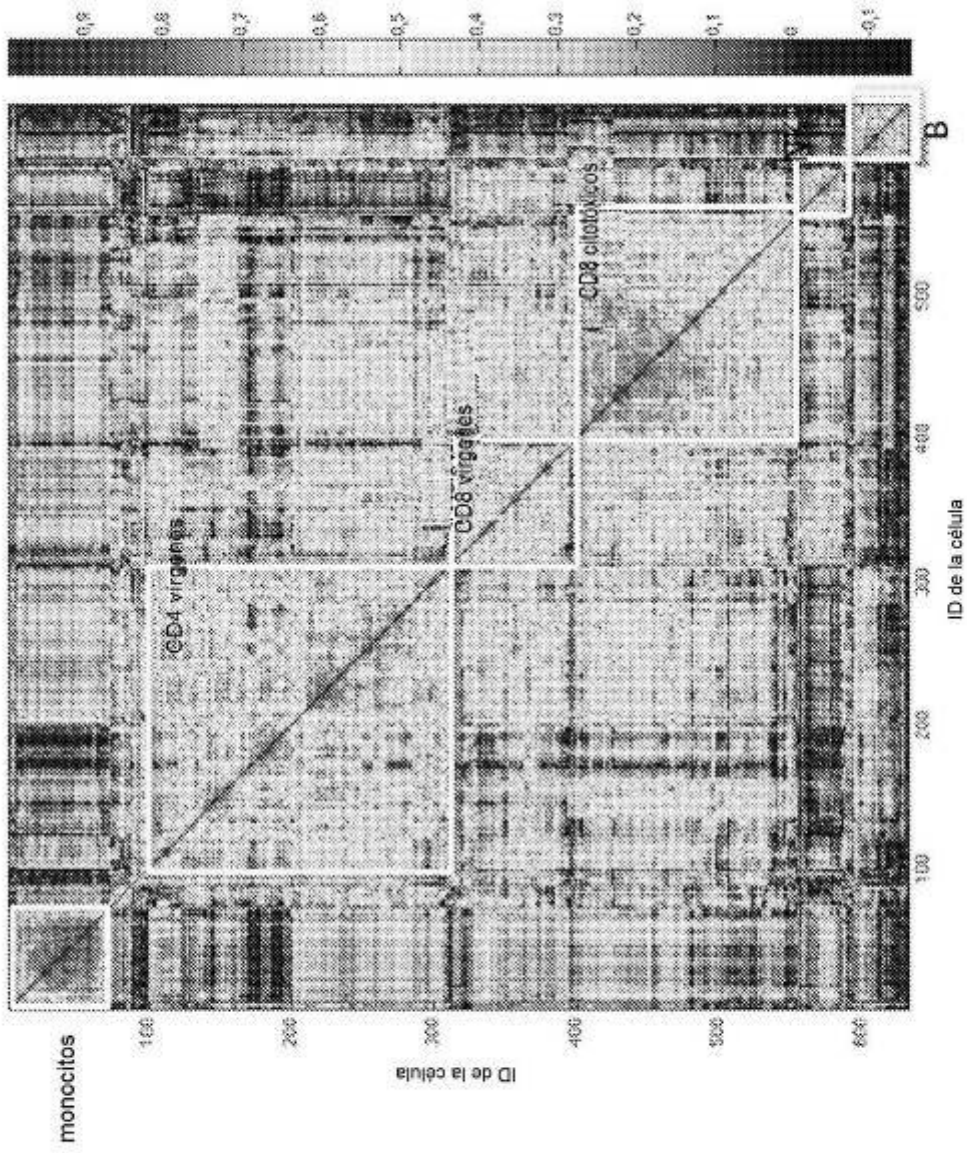


FIG. 40A

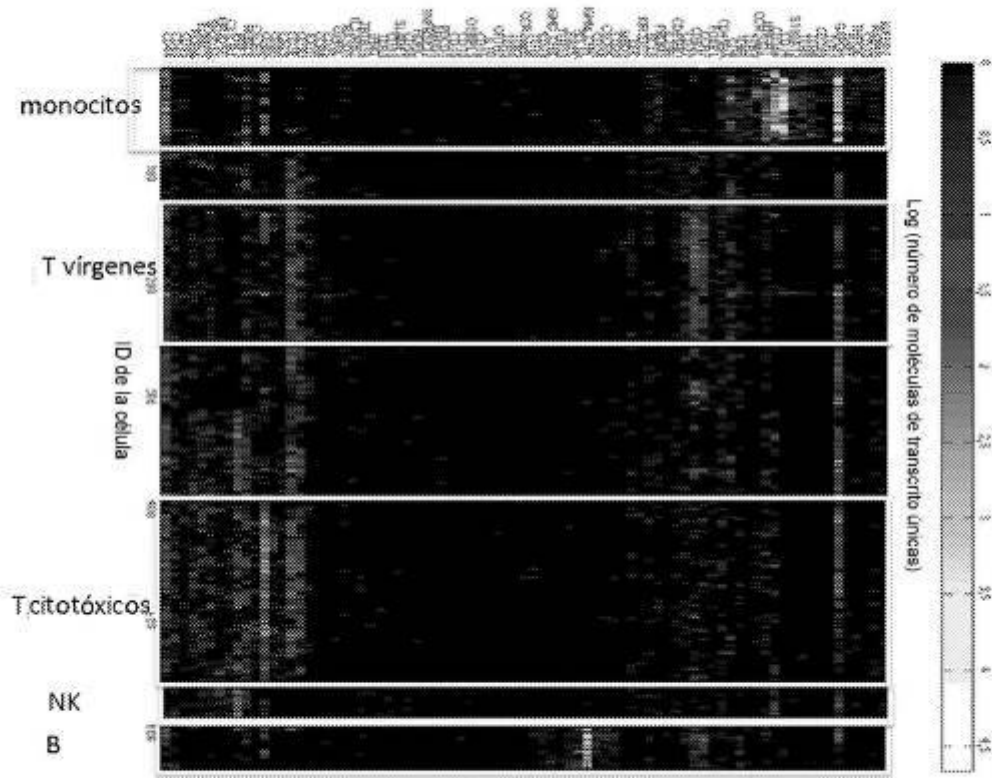


FIG. 40B

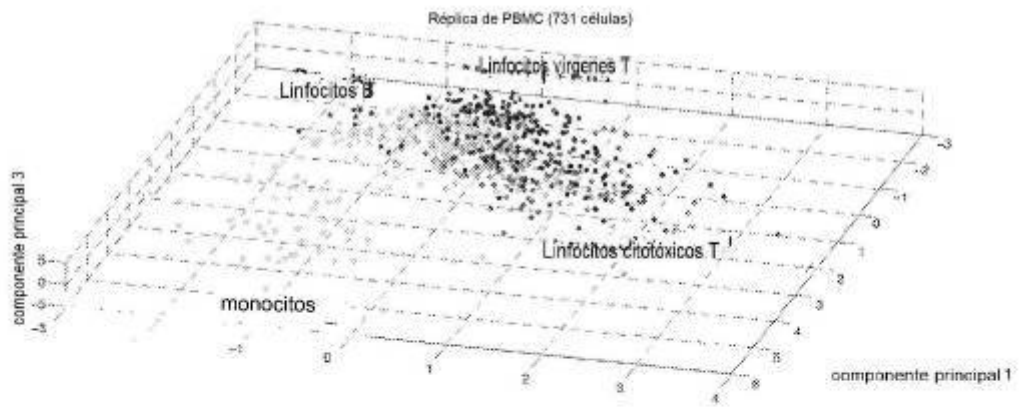


FIG. 41

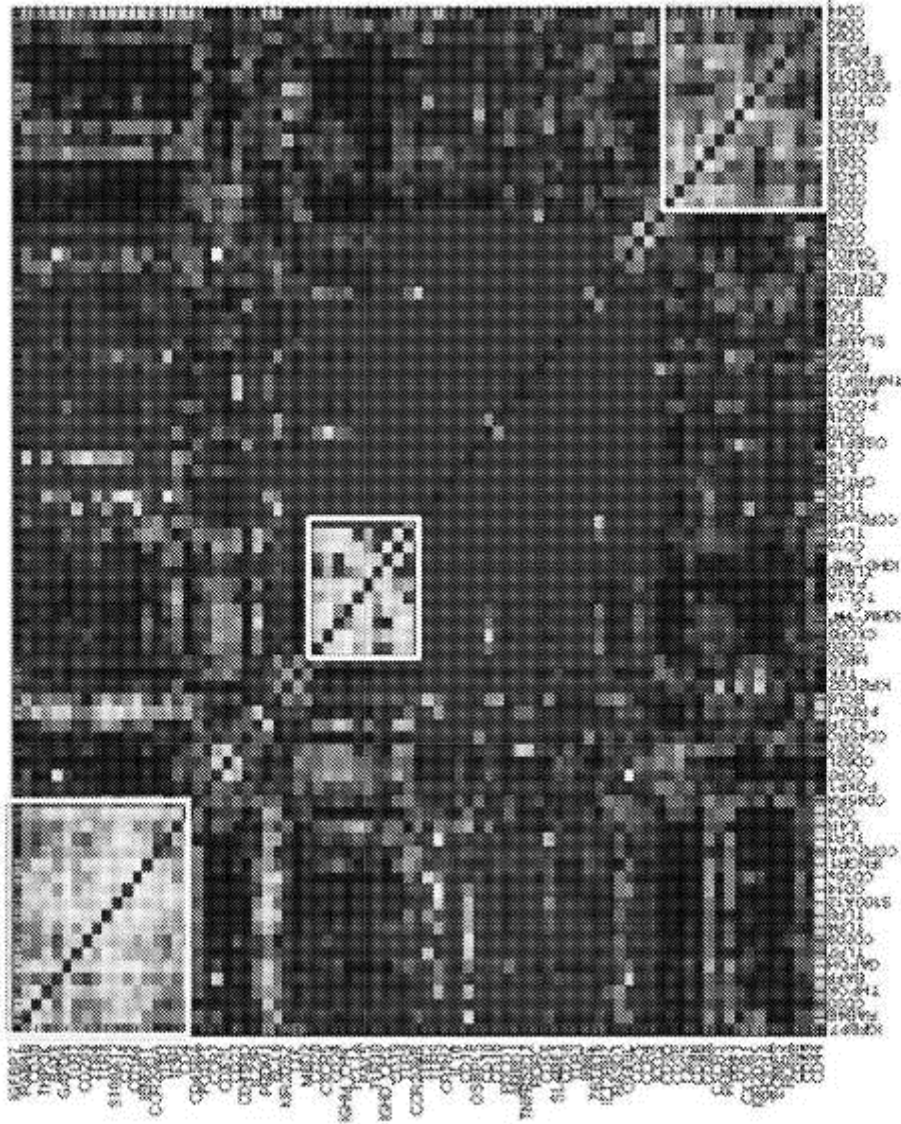


FIG. 42

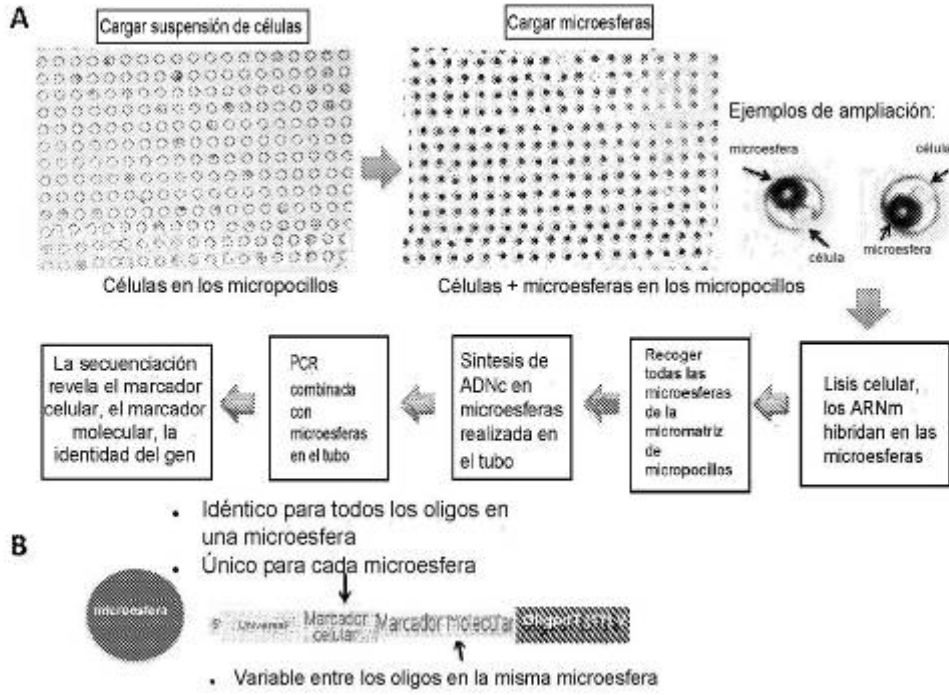


FIG. 43

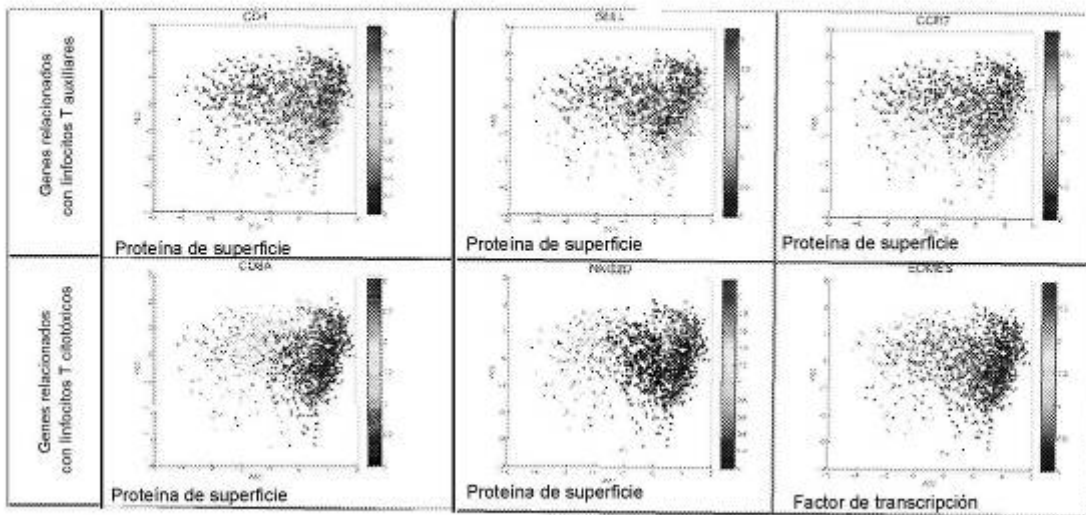


FIG. 44A

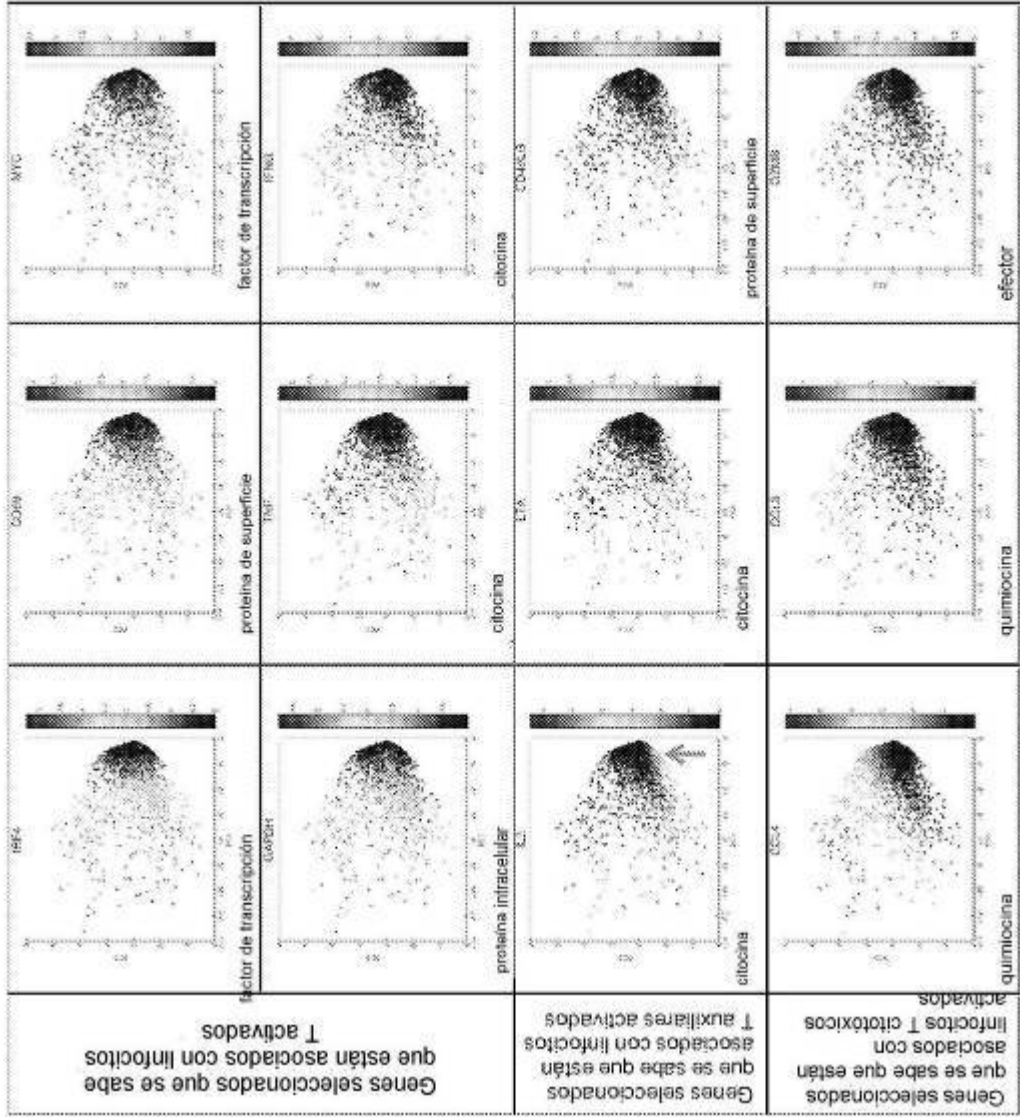


FIG. 44B

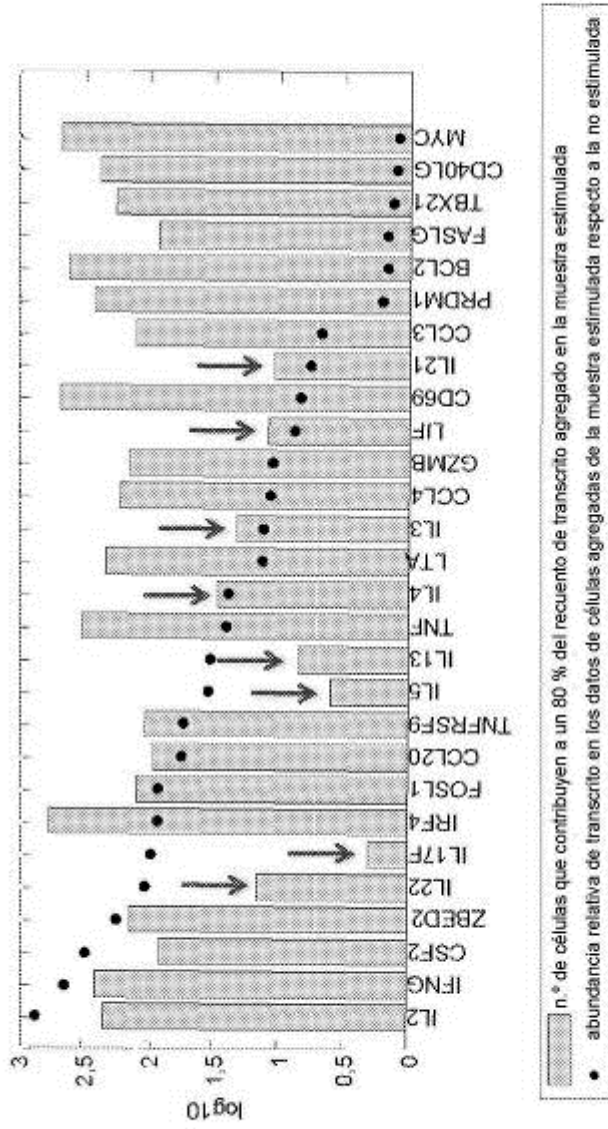


FIG. 44C

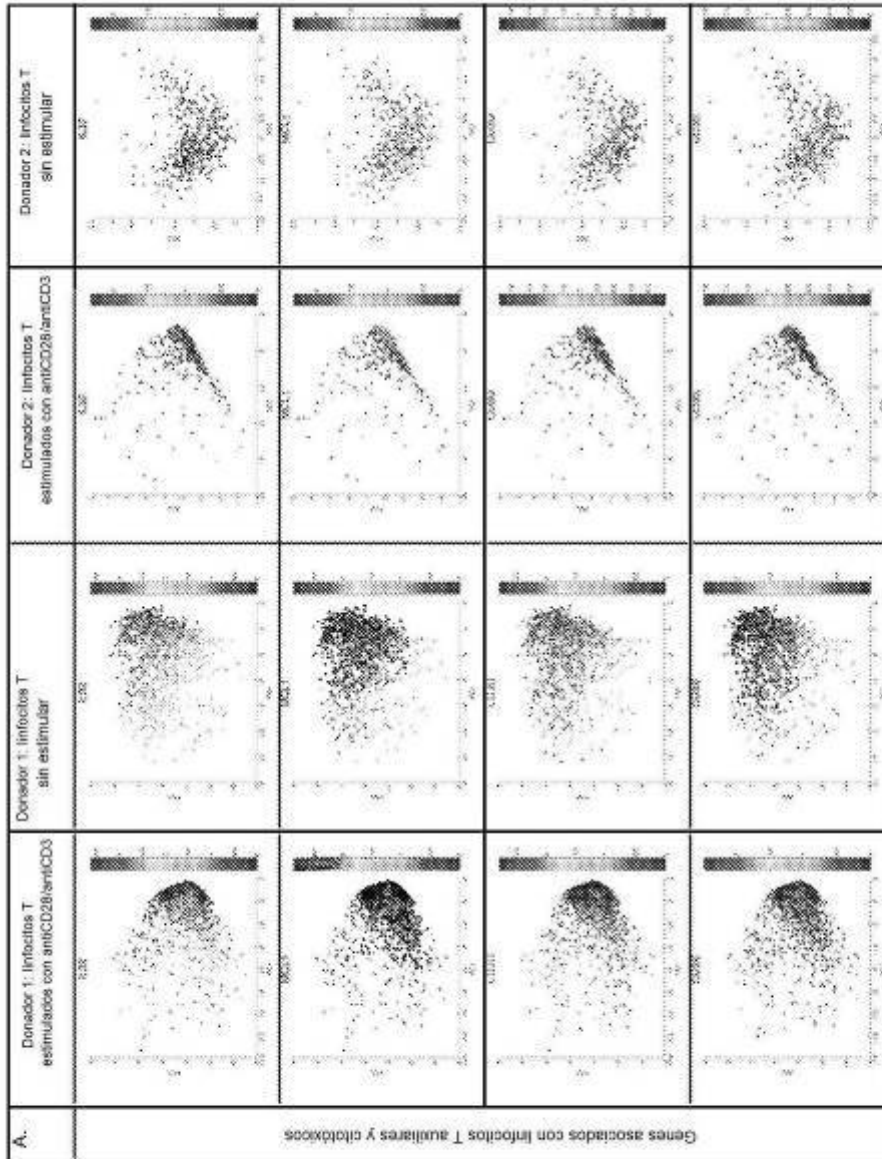


FIG. 45A

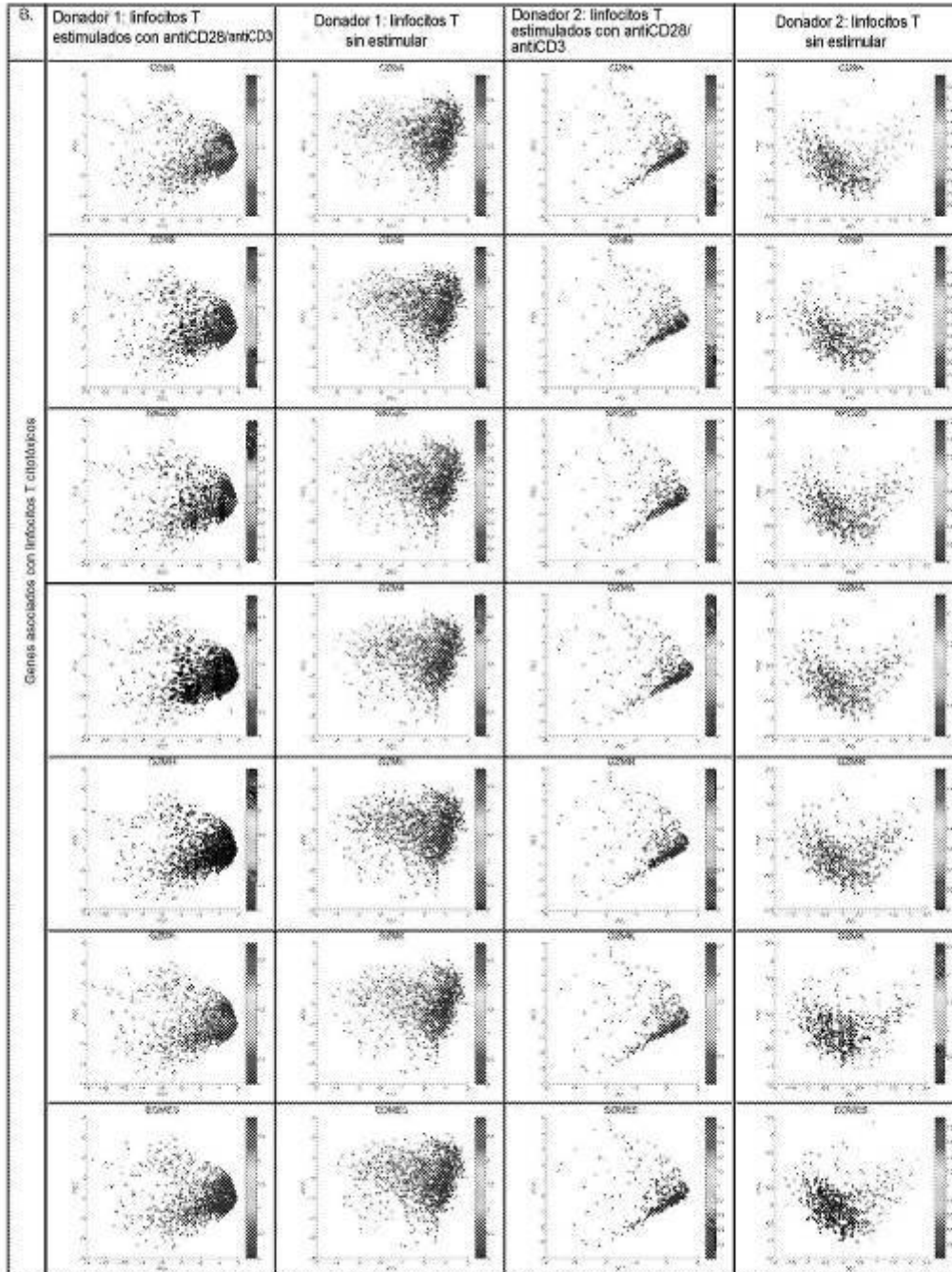


FIG. 45B

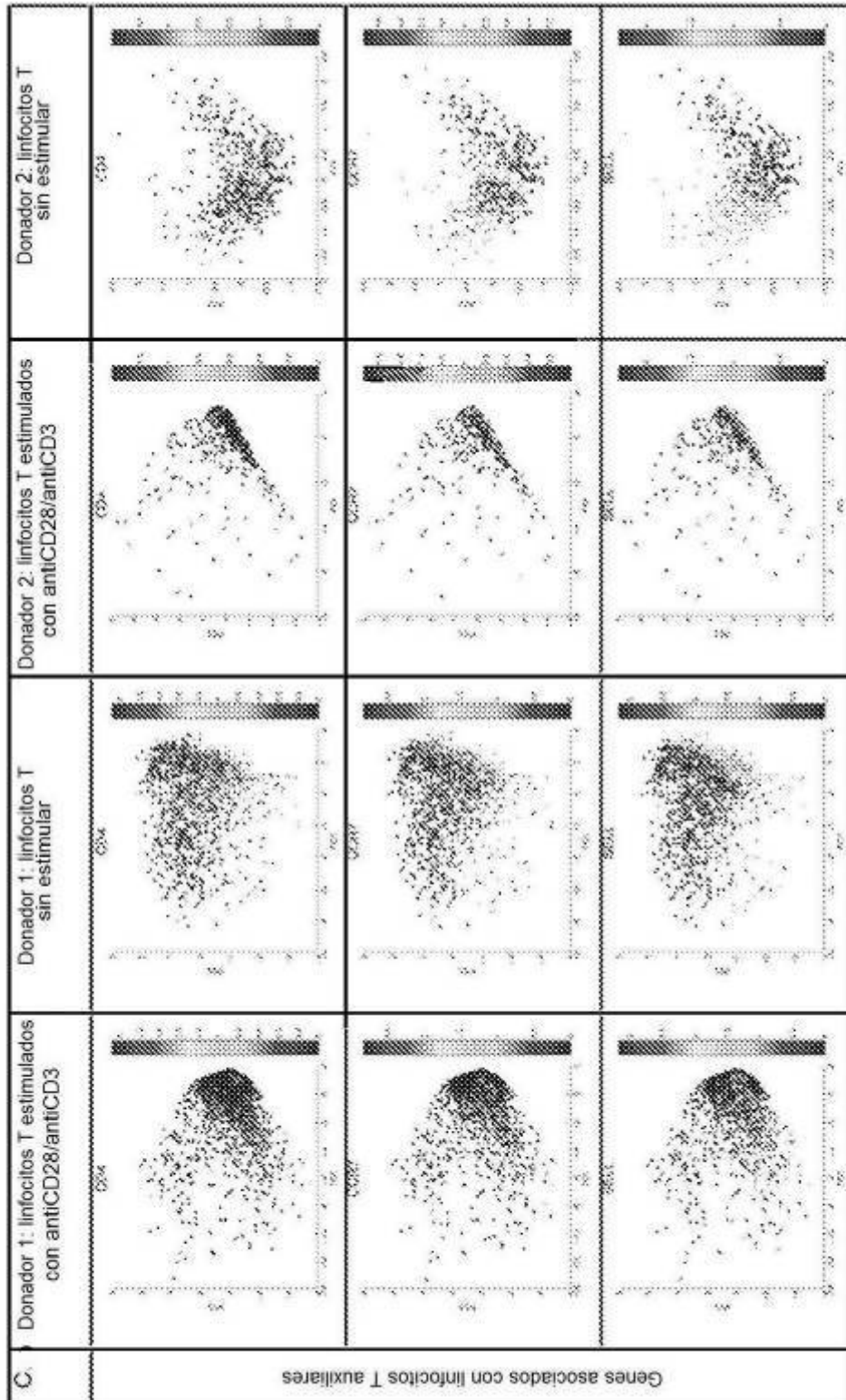


FIG. 45C

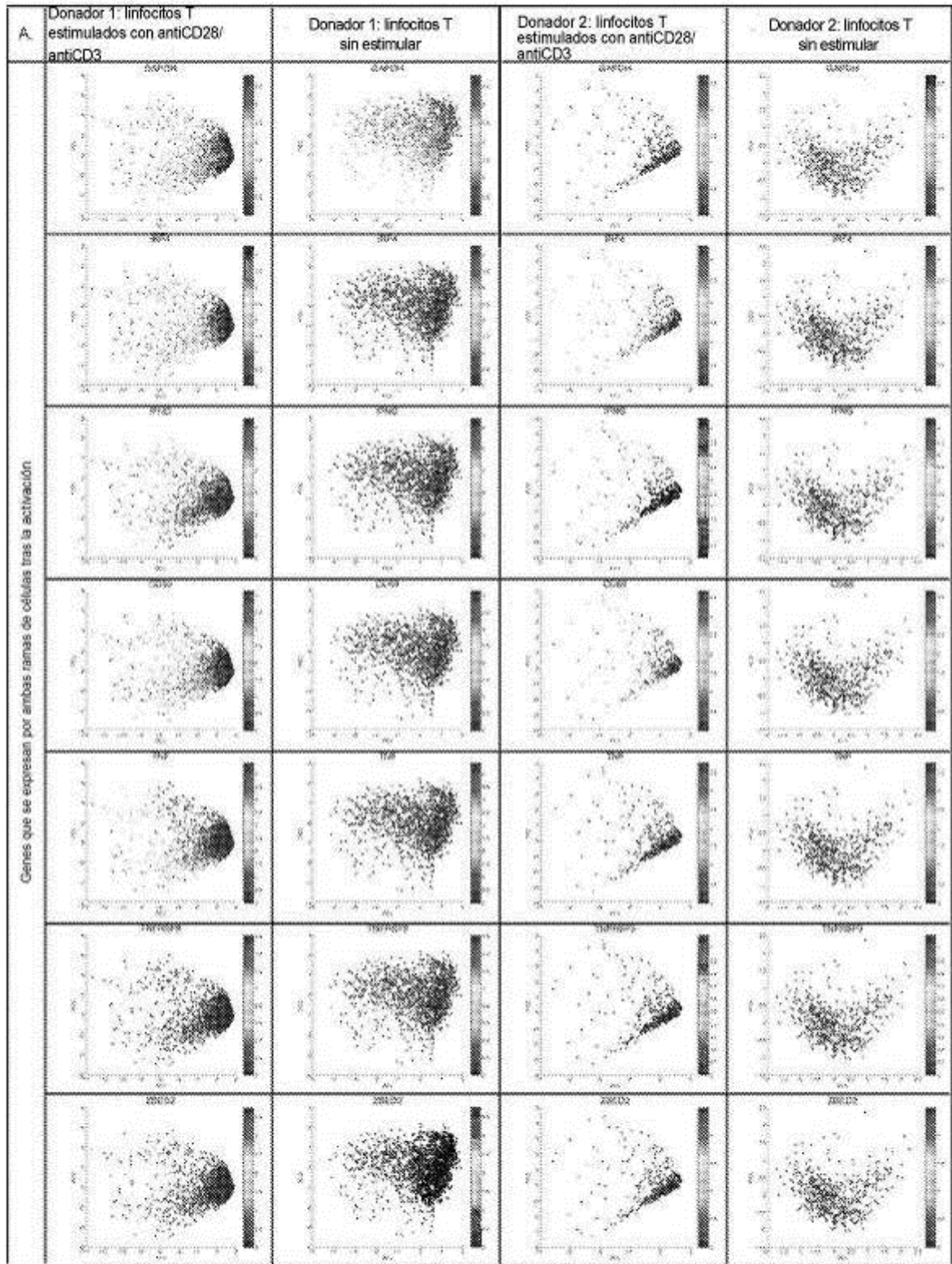


FIG. 46A

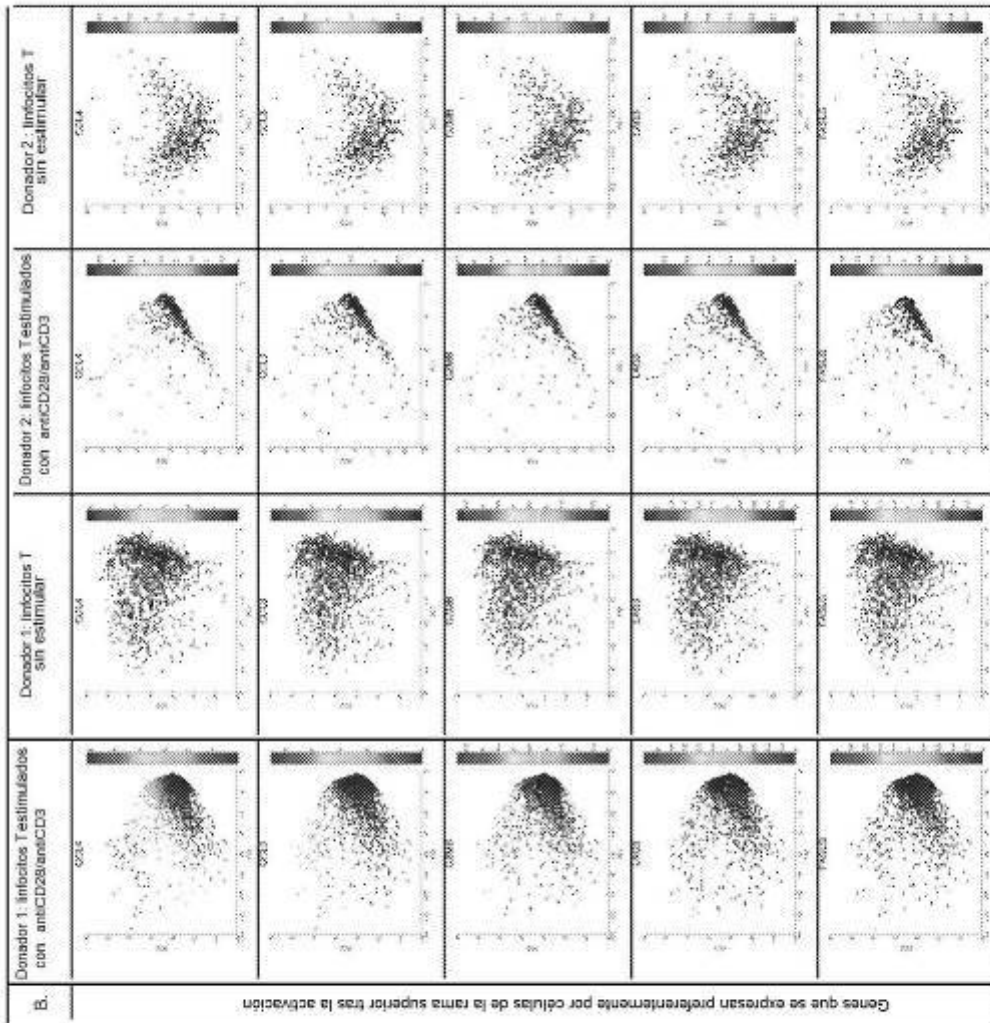


FIG. 46B

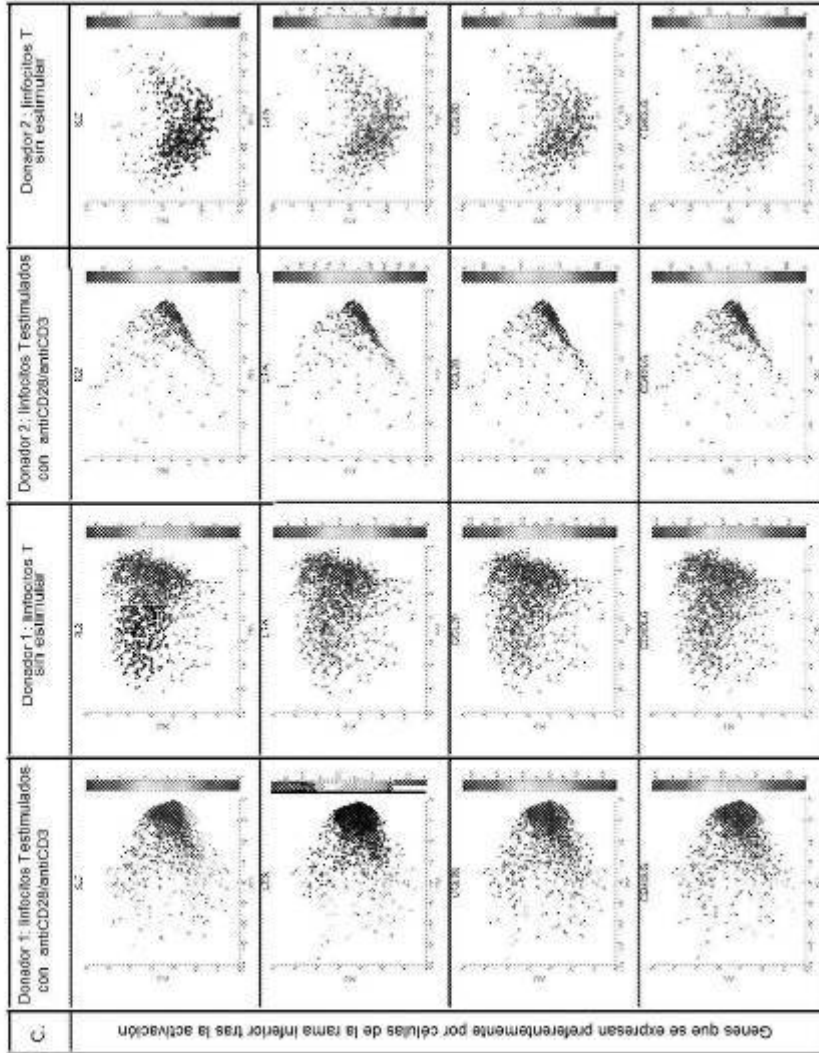


FIG. 46C

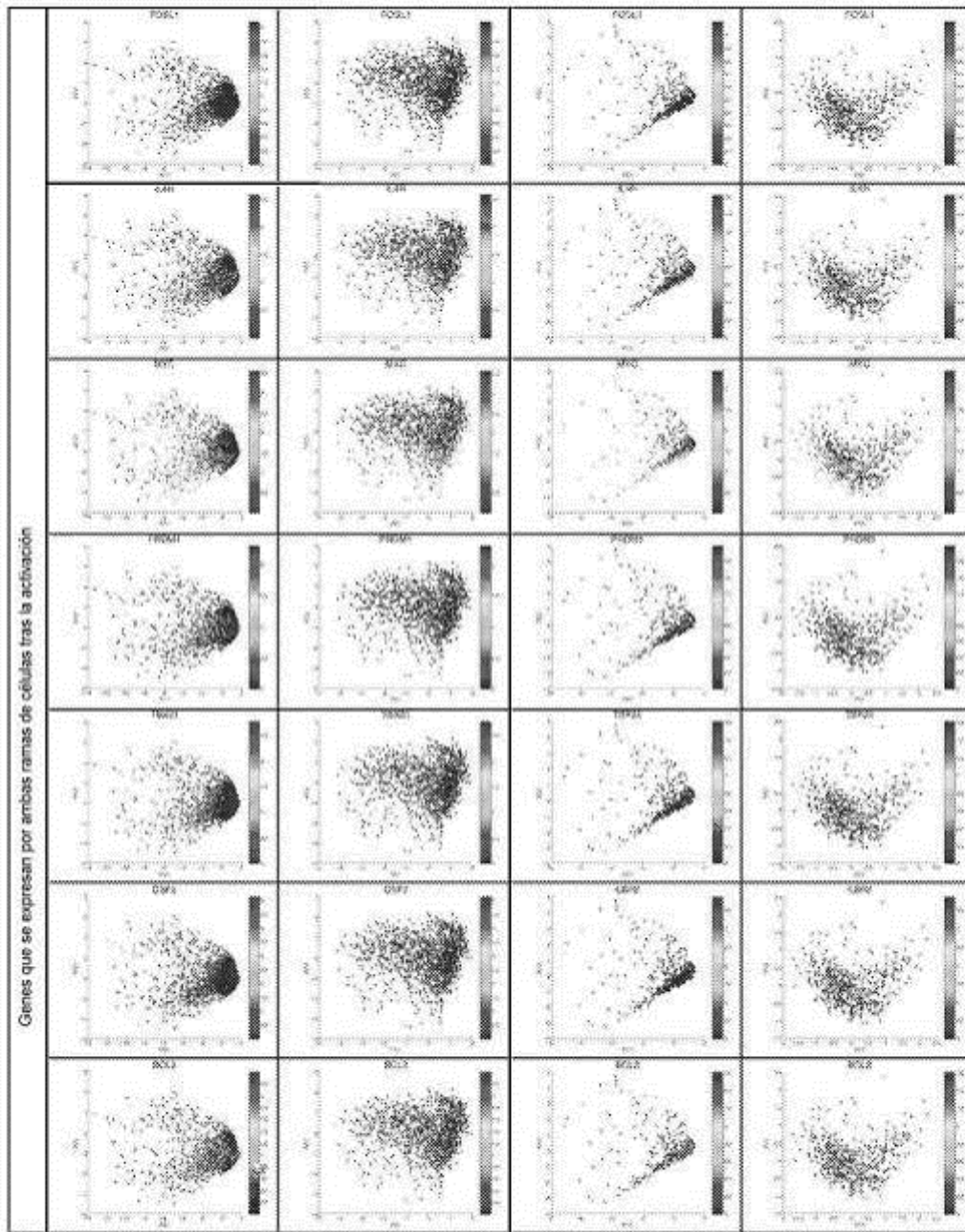


FIG. 46D

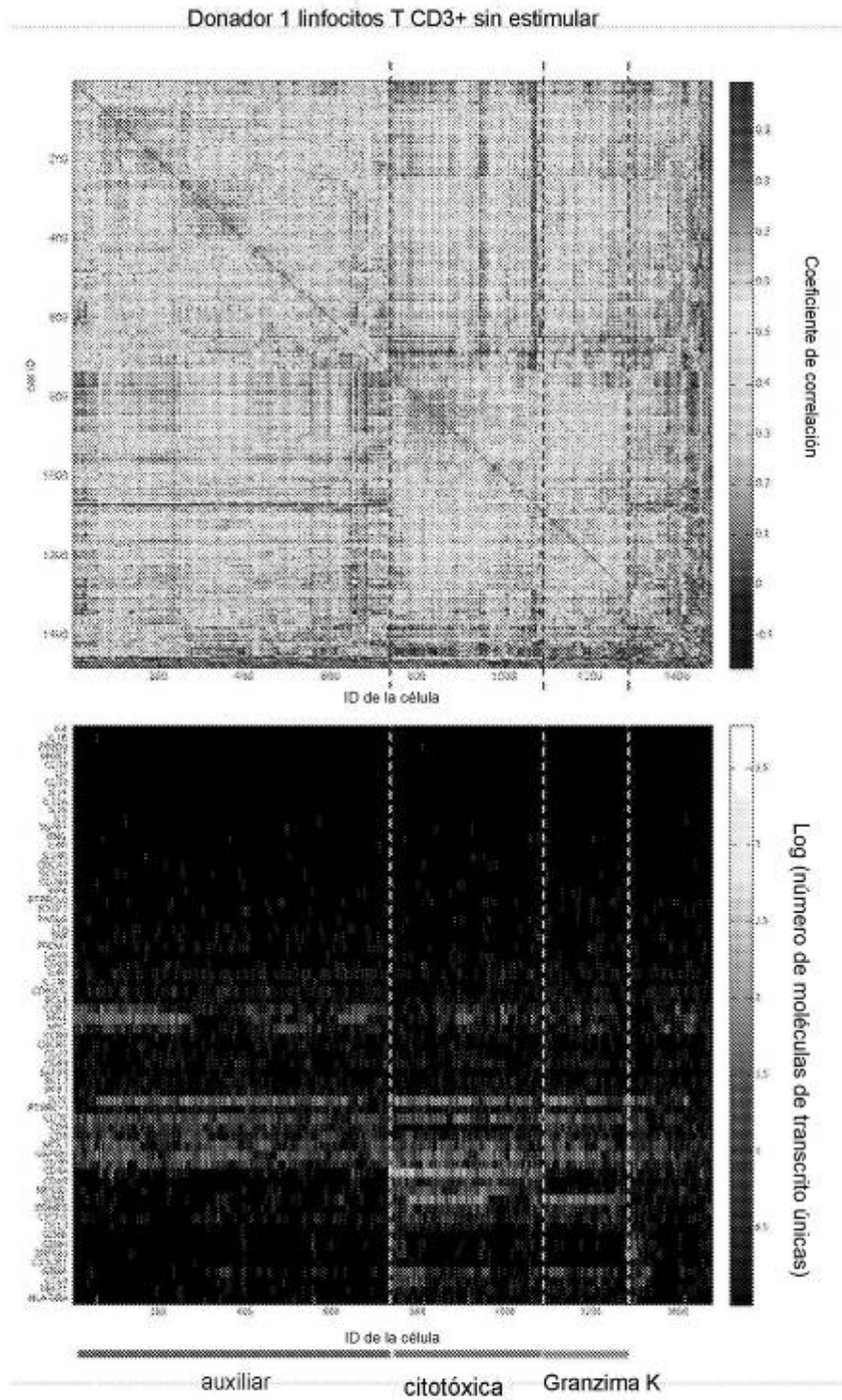


FIG. 47

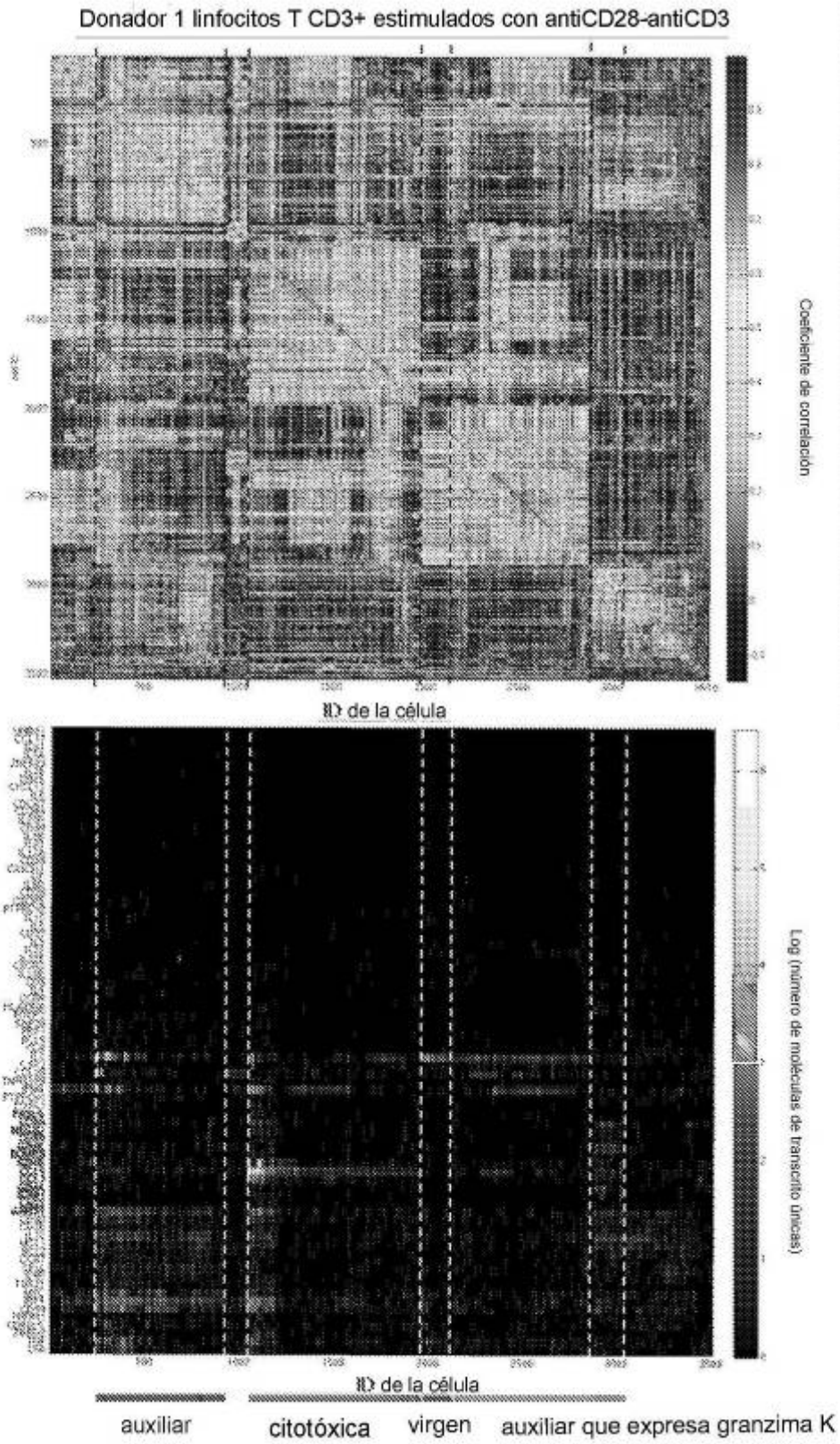


FIG. 48

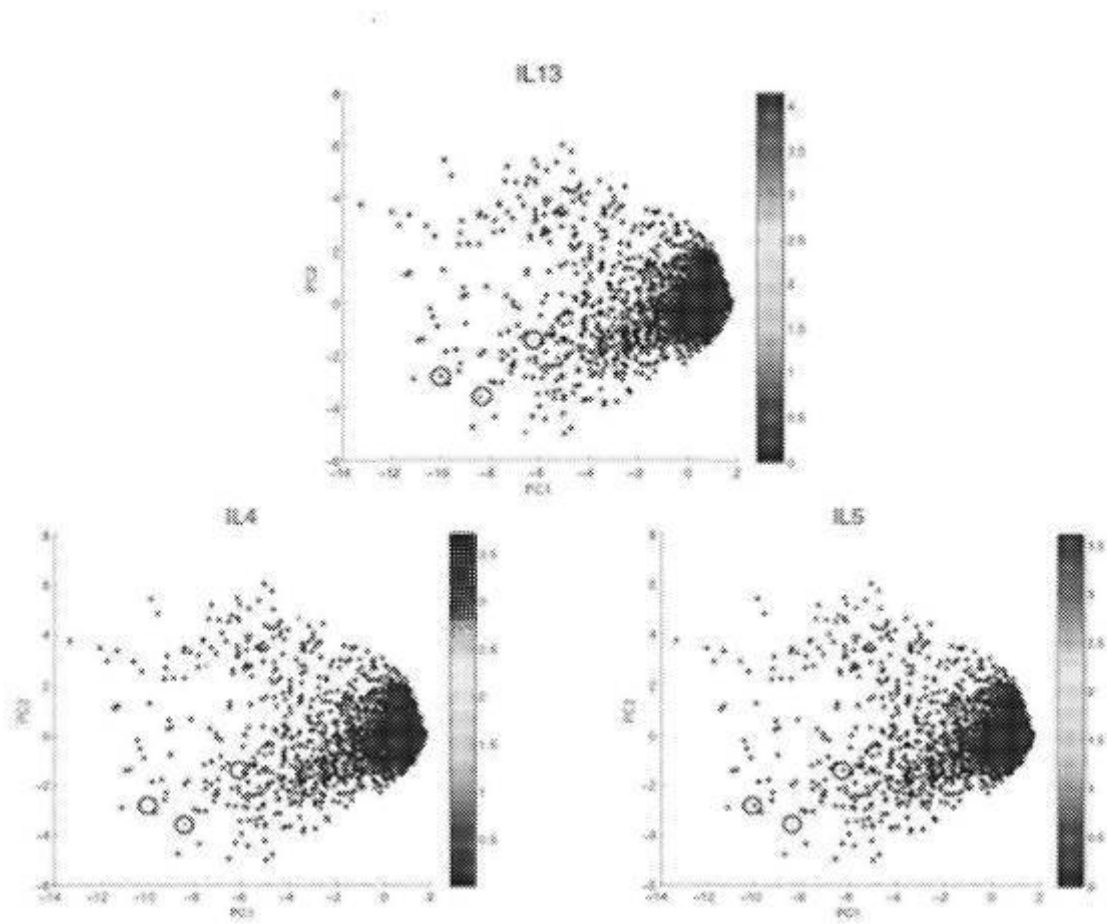
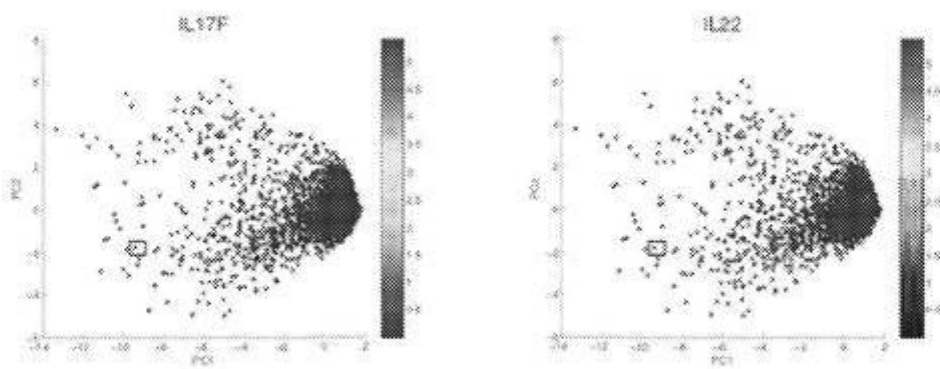


FIG. 49A ○ Células que coexpresaban IL4, IL5, IL13



□ Célula individual que coexpresaba IL17F e IL22

FIG. 49B

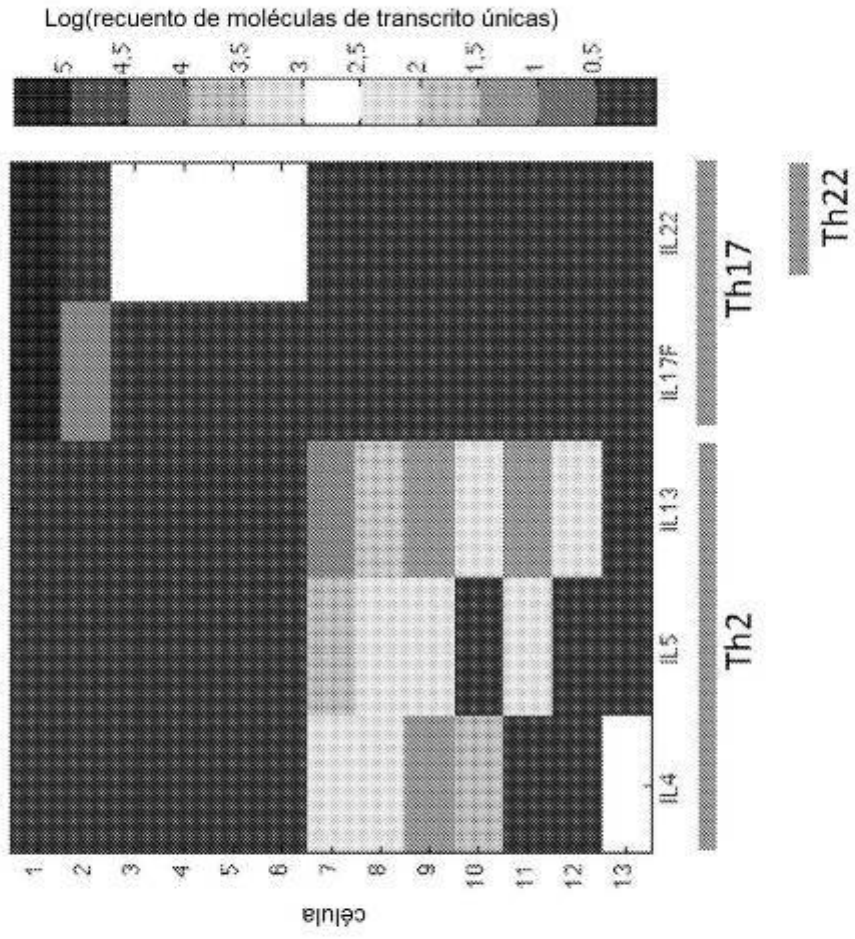


FIG. 49C

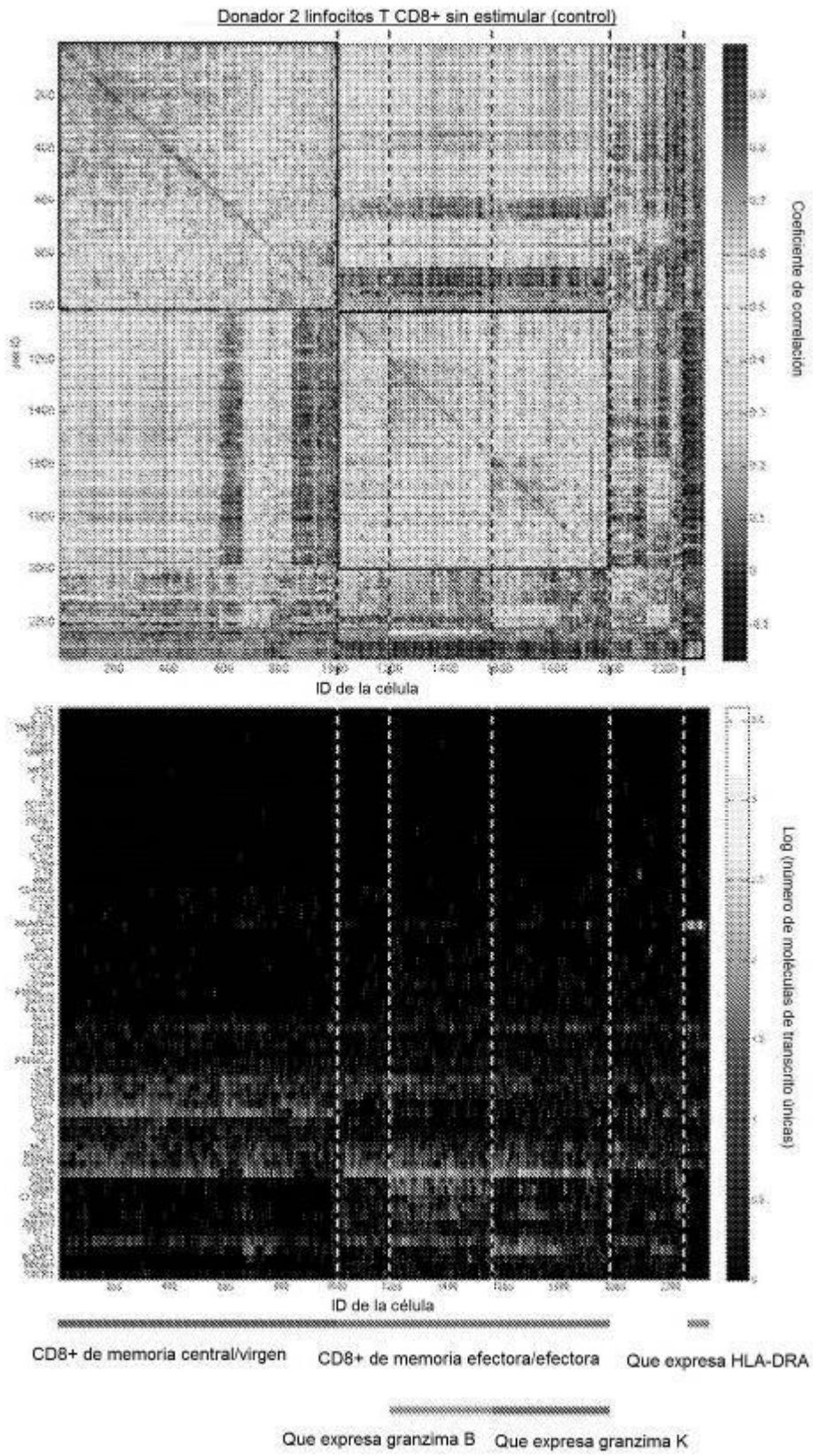


FIG. 50A

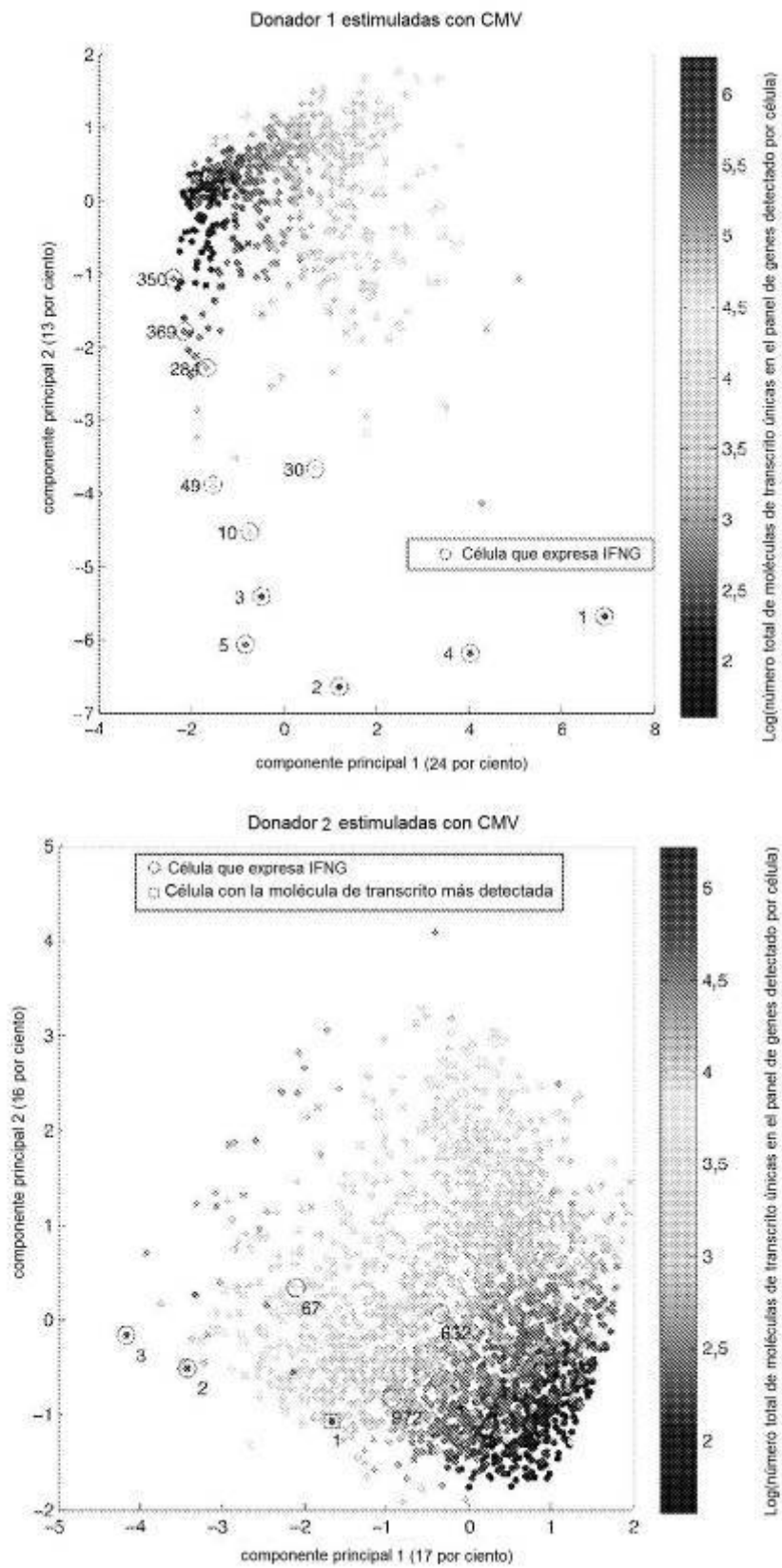


FIG. 50B

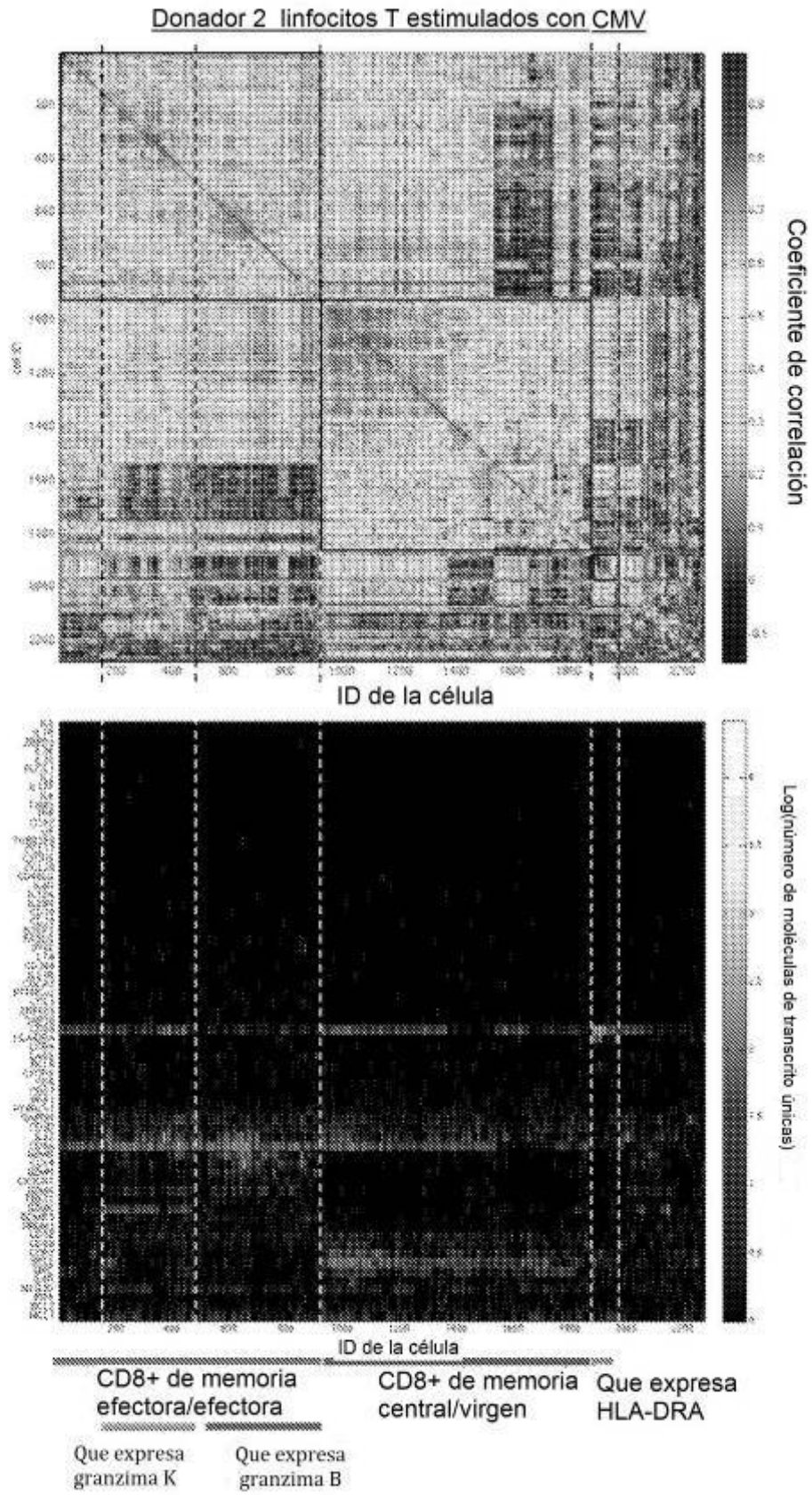


FIG. 51

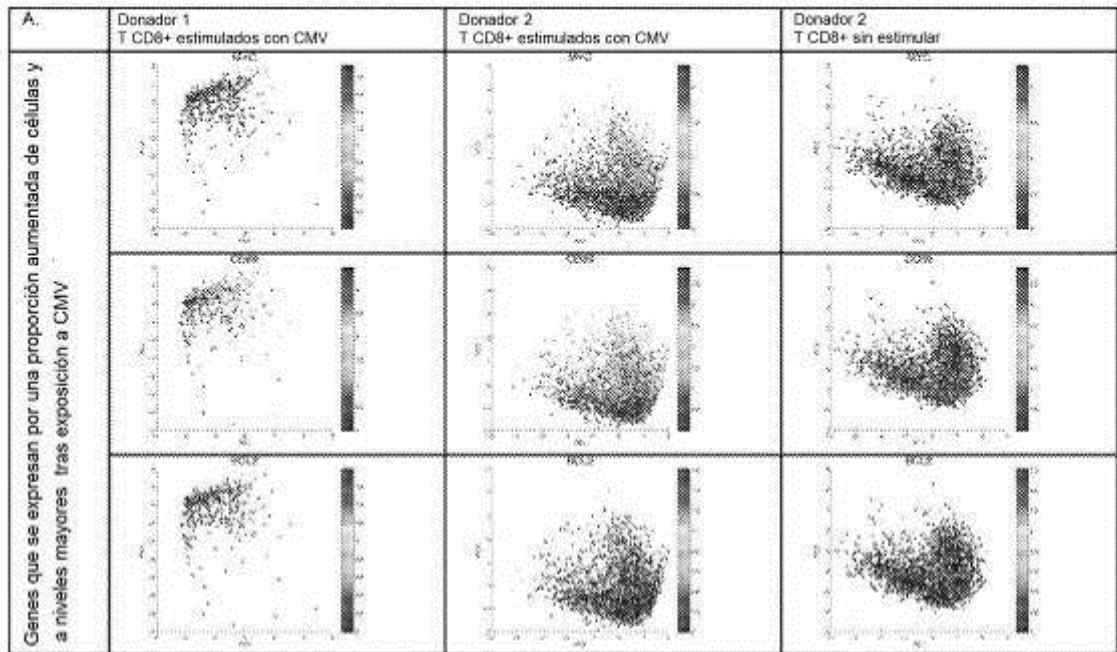


FIG. 52A

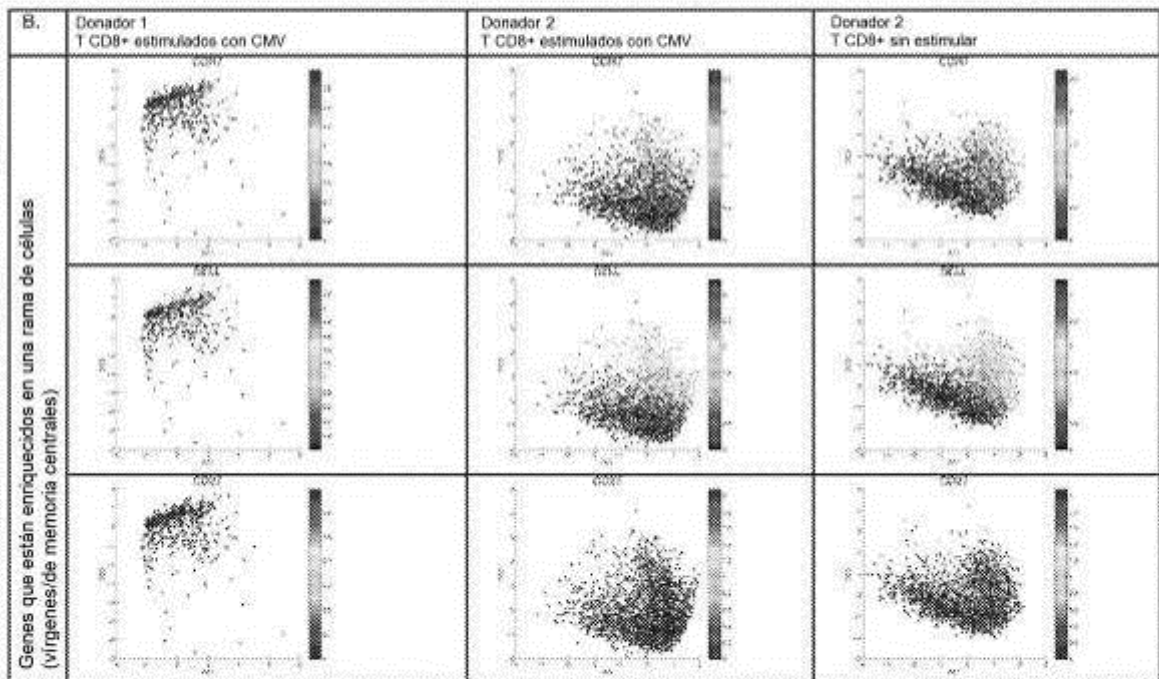


FIG. 52B

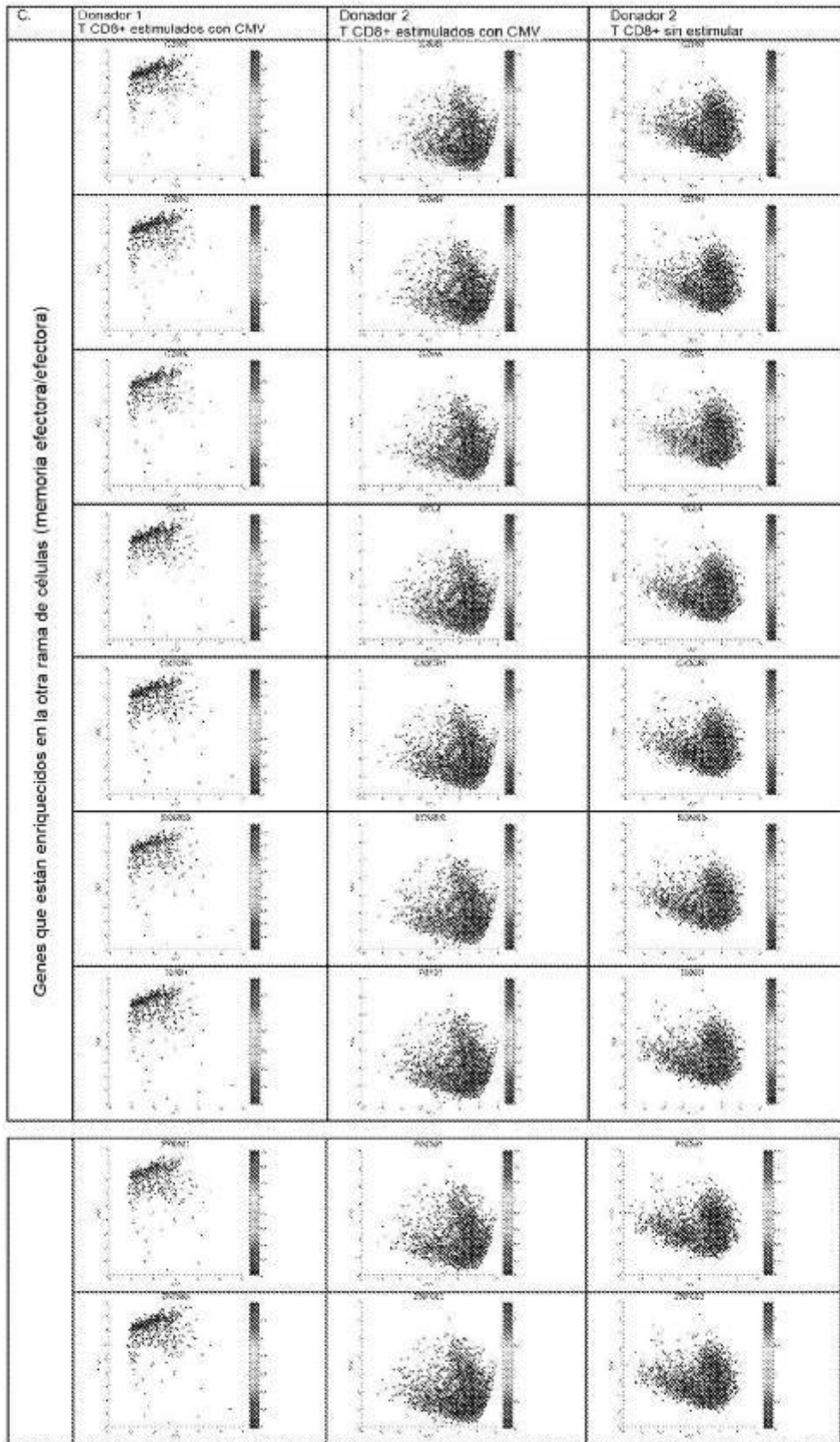


FIG. 52C

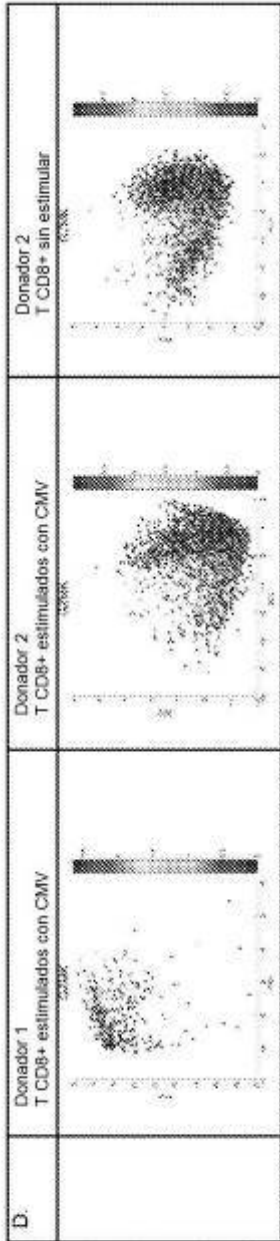


FIG. 52D

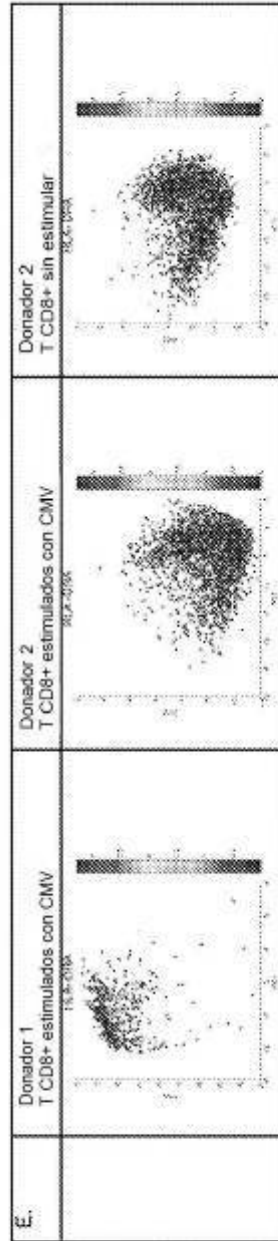


FIG. 52E

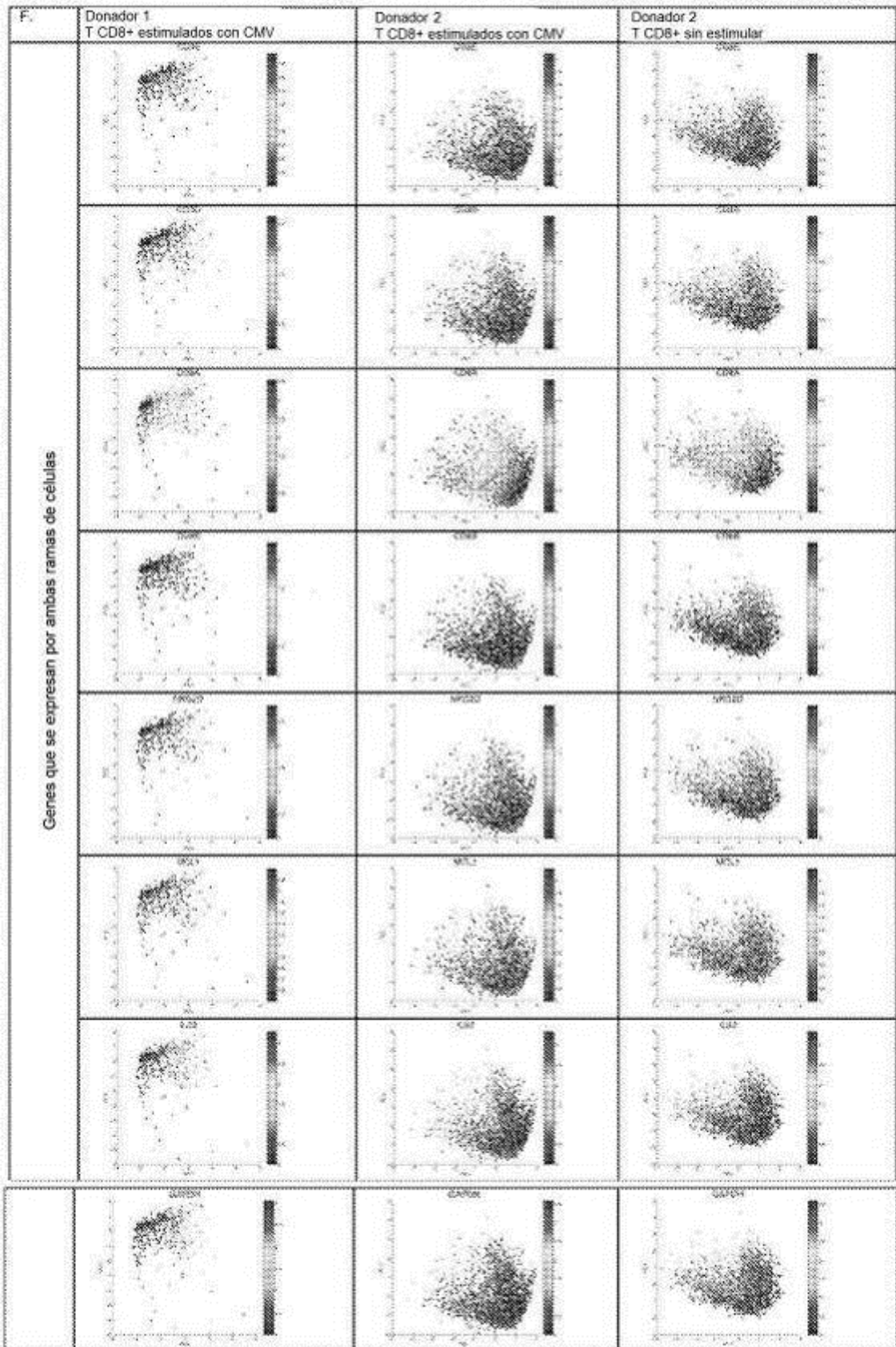


FIG. 52F

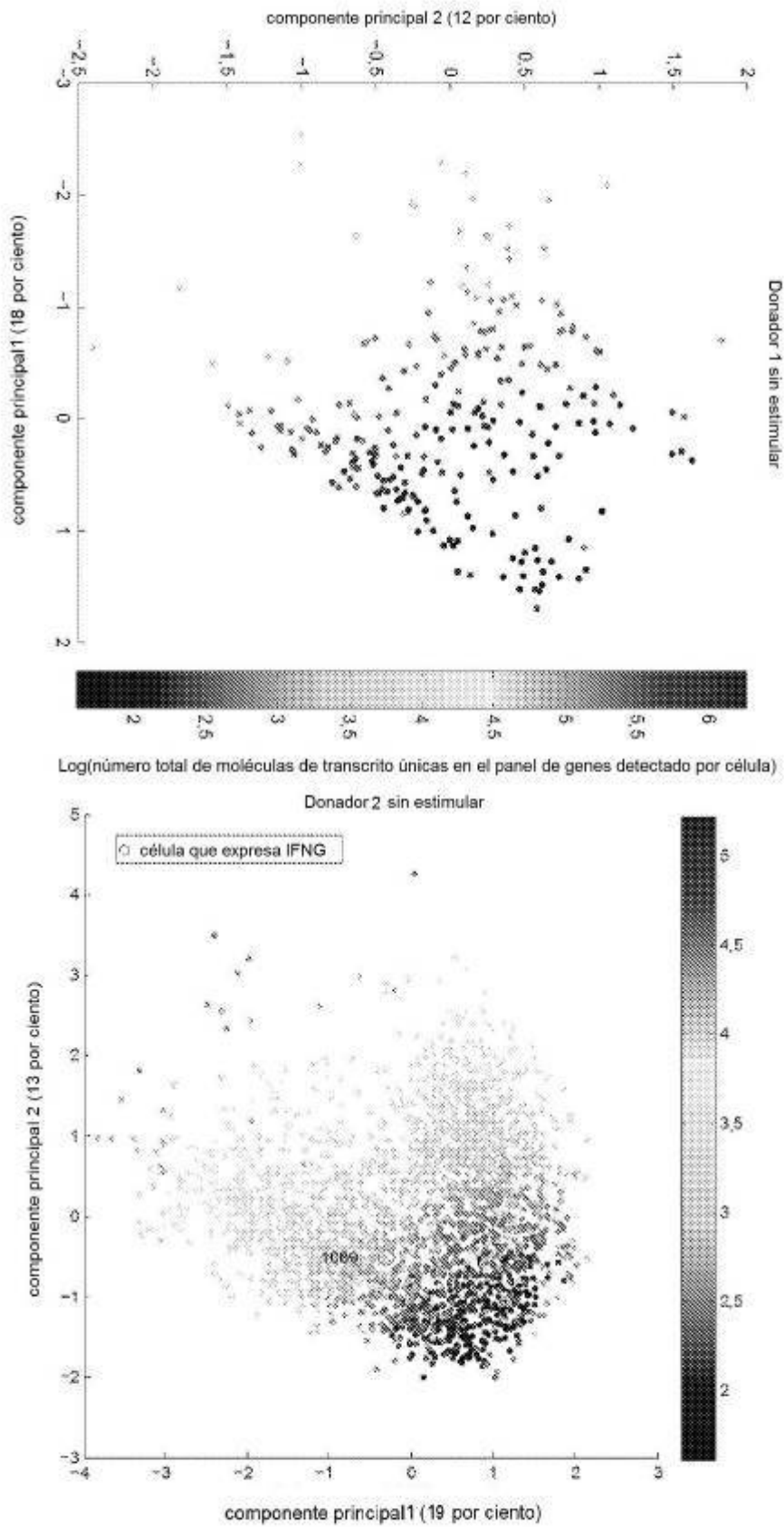


FIG. 53

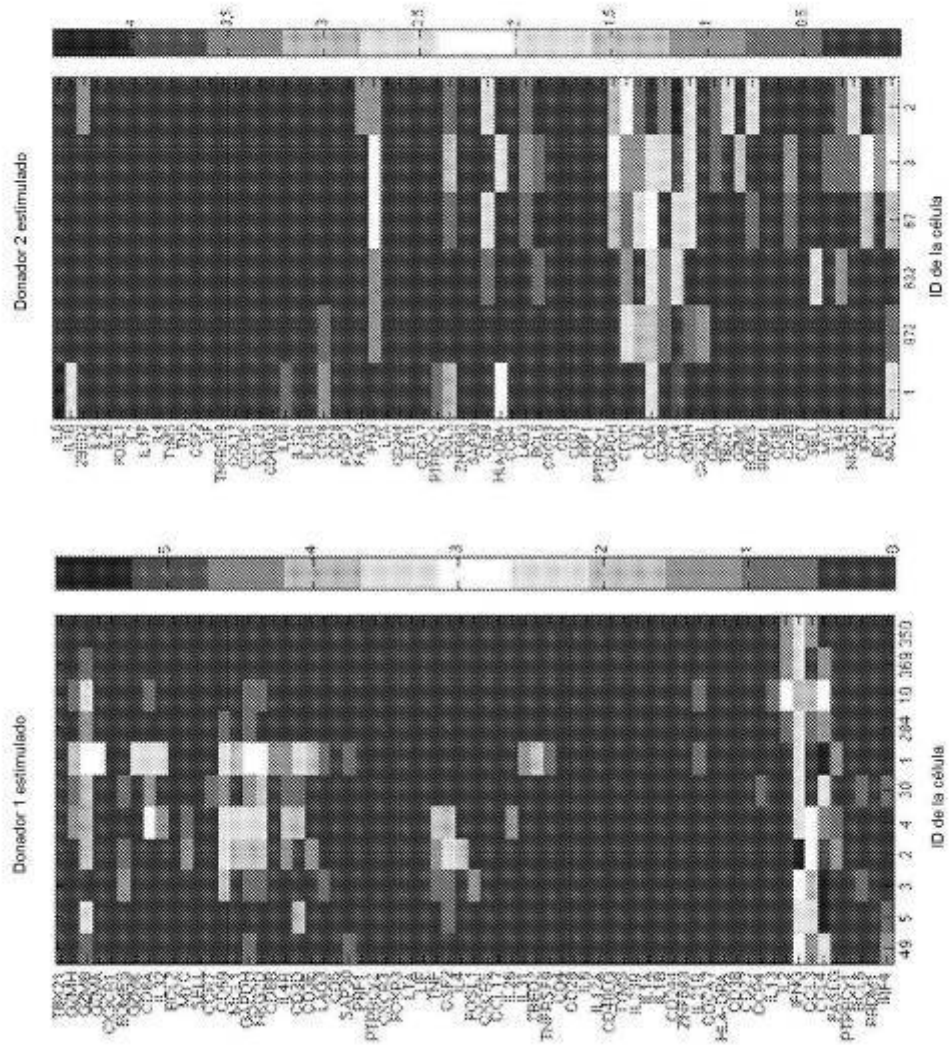


FIG. 54

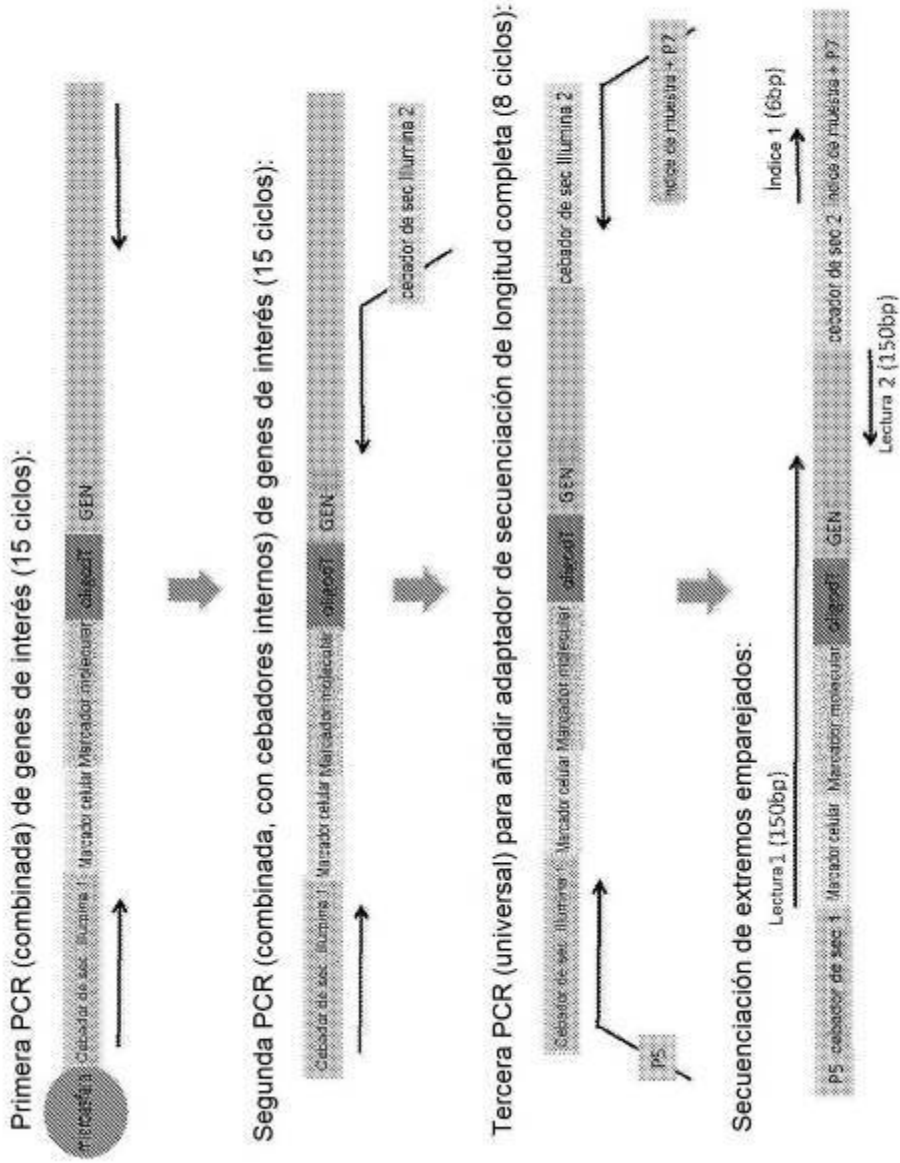


FIG. 55

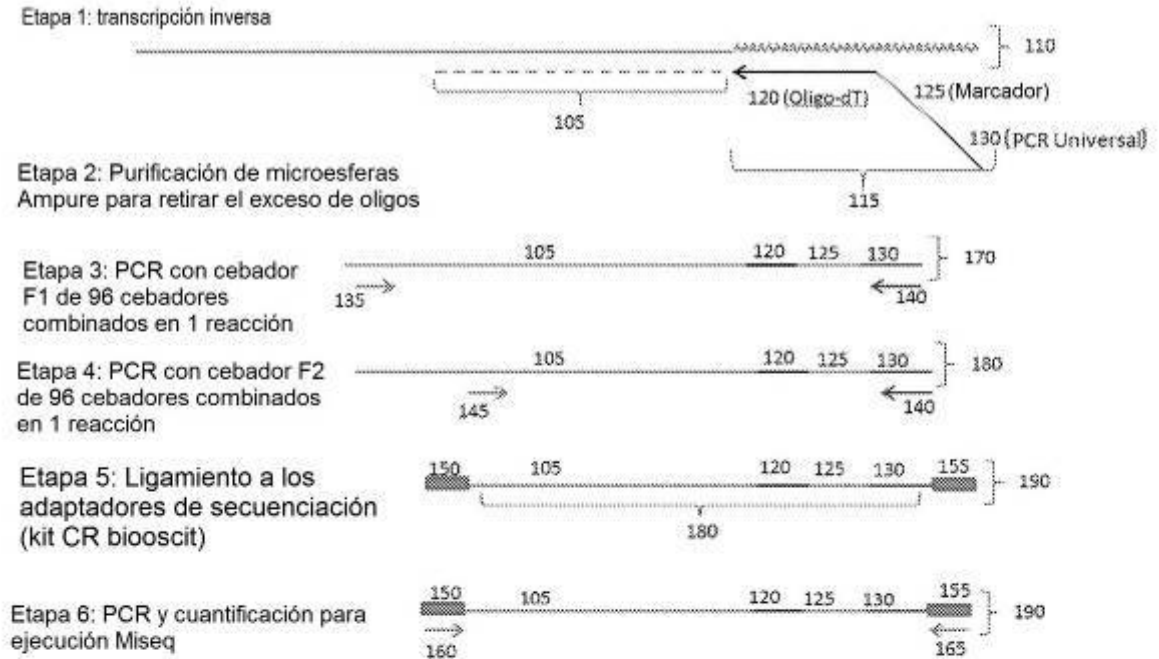


FIG. 56

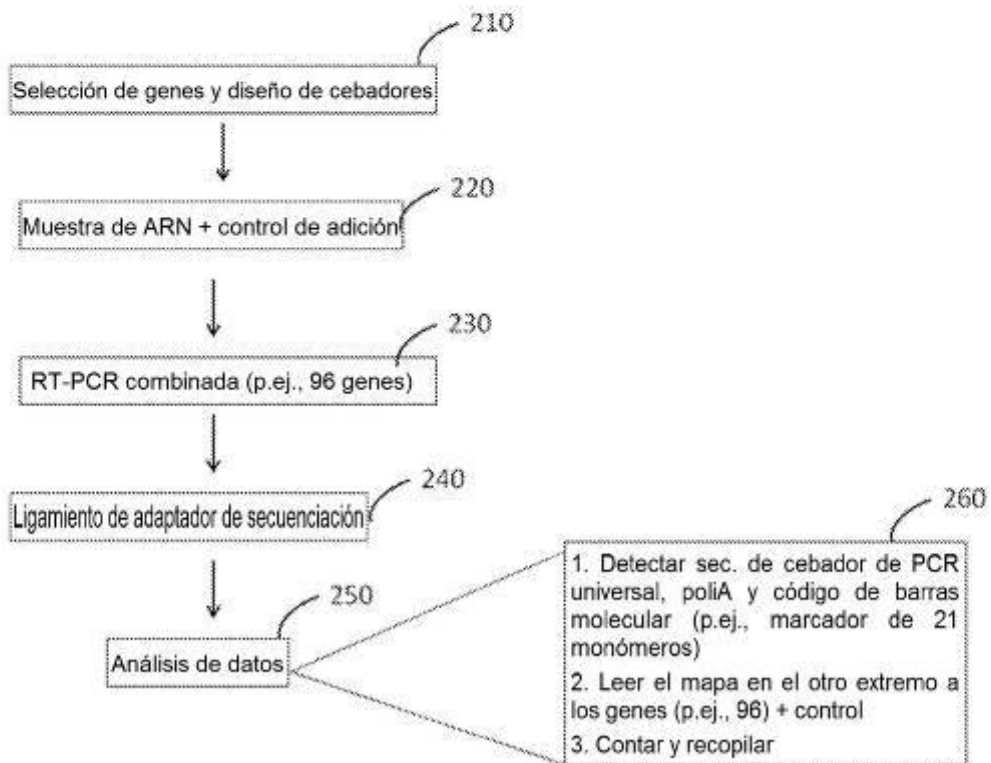


FIG. 57

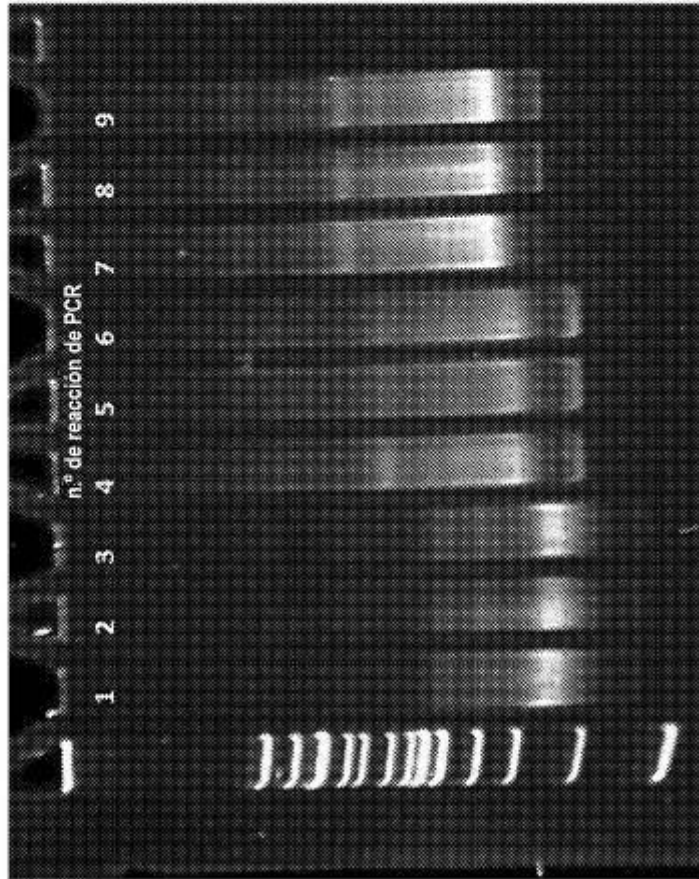


FIG. 58A

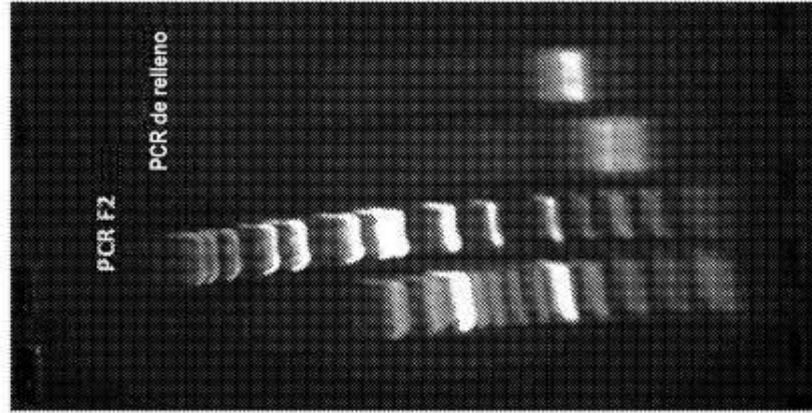
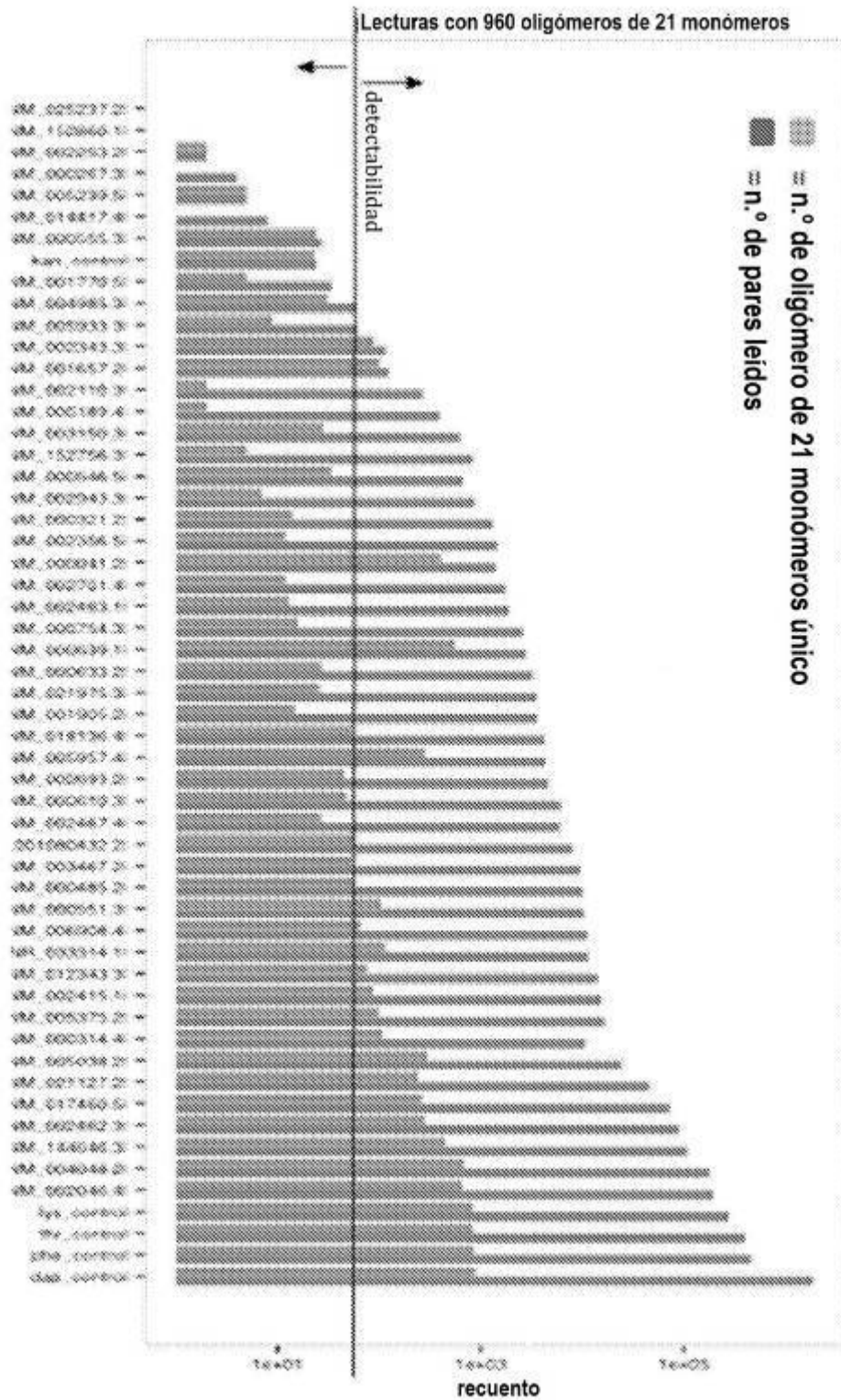


FIG. 58B



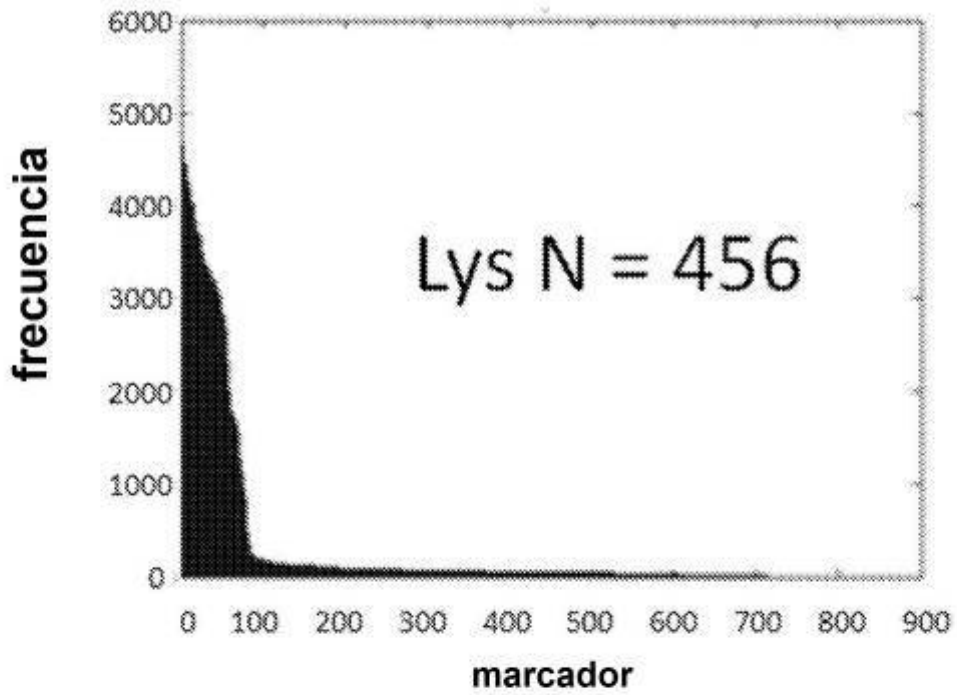


FIG. 60A

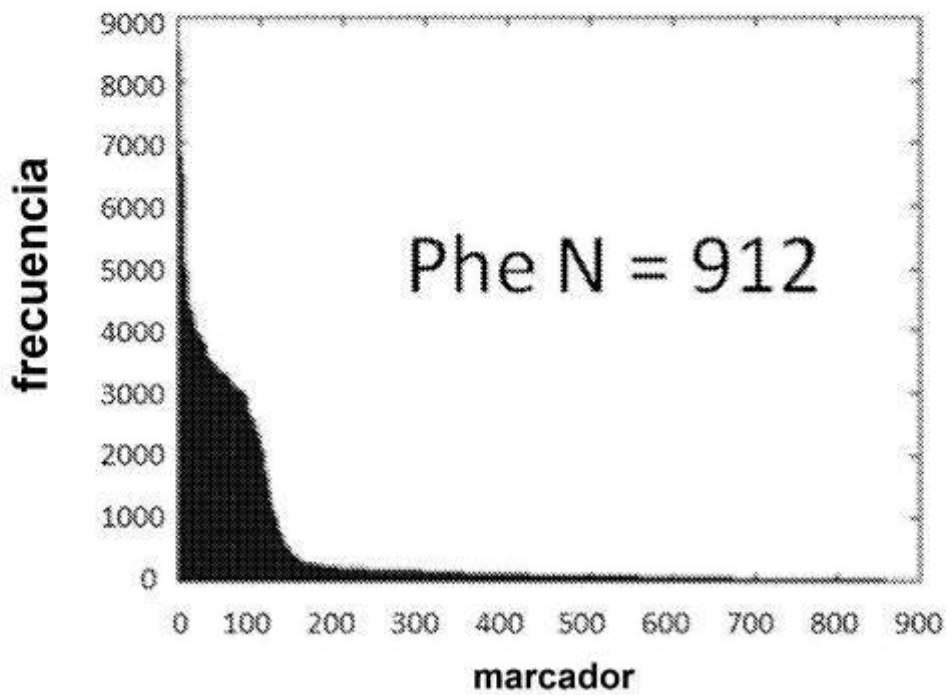


FIG. 60B

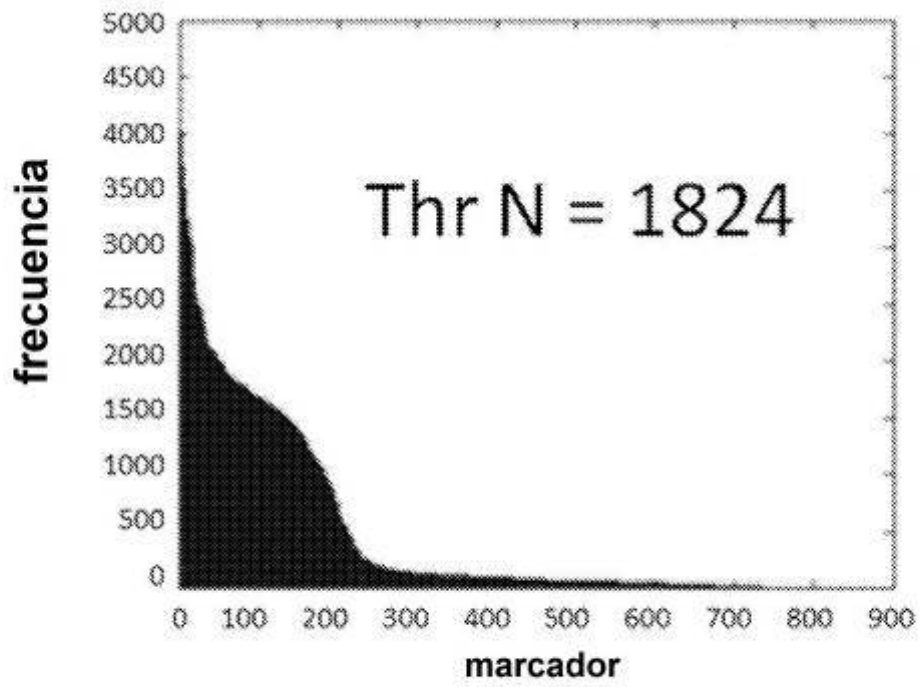


FIG. 60C

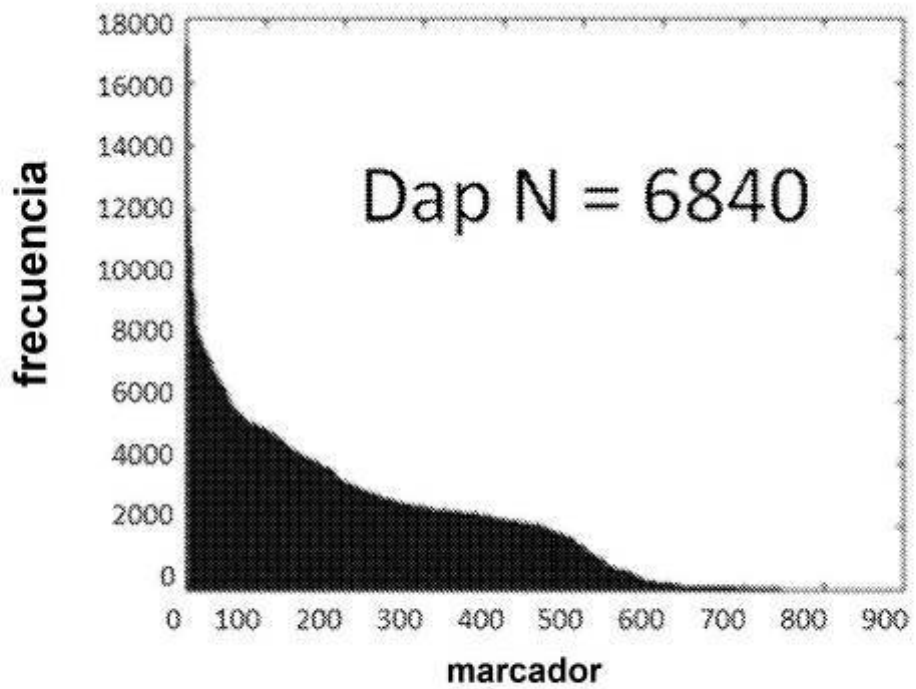


FIG. 60D

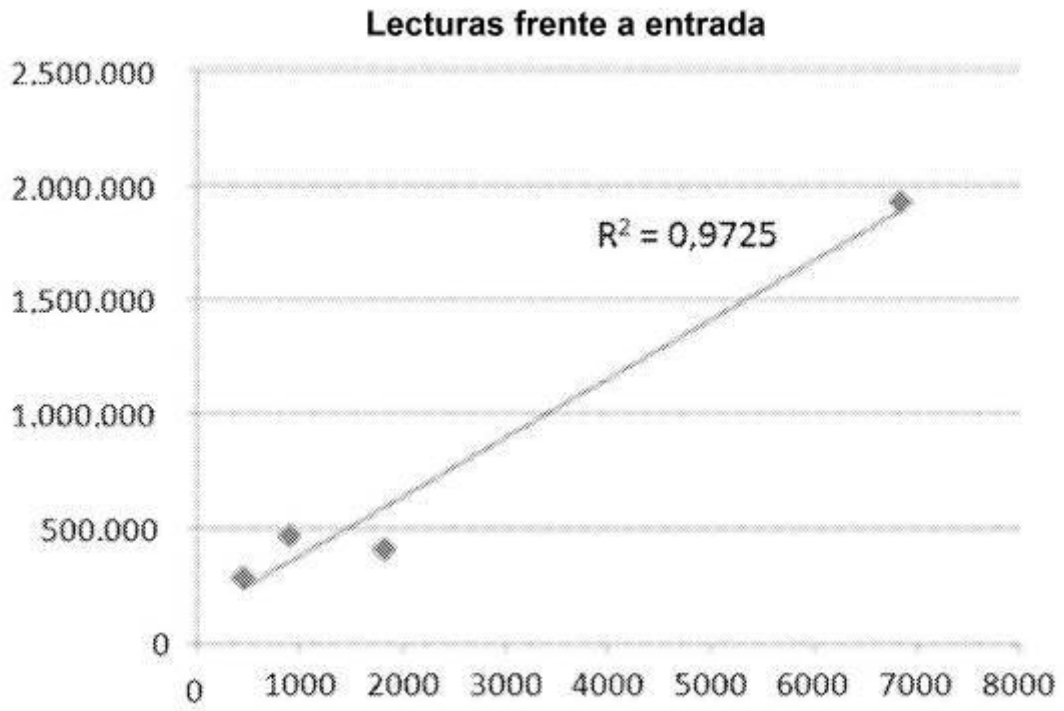


FIG. 60E

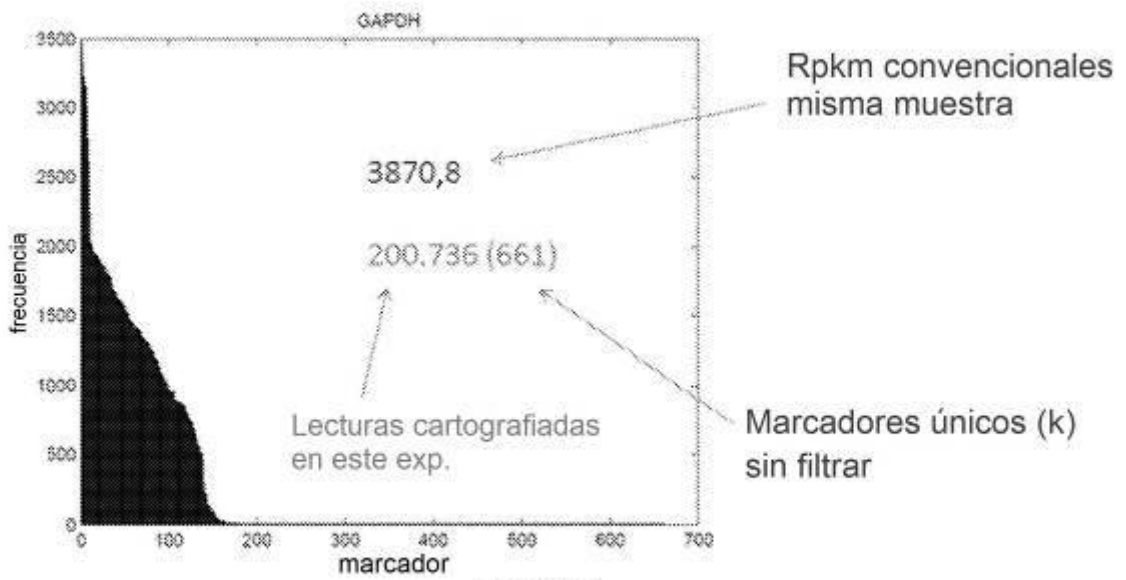


FIG. 61

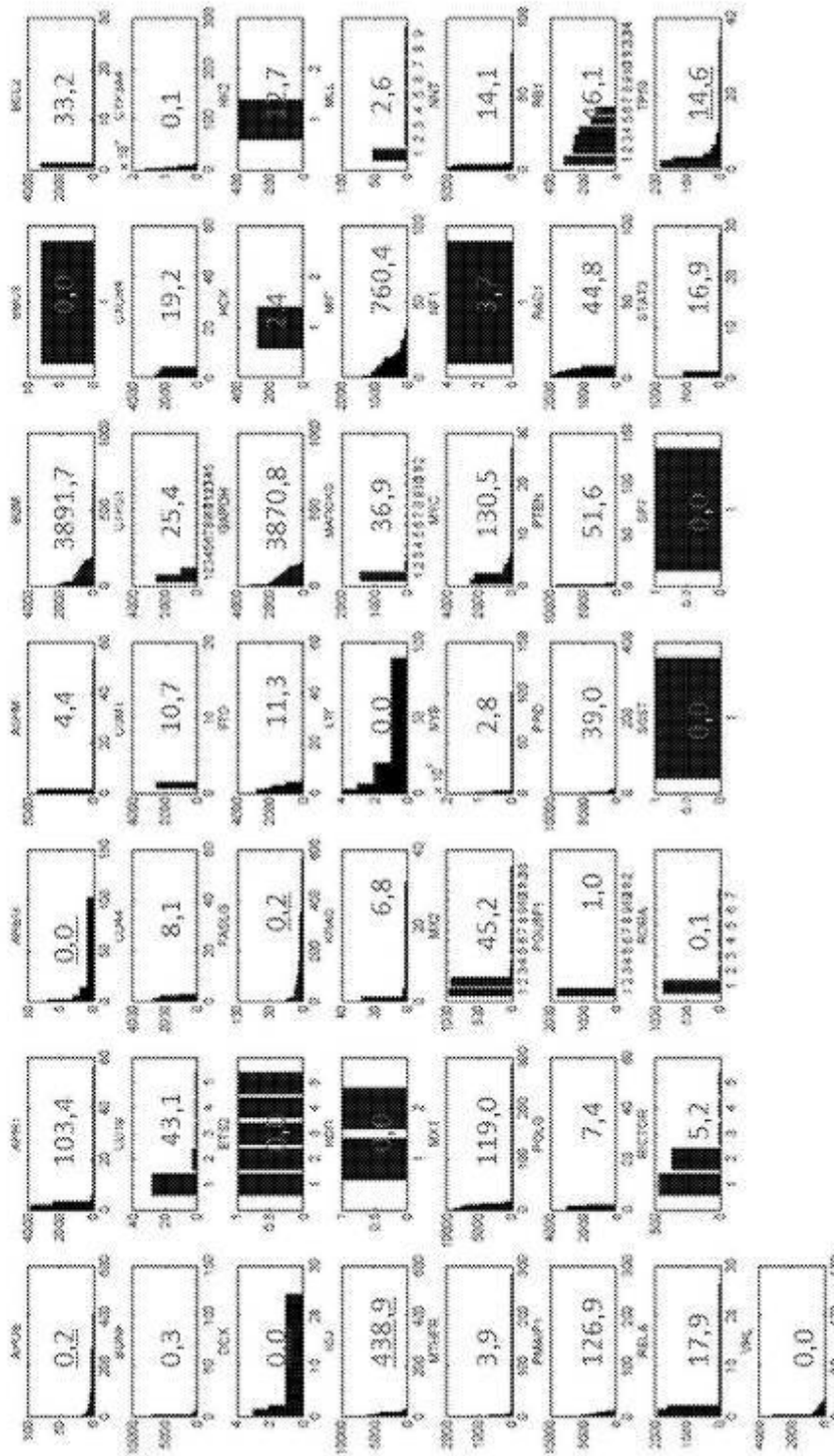


FIG. 62

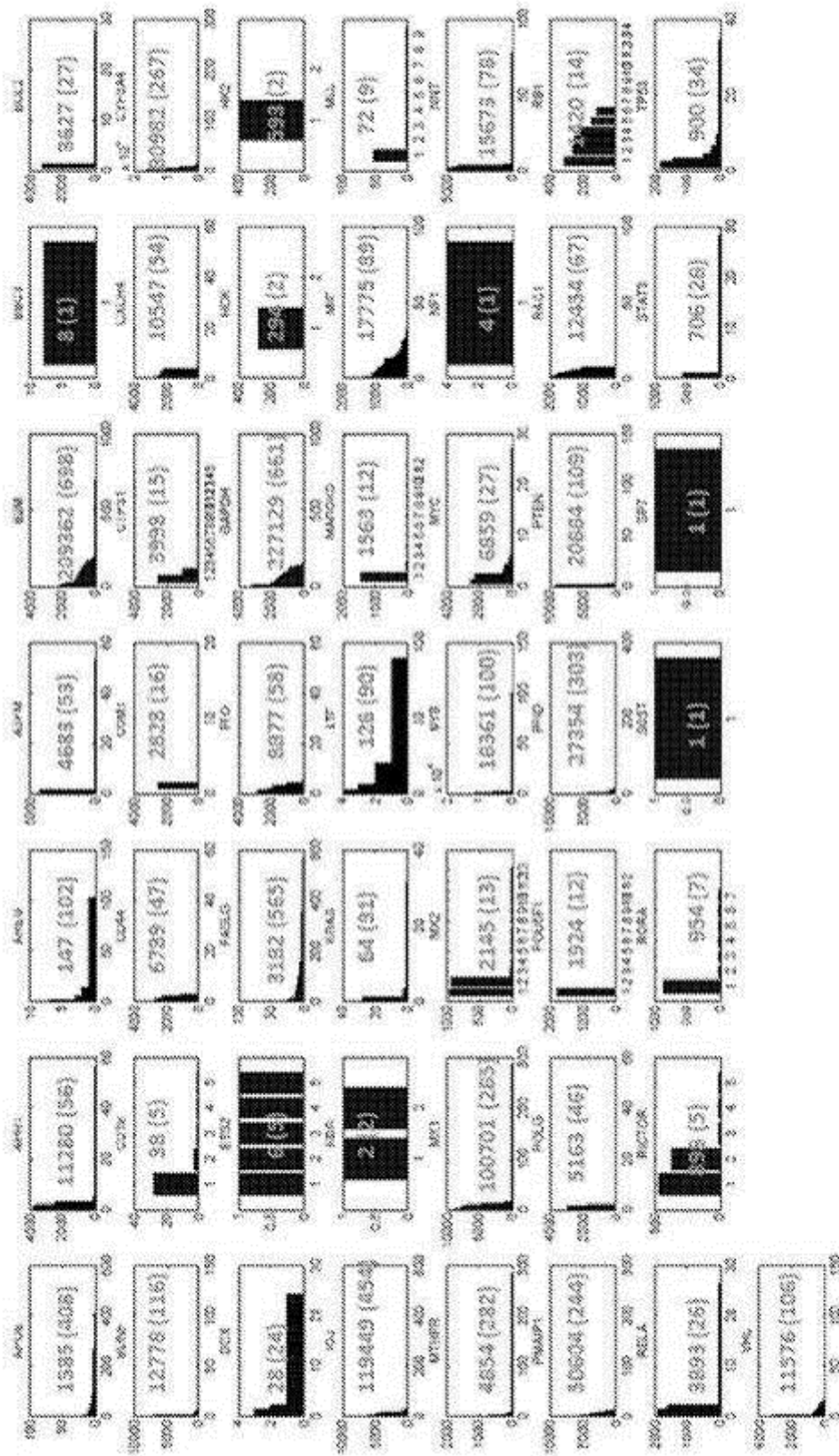


FIG. 63

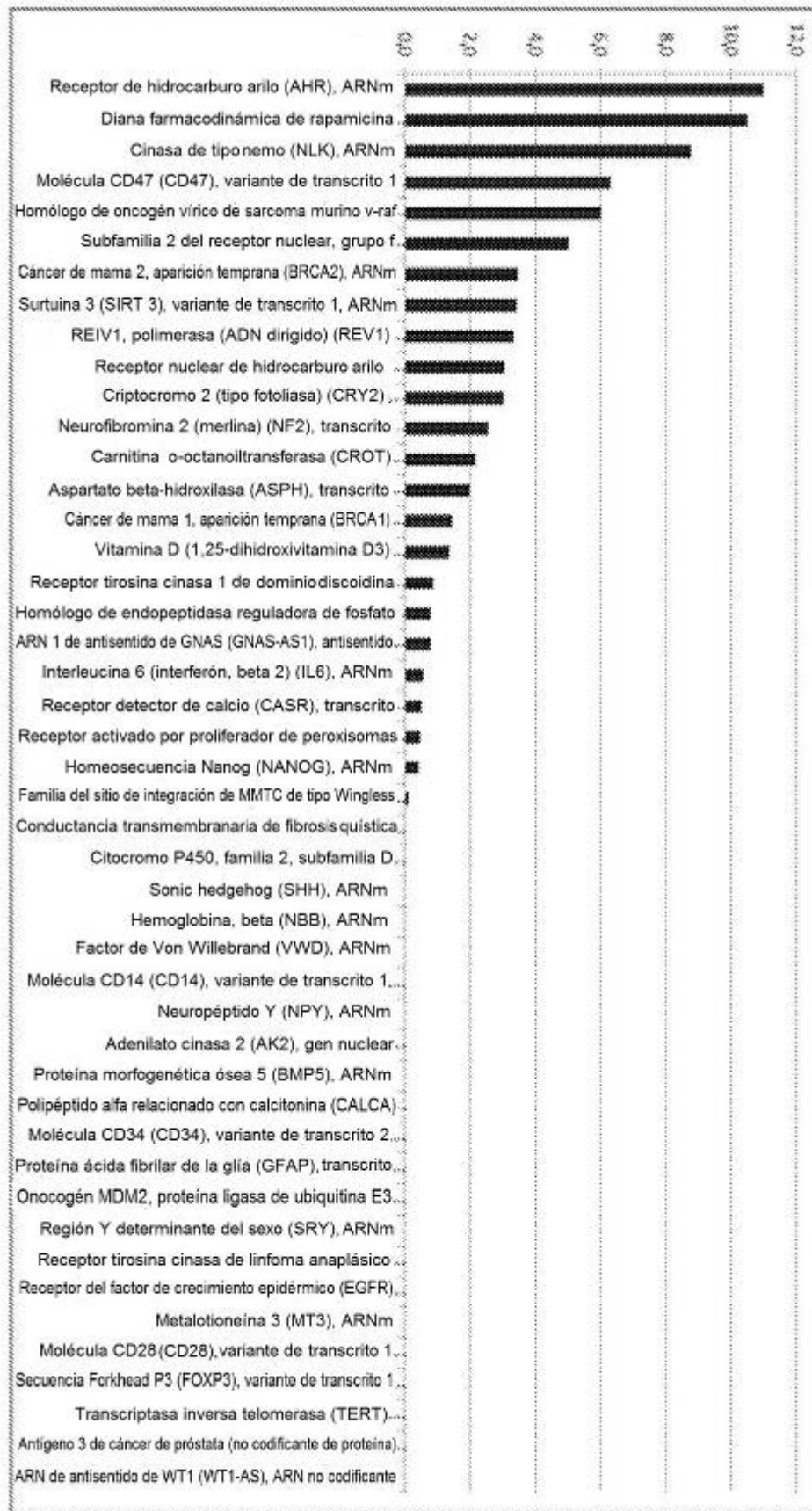


FIG. 64

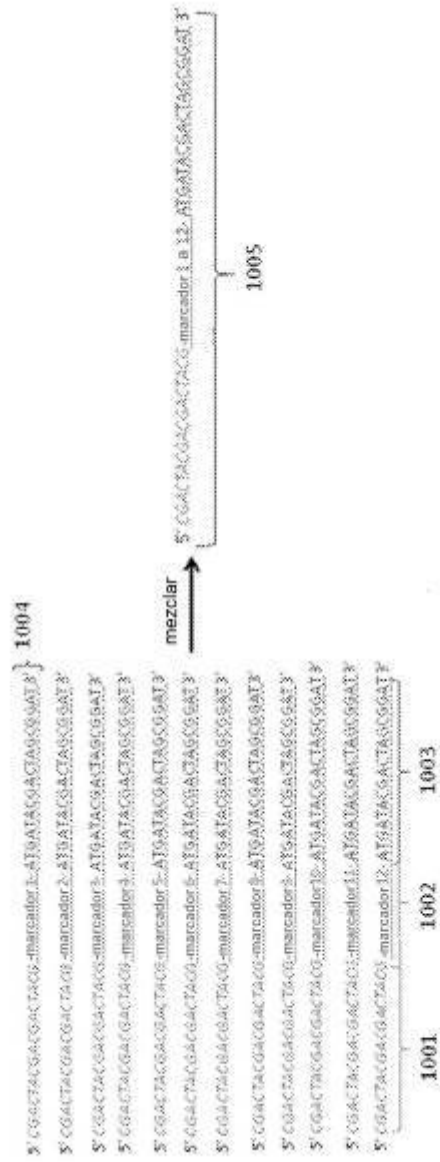


FIG. 65

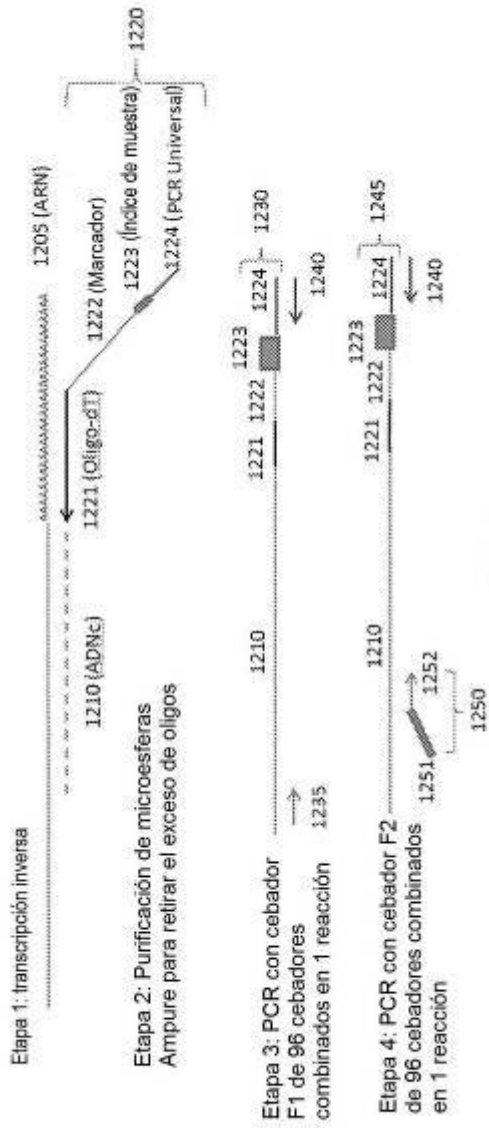


FIG. 67

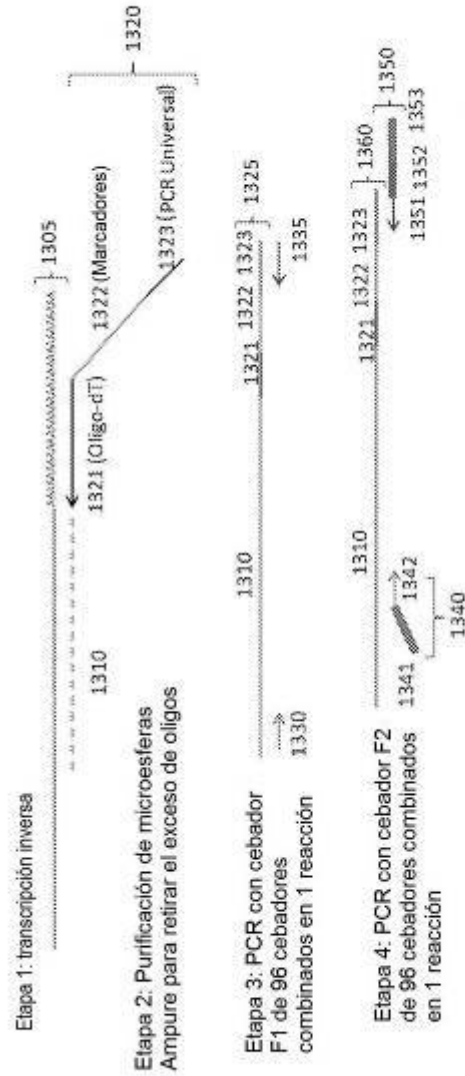


FIG. 68

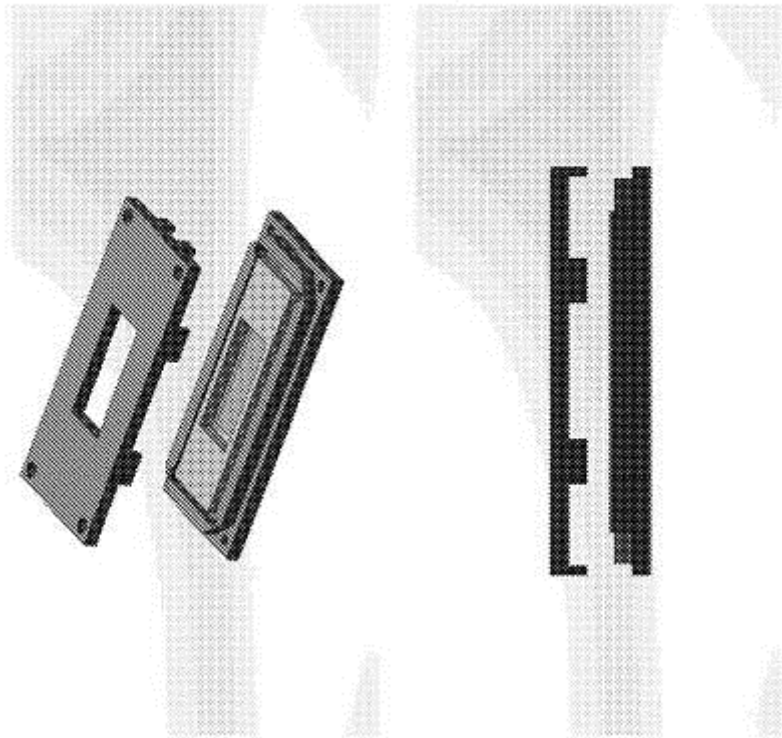


FIG. 69

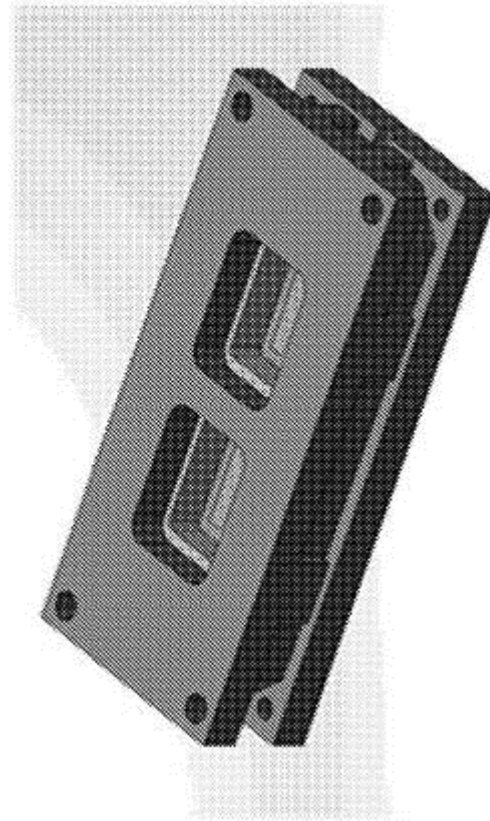


FIG. 70

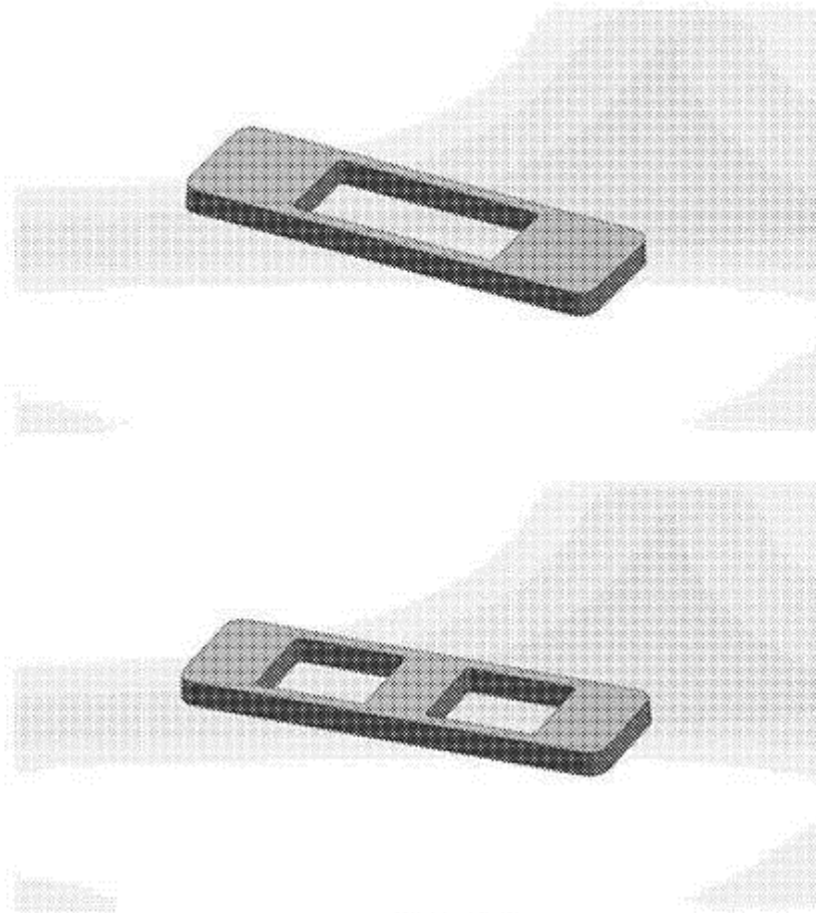


FIG. 71

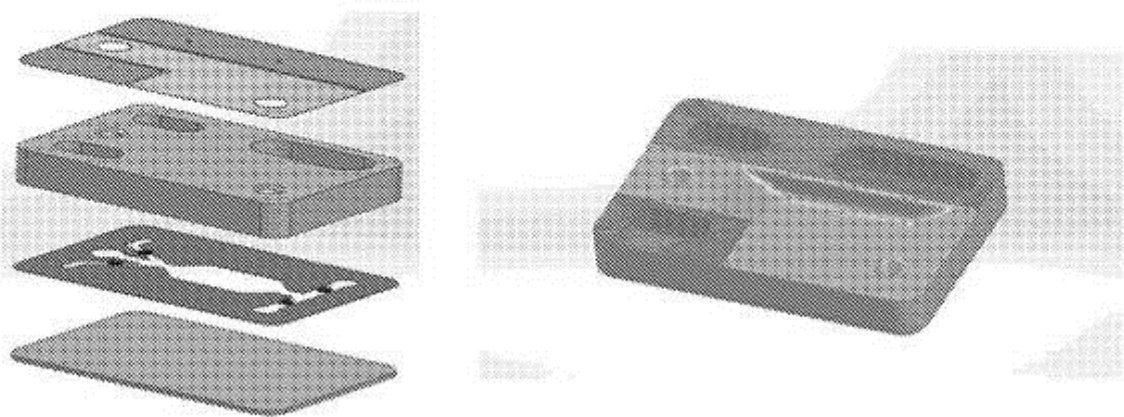


FIG. 72

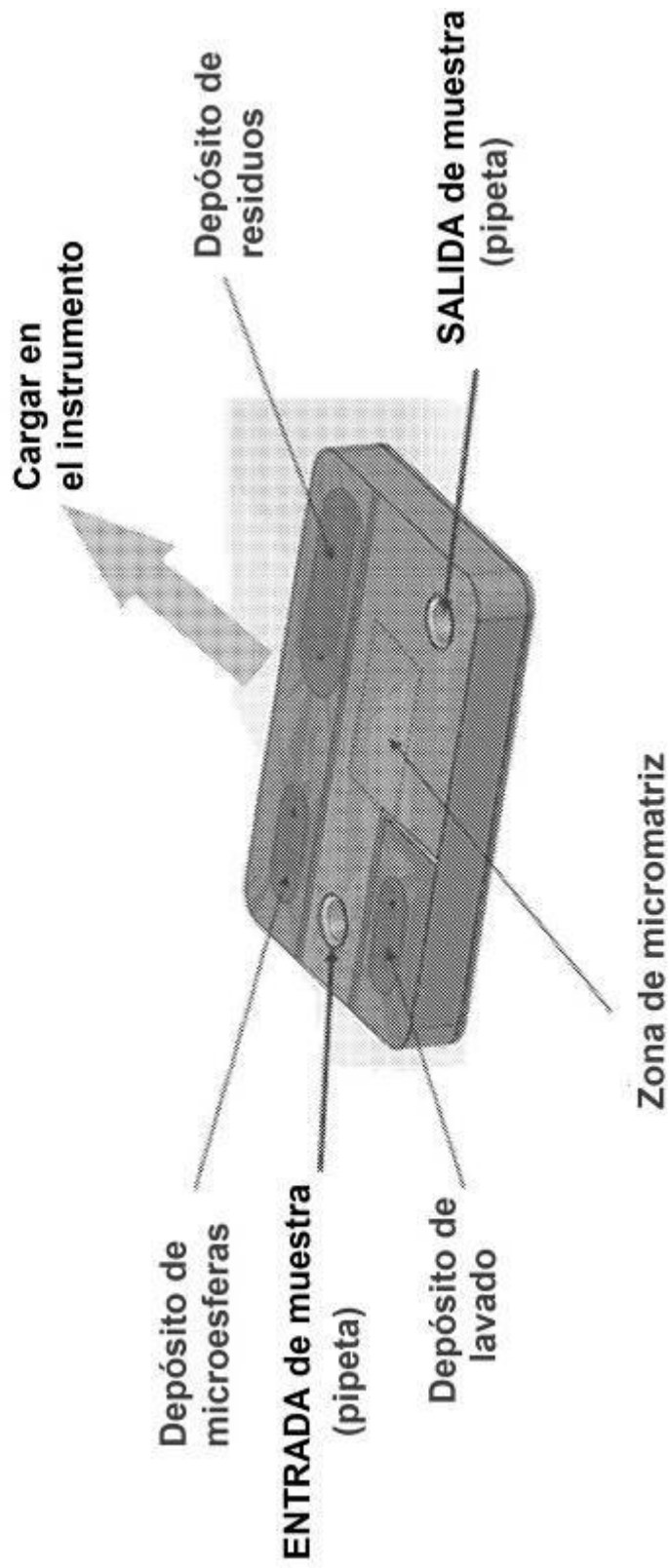


FIG. 73

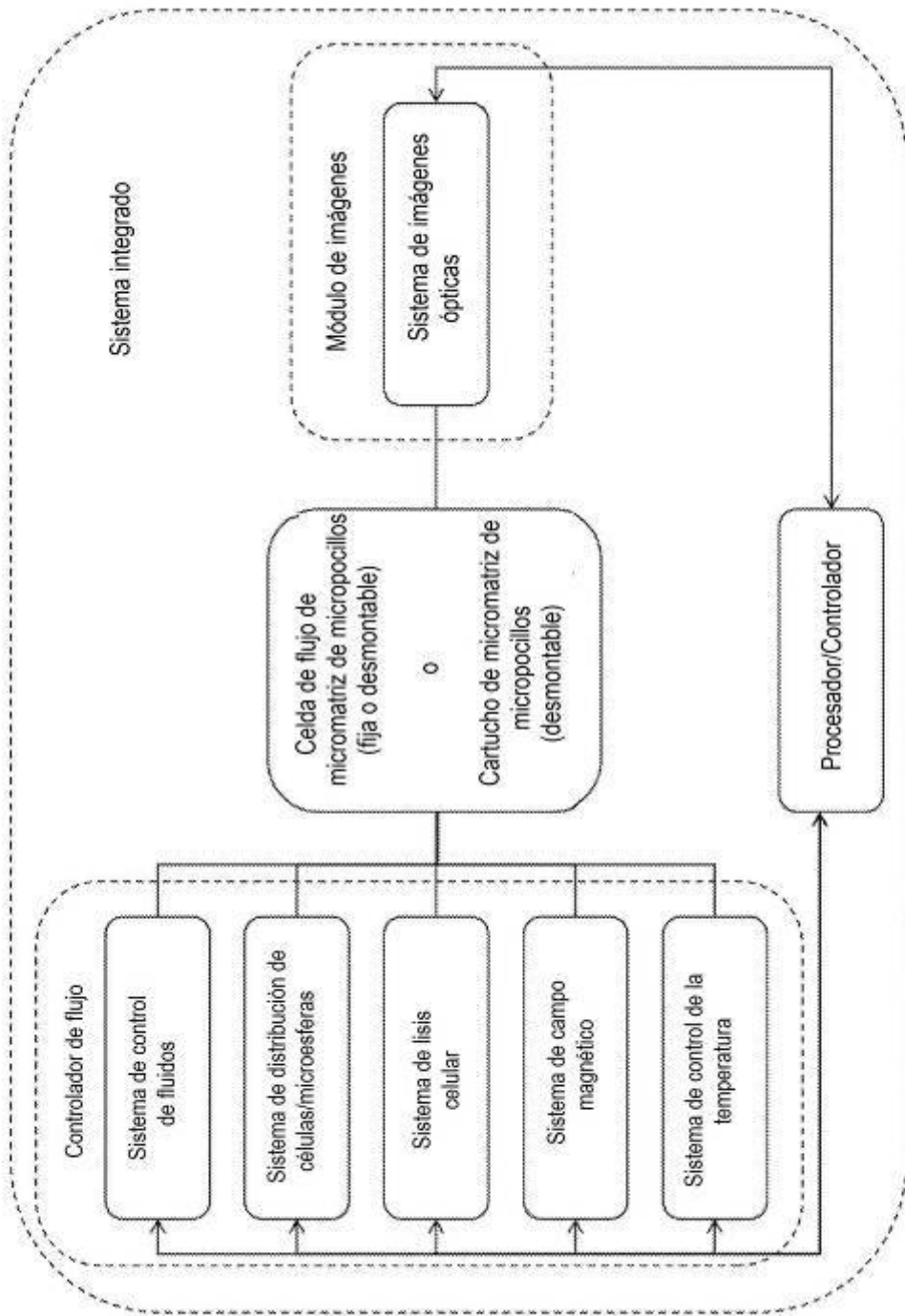


FIG. 74

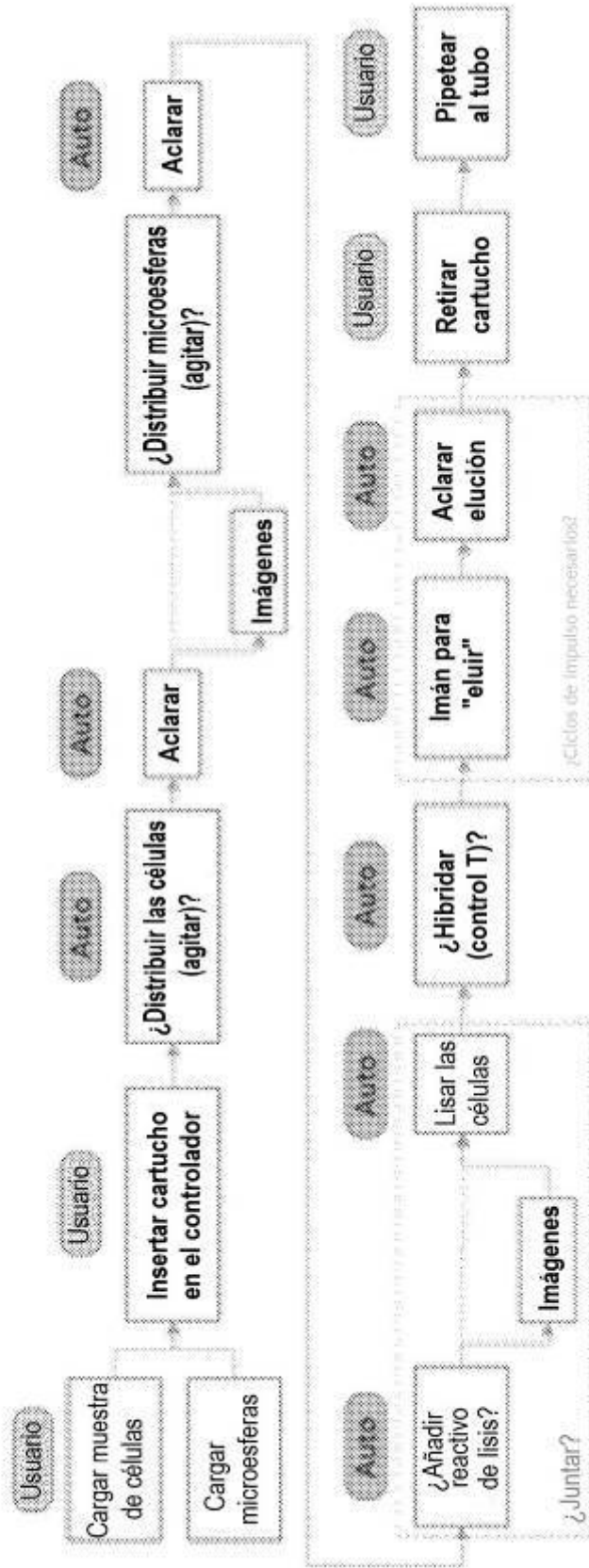


FIG. 75

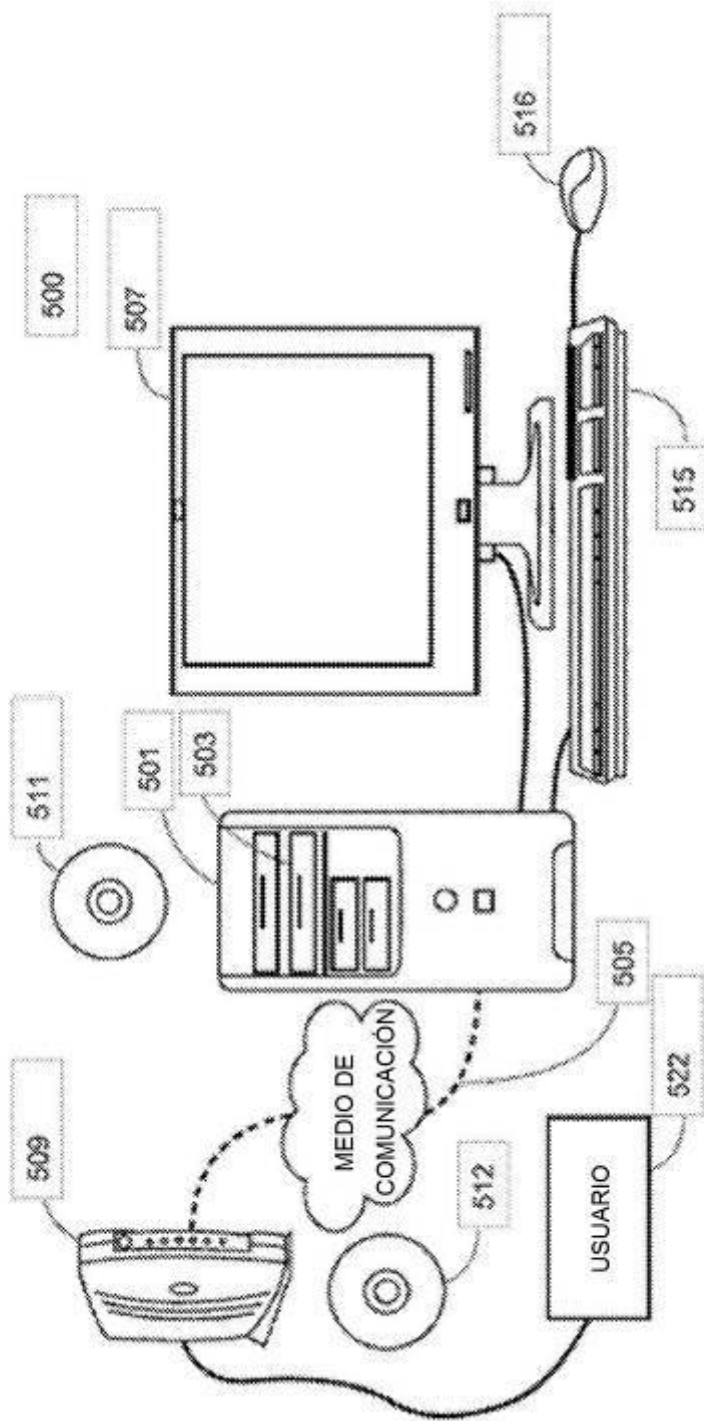


FIG. 76

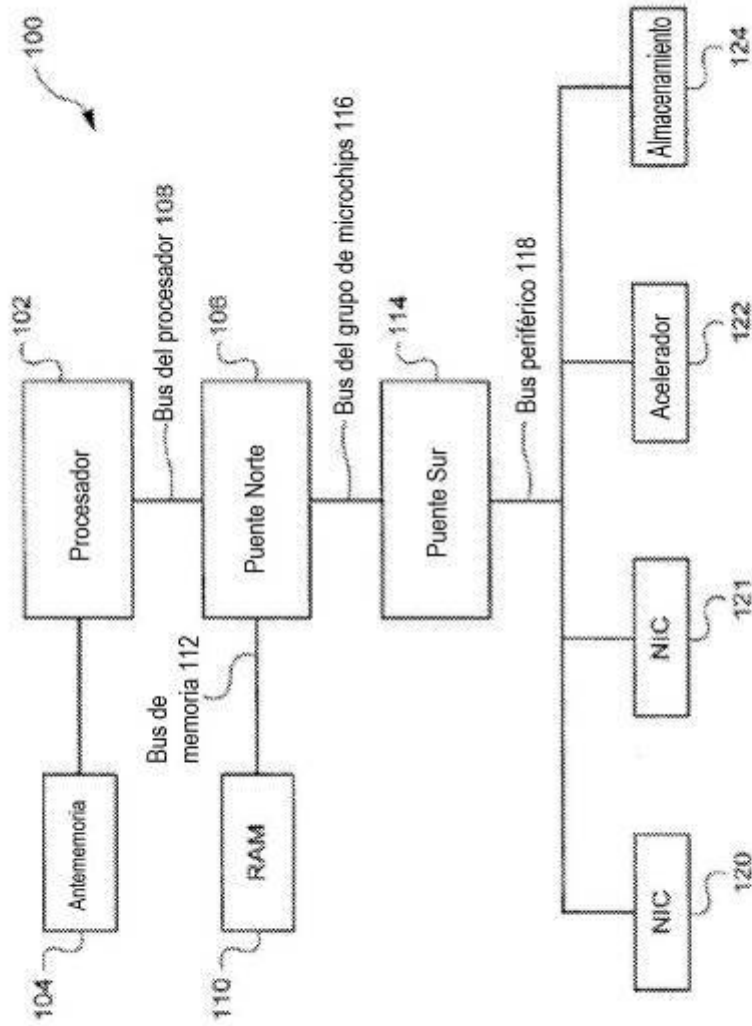


FIG. 77

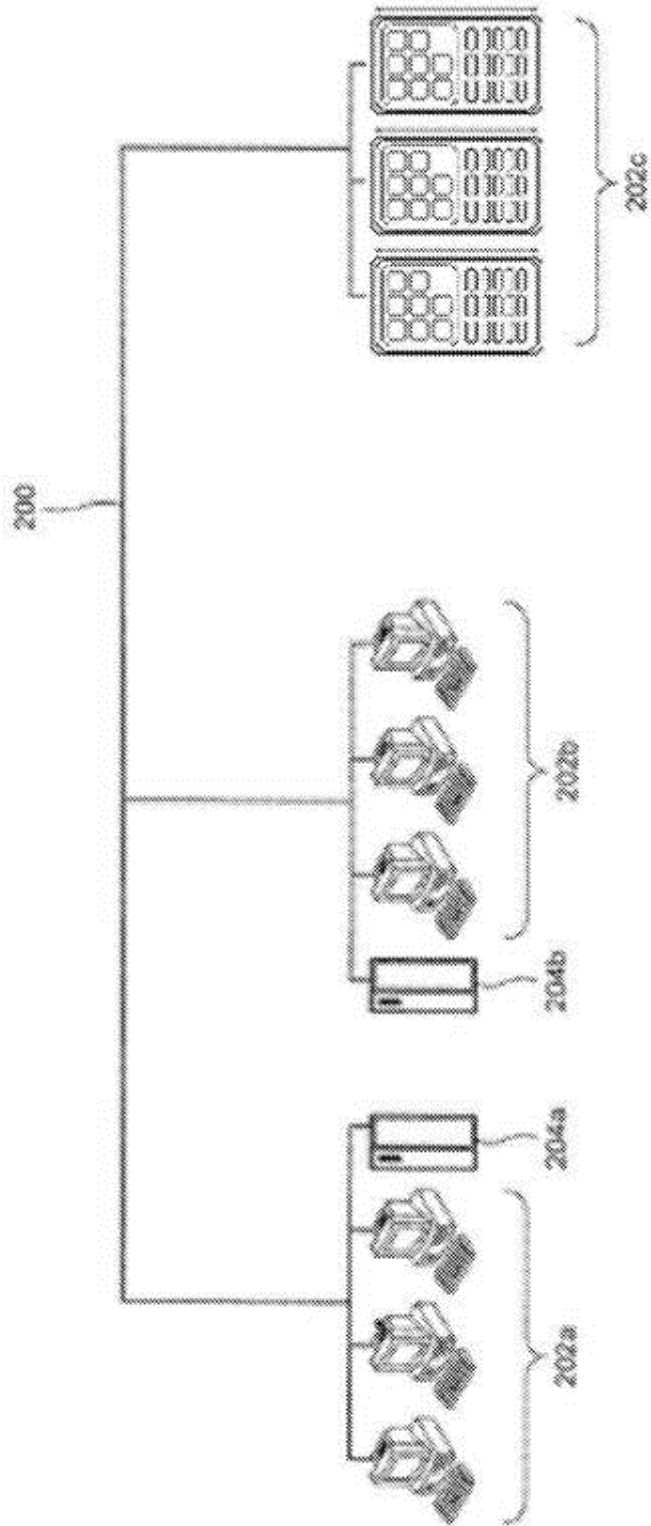


FIG. 78

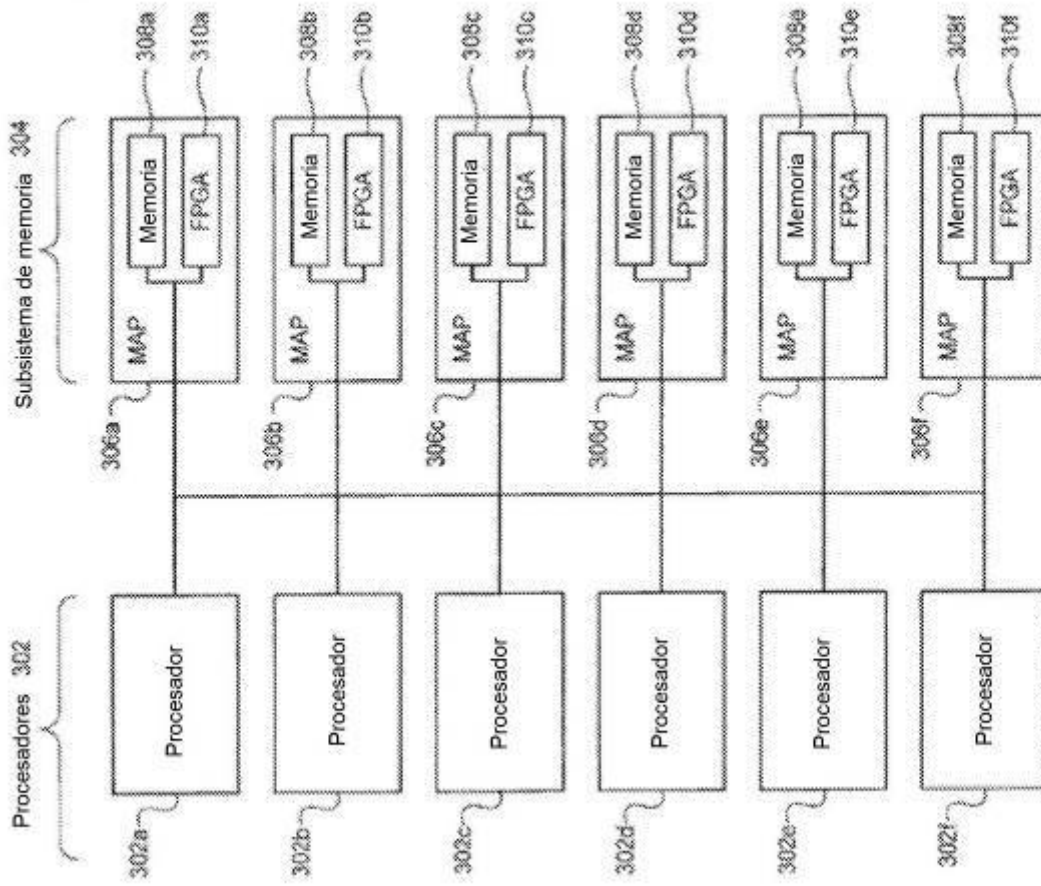


FIG. 79

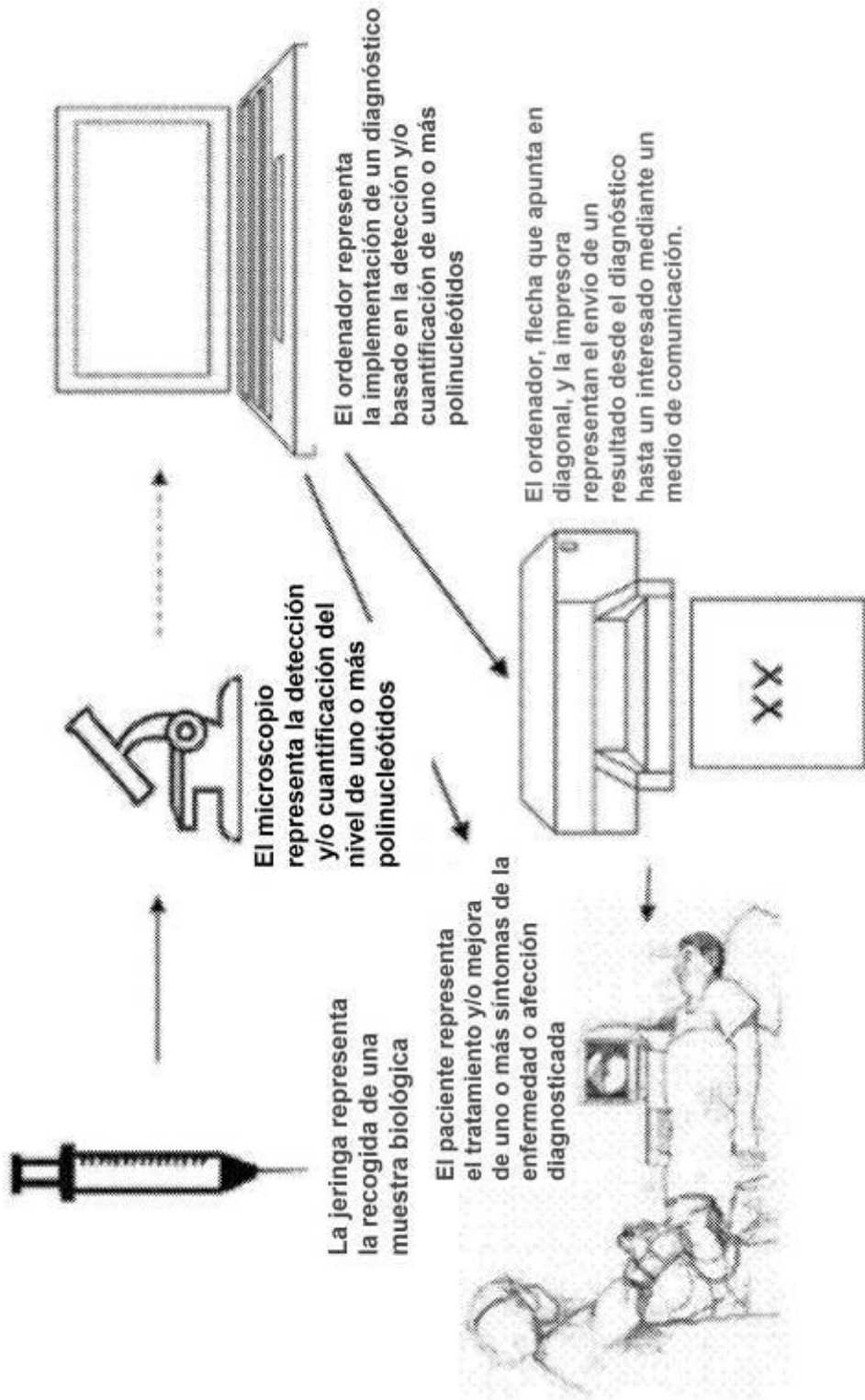


FIG. 80