

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 774 065**

51 Int. Cl.:

C12Q 1/6881 (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **01.04.2015 PCT/US2015/023915**

87 Fecha y número de publicación internacional: **08.10.2015 WO15153788**

96 Fecha de presentación y número de la solicitud europea: **01.04.2015 E 15772627 (4)**

97 Fecha y número de publicación de la concesión europea: **19.02.2020 EP 3126522**

54 Título: **Determinación de células T y células B específicas de antígeno**

30 Prioridad:

01.04.2014 US 201414242520

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

16.07.2020

73 Titular/es:

**ADAPTIVE BIOTECHNOLOGIES CORP. (100.0%)
1551 Eastlake Avenue East, Suite 200
Seattle, Washington 98102, US**

72 Inventor/es:

**KLINGER, MARK y
FAHAM, MALEK**

74 Agente/Representante:

SÁEZ MAESO, Ana

ES 2 774 065 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

Determinación de células T y células B específicas de antígeno

Antecedentes de la invención

5 Muchas funciones inmunes cruciales están mediadas por receptores de células T (TCR), que comprenden subunidades α y β que juntas se unen a un complejo que consiste en un péptido antigénico y moléculas del complejo mayor de histocompatibilidad (MHC). Se cree que varias enfermedades importantes surgen de la función aberrante de las células T: por ejemplo, se cree que los cánceres surgen de una falla de la vigilancia inmune, es decir, la función de las células T para detectar y destruir clones de células transformadas antes de que se conviertan en tumores; y se cree que las enfermedades autoinmunes surgen de una respuesta demasiado activa o aberrante de las células T a los antígenos propios, Abbas et al., *Cellular and Molecular Immunology*, cuarta edición (W.B. Saunders Company, 2000). En consecuencia, ha habido interés tanto en identificar y rastrear células T específicas de antígeno como en aprovechar las funciones de las células T en varios enfoques terapéuticos para el tratamiento tanto del cáncer como de enfermedades autoinmunes, por ejemplo, Molloy et al., *Current Opinion in Pharmacology*, 5: 438-443 (2005); Morgan et al., *Science*, 314: 126-129 (2006); Turcotte y Rosenberg, *Adv. Surg.*, 45: 341-360 (2011). Varios retos son planteados por intereses: las técnicas actuales para identificar y rastrear las células T específicas de antígeno, especialmente a gran escala, son difíciles y costosas, y asimismo, las técnicas actuales para identificar y aislar las subunidades TCR α y TCR β emparejadas que forman un receptor funcional son difíciles y costosas, por ejemplo, Sharma et al., *Human Immunology*, 75: 514-519 (2014). Con respecto a la detección de células T específicas de antígeno, el uso de tinción multimérica directa requiere el desarrollo laborioso de reactivos específicos restringidos a HLA, y otros ensayos, tales como ELISPOT, tinción de citocinas intracelulares y ensayos de proliferación, enumerar células T específicas de antígeno basadas en la detección de la activación después de la estimulación de las células T *in vitro* con antígeno, por ejemplo, Gratama et al., *Cytometry A*, 73A: 971-974 (2008). Con respecto al aislamiento de pares funcionales de cadenas de TCR, típicamente se identifica una célula T de interés y se expande clonalmente para permitir el aislamiento y análisis de ácidos nucleicos que codifican cada subunidad. Incluso para un antígeno de enfermedad común, tal como MART-1 en melanoma, el proceso de análisis de células individuales, clonación y aislamiento del receptor debe repetirse para cada paciente.

Del mismo modo, la determinación de alto rendimiento de las cadenas pesadas y ligeras de los receptores de células B es de gran interés para desarrollar nuevas investigaciones y compuestos de unión terapéutica, por ejemplo, Reddy et al., *Nature Biotechnology*, 28 (9): 965-969 (2010), y para determinar clonotipos específicos del paciente y/o de toda la población correlacionados con estados de enfermedad, por ejemplo, Faham y Willis, patentes estadounidenses Nos. 8.236.503; 8.748.103; y similares.

Recientemente, se han propuesto aplicaciones de diagnóstico y pronóstico que usan secuenciación de ADN a gran escala ya que el costo por base de secuenciación de ADN ha disminuido y las técnicas de secuenciación se han vuelto más convenientes, por ejemplo, Welch et al., *Hematology Am. Soc. Hematol. Educ. Program*, 2011: 30-35; Cronin et al., *Biomark Med.*, 5: 293-305 (2011); Palomaki et al., *Genetics in Medicine* (publicación en línea del 2 de febrero de 2012). En particular, los perfiles de ácidos nucleicos que codifican moléculas inmunes, tal como los receptores de células T o células B, o sus componentes, contienen una gran cantidad de información sobre el estado de salud o enfermedad de un organismo, de modo que los indicadores de diagnóstico y pronóstico basados en el uso de tales perfiles se están desarrollando para una amplia variedad de condiciones, Faham y Willis (citado anteriormente); Freeman et al., *Genome Research*, 19: 1817-1824 (2009); Boyd et al., *Sci. Transl. Med.*, 1 (12): 12ra23 (2009); He et al., *Oncotarget* (8 de marzo de 2011). Se han descrito otras aplicaciones de secuenciación de repertorios inmunes (Klinger et al., *PlosOne* 2013; 8 (9); Robins H, *Curr. Opin. Immunol.* 2013 25(S) 646-52; Van Heijst et al., *Nat. Med.* 2013, 19 (3): 372-7). Los perfiles actuales basados en la secuencia de los repertorios inmunes consisten en ácidos nucleicos que codifican solo cadenas receptoras únicas; por lo tanto, la información potencialmente útil de cadenas de TCR α y TCR β correctamente emparejadas no está disponible.

En vista de lo anterior, sería muy útil para el tratamiento del cáncer, enfermedades infecciosas y enfermedades autoinmunes si hubiera métodos convenientes disponibles para determinar los receptores inmunes funcionales de ácidos nucleicos que codifican subunidades que se han extraído y secuenciado por separado, así como para determinar antígenos específicos a los que se unen cadenas de receptores emparejadas.

50 Sumario de la invención

La invención proporciona un método para identificar clonotipos de células T específicas de antígeno en una muestra que comprende células T, comprendiendo el método las etapas de:

(a) exponer la muestra a una pluralidad de antígenos en una pluralidad de mezclas de reacción diferentes, en las que cada antígeno de dicha pluralidad de antígenos está presente en una subpluralidad única predeterminada de la pluralidad de mezclas de reacción;

(b) seleccionar células T específicas de antígeno de cada una de la pluralidad de mezclas de reacción;

(c) secuenciar ácidos nucleicos recombinados que codifican clonotipos del mismo segmento del receptor de células T

a partir de:

(i) las células T específicas de antígeno seleccionadas en la etapa (b) de cada una de la pluralidad de mezclas de reacción, en la que la secuenciación produce una pluralidad de lecturas de secuencia para cada mezcla de reacción a partir de la cual se determinan los clonotipos y las frecuencias de clonotipos, y

5 (ii) células T obtenidas de la muestra antes de la etapa (a), en la que la secuenciación produce una pluralidad de lecturas de secuencia a partir de las cuales se determinan los clonotipos y las frecuencias de clonotipos;

(d) para cada mezcla de reacción, comparar las frecuencias de clonotipos determinadas en la etapa (c) (i) con las frecuencias de clonotipos determinadas en la etapa (c) (ii); y

10 (e) para cada uno de la pluralidad de antígenos, identificar un clonotipo de una célula T específica de antígeno que tiene una mayor frecuencia aumentada en cada una de las mezclas de reacción en las que estaba presente el antígeno y no tiene una mayor frecuencia en cada una de las mezclas de reacción en las que el antígeno estaba ausente.

15 En el presente documento se describen métodos para determinar receptores de células T o receptores de células B de subunidades seleccionadas de bibliotecas separadas, particularmente receptores de células T o receptores de células B específicos de antígeno. La invención se ejemplifica en varias implementaciones y aplicaciones, algunas de las cuales se resumen a continuación y a lo largo de la memoria descriptiva.

20 En este documento se describen métodos para determinar linfocitos específicos de antígeno en una muestra de tejido que comprenden etapas de (a) hacer reaccionar bajo condiciones de activación en una mezcla de reacción, una muestra de tejido que comprende linfocitos con un antígeno, teniendo los linfocitos receptores de superficie capaces de unirse al antígeno, y comprendiendo los receptores una primera cadena y una segunda cadena; (b) clasificar linfocitos de la mezcla de reacción en un subconjunto de linfocitos específicos de antígeno y/o linfocitos activados y un subconjunto de linfocitos no específicos de antígeno y/o linfocitos no activados; (c) secuenciar ácidos nucleicos recombinados que codifican una cadena de receptores de linfocitos o una porción de la misma a partir de una muestra de linfocitos de los linfocitos específicos de antígeno y/o linfocitos activados para proporcionar lecturas de secuencia a partir de las cuales se determinan los clonotipos; (d) secuenciar ácidos nucleicos recombinados que codifican una 25 cadena del receptor de linfocitos o una porción de la misma a partir de una muestra de linfocitos de linfocitos no específicos de antígeno y/o linfocitos inactivados para proporcionar lecturas de secuencia a partir de las cuales se determinan los clonotipos; y (e) determinar los linfocitos específicos de antígeno en la muestra de tejido como linfocitos cuyas frecuencias de clonotipo aumentan en el subconjunto de linfocitos específicos y/o antígenos clasificados en relación con las frecuencias de los mismos clonotipos en la mezcla de reacción o el subconjunto de linfocitos clasificados no específicos de antígeno y/o no activados. En algunos casos, los linfocitos son células T y la primera 30 cadena del receptor es una cadena $TCR\alpha$ y la segunda cadena del receptor es una cadena $TCR\beta$. En otros casos, la primera cadena del receptor es una cadena pesada de inmunoglobulina y la segunda cadena del receptor es una cadena ligera de inmunoglobulina.

35 En este documento se describen métodos para determinar receptores de linfocitos específicos de antígeno en una muestra de tejido que comprende las siguientes etapas: (a) formar una pluralidad de subconjuntos a partir de una muestra de tejido que contiene linfocitos, teniendo los linfocitos receptores de superficie capaces de unirse al antígeno, comprendiendo los receptores una primera cadena y una segunda cadena; (b) hacer reaccionar bajo condiciones de activación los linfocitos de cada subconjunto con un antígeno; (c) aislar los linfocitos específicos de antígeno de cada subconjunto; (d) secuenciar ácidos nucleicos recombinados que codifican primeras cadenas de linfocitos en cada 40 subconjunto para proporcionar lecturas de secuencia a partir de las cuales se determinan los primeros clonotipos de cadena; (e) secuenciar ácidos nucleicos recombinados que codifican segundas cadenas de linfocitos en cada subconjunto para proporcionar lecturas de secuencia a partir de las cuales se determinan los segundos clonotipos de cadena; y (f) identificar como receptores de linfocitos específicos de antígeno con esos pares de los primeros clonotipos de cadena y segundos clonotipos de cadena que para cada subconjunto (i) tanto el primer clonotipo de 45 cadena como el segundo clonotipo de cadena están presentes en un subconjunto o ninguno está presente en un subconjunto, y (ii) tanto el primer clonotipo de cadena como el segundo clonotipo de cadena están presentes en al menos un subconjunto y el primer clonotipo de cadena y el segundo clonotipo de cadena no están presentes en al menos un subconjunto. En algunos casos, los linfocitos son células T y la primera cadena es una cadena $TCR\alpha$ y la segunda cadena es una cadena $TCR\beta$. En algunos casos, los linfocitos son células B y la primera cadena es una 50 cadena pesada de inmunoglobulina y la segunda cadena es una cadena ligera de inmunoglobulina.

55 En este documento se describe un método de alto rendimiento para identificar pares de ácidos nucleicos que codifican receptores inmunes funcionales y antígenos a los que se unen específicamente. En algunos casos, los métodos de la invención pueden usarse para identificar antígenos de cáncer específicos del paciente a partir de una biblioteca de antígenos de cáncer, en la que la biblioteca puede basarse en mediciones específicas del paciente, por ejemplo, secuenciación del genoma completo o secuenciación del exoma completo, mediciones SEREX (por ejemplo, Pfreundschuh, patente de los Estados Unidos No. 5.698.396, en la que se identifican ADNc tumorales que codifican antígenos cancerosos), o similares, o pueden basarse en colecciones basadas en poblaciones de antígenos de cáncer (es decir, basadas en antígenos de cáncer compartidos por múltiples pacientes debido a la ocurrencia común en tipos

de cáncer particulares). En otros casos, los métodos de la invención pueden usarse para identificar y cuantificar la inmunogenicidad de los epítomos de proteínas de agentes infecciosos.

Estos aspectos caracterizados anteriormente, así como otros aspectos, de la presente invención se ejemplifican en una serie de implementaciones y aplicaciones ilustradas, algunas de las cuales se muestran en las figuras y se caracterizan en la sección de reivindicaciones que sigue. Sin embargo, el resumen anterior no pretende describir cada realización ilustrada o cada implementación de la invención.

Breve descripción de los dibujos

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

La Figura 1A ilustra esquemáticamente las etapas para aparear cadenas de TCR α y TCR β de moléculas secuenciadas por separado.

La Figura 1B ilustra esquemáticamente las etapas para determinar las cadenas TCR α y TCR β que se originan a partir de la misma célula T o inmunoglobulinas de cadena pesada y ligera que se originan de la misma célula B.

La Figura 1C ilustra esquemáticamente la identificación de células T específicas de antígeno que interactúan con un solo antígeno.

La Figura 1D ilustra esquemáticamente la identificación de células T específicas de antígeno que interactúan con una pluralidad de antígenos.

La Figura 1E ilustra las etapas para la identificación física de células T específicas de antígeno usando la metodología de células individuales.

La Figura 1F ilustra un esquema de PCA para unir secuencias objetivo en las que los pares de cebadores internos tienen complementariedad.

Las Figuras 2A-2C muestran un esquema de PCR de dos etapas para amplificar genes de TCR β .

La Figura 3A ilustra detalles de la determinación de una secuencia de nucleótidos del producto de PCR de la Figura 2C. La Figura 3B ilustra detalles de la determinación de una secuencia de nucleótidos del producto de PCR de la Figura 2C.

La Figura 4 ilustra un ejemplo de una muestra de tejido dividida o porcionada en alícuotas en una pluralidad de subconjuntos 1 a 10 y ejemplos de diferentes subpluralidades de subconjuntos de la pluralidad.

Las Figuras 5A y 5B ilustran una realización de la invención para determinar células T y los TCR específicos para una pluralidad de antígenos.

La Figura 6 muestra datos para la identificación de clonotipos de células T específicos de pp65₄₉₅ de CMV a partir de células T clasificadas pentámero+. El panel (A) muestra las frecuencias de clonotipos de células T CD8+ pentámero+ frente a pentámero- de pp65₄₉₅ de CMV de un respondedor de CMV caracterizado. Los 8 puntos (encerrados en la elipse 600 punteada) indican clonotipos mayores enriquecidos más de 10 veces y que exceden un umbral de frecuencia mínimo equivalente a 20 células en la población clasificada (pentámero+). (B) Todos los 8 clonotipos identificados en el panel A están enriquecidos en PBMC (sin clasificar) del mismo individuo. Los puntos encerrados por la elipse (602) punteada indican clonotipos identificados en el panel A.

La Figura 7 muestra los datos para la identificación de clonotipos de células T específicos de pp65₄₉₅ de CMV de células clasificadas que responden después de la incubación del péptido. Las frecuencias de clonotipos de células CD137+ clasificadas que responden después de la incubación del péptido pp65₄₉₅ de CMV frente a células CD137- clasificadas que no responden (panel A) o PBMC no clasificadas (panel B). Los 9 puntos de datos encerrados por la elipse (700) punteada en el panel A indican clonotipos enriquecidos más de 10 veces y que exceden un umbral de frecuencia mínima equivalente a 20 células en la población clasificada (CD137+). Los puntos de datos encerrados en la elipse (702) punteada en el panel B indican aquellos clonotipos identificados en el panel A. Los clonotipos identificados en el panel A no están enriquecidos en células CD137+ clasificadas frente a las células T CD137- (las flechas del panel C que indican los puntos de datos correspondientes a aquellos encerrados por las elipses en los paneles A y B) después de la incubación sin péptido.

La Figura 8 ilustra la superposición entre clonotipos identificados en ensayos con base en pentámero y con base en CD137. En el panel A, el gráfico muestra las frecuencias de clonotipos de los 8 clonotipos (encerrados por la elipse 800 punteada) identificados en los análisis de pentámeros en los perfiles de clonotipos de las células que responden a CD137+ después de la incubación del péptido pp65₄₉₅ de CMV frente a las células CD137- clasificadas que no responden. En el panel B, los gráficos muestran las frecuencias de clonotipos de los 9 clonotipos (encerrados por la

elipse 802 punteada) identificados en los análisis del ensayo de CD137 en los perfiles de clonotipos de células clasificadas pentámero+ frente a células pentámero- pp65₄₉₅ de CMV. 8/9 de estos clonotipos se superponen con los identificados en el panel A.

5 La Figura 9 muestra datos para la identificación de clonotipos de células T específicos de pp65₄₉₅ de CMV de baja frecuencia después de la incubación y proliferación de péptidos. En el panel A, las frecuencias de clonotipos de células T CD8+ clasificadas que proliferan después de la incubación del péptido pp65₄₉₅ de CMV en el día 6 frente a las PBMC frescas sin clasificar. Los 16 puntos de datos (encerrados por la elipse 900 punteada) indican clonotipos enriquecidos más de 10 veces y que exceden el umbral de frecuencia mínima de 1/10.000 en las células proliferantes clasificadas. El panel B muestra datos de frecuencias de clonotipos de células T CD8+ proliferantes clasificadas después de la incubación sin péptido en el día 6 frente a PBMC frescas sin clasificar. Los clonotipos (representados por puntos de datos encerrados por las elipses (902) punteadas) son aquellos identificados en A. El panel C muestra datos de las frecuencias de clonotipos de células T CD8+ pentámero+ frente a pentámero- de pp65₄₉₅ de CMV. Las elipses (904) punteadas y la flecha (906) indican los 16 clonotipos identificados en el ensayo de proliferación cuyos resultados se representan en el panel A con aquellos clonotipos identificados en la comparación de células T CD8+ pentámero+ frente a pentámero- de pp65₄₉₅ de CMV. El panel D muestra datos de frecuencias de clonotipos de células T CD8+ pentámero+ frente a pentámero- de pp65₄₉₅ DE CMV. Los puntos de datos encerrados por las elipses (908) punteadas y designadas con flechas (910) indican los 25 clonotipos identificados en una variante del ensayo de proliferación descrito anteriormente con los clonotipos identificados en la comparación de células T CD8+ pentámero+ frente a pentámero- de pp65₄₉₅ de CMV. En este ensayo, un grupo de 138 péptidos superpuestos de pp65 se utilizó en lugar del péptido pp65₄₉₅ único.

La Figura 10 ilustra los conceptos de clonotipos en el espacio de secuencia y las distancias entre clonotipos estrechamente relacionados.

La Figura 10B es un diagrama de flujo que ilustra una realización de un método para distinguir clonotipos genuinamente diferentes de los clonotipos que difieren únicamente por errores de secuenciación (que deberían coalescer).

25 La Figura 10C ilustra la forma de una función numérica utilizada para determinar si se unen o no los clonotipos relacionados.

Las Figuras 10D y 10E ilustran el uso de árboles de secuencia en un método de lectura de secuencias que coalecen.

Descripción detallada de la invención

30 La práctica de la presente invención puede emplear, a menos que se indique lo contrario, técnicas convencionales y descripciones de biología molecular (incluyendo técnicas recombinantes), bioinformática, biología celular y bioquímica, que están dentro de la habilidad de la técnica. Dichas técnicas convencionales incluyen, pero no se limitan a, muestreo y análisis de células sanguíneas, secuenciación y análisis de ácidos nucleicos y similares. Pueden obtenerse ilustraciones específicas de técnicas adecuadas haciendo referencia al siguiente ejemplo a continuación. Sin embargo, también se pueden utilizar otros procedimientos convencionales equivalentes. Dichas técnicas y descripciones convencionales se pueden encontrar en los manuales estándar de laboratorio. En particular, las publicaciones de las solicitudes internacionales Nos. WO2013/188831 y WO2014/145992.

Identificación de cadenas emparejadas de receptores de células T sin selección específica de antígeno

En el presente documento se describen métodos para emparejar pares de cadenas de receptores inmunes de poblaciones de sus ácidos nucleicos codificadores que se han secuenciado. De acuerdo con un caso, las poblaciones de ácido nucleico que codifican repertorios de regiones variables de cadena pesada y regiones variables de cadena ligera se secuencian de modo que se formen dos listas separadas de secuencias sin ninguna correspondencia entre los miembros de cada lista. Esto puede lograrse llevando a cabo operaciones de secuenciación separadas, o proceso, para cada cadena, o puede lograrse llevando a cabo un solo proceso de secuenciación con los ácidos nucleicos marcados de acuerdo con la identidad del tipo de cadena que codifica. De acuerdo con otro caso, se secuencian las poblaciones de ácido nucleico que codifican repertorios de cadenas alfa de receptores de células T (TCR α) y cadenas beta de receptores de células T (TCR β), de modo que se forman dos listas separadas de secuencias sin ninguna correspondencia entre los miembros de cada lista. De acuerdo con otro caso, se secuencian poblaciones de ácido nucleico que codifican los repertorios de las cadenas gamma de receptores de células T (TCR γ) y las cadenas delta de receptores de células T (TCR δ), de modo que se forman dos listas separadas de secuencias sin ninguna correspondencia entre los miembros de cada lista. Como se indicó anteriormente, esto puede lograrse llevando a cabo procesos de secuenciación separados para cada cadena, o puede lograrse llevando a cabo un solo proceso de secuenciación con los ácidos nucleicos marcados de acuerdo con la identidad del tipo de cadena que codifica (es decir, TCR α y TCR β , o TCR γ y TCR δ , respectivamente). En los últimos casos, se pueden seguir dos enfoques para aparear o emparejar cadenas TCR α y TCR β (o TCR γ y TCR δ) en cadenas que son funcionales, por ejemplo, porque se originan a partir de la misma célula T. En un primer enfoque, se determinan las frecuencias de cada ácido nucleico codificador y las cadenas TCR α y cadenas TCR β cuyas secuencias de nucleótidos codificantes tienen las mismas frecuencias se emparejan para formar una TCR funcional o reconstituida. Las cadenas TCR γ y TCR δ pueden emparejarse mediante el mismo proceso. En un segundo enfoque, que es aplicable para emparejar los tres tipos de

pares de receptores inmunes, una población de linfocitos se divide repetidamente en una pluralidad de subconjuntos. Dichos subconjuntos se pueden obtener colocando alícuotas de una muestra de tejido en recipientes o cámaras de reacción separados. Se extraen y secuencian por separado de cada porción, o subpoblación, de los subconjuntos, los ácidos nucleicos que codifican las dos cadenas de receptores inmunes diferentes, de modo que se forman dos listas separadas de secuencias sin ninguna correspondencia entre los miembros de cada lista. Como se describió anteriormente, esto puede lograrse llevando a cabo procesos de secuenciación separados para cada cadena, o puede lograrse llevando a cabo un solo proceso de secuenciación con los ácidos nucleicos marcados de acuerdo con la identidad del tipo de cadena que codifica. Para ilustración mediante un ejemplo, si una muestra que contiene células T o células B se divide en alícuotas en 100 submuestras, de modo que, en promedio, cada alícuota contiene un subconjunto que consta de aproximadamente 1/100 del número total de células T o células B en la muestra original, entonces 20 de estos subconjuntos pueden seleccionarse al azar como una porción del número total de subconjuntos. (Tal porción podría ser cualquier número mayor que uno y menor que 100, aunque como se describe más detalladamente a continuación, un número en el intervalo de 10 a 20 es una buena compensación entre la cantidad de secuenciación requerida y la probabilidad de identificar pares de receptores presentes en una frecuencia de interés). En un caso, una pluralidad de subconjuntos está en el intervalo de 20 a 2000 y una porción de subconjuntos del mismo está en el intervalo de 10 a 50. En otra realización, una porción de subconjuntos está en el intervalo de 10 a 20. Los ejemplos de las realizaciones anteriores se ilustran en las Figuras 1A y 1B.

Como se ilustra en la Figura 1A, el ácido nucleico (que puede ser ADN o ARN) se extrae de una muestra que contiene células T (100), después de lo cual en volúmenes de reacción separados, se combinan cebadores (102) específicos para ácidos nucleicos que codifican TCR α (o una porción de los mismos) y cebadores (104) específicos para ácidos nucleicos que codifican TCR β (o una porción de los mismos) bajo condiciones que permiten amplificar las respectivas poblaciones de ácido nucleico, por ejemplo, mediante una reacción en cadena de la polimerasa (PCR) en dos etapas, como la descrita por Faham y Willis (citada anteriormente). La guía y las descripciones para seleccionar tales cebadores y llevar a cabo tales reacciones se describen ampliamente en la literatura de inmunología molecular y más abajo (para TCR β e IgH) y en referencias tales como Yao et al., *Cellular and Molecular Immunology*, 4: 215-220 (2007) (para TCR α). En un caso, los amplicones (106) y (108) producidos por una PCR de dos etapas están listos para el análisis de secuencia utilizando un secuenciador de próxima generación disponible comercialmente, tal como el secuenciador personal MiSeq (Illumina, San Diego, CA). Después de determinar las secuencias de nucleótidos (107) y (109), se obtienen bases de datos o tablas (110 y 112, respectivamente). Se pueden contar secuencias similares y construir gráficos de frecuencia contra secuencia (114 y 116). Los TCR reconstituidos pueden determinarse haciendo coincidir (118) los TCR α y TCR β con frecuencias idénticas o con frecuencias que tengan el mismo orden de intervalo. Claramente, este método funciona de manera más eficiente cuando las frecuencias de diferentes TCR α y TCR β no están demasiado juntas, es decir, son distintas, incluso teniendo en cuenta el error experimental.

Una vez que se identifica un par de secuencias de clonotipo que tienen frecuencias iguales (o igualmente clasificadas), las secuencias de longitud completa que codifican cada cadena pueden reconstruirse a partir de las regiones constantes y variables conocidas utilizando técnicas convencionales para la manipulación y expresión genética, por ejemplo, Walchli et al., *PLoSOne*, 6 (11): e27930 (2011); o similares.

Se puede obtener una mayor precisión en la determinación de las frecuencias de la cadena del receptor en una variación de la realización anterior, que se puede ver en referencia a las Figuras 2A y 2B en las que el ARN que codifica TCR β se amplifica en una PCR de dos etapas. Como se describe más detalladamente a continuación, el cebador (202) y el conjunto de cebadores (212) se usan en una primera etapa de amplificación para unir el sitio (214) de unión al cebador común a todos los ácidos nucleicos que codifican los TCR β . La Figura 2B ilustra los componentes de una segunda etapa de amplificación para generar más material y para unir los sitios de unión del cebador P5 (222) y P7 (220) que se usan en la formación de conglomerados (mediante PCR de puente) en el protocolo de secuenciación basado en Solexa. El cebador P7 (220) también puede incluir una etiqueta (221) de muestra para multiplexar hasta 96 muestras para secuenciación concurrente en el mismo proceso, por ejemplo, Nota de aplicación Illumina 770-2008-011 (2008). Se puede usar un tipo diferente de etiqueta en el mismo cebador para aumentar la precisión de la determinación de las frecuencias de la cadena del receptor. En este caso, el cebador P7 se modifica para incluir un conjunto de etiquetas muy diverso, de modo que en lugar de 96 etiquetas, el cebador P7 se modifica para tener 10.000 etiquetas distintas, o más. En otras palabras, el cebador P7 es una mezcla de 10.000 o más oligonucleótidos distintos, cada uno de los cuales tiene una región de unión de plantilla idéntica, una secuencia de etiqueta distinta, y una porción de cola 5' idéntica (por ejemplo, (223) en la Figura 2B). Con esta disposición, cualquier subconjunto de ácidos nucleicos que codifique la misma cadena del receptor (por ejemplo, menos de 100) recibirá una etiqueta diferente con alta probabilidad. Tal proceso de emparejamiento de miembros de un pequeño conjunto de ácidos nucleicos con un conjunto mucho más grande de etiquetas para fines de conteo, etiquetado y clasificación es bien conocido y se describe en varias formas en las siguientes referencias, Brenner, patente de los Estados Unidos No. 6.172.214; Brenner et al., patente de los Estados Unidos No. 7.537.897; y Macevicz, publicación Internacional de patente WO US 2005/111242; Brenner et al., *Proc. Natl. Acad. Sci.*, 97: 1665-1670 (2000); Casbon et al., *Nucleic Acids Research*, 39 (12): e81 (2011); Fu et al., *Proc. Natl. Acad. Sci.*, 108: 9026-9031 (2011). La construcción de conjuntos de etiquetas de oligonucleótidos de hibridación cruzada mínimamente, o etiquetas con otras propiedades útiles, se describen en los siguientes ejemplos de referencias: Brenner, patente de los Estados Unidos No. 6.172.214; Morris et al., publicación de la patente de los Estados Unidos 2004/0146901; Mao et al., publicación de la patente de los Estados Unidos No. 2005/0260570; y similares. Preferiblemente, el conjunto de etiquetas debe ser al menos 100 veces (o más)

del tamaño del conjunto de ácidos nucleicos a marcar si todos los ácidos nucleicos van a recibir una etiqueta única con alta probabilidad. Para cadenas de receptores inmunes, en una realización, el número de etiquetas distintas está en el intervalo de 10.000 a 100.000; en otra realización, el número de etiquetas distintas está en el intervalo de 10.000 a 50.000; y en otra realización, el número de etiquetas distintas está en el intervalo de 10.000 a 20.000. Como se describe en Brenner, patente de los Estados Unidos No. 6.172.214, tales combinaciones grandes de etiquetas de oligonucleótidos pueden sintetizarse mediante métodos combinatorios; alternativamente, los cebadores que contienen etiquetas únicas pueden ser sintetizados individualmente mediante métodos no combinatorios, tales como los descritos por Cleary et al., *Nature Methods*, 1: 241-248 (2004); York et al., *Nucleic Acids Research*, 40 (1): e4 (2012); LeProust et al., *Nucleic Acids Research*, 38 (8): 2522-2540 (2010); y similares.

En un caso, el método anterior puede llevarse a cabo mediante las siguientes etapas: (a) obtención de una muestra que contenga células T; (b) determinación de secuencias de nucleótidos de cadenas de TCR α de células T a partir de la muestra, teniendo cada cadena de TCR α una frecuencia de aparición en la muestra; (c) determinación de secuencias de nucleótidos de cadenas de TCR β de células T a partir de la muestra, teniendo cada cadena de TCR β una frecuencia de aparición en la muestra; y (d) identificación de cadenas de TCR α y cadenas de TCR β emparejadas como aquellas que tienen la misma frecuencia dentro de la muestra. Las frecuencias de las respectivas cadenas de TCR α y cadenas de TCR β pueden determinarse a partir de las tabulaciones de los ácidos nucleicos codificadores, o clonotipos. Alternativamente, las frecuencias de las respectivas cadenas de TCR α y cadenas de TCR β pueden determinarse a partir de las tabulaciones de polipéptidos codificados por los clonotipos. Como se mencionó anteriormente, las frecuencias de clonotipos pueden determinarse contando clonotipos directa o indirectamente usando un esquema de etiquetado como se describió anteriormente.

La Figura 1B ilustra la identificación de subunidades de receptores de emparejamiento que pueden aplicarse a TCR o BCR y que pueden usarse incluso cuando las frecuencias de receptores entre cadenas de subunidades son cercanas o indistinguibles, ya sea por error experimental o de otro tipo. Comenzando con una muestra que contiene linfocitos (149), que pueden ser células T o células B, los subconjuntos se forman separando o dividiendo la muestra en una pluralidad de subconjuntos (152), 1 a K (en la figura). En algunas realizaciones, solo se analiza una porción del subconjunto K; por lo tanto, no es necesario formar realmente todos los subconjuntos de K. Se pueden formar subconjuntos de solo la porción que realmente se está analizando. Por ejemplo, si la muestra tiene un volumen de 100 μ L y K = 10, pero solo se va a analizar una porción que consta de 20 subconjuntos, entonces solo se deben formar veinte subconjuntos de 1 μ L. De cada subconjunto (152) se secuencian los ácidos nucleicos que codifican cada cadena diferente del receptor inmune (TCR α y TCR β que se muestran en el subconjunto 1), formando así pares de listas, por ejemplo, (162), (164), (166) y (168) para los subconjuntos 1, 2 ... K-1, K, respectivamente. Cada par de tales listas contiene una primera lista de secuencias de nucleótidos de una primera cadena del receptor inmune, por ejemplo, lista (154) para TCR α del subconjunto 1, y una segunda lista de secuencias de nucleótidos de una segunda cadena del receptor inmune, por ejemplo, la lista (156) para TCR β del subconjunto 1. En un caso, el número de subconjuntos, K, es un número en el intervalo de 5 a 500; en otra realización, K es un número en el intervalo de 10 a 100; en otra realización, K es un número en el intervalo de 20 a 50. En algunos casos, una porción de subconjuntos analizados es de 10 o menos subconjuntos; en otras realizaciones, una porción de subconjuntos analizados es de 20 o menos subconjuntos; en otras realizaciones, una porción de los subconjuntos analizados es al menos cinco por ciento de los subconjuntos; en otras realizaciones, una porción de los subconjuntos analizados es al menos diez por ciento de los subconjuntos; en otras realizaciones, una porción de los subconjuntos analizados es al menos veinte por ciento de los subconjuntos.

Cada tipo de linfocito en la muestra, por ejemplo, linfocito (150), está presente en la muestra en una frecuencia particular. La distribución de linfocitos en los subconjuntos se aproxima fácilmente mediante un modelo binomial; por lo tanto, para un linfocito arbitrario (por ejemplo (150)) que tiene un clonotipo particular, (a) su frecuencia en la muestra, (b) el número total de linfocitos en la muestra y (c) el número de subconjuntos puede estar relacionado con la expectativa de encontrar al menos uno de los linfocitos particulares en una fracción predeterminada de subconjuntos. Esta relación se puede expresar de la siguiente manera: $r = (1-f)^{(N/K)}$, en la que r es la fracción de subconjuntos que contiene al menos uno de los linfocitos particulares, f es la frecuencia del linfocito particular en la muestra, N es el número total de linfocitos en la muestra, y K es el número de subconjuntos. Por lo tanto, si se establece $r = \frac{1}{2}$ y se toma N como una constante, entonces se pueden seleccionar valores sucesivos de K para que los linfocitos de diferentes frecuencias estén presentes en aproximadamente la mitad de los subconjuntos. Se podrían seleccionar otros valores de r, pero $r = \frac{1}{2}$ proporciona resultados con la potencia estadística más alta, por lo que se prefiere el valor $r \sim \frac{1}{2}$. Una vez que se obtienen dichas listas, se examinan para identificar pares de primera y segunda secuencias de nucleótidos que ocurren juntas en un subconjunto o están ausentes en un subconjunto. A modo de ejemplo, los miembros del par (158) aparecen en las listas (164) del subconjunto 2 y en las listas (166) del subconjunto K-1, pero ninguno de los miembros del par aparece en las listas (162) o (168) de subconjuntos 1 y K, ya sea solos o juntos. Esto, por supuesto, refleja la presencia o ausencia del linfocito particular que está en los subconjuntos 2 y K-1, pero está ausente en los subconjuntos 1 y K, tal como el linfocito (150). Tal patrón confirma que los miembros del par (158) van juntos y corresponden a las cadenas de un receptor inmune funcional. Otros linfocitos en la muestra (149) pueden estar presentes en aproximadamente la misma frecuencia, tal como los linfocitos (153). Sin embargo, la probabilidad de que ocurra al menos uno de los linfocitos (153) en exactamente los mismos subconjuntos que los linfocitos (150) es extremadamente baja, especialmente si r es aproximadamente la mitad y la porción de los subconjuntos K analizados está en el intervalo de 10 a 20 o más.

En un caso, la primera y segunda cadenas de linfocitos coincidentes de una sucesión de clases de frecuencia pueden determinarse llevando a cabo el proceso anterior repetidamente para diferentes valores de K. Por ejemplo, una muestra de 1 mL de sangre periférica de un individuo normal contiene aproximadamente $1-4,8 \times 10^6$ linfocitos de los cuales aproximadamente 10-15 por ciento son células B, aproximadamente 70-85 por ciento son células T y aproximadamente 10 por ciento son células NK; por lo tanto, la muestra de 1 mL puede contener de aproximadamente 7×10^5 a aproximadamente 4×10^6 células T. Si el número de linfocitos T en una muestra de 1 mL es $N = 10^6$, entonces las cadenas de los TCR coincidentes de las células T de las siguientes frecuencias se emparejan identificando las que aparecen juntas en el cincuenta por ciento de los subconjuntos y en absoluto en el otro cincuenta por ciento de los subconjuntos :

Frecuencia	Número de subconjuntos	Volumen (μL)
0,001	1432	0,7
0,0005	722	1,4
0,0001	144	6,9
0,00005	72	13,9

Como se mencionó anteriormente, no todos los subconjuntos en una frecuencia particular necesitan ser analizados. Si hay una gran cantidad de linfocitos que tienen frecuencias cercanas o cercanas a una frecuencia seleccionada, por ejemplo, $f = 0,001$, todos pueden resolverse tomando una porción cada vez más grande del número total de subconjuntos hasta que cada par que aparecen juntos en el cincuenta por ciento de los subconjuntos pueda distinguirse de todos los demás pares con la misma frecuencia. Esto se debe a que la probabilidad de que dos linfocitos diferentes ocurran exactamente en los mismos subconjuntos del cincuenta por ciento se vuelve infinitesimal a medida que aumenta la porción de los subconjuntos.

Identificación de cadenas de receptores de células T emparejadas y no emparejadas con selección específica de antígeno

En algunos casos, el método está dirigido a identificar linfocitos específicos de antígeno por uno o un par de cadenas de receptores inmunes. Las células T específicas de antígeno pueden identificarse mediante por uno o un par de cadenas de receptores inmunes, tales como $\text{TCR}\alpha$ o $\text{TCR}\beta$, o $\text{TCR}\alpha$ y $\text{TCR}\beta$ juntas; o $\text{TCR}\delta$, o $\text{TCR}\gamma$, o $\text{TCR}\delta$ y $\text{TCR}\gamma$ juntas. En algunas realizaciones, la secuencia de nucleótidos que codifica una única cadena del receptor inmune, tal como $\text{TCR}\beta$, se usa para identificar células T específicas de antígeno. Del mismo modo, las células B específicas de antígeno pueden identificarse mediante una o un par de cadenas de receptores inmunes, tales como la cadena pesada de inmunoglobulina (IgH), la cadena ligera de inmunoglobulina (IgL) o IgH e IgL juntas; o IgH e IgL lambda ($\text{IgL}\lambda$) juntas; o IgH e IgL kappa ($\text{IgL}\kappa$) juntas. Algunas veces, tales secuencias de nucleótidos se denominan en el presente documento "clonotipo", aunque los clonotipos también pueden ser pares ordenados de secuencias de nucleótidos específicas para una célula T particular o unan célula B particular, tales como las secuencias de nucleótidos que codifican las cadenas $\text{TCR}\alpha$ y $\text{TCR}\beta$ de la célula T, que puede representarse (por ejemplo) como (S_α , S_β) o una notación similar, en la que S_α es una secuencia de un segmento de $\text{TCR}\alpha$ y S_β es una secuencia de un segmento de $\text{TCR}\beta$, y como un par son un clonotipo de la célula de la que se originan. Del mismo modo, se puede usar una notación similar para pares ordenados de cadenas de receptores de células B, por ejemplo, (S_H , S_L) para un par de secuencias que forman un clonotipo de células B.

Las características de algunos casos se ilustran en la Figura 1C, que aunque ilustrada en términos de cadenas receptoras de células T, los expertos en la materia entenderán que es aplicable a las cadenas receptoras de células B de acuerdo con lo previsto por los inventores. A una muestra de tejido (170) que comprende células T (173) se le agrega antígeno (171) en condiciones de interacción para que las células T específicas para antígeno (171) puedan interactuar con el antígeno (171). Dicha interacción puede ser directa o indirecta. Las interacciones directas incluyen la unión del antígeno (171) a células T específicas de antígeno, la unión de conjugados péptido antigénico-multímero a células T específicas de antígeno, y similares. Los conjugados de péptido-multímero, tal como los tetrámeros, son reactivos bien conocidos para los expertos en la materia, por ejemplo, Bousso, *Microbes Infect.* 2 (4): 425-429 (2000); Klenerman et al., *Nature Reviews Immunol.*, 2 (4): 263-272 (2002); y similares. Las interacciones indirectas incluyen la presentación de antígeno o péptidos antigénicos a células T específicas de antígeno por células presentadoras de antígeno, tales como células dendríticas, APC artificiales y similares. En algunas interacciones, las células T específicas de antígeno pueden convertirse en células T activadas que pueden proliferar y/o desarrollar o expresar marcadores de activación, los cuales proporcionan medios para seleccionar y/o enriquecer células T específicas de antígeno usando técnicas convencionales. El antígeno (171) puede comprender una amplia variedad de compuestos o composiciones como se discute más detalladamente a continuación. Las proteínas y péptidos derivados de una o más proteínas son de especial interés, particularmente cuando las proteínas están asociadas con cánceres o

enfermedades infecciosas, tales como infecciones bacterianas o virales. El antígeno (171) puede combinarse, exponerse o agregarse a una muestra de tejido (170) en una variedad de formas conocidas en la técnica, por ejemplo, Berzofsky et al., J. Clin. Investigación, 113: 1515-1525 (2004). Después de combinar el antígeno (171) con la muestra de tejido (170) en una mezcla de reacción, las células T específicas de antígeno (173) y las células T no específicas de antígeno se exponen por igual al antígeno (171) con el que interactúan directa o indirectamente.

En algunas realizaciones, las células T específicas de antígeno (173) se activan, posiblemente después de un período de incubación con el antígeno (171). Un período de incubación puede variar ampliamente. En algunas realizaciones, la incubación puede ser durante un intervalo de unos pocos minutos (por ejemplo, 10 minutos) a una hora o más; en otras realizaciones, la incubación puede ser durante un intervalo de unas pocas horas (por ejemplo, 2 horas) a 8 o más horas. En otras realizaciones, las células T específicas de antígeno (173) interactúan con el antígeno uniéndose a o formando complejos con antígeno o reactivos antigénicos, tales como conjugados péptido antigénico-multímero, de modo que la activación puede no tener lugar. Una etapa de exposición puede incluir la etapa de incubar una muestra de tejido con un antígeno. Por ejemplo, en el caso de un antígeno proteico y una muestra de tejido que comprende PBMC, una etapa de exposición puede incluir combinar la muestra de tejido con péptidos derivados del antígeno proteico de modo que las células dendríticas en la muestra de tejido presenten los péptidos a células T específicas de antígeno en la muestra de tejido que, a su vez, interactúan con las células dendríticas presentadoras de antígeno y se activan. Después de exponer las células T (173) al antígeno para que las células T específicas del antígeno interactúen con el antígeno, se seleccionan las células T (172) específicas del antígeno en función de alguna característica resultante de la interacción, tal como la unión péptido antigénico-multímero, marcadores de activación inducidos, proliferación de células T, o similares. Como se mencionó anteriormente, la etapa de seleccionar células T (172) específicas de antígeno puede ser alternativamente una etapa de enriquecer células T específicas de antígeno a partir de la mezcla de reacción, y/o una etapa de separación de células T específicas de antígeno de la mezcla de reacción, y/o una etapa de aislar células T específicas de antígeno de la mezcla de reacción. Después de que las células T (172) específicas de antígeno se enriquecen, separan y/o aíslan, sus clonotipos se determinan secuenciando un segmento predeterminado de un ácido nucleico recombinado que codifica un porción de un receptor inmune, tal como TCR β y/o TCR α .

Un segmento predeterminado elegido puede variar ampliamente; en algunas realizaciones, abarca toda o una parte de una región V (D)J, de modo que los clonotipos basados en ella tienen una diversidad máxima para la identificación única de clones de células T. La determinación de los clonotipos se describe más completamente a continuación, pero brevemente, los ácidos nucleicos recombinados que codifican uno o más receptores inmunes seleccionados (tales como TCR β como se muestra en la Figura 1C) se secuencian (por ejemplo, aislando espacialmente moléculas de los mismos, amplificando tales moléculas, y llevando a cabo etapas de secuenciación mediante una química de secuenciación de alto rendimiento, tal como la disponible con secuenciadores de ADN comerciales de próxima generación). Como resultado de estas etapas de secuenciación, se producen lecturas de secuencia (176) que se utilizan para determinar los clonotipos y las frecuencias de clonotipos de las células T específicas de antígeno. Los clonotipos y las frecuencias de clonotipos también se determinan para células T de la muestra (174) de tejido a partir de lecturas de secuencia (178) o para células T (175) no específicas de antígeno a partir de lecturas de secuencia (177). Las células T no específicas de antígeno pueden obtenerse a partir de un procedimiento de clasificación bidireccional (por ejemplo, usando FACS o MACS) basado en células T marcadas de acuerdo con una interacción, tal como, una interacción de células T específicas de antígeno con marcadores fluorescentes multímeros de péptidos antigénicos. Estos datos pueden analizarse luego para identificar clonotipos asociados con células T específicas de antígeno, por ejemplo, como se describe en el Ejemplo a continuación y las Figuras 6-9. Brevemente, las células T específicas de antígeno pueden estar asociadas con frecuencias de clonotipo que aumentan en la población seleccionada de células T en relación con las frecuencias del mismo clonotipo en poblaciones de células T no específicas de antígeno o en la población de células T en una muestra de tejido (170).

Los ejemplos de etapas para implementar este método (es decir, para determinar los clonotipos asociados con células T específicas de antígeno en una muestra de tejido) pueden incluir lo siguiente: (a) exponer las células T de la muestra a un antígeno para que las células T específicas para el antígeno interactúen con el antígeno; (b) secuenciar ácidos nucleicos recombinados que codifican una cadena del receptor de células T o una porción de la misma a partir de una muestra de células T de la muestra de tejido para proporcionar lecturas de secuencia a partir de las cuales se determinan los clonotipos; (c) aislar células T específicas de antígeno de la muestra de tejido basándose en su interacción con el antígeno; (d) secuenciar ácidos nucleicos recombinados que codifican una cadena del receptor de células T o una porción de la misma a partir de una muestra de células T específicas de antígeno aisladas para proporcionar lecturas de secuencia a partir de las cuales se determinan los clonotipos; y (e) determinar las células T específicas de antígeno en la muestra de tejido como células T cuyas frecuencias de clonotipo aumentan en la muestra de células T aisladas en relación con las frecuencias de los mismos clonotipos en una muestra de células T en la muestra de tejido. En algunas realizaciones, una etapa de exposición puede llevarse a cabo haciendo reaccionar en condiciones de interacción un antígeno con una muestra de tejido; en otras realizaciones más, una etapa de exposición puede llevarse a cabo mediante reacción en condiciones de activación de un antígeno con una muestra de tejido. Como se mencionó anteriormente, la etapa de exponer para esta y otras realizaciones puede variar ampliamente, y su implementación puede depender de la naturaleza de la muestra de tejido y la naturaleza del antígeno, así como de otros factores. Por ejemplo, si una muestra de tejido incluye células presentadoras de antígeno, como las células dendríticas, entonces la exposición puede incluir la adición de un antígeno, tal como una proteína, directamente a la

muestra de tejido, o puede incluir la producción de material antigénico a partir de un antígeno de interés seguido de la adición del material antigénico. Una activación de células T más eficiente con un antígeno proteico, por ejemplo, se puede lograr exponiendo una muestra de tejido a un conjunto de péptidos superpuestos derivados del antígeno proteico de interés, utilizando técnicas convencionales. Alternativamente, se pueden usar composiciones presentadoras de antígeno artificiales en la etapa de exposición o su equivalente, por ejemplo, Oelke et al., *Nature Medicine*, 9 (5): 619-624 (2003). La etapa de exponer las células T en una muestra de tejido puede incluir la exposición de tales células T a células enteras que contienen antígeno, a células modificadas por genes que expresan antígeno, a proteína completa, a péptidos derivados de un antígeno proteico, a vectores virales que expresan un antígeno, a células dendríticas modificadas con antígeno o cargadas. En algunas realizaciones, una muestra de tejido es una muestra de sangre; en otras realizaciones, una muestra de tejido es una muestra de células mononucleares de sangre periférica (PBMC) derivadas de sangre periférica usando técnicas convencionales. En algunas realizaciones, la etapa de exposición puede llevarse a cabo haciendo reaccionar en condiciones de activación una muestra de tejido que comprende células T con un antígeno, en la que se describen anteriormente diversas condiciones de activación. En vista de la amplia variedad de muestras de tejido y antígenos, la etapa de exposición puede llevarse a cabo alternativamente mediante una etapa de reacción en condiciones de activación de una muestra de tejido que comprende células T con un antígeno.

Los ejemplos de etapas adicionales para implementar el método anterior pueden comprender: (a) hacer reaccionar en condiciones de activación una muestra de tejido que comprende células T con un antígeno; (b) clasificar la muestra de tejido de células T activadas y células T no activadas; (c) secuenciar ácidos nucleicos recombinados que codifican una cadena del receptor de células T o una porción de la misma a partir de una muestra de células T de las células T activadas para proporcionar lecturas de secuencia a partir de las cuales se determinan los clonotipos; (d) secuenciar ácidos nucleicos recombinados que codifican una cadena del receptor de células T o una porción de la misma a partir de una muestra de células T de las células T no activadas para proporcionar lecturas de secuencia a partir de las cuales se determinan los clonotipos; y (e) determinar las células T específicas de antígeno en la muestra de tejido como células T cuyas frecuencias de clonotipo aumentan en la muestra de células T activadas en relación con las frecuencias de los mismos clonotipos en la muestra de tejido o en una muestra de células T no activadas. Asimismo, los ejemplos de etapas para implementar el método anterior pueden comprender: (a) hacer reaccionar en condiciones de interacción una muestra de tejido que comprende células T con un antígeno; (b) clasificar las células T de la muestra de tejido en un primer subconjunto de células T que forman complejos con el antígeno o sus reactivos antigénicos y en un segundo subconjunto de células T que no forman complejos con el antígeno o sus reactivos antigénicos; (c) secuenciar ácidos nucleicos recombinados que codifican una cadena del receptor de células T o una porción de la misma a partir de una muestra del primer subconjunto para proporcionar lecturas de secuencia a partir de las cuales se determinan los clonotipos; (d) secuenciar ácidos nucleicos recombinados que codifican una cadena del receptor de células T o una porción de la misma a partir de una muestra de células T de la muestra de tejido o el segundo subconjunto para proporcionar lecturas de secuencia a partir de las cuales se determinan los clonotipos; y (e) determinar las células T específicas de antígeno en la muestra de tejido tal como las células T cuyas frecuencias de clonotipo aumentan en la muestra de células T del primer subconjunto en relación con las frecuencias de los mismos clonotipos en la muestra de tejido o en una muestra de células T del segundo subconjunto. Como se usa en el presente documento, el término "reactivos antigénicos" significa reactivos derivados de un antígeno diseñado para unirse o formar complejos con células T cuyos TCR son específicos para el antígeno. Los ejemplos de reactivos antigénicos incluyen, pero no se limitan a, multímeros conjugados con péptidos derivados de un antígeno.

En algunos casos, el método anterior para determinar las células T específicas de antígeno en una muestra de tejido puede llevarse a cabo mediante etapas que comprenden: (a) hacer reaccionar, bajo condiciones de activación en una mezcla de reacción, una muestra de tejido que comprende células T con un antígeno o reactivos antigénicos de los mismos; (b) secuenciar ácidos nucleicos recombinados que codifican una cadena del receptor de células T o una porción de la misma a partir de una muestra de células T de la mezcla de reacción antes de la adición del antígeno a la mezcla de reacción para proporcionar lecturas de secuencia a partir de las cuales se determinan los clonotipos; (c) incubar la mezcla de reacción después de la adición del antígeno o reactivo antigénico de la misma durante un intervalo predeterminado; (d) secuenciar ácidos nucleicos recombinados que codifican una cadena del receptor de células T o una porción de la misma a partir de una muestra de células T de la mezcla de reacción incubada para proporcionar lecturas de secuencia a partir de las cuales se determinan los clonotipos; (e) determinar las células T específicas de antígeno en la muestra de tejido como células T cuyas frecuencias de clonotipo aumentan en la mezcla de reacción incubada en relación con las frecuencias de los mismos clonotipos en la mezcla de reacción antes de la adición del antígeno. En algunos casos, un intervalo predeterminado para la incubación suele ser mayor de ocho horas; en otros casos, un intervalo predeterminado puede ser mayor de 24 horas; en realizaciones adicionales, un intervalo predeterminado puede estar dentro de un intervalo de 8 horas a 72 horas.

Los ejemplos de etapas para implementar este método (es decir, para determinar clonotipos asociados con células B específicas de antígeno en una muestra de tejido) pueden incluir lo siguientes: (a) exponer las células T de la muestra a un antígeno para que las células T específicas para el antígeno interactúen con el antígeno; (b) secuenciar ácidos nucleicos recombinados que codifican una cadena del receptor de células T o una porción de la misma a partir de una muestra de células B de la muestra de tejido para proporcionar lecturas de secuencia a partir de las cuales se determinan los clonotipos; (c) aislar células B específicas de antígeno de la muestra de tejido basándose en su interacción con el antígeno; (d) secuenciar ácidos nucleicos recombinados que codifican una cadena del receptor de

- células T o una porción de la misma a partir de una muestra de células B específicas de antígeno aisladas para proporcionar lecturas de secuencia a partir de las cuales se determinan los clonotipos; y (e) determinar las células B específicas de antígeno en la muestra de tejido como células B cuyas frecuencias de clonotipo aumentan en la muestra de células B aisladas en relación con las frecuencias de los mismos clonotipos en una muestra de células B en la muestra de tejido. En algunos casos, una etapa de exposición puede llevarse a cabo haciendo reaccionar en condiciones de interacción un antígeno con una muestra de tejido; en otros casos, una etapa de exposición puede llevarse a cabo mediante reacción en condiciones de activación de un antígeno con una muestra de tejido. Como se mencionó anteriormente, la etapa de exposición puede variar ampliamente, y su implementación puede depender de la naturaleza de la muestra de tejido y la naturaleza del antígeno, así como de otros factores.
- Los ejemplos de etapas adicionales para implementar el método anterior para células B pueden comprender: (a) hacer reaccionar en condiciones de unión una muestra de tejido que comprende células B con un antígeno; (b) clasificar a partir de muestras de tejido, células B unidas a antígeno y células B no unidas; (b) secuenciar ácidos nucleicos recombinados que codifican una cadena del receptor de células B o una porción de la misma a partir de una muestra de células B de las células B unidas a antígeno para proporcionar lecturas de secuencia a partir de las cuales se determinan los clonotipos; (c) secuenciar ácidos nucleicos recombinados que codifican una cadena del receptor de células B o una porción de la misma a partir de una muestra de células B de las células B no unidas para proporcionar lecturas de secuencia a partir de las cuales se determinan los clonotipos; y (d) determinar las células B específicas de antígeno en la muestra de tejido como células B cuyas frecuencias de clonotipo aumentan en la muestra de células B unidas a antígeno en relación con las frecuencias de los mismos clonotipos en la muestra de tejido o en una muestra de células B no unidas. Asimismo, los ejemplos de etapas para implementar el método anterior pueden comprender: (a) hacer reaccionar bajo condiciones de interacción una muestra de tejido que comprende células B con un antígeno; (b) clasificar las células B de la muestra de tejido en un primer subconjunto de células B que forman complejos con el antígeno o reactivos antigénicos del mismo y en un segundo subconjunto de células B que no forman complejos con el antígeno o reactivos antigénicos del mismo; (b) secuenciar ácidos nucleicos recombinados que codifican una cadena del receptor de células B o una porción de la misma a partir de una muestra del primer subconjunto para proporcionar lecturas de secuencia a partir de las cuales se determinan los clonotipos; (c) secuenciar ácidos nucleicos recombinados que codifican una cadena del receptor de células B o una porción de la misma a partir de una muestra de células B de la muestra de tejido o el segundo subconjunto para proporcionar lecturas de secuencia a partir de las cuales se determinan los clonotipos; y (d) determinar las células B específicas de antígeno en la muestra de tejido como células B cuyas frecuencias de clonotipo aumentan en la muestra de células B del primer subconjunto en relación con las frecuencias de los mismos clonotipos en la muestra de tejido o en una muestra de células B del segundo subconjunto. Como se usa en el presente documento, el término "reactivos antigénicos" significa reactivos derivados de un antígeno diseñado para unirse a, o formar complejos con células B cuyos BCR son específicos para el antígeno.
- En algunos casos, el método anterior para determinar las células B específicas de antígeno en una muestra de tejido puede llevarse a cabo mediante las etapas que comprenden: (a) hacer reaccionar bajo condiciones de unión en una mezcla de reacción una muestra de tejido que comprende células B con un antígeno o reactivos antigénicos del mismo; (b) secuenciar ácidos nucleicos recombinados que codifican una cadena del receptor de células B o una porción de la misma a partir de una muestra de células B de la mezcla de reacción antes de la adición del antígeno a la mezcla de reacción para proporcionar lecturas de secuencia a partir de las cuales se determinan los clonotipos; (c) incubar la mezcla de reacción después de la adición del antígeno o reactivo antigénico del mismo durante un intervalo predeterminado; (d) secuenciar los ácidos nucleicos recombinados que codifican una cadena del receptor de células B o una porción de la misma a partir de una muestra de células B de la mezcla de reacción incubada para proporcionar lecturas de secuencia a partir de las cuales se determinan los clonotipos; (d) determinar las células B específicas de antígeno en la muestra de tejido como células B cuyas frecuencias de clonotipo aumentan en la mezcla de reacción incubada en relación con las frecuencias de los mismos clonotipos en la mezcla de reacción antes de la adición de antígeno. En algunos casos, un intervalo predeterminado para la incubación suele ser mayor de ocho horas; en otros casos, un intervalo predeterminado puede ser mayor de 24 horas; en otros casos, un intervalo predeterminado puede estar dentro de un intervalo de 8 horas a 72 horas.
- En algunos casos, la etapa de aislar células T específicas de antígeno se puede sustituir con una etapa de separar una muestra de células T específicas de antígeno de la muestra de tejido después de la exposición a un antígeno de interés o una etapa de recuperación específica de células T específicas de antígeno a partir de la muestra de tejido después de la exposición a un antígeno de interés. En algunos casos, dicha etapa de aislamiento puede llevarse a cabo clasificando las células T activadas o que interactúan con el antígeno de una muestra de tejido; asimismo, las células T no específicas de antígeno y/o las células T inactivadas pueden clasificarse a partir de una muestra de tejido. Dichas etapas pueden llevarse a cabo mediante una variedad de métodos que incluyen, pero no se limitan a, (i) reactivos de tinción del multímero péptido-MHC (tales como tetrámeros, pentámeros o similares), seguido de clasificación, paneo o bien captura de complejos entre dichos reactivos y células T específicas de antígeno, (ii) clasificación o paneo en función de marcadores de activación, tales como CD137, CD154 u otros (descritos más detalladamente a continuación), o (iii) proliferación (y, por lo tanto, por ejemplo, un aumento en la frecuencia) de células T específicas de antígeno sobre las células T no específicas de antígeno. Por lo tanto, dicha etapa de aislamiento puede comprender una etapa de aislamiento de células T activadas; o una etapa para separar las células T activadas de la muestra de tejido. Los marcadores de activación de células T, como se indicó anteriormente, se pueden usar para clasificar, panear o bien capturar células T activadas, utilizando técnicas convencionales. En general,

se presenta una etapa para obtener una muestra de células T a partir de una agrupación de células T derivadas de la muestra de tejido, cuya agrupación se enriquece en células T específicas de antígeno y/o células T activadas. En algunos casos, las células T con un marcador de activación pueden clasificarse o aislarse usando un compuesto de unión, tal como un anticuerpo, que se une específicamente al marcador de activación y que puede marcarse directa o indirectamente de acuerdo con métodos convencionales, por ejemplo, FACS, separación basada en perlas magnéticas o técnicas similares.

En otra aplicación, la inmunogenicidad de las células T puede medirse en las siguientes etapas: (a) hacer reaccionar en condiciones de activación una muestra de tejido que comprende células T con un antígeno o un reactivo antigénico del mismo; (b) secuenciar ácidos nucleicos recombinados que codifican una cadena del receptor de células T o una porción de la misma a partir de una muestra de células T de la muestra de tejido expuesta al antígeno o reactivos antigénicos del mismo para proporcionar lecturas de secuencia a partir de las cuales se determinan los clonotipos; (c) aislar células T activadas de la muestra de tejido; (d) secuenciar ácidos nucleicos recombinados que codifican una cadena del receptor de células T o una porción de la misma a partir de una muestra de células T activadas aisladas de la muestra de tejido para proporcionar lecturas de secuencia a partir de las cuales se determinan los clonotipos; y (e) cuantificar la inmunogenicidad del antígeno en función de los aumentos en las frecuencias de clonotipos en la muestra de células T aisladas expuestas al antígeno con respecto a las frecuencias de los mismos clonotipos en la muestra de tejido antes de dicha etapa de aislamiento. Los ejemplos de funciones de aumentos en las frecuencias de clonotipos incluyen un promedio de aumentos entre las células T específicas de antígeno aisladas; otro ejemplo de función de los aumentos incluye una distancia promedio de puntos de datos de clonotipos que registran aumentos en la frecuencia desde la diagonal en gráficos tales como aquellos de las Figuras 6, 7 u 8. Aún otra medida de inmunogenicidad de células T incluye cualquiera de varias medidas de similitud de un perfil de clonotipo de células T de la muestra de tejido expuesta antes del aislamiento y un perfil de clonotipo de células T de una muestra de células T aisladas (o separadas) de la muestra de tejido, tal como se describe en Faham et al., publicación internacional de patente WO/2013/036459. En este caso, los antígenos de particular interés son las proteínas terapéuticas, tales como los anticuerpos terapéuticos.

En un caso, una medida de similitud para usar con estas realizaciones de la invención es una función que varía monotónicamente que mapea (o es capaz de mapear mediante una transformación simple) al menos dos conjuntos de mediciones de frecuencia de clonotipo (por ejemplo, dos perfiles de clonotipo basados en secuencias) con el intervalo unitario $[0,1]$. Las transformaciones simples incluyen, pero no se limitan a, cualquier transformación lineal de variables dependientes, transformaciones logarítmicas, tales como $y_{ij} = \ln(n_{ij} + 1)$ (en la que n_{ij} es el número de clonotipo i en la muestra j), o similares. Un valor de cero significa que no hay similitud entre los perfiles de clonotipo y un valor de uno significa que dos perfiles de clonotipo son estadísticamente idénticos. Se describen ejemplos de medidas de similitud que pueden implementarse en estas realizaciones en Legendre y Legendre, *Numerical Ecology* (Elsevier, 1998); Magurran, *Measurement of Biological Diversity* (Wiley-Blackwell, 2003); Wolda, *Oecologia* (Berl), 50: 296-302 (1981); y referencias similares. Tales medidas de similitud incluyen, pero no se limitan a, el índice de Czekanowski, el coeficiente de Dice, el índice de teoría de la información de Horn, la métrica de Canberra, el índice de Morisita, el índice de similitud de Kaczenski, el índice de Sorensen, el índice de Jacquard, el índice de Bray-Curtis y similares. En un caso, las medidas de similitud son métricas de similitud; o en otras palabras, las medidas de similitud empleadas tienen propiedades de una medida de distancia, tal como, (i) el valor de la medida es siempre no negativo, (ii) la medida es cero si y solo si las medidas del perfil de clonotipo son idénticas, (iii) el valor de la medida es invariante con respecto al orden de las mediciones del perfil de clonotipo (a veces expresado como $d(x, y) = d(y, x)$), (iv) la desigualdad del triángulo se mantiene con respecto a tres medidas diferentes del perfil de clonotipo. En otro caso, las medidas de similitud pueden ser coeficientes de correlación (sujetos a una transformación simple, por ejemplo, tomando su valor absoluto, ajustando su valor o similar, de modo que su valor esté restringido al intervalo de la unidad). Los ejemplos de coeficientes de correlación incluyen, pero no se limitan a, el coeficiente de correlación de momento del producto de Pearson y correlaciones de intervalo, tales como el coeficiente de correlación de intervalo de Spearman, el coeficiente de correlación de intervalo tau de Kendall y similares. En una realización, se emplea un índice de Morisita-Horn (C_{12}) (que incluye el índice Morisita-Horn con una transformación logarítmica), como se describe en Wolda (citado anteriormente), con las realizaciones.

Otra instancia para identificar pares de cadenas de receptores inmunes de linfocitos específicos de antígeno se ilustra en la Figura 1D, en la que las mezclas de reacción que contienen células T están expuestas a un único antígeno. Como se indicó anteriormente, aunque ilustrado en términos de cadenas de receptores de células T, los expertos en la materia entenderán que el método ilustrado en la Figura 1D es aplicable a las cadenas de receptores de células B. De manera similar a la realización de la Figura 1B, una muestra de tejido se divide en subconjuntos (152) de 1 a K y se puede seleccionar una porción de los subconjuntos para su análisis. Los intervalos en los valores de K y la porción seleccionada pueden ser los mismos que para la realización de la Figura 1B. En una realización, como anteriormente, las particiones pueden ser partes alícuotas de la muestra de tejido, en las que se proporcionan cantidades aproximadamente iguales de muestra de tejido a cada subconjunto, por ejemplo, distribuyendo cantidades iguales de muestra de tejido a cada una de las mezclas de reacción de K, que pueden estar contenidas por recipientes o reactores, tales como pozos en una placa de múltiples pozos. Las muestras de tejido también se pueden distribuir a una pluralidad de cámaras de separación de K de un dispositivo de microfluído en relación con esto y/o las realizaciones descritas anteriormente. Las células T de cada subconjunto se exponen al antígeno (179) después de lo cual las mezclas de reacción en los recipientes de K se incuban durante un tiempo (por ejemplo, un intervalo

predeterminado) suficiente para que las células T respondan o interactúen con el antígeno, ya sea directamente o en una forma procesada (por ejemplo, un reactivo de antigénico). Dicha respuesta puede incluir formar un complejo estable con antígeno o una forma procesada del mismo, o puede incluir el desarrollo y/o la expresión de marcadores de activación por células T, o puede incluir proliferación por células T específicas para el antígeno. Las células T (180) que interactúan con el antígeno o que responden al antígeno se seleccionan luego y se aíslan (por ejemplo, clasificadas) de cada uno de las cámaras de K, después de lo cual los ácidos nucleicos recombinados que codifican porciones predeterminadas de una o ambas cadenas de TCR se secuencian para proporcionar lecturas de secuencia a partir de las cuales se forman los clonotipos y los perfiles de clonotipos (por ejemplo, 154 y 156). Como anteriormente con la Figura 1B, una vez que se obtienen dichos perfiles, se examinan para identificar pares de la primera y segunda secuencias de nucleótidos da se producen juntas en un subconjunto o están ausentes de un subconjunto. A modo de ejemplo, los miembros de la pareja (181) aparecen en las listas (184) del subconjunto 2 y en las listas (186) del subconjunto K-1, pero ninguno de los miembros de la pareja aparece en las listas (182) o (188) de los subconjuntos 1 y K, ya sea solos o juntos. Como se indicó anteriormente, esto refleja la presencia o ausencia de un linfocito particular, que en esta ilustración se encuentra en los subconjuntos 2 y K-1, pero está ausente de los subconjuntos 1 y K. Tal patrón confirma que los miembros de la pareja (181) van juntos y corresponden a las cadenas de un receptor inmune funcional que es específico para el antígeno, Ag₁ (179).

En algunos casos, el método anterior de determinación de receptores de células T específicas de antígeno en una muestra de tejido puede comprender las siguientes etapas: (a) dividir una muestra de tejido que contiene células T en una pluralidad de subconjuntos; (b) exponer las células T de cada porción de subconjuntos a un antígeno para que se activen las células T específicas para el antígeno; (c) aislar las células T activadas de cada subconjunto de la porción; (d) secuenciar ácidos nucleicos recombinados que codifican las cadenas α del receptor de células T en cada subconjunto de la porción para proporcionar lecturas de secuencia a partir de las cuales se determinan los clonotipos de cadena α ; (e) secuenciar ácidos nucleicos recombinados que codifican cadenas β del receptor de células T en cada subconjunto de la porción para proporcionar lecturas de secuencia a partir de las cuales se determinan los clonotipos de cadena β ; y (f) identificar receptores de células T específicas de antígeno con aquellos pares de clonotipos de cadena α y clonotipos de cadena β para cada subconjunto de la porción (i) ya sea que estén presentes tanto el clonotipo de cadena α como el clonotipo de cadena β en un subconjunto o ninguno está presente en un subconjunto, y (ii) tanto el clonotipo de cadena α como el clonotipo de cadena β están presentes en al menos un subconjunto de la porción y el clonotipo de cadena α y el clonotipo de cadena β no están presentes en al menos un subconjunto de la porción. $\alpha\beta$

Como alternativa, en algunos casos, el método anterior de determinación de receptores de células T específicas de antígeno en una muestra de tejido puede comprender las siguientes etapas: (a) formar una pluralidad de subconjuntos a partir de una muestra de tejido que contiene células T; (b) hacer reaccionar en condiciones de activación las células T de cada subconjunto con un antígeno; (c) aislar las células T específicas de antígeno de cada subconjunto; (d) secuenciar ácidos nucleicos recombinados que codifican cadenas α del receptor de células T en cada subconjunto para proporcionar lecturas de secuencia a partir de las cuales se determinan los clonotipos de cadena α ; (e) secuenciar ácidos nucleicos recombinados que codifican cadenas β del receptor de células T en cada subconjunto para proporcionar lecturas de secuencia a partir de las cuales se determinan los clonotipos de cadena β ; (d) identificar receptores de células T específicas de antígeno con aquellos pares de clonotipos de cadena α y clonotipos de cadena β para cada subconjunto (i) ya sea que estén presentes tanto el clonotipo de cadena α como el clonotipo de cadena β en un subconjunto o ninguno está presente en un subconjunto, y (ii) tanto el clonotipo de cadena α como el clonotipo de cadena β están presentes en al menos un subconjunto y no están presentes el clonotipo de cadena α y el clonotipo de cadena β en la menos un subconjunto. En algunos de estos últimos casos, la pluralidad de subconjuntos formados puede corresponder a una porción de la pluralidad en la que se divide una muestra de tejido en las realizaciones anteriores. En algunos casos, la etapa de formar una pluralidad de subconjuntos puede comprender colocar porciones alícuotas de una muestra de tejido en recipientes de reacción separados. En algunos casos, tales porciones son porciones iguales.

En algunos casos, el método anterior de determinación de receptores de células B específicas de antígeno en una muestra de tejido puede comprender las siguientes etapas: (a) dividir una muestra de tejido que contiene células B en una pluralidad de subconjuntos; (b) exponer las células B de cada uno de una porción de subconjuntos a un antígeno para que las células B específicas para el antígeno se unan al mismo; (c) aislar las células B unidas a cada subconjunto de la porción; (d) secuenciar ácidos nucleicos recombinados que codifican cadenas de IgH en cada subconjunto de la porción para proporcionar lecturas de secuencia a partir de las cuales se determinan los clonotipos de cadena de IgH; (e) secuenciar de ácidos nucleicos recombinados que codifican cadenas de IgL en cada subconjunto de la porción para proporcionar lecturas de secuencia a partir de las cuales se determinan los clonotipos de la cadena de IgL; y (f) identificar receptores de células B específicas de antígeno con aquellos pares de clonotipos de cadena de IgH y clonotipos de cadena de IgL para cada subconjunto de la porción (i) ya sea que estén presentes tanto el clonotipo de cadena de IgH como el clonotipo de cadena de IgL en un subconjunto o ninguno está presente en un subconjunto, y (ii) tanto el clonotipo de cadena de IgH como el clonotipo de cadena de IgL están presentes en al menos un subconjunto de la porción y el clonotipo de la cadena de IgH y el clonotipo de la cadena de IgL no están presentes en al menos un subconjunto de la porción.

Como alternativa, en algunos casos, el método anterior de determinación de receptores de células B específicas de antígeno en una muestra de tejido puede comprender las siguientes etapas: (a) formar una pluralidad de subconjuntos a partir de una muestra de tejido que contiene células B; (b) hacer reaccionar en condiciones de unión a antígeno las células B de cada subconjunto con un antígeno; (c) aislar las células B específicas de antígeno de cada subconjunto (es decir, células B de unión a antígeno); (d) secuenciar ácidos nucleicos recombinados que codifican cadenas de IgH en cada subconjunto para proporcionar lecturas de secuencia a partir de las cuales se determinan los clonotipos de cadena de IgH; (e) secuenciar ácidos nucleicos recombinados que codifican cadenas de IgL en cada subconjunto para proporcionar lecturas de secuencia a partir de las cuales se determinan los clonotipos de la cadena de IgL; (d) identificar los receptores de células B específicas de antígeno con aquellos pares de clonotipos de cadena de IgH y clonotipos de cadena de IgL para cada subconjunto (i) ya sea que estén presentes tanto el clonotipo de cadena de IgH como el clonotipo de cadena de IgL en un subconjunto o ninguno está presente en un subconjunto, y (ii) tanto el clonotipo de cadena de IgH como el clonotipo de cadena de IgL están presentes en al menos un subconjunto y el clonotipo de cadena de IgH y el clonotipo de cadena de IgL no están presentes en al menos un subconjunto. En algunos de estos últimos casos, la pluralidad de subconjuntos formados puede corresponder a una porción de la pluralidad en la que se divide una muestra de tejido en las realizaciones anteriores. En algunos casos, la etapa de formar una pluralidad de subconjuntos puede comprender colocar porciones alícuotas de una muestra de tejido en recipientes de reacción separados. En algunos casos, tales porciones son porciones iguales.

Otra realización para identificar clonotipos de linfocitos específicos de antígeno se ilustra en las Figuras 5A-5B. Como anteriormente, aunque se ilustran en términos de cadenas de receptores de células T, la realización ilustrada en las Figuras 5A-5B los expertos en la materia entenderán que es aplicable a las cadenas de receptores de células B. En las Figuras 5A-5B, se expone una pluralidad de antígenos (500) a las células T en una pluralidad de diferentes mezclas de reacción. En un aspecto, esta realización permite la identificación de células T específicas de antígeno de muestras de tejido escaso, tal como el de un paciente con cáncer cuya muestra de tejido se utilizará para identificar clonotipos para el monitoreo de la enfermedad (por ejemplo, análisis de enfermedad residual mínima (MRD)), para producir un reactivo inmunoterapéutico específico del paciente utilizando células T específicas de antígeno de cáncer, o similares. Los subconjuntos (o mezclas de reacción) (502) desde 1 hasta K (mostrados en las Figuras 5A-5B como 1-8) se forman a partir de una muestra de tejido. El número de antígenos diferentes empleados puede variar ampliamente y, en algunas realizaciones, el número depende de la naturaleza de los antígenos. Para antígenos proteicos, en algunas realizaciones, se puede emplear una pluralidad de antígenos proteicos; en otras realizaciones, una pluralidad de antígenos proteicos puede estar en el intervalo de 2 a 100 antígenos proteicos; en otras realizaciones, una pluralidad puede estar en el intervalo de 2 a 50 antígenos proteicos; en otras realizaciones, una pluralidad puede estar en el intervalo de 2 a 10 antígenos proteicos; en otras realizaciones más, una pluralidad puede estar en el intervalo de 2 a 1000 antígenos proteicos.

Cada antígeno de pluralidad (500) se expone a (o se presenta a) células T de una subpluralidad de mezclas de reacción (502) menor que la pluralidad total de mezclas de reacción de K (en esta ilustración, las subpluralidades son cada una 4). Las selecciones de los subconjuntos de mezclas de reacción en las que se colocan los antígenos están predeterminadas para cada antígeno. En algunas realizaciones, cada antígeno se aplica o se expone a una subpluralidad única de subconjuntos. Es decir, la selección de subconjuntos que componen una subpluralidad correspondiente a un antígeno particular es única para ese antígeno. El tamaño de las subpluralidades puede ser igual o diferente para cada antígeno; pero en algunas realizaciones, el tamaño de las subpluralidades (es decir, el número de subconjuntos en cada una) es el mismo para cada antígeno (que es igual a 4 en las Figuras 5A-5B). En algunas realizaciones, las subpluralidades de los subconjuntos corresponden a una combinación diferente de subconjuntos fuera de la pluralidad (en este caso 8), como se mencionó anteriormente. Por lo tanto, para algunas realizaciones, el número de subconjuntos posibles es el mismo que el número de diferentes combinaciones de subconjuntos R seleccionados del número total de subconjuntos, K. (Por ejemplo, para $R = 4$ y $K = 8$, el número de combinaciones diferentes es $K!/R!$ ($KR!$)).

Una selección de diferentes combinaciones (o subpluralidades) para los antígenos (500) se indica mediante la matriz (506) de + y - que indican qué antígeno está expuesto a las células T de dichos subconjuntos. Como se mencionó anteriormente, la selección de subconjuntos en los que se aplica (o expone) un antígeno está predeterminada; por lo tanto, por ejemplo, el antígeno Ag_1 se aplica a la subpluralidad de los subconjuntos 1, 4, 5 y 7. Una subpluralidad de subconjuntos que están expuestos al antígeno puede variar entre 2 y $K-1$; sin embargo, en algunas realizaciones, el tamaño de la subpluralidad es un número entero igual o más cercano a $K/2$. Como se indicó anteriormente, después de la exposición al antígeno y la incubación opcional, se seleccionan células T específicas de antígeno (504) (por ejemplo, con base en la interacción con un antígeno en la mezcla de reacción) y se generan perfiles de clonotipo para ácidos nucleicos recombinados que codifican una cadena de TCR seleccionada o una porción de la misma (como se ilustra para el subconjunto 1), que permite identificar y/o aislar su célula T correspondiente. Antes de la exposición, se puede tomar una muestra de células T de los subconjuntos de muestras de tejido (por ejemplo, 510). Los ácidos nucleicos recombinados que codifican clonotipos del mismo segmento de TCR se secuencian tanto en la muestra (510) como en la muestra (511) para producir lecturas de secuencia (514) y (515) a partir de las cuales se determinan los clonotipos y las frecuencias de clonotipos. Las frecuencias de clonotipos que aumentan en los grupos seleccionados de células T (ilustradas en las listas (520) en la Figura 10B) corresponden a células T específicas para antígenos (por ejemplo, Ag_1 o Ag_4 en la mezcla de reacción 1). Puede ser un clonotipo específico de antígeno identificado observando un clonotipo que aumenta en frecuencia en cada mezcla de reacción de un antígeno dado.

Por ejemplo, en las Figuras 5A-5B, se observa que el mismo clonotipo (518a, 518b, 518c y 518d) ha aumentado en frecuencia dentro de las mezclas de reacción 1, 4, 5 y 7, que corresponde a la subpluralidad única de los subconjuntos en los que se agregó el antígeno 1, pero no ha aumentado en las otras mezclas de reacción en las que el antígeno 1 estaba ausente; por lo tanto, el clonotipo (518) identifica una célula T con un TCR específico para el antígeno 1. Asimismo, se observa que el mismo clonotipo (522a, 522b, 522c y 522d) ha aumentado en frecuencia dentro de las mezclas de reacción 1, 2, 3 y 8, lo que corresponde a la subpluralidad única de los subconjuntos en los que se agregó el antígeno 4, pero no a han aumentado en frecuencia en las otras mezclas de reacción en las el antígeno 4 estaba ausente; por lo tanto, el clonotipo (522) identifica una célula T con un TCR específico para el antígeno 4. Dado que cada antígeno está expuesto a las células T en una subpluralidad única de mezclas de reacción (o subconjuntos), siempre que se observe el mismo clonotipo en cada mezcla de reacción de la única subpluralidad, entonces el clonotipo corresponde a un TCR específico para el antígeno correspondiente a la subpluralidad.

En un caso, los métodos para determinar los clonotipos de células T específicas de antígeno en una muestra de tejido se pueden llevar a cabo con las siguientes etapas: (a) formar una pluralidad de subconjuntos a partir de una muestra de tejido que contiene células T; (b) exponer en condiciones de interacción las células T en una subpluralidad de subconjuntos a uno o más antígenos, de modo que las células T específicas para uno o más antígenos sean capaces de interactuar con ellas, y en la que cada antígeno diferente se expone a las células T en un subpluralidad diferente; (c) enriquecer las células T que interactúan con el antígeno de cada subconjunto de una subpluralidad; (d) secuenciar ácidos nucleicos recombinados que codifican una cadena del receptor de células T o una porción de la misma a partir de dichas células T enriquecidas en cada subconjunto de la subpluralidad para proporcionar lecturas de secuencia a partir de las cuales se determinan los clonotipos; (e) secuenciar ácidos nucleicos recombinados que codifican la cadena del receptor de células T o una porción de la misma a partir de dichas células T en cada subconjunto de la subpluralidad antes de dicha etapa de enriquecimiento o de células T no enriquecidas en cada subconjunto de la subpluralidad para proporcionar lecturas de secuencia a partir de las cuales se determinan los clonotipos; y (f) identificar un clonotipo de una célula T específica para un antígeno de uno o más antígenos como un clonotipo cuya frecuencia aumenta en cada subconjunto de una subpluralidad correspondiente al antígeno y no aumenta en subconjuntos fuera de dicha subpluralidad. Es decir, en algunos casos, tales clonotipos se identifican observando los clonotipos en todas las mezclas de reacción que aumentan en frecuencia (520 en la Figura 5B) e identificar los clonotipos que aparecen en cada uno de los subconjuntos de la subpluralidad correspondiente a un antígeno dado y que está ausente en todos los otros subconjuntos. En otras palabras, un clonotipo de una célula T específica para un antígeno aumenta en frecuencia solo en los subconjuntos o mezclas de reacción a los que se agregó el antígeno y no en los demás. En algunos casos, se puede identificar un clonotipo de una célula T específica de antígeno cada vez que la frecuencia de dicho clonotipo aumenta sustancialmente en cada subconjunto de una subpluralidad correspondiente al antígeno y no aumenta sustancialmente en todos los demás subconjuntos (cada subconjunto no forma parte de la subpluralidad).

En un caso, el método anterior para determinar clonotipos de células B específicas de antígeno en una muestra de tejido se puede llevar a cabo con las siguientes etapas: (a) formar una pluralidad de subconjuntos a partir de una muestra de tejido que contiene células B; (b) exponer en condiciones de interacción las células B en una subpluralidad de subconjuntos a uno o más antígenos, de modo que las células B específicas para uno o más antígenos sean capaces de interactuar con ellas, y en las que cada antígeno diferente se expone a las células B en un subpluralidad diferente; (c) enriquecer las células B que interactúan con el antígeno de cada subconjunto de una subpluralidad; (d) secuenciar ácidos nucleicos recombinados que codifican una cadena del receptor de células B o una porción de la misma a partir de dichas células B enriquecidas en cada subconjunto de la subpluralidad para proporcionar lecturas de secuencia a partir de las cuales se determinan los clonotipos; (e) secuenciar ácidos nucleicos recombinados que codifican la cadena del receptor de células B o una porción de la misma de dichas células B en cada subconjunto de la subpluralidad antes de dicha etapa de enriquecimiento o de las células B no enriquecidas en cada subconjunto de la subpluralidad para proporcionar lecturas de secuencia a partir de las cuales se determinan los clonotipos; y (f) identificar un clonotipo de una célula B específica para un antígeno de uno o más antígenos como un clonotipo cuya frecuencia aumenta en cada subconjunto de una subpluralidad correspondiente al antígeno y no aumenta en subconjuntos fuera de dicha subpluralidad. Es decir, en algunos casos, dichos clonotipos se identifican observando los clonotipos en todas las mezclas de reacción que aumentan en frecuencia e identificar los clonotipos que aparecen en cada uno de los subconjuntos de la subpluralidad correspondiente a un antígeno dado y que está ausente en todos los demás subconjuntos. En otras palabras, un clonotipo de una célula B específica para un antígeno aumenta en frecuencia solo en los subconjuntos o mezclas de reacción a los que se agregó el antígeno y no en los demás. En algunos casos, se puede identificar un clonotipo de una célula B específica de antígeno siempre que la frecuencia de dicho clonotipo aumente sustancialmente en cada subconjunto de una subpluralidad correspondiente al antígeno y no aumente sustancialmente en cada otro subconjunto (cada subconjunto no forma parte de la subpluralidad).

Para mayor claridad, la Figura 4 ilustra el proceso de selección de subpluralidades de una pluralidad de subconjuntos de acuerdo con algunos de los ejemplos de la invención. La muestra de tejido (400) se separa en una pluralidad de subconjuntos (402), por ejemplo, 10 como se muestra en la Figura 4. La muestra de tejido (400) también se puede dividir en alícuotas en una pluralidad de subconjuntos, o se puede formar una pluralidad de subconjuntos a partir de ella, que puede o no usar toda la cantidad de muestra de tejido (400). Una subpluralidad de la pluralidad (402) es una selección de dos a nueve subconjuntos de pluralidad (402). En algunas realizaciones, se seleccionan varias subpluralidades que tienen cada una el mismo número de subconjuntos, como se ilustra en la Figura 4, en la que cada

subpluralidad consta de cinco subconjuntos. En algunas realizaciones de la invención, se expone un antígeno diferente a las células T en subconjuntos de una subpluralidad diferente. Por lo tanto, por ejemplo, la subpluralidad 1 puede estar expuesta al antígeno 1, la subpluralidad 2 expuesta al antígeno 2, y así sucesivamente. En consecuencia, en la Figura 4, el subconjunto 1 está expuesto al antígeno 1, antígeno 3 y antígeno 4; asimismo, el subconjunto 2 está expuesto al antígeno 1 y al antígeno 2; etc.

Identificación de clonotipos de células T específicas de antígeno utilizando técnicas de células individuales

En algunos casos, los clonotipos de linfocitos específicos de antígeno pueden identificarse usando técnicas de células individuales, tal como se describe en Faham y Willis, patentes de Estados Unidos Nos. 8.236.503 y 8.507.205. Aunque se ilustra en términos de cadenas de receptores de células T, los expertos en la materia entenderán que el método ilustrado en el presente documento es aplicable a las cadenas de receptores de células B. En un aspecto, la etapa de seleccionar las células T que interactúan con el antígeno se lleva a cabo mediante la disposición de las células T expuestas al antígeno o antígenos en los reactores para que una fracción sustancial de los reactores contenga una sola célula T y un solo reactivo antigénico marcado, generalmente unido a un TCR de la célula T. Un objetivo de estos métodos es llevar a cabo una reacción de ensamblaje del ciclo de la polimerasa (PCA) (también conocida como una "PCR de enlace") en células individuales en los reactores para unir sus secuencias de ácido nucleico recombinadas (por ejemplo, que codifica una porción de un TCR) a una etiqueta de secuencia que está asociada con, o etiquetas, un reactivo antigénico presente en el reactor con la célula T única. Los productos de dicho enlace se denominan en el presente documento "productos de coalescencia". Después de su generación, los productos de coalescencia se pueden secuenciar para identificar tanto el clonotipo del TCR como la etiqueta de secuencia que, a su vez, identifica el reactivo antigénico. La Figura 1E ofrece una visión general. Las células linfoides (1010) (que se muestran combinadas con reactivos antigénicos (1000)) tienen cada una un ácido nucleico de identificación distinto (1012), que en la figura están simplificadas (sin ninguna intención de ser limitantes) como ARN mensajeros (ARNm) (1012), que en las tres células ilustradas en la figura están etiquetadas "C₁", "C₂" y "C₃", para indicar que son tres ácidos nucleicos recombinados diferentes únicos para cada célula, respectivamente. Estos ácidos nucleicos recombinados codifican TCR (por ejemplo, 1001) expresados en la superficie de las respectivas células T. Como se mencionó anteriormente, las células T (1010) se muestran combinadas con reactivos antigénicos (1000), que pueden ser multímeros convencionales, tales como tetrámeros, que están marcados con etiquetas de secuencia (por ejemplo, 1003) que identifican las porciones del MHC y el péptido del reactivo antigénico (por ejemplo, como se muestra en el recuadro en la caja punteada 1004). El reactivo antigénico (1000) se ejemplifica con una estructura convencional que comprende un componente marco (1002), tal como una molécula de estreptavidina; fracciones de unión al MHC (tales como péptidos biotinilados (por ejemplo, 1006)); y complejos del MHC-péptido (1004).

El reactivo antigénico (1000) también puede incluir etiquetas de secuencia (tales como, 1003), que pueden producirse de acuerdo con lo enseñado por Kwong et al., patente de Estados Unidos No. 8.394.590. El MHC y la porción del péptido determinan la especificidad del reactivo para un TCR y viceversa. Los reactivos antigénicos (1000) se producen de modo que sustancialmente cada reactivo antigénico diferente (por ejemplo, cada multímero diferente) tenga una etiqueta de secuencia diferente. En algunos casos, las etiquetas de secuencia y las porciones del MHC-péptido se seleccionan de modo que, con el conocimiento de la secuencia de una etiqueta, se pueda determinar de manera única la identidad de la porción del MHC-péptido. Es decir, por ejemplo, existe una correspondencia uno a uno entre una etiqueta de secuencia y un complejo de MHC-péptido, de modo que (por ejemplo), una etiqueta de secuencia "X" indica la presencia de ácido nucleico recombinado "C₁", una etiqueta de secuencia "Y" indica la presencia de ácido nucleico recombinado "C₂", y una etiqueta de secuencia "Z" indica la presencia de ácido nucleico recombinado "C₃". Los reactivos antigénicos (1000) se combinan (1008) con células T (1010) en una mezcla de reacción y se incuban en condiciones de interacción antigénica que permiten la formación de complejos de reactivos de células T cada vez que un TCR es específico para un reactivo antigénico. Después de dicha incubación, las células se disponen (1016) en reactores unicelulares, que pueden variar ampliamente y pueden incluir, entre otros, placas con matrices de pozos con volúmenes de nano litros, dispositivos microfluídicos y similares, como se describe más detalladamente a continuación. En algunos casos, los reactores unicelulares son micelas acuosas en una emulsión, tal como se ilustra (1017) en la Figura 1E, en la que una fracción sustancial de micelas en la emulsión contiene una sola célula T junto con un solo reactivo antigénico. En un caso, la emulsión unicelular (126) se genera usando un generador de emulsión microfluídica, tal como se describe por Zeng et al., *Anal. Chem.*, 82: 3183-3190 (2010), o similares.

Los reactores (1018) contienen una mezcla de reacción de PCA que, por ejemplo, puede comprender una polimerasa de ácido nucleico, cebadores externos y cebadores de enlace (descritos más completamente a continuación), nucleósidos trifosfato, una solución tampón y similares. En algunos casos, una mezcla de reacción de PCA también puede incluir uno o más reactivos de lisis celular, para lograr el acceso de los reactivos anteriores a los ácidos nucleicos recombinados intracelulares, tales como ARNm. Para cada reactor (1018) que contiene una célula, la reacción de PCA (1020) genera productos de coalescencia (1022) que pueden comprender uno o más pares de secuencias, de modo que un miembro del par es una etiqueta de secuencia y el otro miembro es un ácido nucleico recombinado predeterminado. En otros casos, los productos de coalescencia pueden comprender tripletes de secuencias o concatenaciones de orden superior, por ejemplo, como lo enseñan Faham y Willis, patente de Estados Unidos No. 8.507.205. En algunos casos, se puede generar un producto de coalescencia de un solo tipo para cada célula (o por reactor) o se puede generar una pluralidad de diferentes tipos de productos de coalescencia para cada célula (o por reactor). Tal pluralidad puede ser al menos 2, o puede estar en el intervalo de 2 a 500, o de 2 a 200, o de 2 a 100, o

de 2 a 20. En un caso, dicha pluralidad puede estar en el intervalo de 2 a 10. En algunos casos, dicha pluralidad es dos.

Después de completar la reacción de PCA (1020), la emulsión (1017) se rompe y los productos de coalescencia (1026) se aíslan (1024). Los productos de coalescencia (1026) se representan en la Figura 1E como conjugados de etiquetas de secuencia (X, Y o Z) y ácidos nucleicos recombinados (por ejemplo, clonotipos) (C₁, C₂ y C₃). Se puede usar una variedad de métodos convencionales para aislar productos de coalescencia (1026) de la mezcla de reacción, que incluyen, pero no se limitan a, cromatografía en columna, precipitación con etanol, purificación por afinidad después del uso de cebadores biotinilados, electroforesis en gel o similares. Como parte de la reacción de PCA (1020) o después del aislamiento (1024), se pueden agregar secuencias adicionales a los productos de coalescencia (1026) según sea necesario para la secuenciación (1028). La secuenciación puede llevarse a cabo utilizando un instrumento convencional de alto rendimiento, por ejemplo, Genome Analyzer Ix (Illumina, Inc., San Diego), o similares.

Las reacciones de ensamblaje del ciclo de la polimerasa (PCA) permiten coalescer una pluralidad de fragmentos de ácido nucleico para formar un producto de coalescencia único en uno o más ciclos de hibridación de fragmentos y extensión de polimerasa, por ejemplo, Xiong et al., *FEBS Micro Biol. Rev.*, 32: 522-540 (2008). Las reacciones de PCA vienen en muchos formatos. En un formato de interés, PCA sigue una pluralidad de reacciones en cadena de la polimerasa (PCR) que tienen lugar en un volumen de reacción común, en el que cada componente de la PCR incluye al menos un cebador de enlace que permite que las cadenas del amplicón resultante se hibriden con cadenas de otro amplicón en la reacción y se extiendan para formar un producto de coalescencia o un precursor de un producto de coalescencia. La PCA en sus diversos formatos (y bajo varios nombres alternativos) es un método bien conocido para el ensamblaje de fragmentos y la síntesis de genes, varias formas de las cuales se describen en las siguientes referencias: Yon et al., *Nucleic Acids Research*, 17: 4895 (1989); Chen et al., *J. Am. Chem. Soc.*, 116: 8799-8800 (1994); Stemmer et al., *Gene*, 164: 49-53 (1995); Hoover et al., *Nucleic Acids Research*, 30: e43 (2002); Xiong et al., *Biotechnology Advances*, 26: 121-134 (2008); Xiong et al., *FEBS Microbiol. Rev.*, 32: 522-540 (2008); y similares.

Un ejemplo de formato de PCA (pero no limitante) útil en los presentes métodos se describe en la Figura 1F, que ilustra un esquema de PCA para unir dos fragmentos A' (1208) y B' (1210) separados en un solo producto de coalescencia (1222). El fragmento A' (1208) se amplifica con los cebadores (1200) y (1202) y el fragmento B' (1210) se amplifica con los cebadores (1206) y (1204) en la misma mezcla de PCR. Los cebadores (1200) y (1206) son cebadores "externos" de la reacción PCA y los cebadores (1202) y (1204) son los cebadores "internos" de la reacción PCA. Los cebadores internos (1202) y (1204) tienen cada uno una cola (1203 y 1205, respectivamente) que no son complementarias a A' o B' (o secuencias adyacentes si A' y B' son segmentos incrustados en una secuencia más larga). Las colas (1203) y (1205) son complementarias entre sí. Generalmente, tales colas de cebador interno se seleccionan para hibridación selectiva con su cebador interno correspondiente (y no en otro lugar); pero de lo contrario, tales colas pueden variar ampliamente en longitud y secuencia. En un caso, tales colas tienen una longitud en el intervalo de 8 a 30 nucleótidos; o una longitud en el intervalo de 14 a 24 nucleótidos. A medida que avanzan las PCR (1212), se producen fragmentos de producto A (1215) y B (1217) que incorporan colas (1203) y (1205) en las regiones finales (1214) y (1216), respectivamente. Durante las PCR, los fragmentos de producto A (1215) y B (1217) se desnaturalizarán y algunos de las cadenas "superiores" (1215a) de A hibridada (1218) con las cadenas inferiores (1217b) de B y los extremos 3' se extienden (1219) para formar (1220) el producto de coalescencia A-B (1222). El producto de coalescencia A-B (1222) puede amplificarse adicionalmente por un exceso de cebadores externos (1200) y (1206). En algunos casos, la región del producto de coalescencia (1222) formado a partir de las colas (1203) y (1205) puede incluir uno o más sitios de unión del cebador para su uso en análisis posteriores, tales como secuenciación de alto rendimiento. Típicamente, en las reacciones de PCA, las concentraciones de cebadores externos son mayores que las concentraciones de cebadores internos, de modo que la amplificación del producto de coalescencia continúa después de la formación inicial. Por ejemplo, en una realización para fusionar dos ácidos nucleicos objetivo, la concentración del cebador externo puede ser de aproximadamente 10 a 100 veces mayor que la de los cebadores internos, por ejemplo, 1 µM para cebadores externos y 0,01 µM para cebadores internos. De lo contrario, una reacción de PCA puede comprender los componentes de una PCR.

Análisis de células individuales. Como se mencionó anteriormente, en algunos casos, las células de una población están dispuestas en reactores que contienen cada uno una célula. Esto se puede lograr mediante una variedad de plataformas de reactores unicelulares a gran escala conocidas en la técnica, por ejemplo, Clarke et al., publicación de la patente de los Estados Unidos No. 2010/0255471; Mathies et al., publicación de la patente de los Estados Unidos No. 2010/0285975; Edd et al., publicación de la patente de los Estados Unidos No. 2010/0021984; Colston et al., publicación de la patente de los Estados Unidos No. 2010/0173394; Love et al., publicación internacional de la patente WO2009/145925; Muraguchi et al., publicación de la patente de los Estados Unidos No. 009/0181859; Novak et al., *Angew. Chem. Int. Ed.*, 50: 390-395 (2011); Chen et al., *Biomed Microdevices*, 11: 1223-1231 (2009); y similares. En un caso, las células están dispuestas en pozos de una matriz de micropozos en la que se producen reacciones, tales como las reacciones de PCA; En otro aspecto, las células están dispuestas en micelas de una emulsión de agua en aceite, en la que las micelas sirven como reactores. Reactores micelares generados por dispositivos microfluídicos, por ejemplo, Mathies et al., (citado anteriormente) o Edd et al., (citado anteriormente) son de particular interés porque se pueden generar micelas de tamaño uniforme con menor cizallamiento y tensión en las células que en los procesos de emulsificación a granel. Las composiciones y técnicas para las emulsificaciones, incluida la realización de reacciones de amplificación, tales como PCR, en micelas se encuentran en las siguientes referencias: Becher, "Emulsions: Theory and Practice" (Oxford University Press, 2001); Griffiths y Tawfik, patente de los Estados Unidos

No. 6.489.103; Tawfik y Griffiths, *Nature Biotechnology*, 16: 652-656 (1998); Nakano et al., *J. Biotechnology*, 102: 117-124 (2003); Dressman et al., *Proc. Natl. Acad. Sci.*, 100: 8817-8822 (2003); Dressman et al., patente de los Estados Unidos No. 8.048.627; Berka et al., patentes de los Estados Unidos Nos. 7.842.457 y 8.012.690; Diehl et al., *Nature Methods*, 3: 551-559 (2006); Williams et al., *Nature Methods*, 3: 545-550 (2006); Zeng et al., *Analytical Chemistry*, 82 (8): 3183-3190 (2010); instrucciones del kit de emulsión y purificación de ADN Micellula (EURx, Gdansk, Polonia, 2011); y similares. En un caso, la mezcla de etiquetas de secuencia homogéneas (por ejemplo, perlas) y la mezcla de reacción se agrega gota a gota en una mezcla giratoria de aceite biocompatible (por ejemplo, aceite mineral ligero, Sigma) y se deja emulsionar. En otra realización, las etiquetas de secuencia homogéneas y la mezcla de reacción se agregan gota a gota en un flujo cruzado de aceite biocompatible. El aceite utilizado puede complementarse con uno o más estabilizadores de emulsión biocompatibles. Estos estabilizadores de emulsión pueden incluir Atlox 4912, Span 80 y otros estabilizadores adecuados reconocidos y disponibles comercialmente. En algunos casos, la emulsión es termoestable para permitir el ciclo térmico, por ejemplo, al menos 94 °C, al menos 95 °C o al menos 96 °C. Preferiblemente, las gotas formadas varían en tamaño desde aproximadamente 5 micrómetros hasta aproximadamente 500 micrómetros, más preferiblemente de aproximadamente 10 micrómetros a aproximadamente 350 micrómetros, incluso más preferiblemente de aproximadamente 50 a 250 micrómetros, y lo más preferiblemente de aproximadamente 100 micrómetros a aproximadamente 200 micrómetros. Ventajosamente, la mezcla de fluidos de flujo cruzado permite el control de la formación de gotas y la uniformidad del tamaño de las gotas.

En algunos casos, se producen micelas que tienen una distribución uniforme de volúmenes de modo que los reactivos disponibles en dichos reactores dan como resultado ácidos nucleicos objetivo y etiquetas de secuencia amplificadas de manera similar. Es decir, volúmenes de reactor ampliamente variables, por ejemplo, volúmenes de micelas, pueden conducir a fallas de amplificación y/o grados de amplificación muy variables. Tales fallas y variaciones impedirían o aumentarían la dificultad de hacer comparaciones cuantitativas de ácidos nucleicos objetivo en células individuales de una población, por ejemplo, diferencias en la expresión génica. En un caso, se producen micelas que tienen una distribución de volúmenes con un coeficiente de variación (CV) del treinta por ciento o menos. En algunos casos, las micelas tienen una distribución de volúmenes con un CV del veinte por ciento o menos.

Las células de una muestra de tejido y reactivo antigénico pueden suspenderse en una mezcla de reacción antes de su disposición en reactores. En un aspecto, una mezcla de reacción es una mezcla de reacción de PCA y es sustancialmente la misma que una mezcla de reacción de PCR con al menos un par de cebadores internos (o de unión) y al menos un par de cebadores externos. En algunos casos, una etapa de lisar las células se puede lograr calentando las células a una temperatura de 95 °C o superior en presencia de un detergente no iónico, por ejemplo, Triton X-100 o Tween-20 al 0,1%, durante un período antes de llevar a cabo una reacción de amplificación. En un caso, dicho período de temperatura elevada puede ser de 10-20 minutos. Alternativamente, una etapa de lisar las células puede realizarse mediante uno o más ciclos de calentamiento y enfriamiento, por ejemplo, 96 °C durante 15 minutos seguido de 10 °C durante 10 minutos, en presencia de un detergente no iónico, por ejemplo, Triton X-100 o Tween-20 al 0,1%. La guía para llevar a cabo una etapa de lisis se describe en Brown et al., *J.R. Soc. Interface* 5: S131-S138 (2008).

Claramente, se pueden emplear muchas configuraciones de dispositivos de microfluidos para generar micelas que contienen células individuales, por ejemplo, Zagoni et al., capítulo 2, *Methods of Cell Biology*, 102: 25-48 (2011); Brouzes, capítulo 10, *Methods of Cell Biology*, 102: 105-139 (2011); Wiklund et al., capítulo 14, *Methods of Cell Biology*, 102: 177-196 (2011); Le Gac et al., capítulo 7, *Methods of Molecular Biology*, 853: 65-82 (2012); y similares.

En algunos casos, este aspecto de la invención para determinar células T específicas de antígeno puede implementarse con las siguientes etapas: (a) exponer en condiciones de interacción una muestra de tejido que contiene células T a reactivos antigénicos marcados con etiquetas de secuencia; (b) disponer en múltiples reactores células T individuales unidas específicamente a al menos un reactivo antigénico, conteniendo cada reactor una mezcla de reacción de ensamblaje del ciclo de polimerasa (PCA) que comprende un par de cebadores externos y uno o más pares de cebadores de unión, al menos un par de tales cebadores externos y de unión que son específicos para un ácido nucleico recombinado que codifica un segmento de una cadena de TCR de la célula T y uno o más pares de tales cebadores externos y de unión que son específicos para una etiqueta de secuencia unida al reactivo antigénico; (c) realizar una reacción de PCA en los reactores para formar productos de coalescencia que comprenden dichos ácidos nucleicos recombinados y dicha etiqueta de secuencia; (d) aislar espacialmente moléculas individuales de productos de coalescencia de los reactores; (e) secuenciar los productos de coalescencia espacialmente aislados de los reactores para generar lecturas de secuencia a partir de las cuales se determinan pares de clonotipos y etiquetas de secuencia; y (f) identificar la especificidad de antígeno de las células T por sus pares de etiquetas de secuencia y clonotipo. En algunos casos, los reactores son micelas acuosas de una emulsión de agua en aceite. En algunos casos, las micelas acuosas son generadas por un dispositivo microfluidico. En algunos casos, los reactores son pozos de nano litros en un sustrato plano. En algunos casos, se lleva a cabo una etapa adicional de lisis de las células T individuales en los reactores antes de realizar la reacción de PCA.

Antígenos

Un antígeno puede ser cualquier compuesto o composición capaz de provocar una respuesta inmune mediada por células (es decir, una respuesta inmune adaptativa), particularmente en un mamífero, tal como un ser humano. En algunas realizaciones, un antígeno puede ser cualquier compuesto que pueda ser reconocido por una célula T en el

contexto de la molécula del MHC. Más particularmente, los antígenos incluyen, entre otros, células, extractos de tejidos, tejidos o lisados celulares, proteínas, individualmente o una mezcla, una pluralidad de proteínas, péptidos, mezclas de péptidos, lípidos, carbohidratos, azúcares y similares. Un antígeno puede ser una célula presentadora de antígeno (APC) transfectada con ácidos nucleicos que codifican el antígeno, por ejemplo, Dornmair et al., publicación de la patente de los Estados Unidos No. 2013/0195900; Siewert et al., *Nature Medicine*, 18 (5): 824-828 (2012). Un antígeno puede ser característico de una enfermedad, tal como una enfermedad infecciosa, una enfermedad autoinmune o un cáncer. El antígeno puede ser, por ejemplo, un antígeno viral, un antígeno bacteriano, un antígeno de cáncer, etc. En algunas realizaciones, un antígeno es un antígeno de cáncer o un antígeno viral. Por "antígeno de cáncer" se entiende cualquier molécula (por ejemplo, proteína, péptido, lípido, carbohidrato, etc.) única o predominantemente expresada o sobreexpresada por una célula tumoral o una célula cancerosa, de modo que el antígeno esté asociado con el tumor o el cáncer. .

Un antígeno de cáncer puede ser un antígeno de cáncer de un solo tipo de cáncer o tumor, de modo que el antígeno de cáncer esté asociado o sea característico de un solo tipo de cáncer o tumor. Alternativamente, un antígeno de cáncer puede ser un antígeno de cáncer (por ejemplo, puede ser característico) de más de un tipo de cáncer o tumor. Por ejemplo, un antígeno de cáncer puede ser expresado por células de cáncer de mama y de próstata y no expresarse en absoluto por células normales, no tumorales o no cancerosas, o expresarse solo mínimamente. Un antígeno de cáncer puede ser un antígeno de cáncer de melanoma o un antígeno de cáncer de mama. Los ejemplos de antígenos de cáncer incluyen aquellos del grupo que consiste en gp100, MART-1, NY-ESO-1, un miembro de la familia de proteínas MAGE, por ejemplo, MAGE-A1, mesotelina, tirosinasa, TRP-1, TRP-2, PMSA, Her-2 y p53.

Un antígeno puede ser un antígeno viral. En algunas realizaciones, "antígeno viral" significa aquellos antígenos codificados por una parte de un genoma viral que pueden detectarse mediante una respuesta inmunológica específica. Los antígenos virales incluyen, por ejemplo, una proteína de cubierta viral, un antígeno viral de la gripe, un antígeno del VIH, un antígeno de la hepatitis B o un antígeno de la hepatitis C.

Un antígeno puede producirse de forma natural, artificial, sintética o recombinante. Por lo tanto, un antígeno puede ser una proteína, polipéptido o péptido sintético, recombinante, aislado y/o purificado. Los métodos para fabricar u obtener tales antígenos son conocidos en la técnica. Por ejemplo, los métodos adecuados para sintetizar de nuevo polipéptidos y proteínas (por ejemplo, polipéptidos y proteínas antigénicos) se describen en Chan et al., *Fmoc Solid Phase Peptide Synthesis*, Oxford University Press, Oxford, Reino Unido, 2005; *Peptide and Protein Drug Analysis*, ed. Reid, R., Marcel Dekker, Inc., 2000; *Epitope Mapping*, ed. Westwood et al., Oxford University Press, Oxford, Reino Unido, 2000; y la patente de los Estados Unidos No. 5,449,752. Además, los polipéptidos y proteínas (por ejemplo, polipéptidos y proteínas antigénicos) pueden producirse de forma recombinante usando ácidos nucleicos que codifican el polipéptido o la proteína usando métodos recombinantes estándar. Véase, por ejemplo, Sambrook et al., *Molecular Cloning: A Laboratory Manual*, 3rd ed., Cold Spring Harbor Press, Cold Spring Harbor, N.Y. 2001; y Ausubel et al., *Current Protocols in Molecular Biology*, Greene Publishing Associates y John Wiley & Sons, NY, 1994. Las secuencias de nucleótidos de muchos antígenos se conocen en la técnica y están disponibles en la base de datos GenBank del sitio web del Centro Nacional de Información Biotecnológica (NCBI). Además, un antígeno puede aislarse y/o purificarse de una fuente, tal como una planta, una bacteria, un insecto, un mamífero, por ejemplo, una rata, un ser humano, etc. Los métodos de aislamiento y purificación son bien conocidos en la técnica.

Un antígeno puede ser un antígeno libre, por ejemplo, un péptido antigénico no unido (por ejemplo, un péptido libre), o puede ser un antígeno unido, por ejemplo, un tetrámero del MHC-péptido o un péptido antigénico presentado por una célula portadora que fue pulsada con el péptido, o un antígeno puede ser una APC transfectada con ácidos nucleicos que codifican el antígeno (como se mencionó anteriormente).

En algunas realizaciones, las células mononucleares de sangre periférica (PBMC) (por ejemplo, que pueden obtenerse de la sangre, por ejemplo, como un producto de leucaféresis) de un sujeto pueden cultivarse directamente en presencia de antígeno, para cargar las células presentadoras de antígeno (APC) entre las PBMC con el antígeno y para activar/estimular las células T específicas de antígeno presentes en las PBMC. A este respecto, las PBMC se pueden recoger de un individuo, en contacto con un antígeno de interés, tal como un antígeno tumoral o un lisado viral, etc. De esta manera, las APC presentes en las PBMC se cargan con el antígeno, que es luego presentado a las células T presentes en la muestra. En algunas realizaciones, las células T específicas de antígeno pueden activarse con tetrámeros del péptido-MHC, véase, por ejemplo, Altman, et al., *Science* 19 de junio de 1998; 280 (5371): 1821. En algunas realizaciones, un antígeno proteico puede exponerse a las células T indirectamente generando un conjunto de péptidos para unirse a las moléculas del MHC, en el que las secuencias de los péptidos se basan en la secuencia de aminoácidos de la proteína, por ejemplo, Stickler et al., *Toxicol. Sci.*, 77 (2): 280-289 (2004). En algunas de tales realizaciones, los péptidos son péptidos superpuestos que cubren la proteína. En algunas realizaciones, los péptidos tienen cada uno un tamaño de 10 a 20 aminoácidos.

Los antígenos pueden incluir APC transfectadas con ácidos nucleicos, tales como ARN transcrito *in vitro* que codifica péptidos ("ARN peptídico"), por ejemplo, como lo enseñan Coughlin et al., *Blood*, 103 (6): 2046-2054 (2004) ; o APC transfectadas con plásmidos, por ejemplo, como lo enseñan Siewert et al., *Nature Medicine*, 18 (5): 824-828 (2012); y Dornmair et al., publicación de la patente de los Estados Unidos No. 2013/0195900. En resumen, el ARN peptídico puede generarse insertando ADN que codifica péptido en un vector de transcripción *in vitro* convencional, tal como pGEM4Z, o similar; generar ARN mensajero mediante transcripción *in vitro*; purificar el ARN transcrito por técnicas

convencionales, por ejemplo, extracción con fenol/cloroforno y/o purificación en columna (RNAeasy, Qiagen); y electroporar el ARNm en las APC. El ADN que codifica el péptido puede ser ADN que codifica un antígeno completo. El ácido nucleico electroporado, transcrito y traducido es luego procesado por las APC en péptidos que se presentan a las células T.

5 Linfocitos

Los linfocitos incluyen células T o células B. Las células T se pueden obtener de varias fuentes, incluidas las células mononucleares de sangre periférica, médula ósea, timo, biopsia de tejido, tumor, tejido de los ganglios linfáticos, tejido linfoide asociado al intestino, tejido linfoide asociado a la mucosa, tejido del bazo o cualquier otro tejido linfoide y tumores. Las células T se pueden obtener a partir de líneas de células T y de fuentes autólogas o alogénicas. Las células T se pueden obtener de un solo individuo o de una población de individuos, por ejemplo, una población de individuos que padecen la misma enfermedad, tal como un cáncer o una enfermedad infecciosa.

En algunas realizaciones, las células de la sangre circulante de un individuo se obtienen por aféresis o leucaféresis. El producto de aféresis generalmente contiene linfocitos, que incluyen células T, monocitos, granulocitos, células B, otros glóbulos blancos nucleados, glóbulos rojos y plaquetas. En una realización, las células recogidas por aféresis o leucaféresis se pueden lavar para eliminar la fracción de plasma y colocar las células en un tampón o medio apropiado para las etapas de procesamiento posteriores. En una realización de la invención, las células se lavan con solución salina tamponada con fosfato (PBS). En una realización alternativa, la solución de lavado carece de calcio y puede carecer de magnesio o puede carecer de muchos, si no todos, los cationes divalentes. Como los expertos en la materia apreciarían fácilmente, una etapa de lavado se puede lograr mediante métodos conocidos por los expertos en la técnica, tal como mediante el uso de una centrífuga semiautomática de "flujo continuo" (por ejemplo, el procesador de células Cobe 2991, Baxter) de acuerdo con las instrucciones del fabricante. Después del lavado, las células se pueden resuspender en una variedad de tampones biocompatibles, tales como, por ejemplo, PBS libre de Ca⁺⁺/Mg⁺⁺. Alternativamente, los componentes indeseables de la muestra de aféresis se pueden eliminar y las células se resuspenden directamente en el medio de cultivo.

En otras realizaciones, las células T se aíslan de los linfocitos de sangre periférica mediante lisis de los glóbulos rojos y mediante centrifugación a través de un gradiente de PERCOLL^{MR}. Una subpoblación específica de células T, como las células T CD28⁺, CD4⁺, CD8⁺, CD45RA⁺ y CD45RO⁺, puede aislarse adicionalmente mediante técnicas de selección positiva o negativa. Por ejemplo, las células T CD3⁺, CD28⁺ pueden seleccionarse positivamente usando perlas magnéticas conjugadas CD3/CD28 (por ejemplo, DYNABEADS[®] M-450 CD3/CD28 T Cell Expander). En algunas realizaciones, el enriquecimiento de una población de células T por selección negativa se puede lograr con una combinación de anticuerpos dirigidos a marcadores de superficie únicos para las células seleccionadas negativamente. Uno de estos métodos es la clasificación y/o selección celular mediante inmunoadherencia magnética negativa o citometría de flujo que utiliza un cóctel de anticuerpos monoclonales dirigidos a marcadores de la superficie celular presentes en las células seleccionadas negativamente. Por ejemplo, para enriquecer en células CD4⁺ por selección negativa, un cóctel de anticuerpos monoclonales típicamente incluye anticuerpos contra CD14, CD20, CD11b, CD16, HLA-DR y CD8.

Otro método para preparar las células T para la estimulación es congelar las células después de la etapa de lavado, lo que no requiere la etapa de eliminación de monocitos. Deseando no estar limitado por la teoría, la etapa de congelación y descongelación posterior proporciona un producto más uniforme al eliminar los granulocitos y, en cierta medida, los monocitos en la población celular. Después de la etapa de lavado que elimina el plasma y las plaquetas, las células pueden suspenderse en una solución de congelación. Si bien muchas soluciones y parámetros de congelación son conocidos en la técnica y serán útiles en este contexto, un método implica el uso de PBS que contiene DMSO al 20% y albúmina de suero humano al 8% (HSA) u otros medios adecuados de congelación celular. Esto luego se diluye 1:1 con medio para que la concentración final de DMSO y HSA sea del 10% y 4%, respectivamente. Las células se congelan luego a -80 °C a una velocidad de 1° por minuto y se almacenan en la fase de vapor de un tanque de almacenamiento de nitrógeno líquido.

Las células B se pueden obtener de varias fuentes, que incluyen células mononucleares de sangre periférica, médula ósea, timo, biopsia de tejido, tumor, tejido de ganglios linfáticos, tejido linfoide asociado al intestino, tejido linfoide asociado a la mucosa, tejido del bazo o cualquier otro tejido linfoide y tumores. Las células B se pueden obtener a partir de líneas de células B y de fuentes autólogas o alogénicas. Las células B se pueden obtener de un solo individuo o una población de individuos, por ejemplo, una población de individuos que padecen la misma enfermedad, tal como un cáncer o una enfermedad infecciosa.

En algunas realizaciones, las células de la sangre circulante de un individuo se obtienen por aféresis o leucaféresis. El producto de aféresis generalmente contiene linfocitos, que incluyen células T, monocitos, granulocitos, células B, otros glóbulos blancos nucleados, glóbulos rojos y plaquetas. En una realización, las células recogidas por aféresis o leucaféresis se pueden lavar para eliminar la fracción de plasma y colocar las células en un tampón o medio apropiado para las etapas de procesamiento posteriores. En una realización de la invención, las células se lavan con solución salina tamponada con fosfato (PBS). En una realización alternativa, la solución de lavado carece de calcio y puede carecer de magnesio o puede carecer de muchos, si no todos, los cationes divalentes. Como los expertos en la materia apreciarían fácilmente, una etapa de lavado se puede lograr mediante métodos conocidos por los expertos en la

técnica, tal como mediante el uso de una centrífuga semiautomática de "flujo continuo" (por ejemplo, el procesador de células Cobe 2991, Baxter) de acuerdo con las instrucciones del fabricante. Después del lavado, las células se pueden resuspender en una variedad de tampones biocompatibles, tales como, por ejemplo, PBS libre de Ca⁺⁺/Mg⁺⁺. Alternativamente, los componentes indeseables de la muestra de aféresis se pueden eliminar y las células se resuspenden directamente en el medio de cultivo.

Las células B también pueden aislarse usando perlas magnéticas unidas a anticuerpos específicos para marcadores de superficie de células B comunes, por ejemplo, CD19 Dynabeads (Life Technologies, Carlsbad, CA).

En algunas realizaciones, los linfocitos infiltrantes de tejido (TIL) son de especial interés, en parte debido a su potencial como agentes antitumorales. Los linfocitos T citotóxicos para su uso en los métodos de la invención pueden obtenerse mediante una variedad de técnicas conocidas en el arte, particularmente el arte de la terapia celular adoptiva, por ejemplo, tal como se describe en Dudley et al., *J. Immunotherapy*, 26 (4): 332-342 (2003); Dudley et al., *Semin. Oncol.*, 34 (6): 524-531 (2007). Los linfocitos T pueden recogerse de acuerdo con técnicas conocidas y enriquecerse o agotarse mediante técnicas conocidas tales como la unión por afinidad a los anticuerpos, o tal como la citometría de flujo y/o la unión por afinidad. Después de las etapas de enriquecimiento y/o agotamiento, la expansión *in vitro* de los linfocitos T deseados se puede llevar a cabo de acuerdo con técnicas conocidas (incluidas, entre otras, las descritas en la patente de los Estados Unidos No. 6.040.177 de Riddell et al.), o variaciones de las mismas que serán evidentes para los expertos en la materia.

Usos de TCR y BCR reconstituidos

Los receptores de células T reconstituidos específicos para un antígeno o conjunto de antígenos dados tienen una variedad de usos tanto individualmente como en grupo, que incluyen, pero no se limitan a, tal como compuestos de unión para inmunoterapia, como componentes de células T transfectadas para inmunoterapia adoptiva, como fuentes de antígeno en las vacunas, y como indicadores del estado inmune. Las cadenas de TCR coincidentes en formato soluble pueden usarse como compuestos de unión de alta afinidad unidos a agentes de captura de células T para terapias anticancerígenas únicas, por ejemplo, tal como lo enseñaron Jakobsen et al., patentes de los Estados Unidos Nos. 7.329.731 y 7.666.604. Las cadenas de TCR coincidentes pueden usarse para construir vectores que, pueden a su vez, usarse para transfectar células T autólogas para la inmunoterapia adoptiva de un paciente. En un caso de esta aplicación, las muestras de las que se analizan los TCR pueden tomarse antes y después de que un paciente haya sido inmunizado con un antígeno de cáncer, de modo que las cadenas elevadas de TCR anticancerígenas se puedan combinar y seleccionar fácilmente. Las referencias que divulgan tales aplicaciones incluyen Turcotte et al., *Adv. Surg.*, 45: 341-360 (2011); Morgan et al., *Science*, 314: 126-129 (2006); Walchli et al., *PLoSOne*, 6: e27930 (2011); Robbins et al., publicación de la patente de los Estados Unidos No. 2010/0034834; y similares.

Una población de TCR emparejados o reconstituidos de una muestra comprende un perfil único del sistema inmune de un individuo, que contiene mucha más información que los perfiles de clonotipos de secuencia única. Es decir, una población de cadenas de TCR emparejadas o inmunoglobulinas de cadenas pesadas y ligeras combinadas comprende un perfil de clonotipo en el que los clonotipos son pares de secuencias de nucleótidos que codifican pares de cadenas de TCR expresadas en la misma célula T o pares de inmunoglobulinas de cadena pesada y ligera expresadas en la misma célula B. En ambos casos, dichos pares pueden estar relacionados directamente con la función de las células T, por ejemplo, mediante la interacción con conjuntos de complejos de péptido-tetrámero MHC, por ejemplo, Palmowski et al., *Immunol. Rev.*, 188: 155-163 (2002); Hadrup et al., *Nature Methods*, 6: 520-526 (2009), o con la función de las células B, por ejemplo, mediante ELISA, por ejemplo, Reddy et al., *Nature Biotechnology*, 28 (9): 965-969 (2010). En un caso, los perfiles del clonotipo de cadenas de receptores inmunes emparejados comprenden al menos 100 pares de clonotipos, en los que cada clonotipo del par comprende una secuencia de 30 a 300 nucleótidos. En otro caso, los perfiles de clonotipo de cadenas de receptores inmunes emparejados comprenden al menos 500 pares de clonotipos, en los que cada clonotipo del par comprende una secuencia de 30 a 300 nucleótidos. En otro caso, los perfiles de clonotipo de cadenas de receptores inmunes coincidentes comprenden al menos 1000 pares de clonotipos, en los que cada clonotipo del par comprende una secuencia de 30 a 300 nucleótidos. En otro caso más, tales perfiles de clonotipos de cadenas de receptores inmunes emparejados comprenden pares de clonotipos de TCR α y TCR β . En otra realización, tales perfiles de clonotipo de cadenas de receptores inmunes emparejados comprenden pares de clonotipos de TCR γ y TCR δ .

Como con los TCR, los receptores de células B reconstituidos específicos para un antígeno dado o un conjunto de antígenos tienen una variedad de usos tanto individualmente como en grupo, que incluyen, entre otros, como indicadores del estado inmunitario, como componentes de células T transfectadas para inmunoterapia adoptiva (por ejemplo, como receptores de antígeno quimérico, Barrett et al., *Ann. Rev. Med.*, 65: 333-347 (2014)), como compuestos de unión para terapias de anticuerpos pasivos para enfermedades infecciosas (Casadevall et al., *Nat Rev. Microbiol.*, 2 (9): 695-702 (2004), o terapias activas o pasivas para el cáncer o el tratamiento autoinmune, o similares.

En algunas realizaciones, el método de las Figuras 5A-5C puede usarse para identificar de una biblioteca de antígenos de cáncer, antígenos específicos del paciente para los cuales el paciente tiene células T reactivas. Una biblioteca de antígenos de cáncer puede derivarse del paciente o puede ser una colección de antígenos que ocurren comúnmente en muchos pacientes que padecen el tipo particular de cáncer. Las células T de cada una de las poblaciones de pacientes que padecen el mismo cáncer pueden examinarse con el método de las Figuras 5A-5C para determinar los

antígenos a los que los pacientes resistentes al cáncer tienen células T y/o células B sensibles y a los que los pacientes no resistentes al cáncer tienen (o carecen) de células T sensibles. A partir de esta información, se pueden seleccionar antígenos que sean objetivos adecuados para compuestos anticancerígenos, tales como anticuerpos terapéuticos y/o terapia celular adoptiva basada en CAR.

5 Muestras

Las muestras, o muestras de tejido, de células T (linfocitos T) pueden incluir, por ejemplo, células T auxiliares (células T efectoras o células Th), células T citotóxicas (CTL), células T de memoria y células T reguladoras, así como otros tipos de células que normalmente se encuentran en una muestra de tejido. En un aspecto, una muestra de células T incluye al menos 1.000 células T; pero más típicamente, una muestra incluye al menos 10.000 células T, y más típicamente, al menos 100.000 células T. En otro aspecto, una muestra incluye al menos 10^6 células T. Las muestras de células B (linfocitos B) pueden incluir células B, así como otros tipos de células que normalmente se encuentran en una muestra de tejido. En un aspecto, una muestra de células B incluye al menos 1.000 células B; pero más típicamente, una muestra incluye al menos 10.000 células B, y más típicamente, al menos 100.000 células B. En otro aspecto, una muestra incluye al menos 10^6 células B.

Las muestras usadas en los métodos de la invención pueden provenir de una variedad de tejidos como se indicó anteriormente, que incluyen, por ejemplo, tejido tumoral, sangre y plasma sanguíneo, líquido linfático, líquido cefalorraquídeo que rodea el cerebro y la médula espinal, líquido sinovial que rodea las articulaciones óseas y similares. En una realización, la muestra es una muestra de sangre. La muestra de sangre puede ser de aproximadamente 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5 o 5.0 mL. La muestra puede ser una biopsia tumoral. La biopsia puede ser, por ejemplo, de un tumor del cerebro, hígado, pulmón, corazón, colon, riñón o médula ósea. Cualquier técnica de biopsia utilizada por los expertos en la técnica puede usarse para aislar una muestra de un sujeto. Por ejemplo, una biopsia puede ser una biopsia abierta, en la que se usa anestesia general. La biopsia puede ser una biopsia cerrada, en la que se realiza un corte más pequeño que en una biopsia abierta. La biopsia puede ser una biopsia por sonda o incisional, en la cual se extrae parte del tejido. La biopsia puede ser una biopsia por escisión, en la que se realizan intentos para extirpar una lesión completa. La biopsia puede ser una biopsia por aspiración con aguja fina, en la que se extrae una muestra de tejido o líquido con una aguja.

La muestra puede ser una biopsia, por ejemplo, una biopsia de piel. La biopsia puede ser de, por ejemplo, cerebro, hígado, pulmón, corazón, colon, riñón o médula ósea. Cualquier técnica de biopsia utilizada por los expertos en la técnica puede usarse para aislar una muestra de un sujeto. Por ejemplo, una biopsia puede ser una biopsia abierta, en la que se usa anestesia general. La biopsia puede ser una biopsia cerrada, en la que se realiza un corte más pequeño que en una biopsia abierta. La biopsia puede ser una biopsia por sonda o incisional, en la que se extrae parte del tejido. La biopsia puede ser una biopsia por escisión, en la que se realizan intentos para extirpar una lesión completa. La biopsia puede ser una biopsia por aspiración con aguja fina, en la que se extrae una muestra de tejido o líquido con una aguja.

Como se discute más completamente a continuación, en algunas realizaciones, una muestra de linfocitos es suficientemente grande para que sustancialmente cada célula T o célula B con un clonotipo distinto esté representada en la misma, formando así un repertorio (como se usa el término en el presente documento). En algunas realizaciones, se toma una muestra que contiene con una probabilidad del noventa y nueve por ciento cada clonotipo de una población presente en una frecuencia de 0,001 por ciento o más. En otra realización, se toma una muestra que contiene con una probabilidad del noventa y nueve por ciento cada clonotipo de una población presente a una frecuencia de 0,0001 por ciento o más. En una realización, una muestra de células T incluye al menos medio millón de células, y en otra realización, dicha muestra incluye al menos un millón de células.

Las muestras de sangre son de particular interés y pueden obtenerse usando técnicas convencionales, por ejemplo, Innis et al., editores, PCR Protocols (Academic Press, 1990); o similares. Por ejemplo, los glóbulos blancos pueden separarse de las muestras de sangre utilizando técnicas convencionales, por ejemplo, Kit RosetteSep (Stem Cell Technologies, Vancouver, Canadá). Las muestras de sangre pueden variar en volumen de 100 μ L a 10 mL; en un aspecto, los volúmenes de muestra de sangre están en el intervalo de 100 μ L a 2 mL. Opcionalmente, subconjuntos de glóbulos blancos, por ejemplo, los linfocitos, pueden aislarse adicionalmente usando técnicas convencionales, por ejemplo, clasificación celular activada por fluorescencia (FACS) (Becton Dickinson, San José, CA), clasificación celular activada magnéticamente (MACS) (Miltenyi Biotec, Auburn, CA), o similares.

Como las recombinaciones de identificación están presentes en el ADN de las células de inmunidad adaptativa de cada individuo, así como en sus transcripciones de ARN asociadas, se puede secuenciar ARN o ADN en los métodos de la invención. Una secuencia recombinada de una célula T que codifica una cadena de una molécula receptora de células T, o una porción de la misma, se denomina clonotipo. Del mismo modo, una secuencia recombinada de una célula B que codifica una cadena de una molécula de inmunoglobulina, o una porción de la misma, también puede denominarse clonotipo. El ADN o ARN puede corresponder a secuencias de genes del receptor de células T (TCR) o de genes de inmunoglobulina. Por ejemplo, el ADN y el ARN pueden corresponder a secuencias que codifican las cadenas α , β , γ o δ de un TCR.

Amplificación de poblaciones de ácido nucleico

Se pueden generar amplicones de poblaciones objetivo de ácidos nucleicos mediante una variedad de técnicas de amplificación. En un aspecto de la invención, la PCR multiplex se usa para amplificar miembros de una mezcla de ácidos nucleicos, particularmente mezclas que comprenden moléculas inmunes recombinadas tales como receptores de células T, o porciones de los mismos. La guía para llevar a cabo las PCR multiplex de tales moléculas inmunes se encuentra en las siguientes referencias: Faham y Willis, patente de los Estados Unidos No. 8.236.503 y patente de los Estados Unidos No. 8.628.927; Morley, patente de los Estados Unidos No. 5.296.351; Gorski, patente de los Estados Unidos No. 5.837.447; Dau, patente de los Estados Unidos No. 6.087.096; Von Dongen et al., publicación de la patente de los Estados Unidos No. 2006/0234234; Publicación de patente europea EP 1544308B1; y similares.

Después de la amplificación de ADN del genoma (o la amplificación de ácido nucleico en forma de ADNc mediante transcripción inversa de ARN), las moléculas individuales de ácido nucleico pueden aislarse, opcionalmente amplificarse nuevamente y luego secuenciarse individualmente. Se pueden encontrar ejemplos de protocolos de amplificación en van Dongen et al., *Leukemia*, 17: 2257-2317 (2003) o van Dongen et al., publicación de la patente de los Estados Unidos No. 2006/0234234. En resumen, un ejemplo de protocolo es el siguiente: Tampón de reacción: tampón ABI II o tampón ABI Gold (Life Technologies, San Diego, CA); 50 μ L de volumen de reacción final; 100 ng de muestra de ADN; 10 pmol de cada cebador (sujeto a ajustes para equilibrar la amplificación como se describe a continuación); dNTP a una concentración final de 200 μ M; $MgCl_2$ a una concentración final de 1,5 mM (sujeto a optimización dependiendo de las secuencias objetivo y la polimerasa); Taq polimerasa (1-2 U/tubo); condiciones del ciclo: activación previa 7 min a 95 °C; hibridación a 60 °C; tiempos del ciclo: 30 s desnaturalización; 30 s hibridación; 30 s extensión. Las polimerasas que pueden usarse para la amplificación en los métodos de la invención están disponibles comercialmente e incluyen, por ejemplo, Taq polimerasa, AccuPrime polimerasa o Pfu. La elección de la polimerasa para usar puede basarse en si se prefiere la fidelidad o la eficiencia.

Se pueden usar mediciones de PCR en tiempo real, tinción PicoGreen, electroforesis de nanofluidos (por ejemplo, LabChip) o absorción UV en una etapa inicial para juzgar la cantidad funcional de material amplificable.

En un aspecto, las amplificaciones multiplex se llevan a cabo de manera que las cantidades relativas de secuencias en una población inicial son sustancialmente las mismas que las de la población amplificada, o amplicón. Es decir, las amplificaciones multiplex se llevan a cabo con un sesgo de amplificación mínimo entre las secuencias miembro de una población de muestra. En una realización, tales cantidades relativas son sustancialmente las mismas si cada cantidad relativa en un amplicón está dentro de cinco veces su valor en la muestra inicial. En otra realización, tales cantidades relativas son sustancialmente las mismas si cada cantidad relativa en un amplicón está dentro de dos veces su valor en la muestra inicial. Como se analiza más detalladamente a continuación, el sesgo de amplificación en la PCR puede detectarse y corregirse utilizando técnicas convencionales para que se pueda seleccionar un conjunto de cebadores de PCR para un repertorio predeterminado que proporcione una amplificación no sesgada de cualquier muestra.

En una realización, se puede evitar el sesgo de amplificación llevando a cabo una amplificación de dos etapas (como se describe en Faham y Willis, citado anteriormente) en la que se implementa un pequeño número de ciclos de amplificación en una primera etapa, o primaria, usando cebadores que tienen colas no complementarias con las secuencias objetivo. Las colas incluyen sitios de unión de cebadores que se agregan a los extremos de las secuencias del amplicón primario para que dichos sitios se usen en una segunda etapa de amplificación usando solo un cebador directo y un cebador inverso, eliminando así una causa primaria de sesgo de amplificación. Preferiblemente, la PCR primaria tendrá un número suficientemente pequeño de ciclos (por ejemplo, 5-10) para minimizar la amplificación diferencial por los diferentes cebadores. La amplificación secundaria se realiza con un par de cebadores y, por lo tanto, el problema de la amplificación diferencial es mínimo. El uno por ciento de la PCR primaria se lleva directamente a la PCR secundaria. Treinta y cinco ciclos (equivalentes a ~ 28 ciclos sin la etapa de dilución de 100 veces) utilizados entre las dos amplificaciones fueron suficientes para mostrar una amplificación robusta independientemente de si la ruptura de los ciclos fue: un ciclo primario y 34 secundarios o 25 primarios y 10 secundarios. Aunque idealmente hacer solo 1 ciclo en la PCR primaria puede disminuir el sesgo de amplificación, existen otras consideraciones. Un aspecto de esto es la representación. Esto juega un papel cuando la cantidad de entrada inicial no excede el número de lecturas obtenidas en última instancia. Por ejemplo, si se obtienen 1.000.000 de lecturas y se comienza con 1.000.000 de moléculas de entrada, entonces tomando solo la representación de 100.000 moléculas para la amplificación secundaria degradaría la precisión de la estimación de la abundancia relativa de las diferentes especies en la muestra original. La dilución de 100 veces entre las 2 etapas significa que la representación se reduce a menos que la amplificación primaria por PCR genere significativamente más de 100 moléculas. Esto indica que se pueden usar un mínimo de 8 ciclos (256 veces), pero más cómodamente, 10 ciclos (~1.000 veces). La alternativa a esto es llevar más del 1% de la PCR primaria a la secundaria, pero debido a la alta concentración del cebador usado en la PCR primaria, se puede usar un factor de dilución grande para garantizar que estos cebadores no interfieran en la amplificación y empeoren el sesgo de amplificación entre las secuencias. Otra alternativa es agregar una etapa de purificación o enzimática para eliminar los cebadores de la PCR primaria para permitir una dilución menor. En este ejemplo, la PCR primaria fue de 10 ciclos y la secundaria de 25 ciclos.

Generación de lecturas de secuencia para clonotipos

Se puede usar cualquier técnica de alto rendimiento para secuenciar ácidos nucleicos en el método de la invención. Preferiblemente, dicha técnica tiene la capacidad de generar de manera rentable un volumen de datos de secuencia

a partir del cual se pueden determinar al menos 1000 clonotipos, y preferiblemente, a partir de los cuales se pueden determinar al menos 10.000 a 1.000.000 de clonotipos. Las técnicas de secuenciación de ADN incluyen reacciones de secuenciación dideoxi clásicas (método de Sanger) usando terminadores o cebadores marcados y separación de gel en placa o capilar, secuenciación por síntesis usando nucleótidos marcados terminados reversiblemente, pirosecuenciación, secuenciación 454, hibridación específica de alelo para una biblioteca de sondas de oligonucleótidos marcados, secuenciación por síntesis usando hibridación específica de alelo para una biblioteca de clones marcados que es seguida por ligadura, monitoreo en tiempo real de la incorporación de nucleótidos marcados durante una etapa de polimerización, secuenciación polony y secuenciación SOLiD. La secuenciación de las moléculas separadas se ha llevado a cabo mediante reacciones secuenciales o de extensión única utilizando polimerasas o ligasas, así como mediante hibridaciones diferenciales simples o secuenciales con bibliotecas de sondas. Estas reacciones se han realizado en muchas secuencias clonales en paralelo, incluidas demostraciones en aplicaciones comerciales actuales de más de 100 millones de secuencias en paralelo. Estos enfoques de secuenciación se pueden utilizar para estudiar el repertorio de receptores de células T (TCR) y/o receptores de células B (BCR). En un aspecto de la invención, se emplean métodos de secuenciación de alto rendimiento que comprenden una etapa de aislar espacialmente moléculas individuales en una superficie sólida donde se secuencian en paralelo. Dichas superficies sólidas pueden incluir superficies no porosas (como en la secuenciación de Solexa, por ejemplo, Bentley et al., *Nature*, 456: 53-59 (2008) o secuenciación del genoma completo, por ejemplo, Drmanac et al., *Science*, 327: 78-81 (2010)), matrices de pozos, que pueden incluir plantillas unidas a perlas o partículas (tal como con 454, por ejemplo, Margulies et al., *Nature*, 437: 376-380 (2005) o secuenciación de Ion Torrent, publicación de las patentes de Estados Unidos Nos. 2010/0137143 o 2010/0304982), membranas micromaquinadas (tales como con secuenciación SMRT, por ejemplo, Eid et al., *Science*, 323: 133 -138 (2009)), o matrices de perlas (como con la secuenciación SOLiD o secuenciación polony, por ejemplo, Kim et al., *Science*, 316: 1481-1414 (2007)). En otro aspecto, dichos métodos comprenden amplificar las moléculas aisladas antes o después de que estén aisladas espacialmente en una superficie sólida. La amplificación previa puede comprender una amplificación con base en emulsión, tal como PCR en emulsión o amplificación de círculo rodante. De particular interés es la secuenciación con base en Solexa, en la que las moléculas de plantilla individuales se aíslan espacialmente en una superficie sólida, después de lo cual se amplifican en paralelo mediante PCR puente para formar poblaciones clonales separadas, o grupos, y luego se secuencian, como se describe en Bentley et al., citado anteriormente) y en las instrucciones del fabricante (por ejemplo, Kit de preparación de muestras y hoja de datos TruSeq^{MR}, Illumina, Inc., San Diego, CA, 2010); y además en las siguientes referencias: patentes de los Estados Unidos Nos. 6.090.592; 6.300.070; 7.115.400; y el documento EP0972081B1. En una realización, las moléculas individuales dispuestas y amplificadas en una superficie sólida forman grupos con una densidad de al menos 10^5 grupos por cm^2 ; o con una densidad de al menos 5×10^5 por cm^2 ; o con una densidad de al menos 10^6 grupos por cm^2 . En una realización, se emplean químicas de secuenciación que tienen tasas de error relativamente altas. En tales realizaciones, los puntajes de calidad promedio producidos por tales químicas son funciones que disminuyen monotónicamente de las longitudes de lectura de secuencia. La secuenciación con base en Solexa también proporciona la capacidad de generar dos lecturas de secuencia de la misma secuencia objetivo (o plantilla) en un grupo, una secuencia leída cada una desde los extremos opuestos de una secuencia objetivo. En algunas realizaciones, dichos pares de lecturas de secuencia se pueden combinar y tratar como una lectura de secuencia única en análisis posteriores, o dichos pares se pueden tratar por separado pero teniendo en cuenta que se originan en el mismo grupo. A veces, el par de lecturas de secuencia de la misma plantilla se conoce como "pares de parejas", y el proceso de secuenciación desde ambos extremos de una plantilla se denomina secuenciación "bidireccional". En algunas realizaciones, una etapa de secuenciación por síntesis usando nucleótidos marcados con terminación reversible incluye la generación de una sola lectura de secuencia para cada grupo o población clonal de plantillas y la generación de una pluralidad de lecturas de secuencia (incluidos, entre otros, pares de parejas) para cada grupo o población clonal de plantillas. En otras realizaciones adicionales, cuando se genera una pluralidad de lecturas de secuencia para cada agrupación o población clonal de plantillas, dicha pluralidad de lecturas de secuencia se puede combinar para formar una lectura de secuencia efectiva única que se usa en análisis posterior, tal como una etapa de coalescencia.

En un caso, se obtiene un perfil de clonotipo basado en secuencia de un individuo usando las siguientes etapas: (a) obtener una muestra de ácido nucleico, por ejemplo, una muestra que contiene células T del individuo; (b) aislar espacialmente moléculas individuales derivadas de dicha muestra de ácido nucleico, comprendiendo las moléculas individuales al menos una plantilla generada a partir de un ácido nucleico en la muestra, cuya plantilla comprende una región reorganizada somáticamente o una porción de la misma, siendo cada molécula individual capaz de producir al menos una lectura de secuencia; (c) secuenciar dichas moléculas individuales aisladas espacialmente para proporcionar lecturas de secuencia; y (d) determinar abundancias de diferentes secuencias de las moléculas de ácido nucleico de la muestra de ácido nucleico para generar el perfil de clonotipo. En una realización, cada una de las regiones reorganizadas somáticamente comprende una región V y una región J. En otra realización, la etapa de secuenciación incluye generar una pluralidad de lecturas de secuencia para cada clonotipo determinado. En otras realizaciones más, la etapa de secuenciación incluye combinar información o datos de una pluralidad de lecturas de secuencia para formar cada clonotipo. En algunas realizaciones, dicha etapa de combinación se puede llevar a cabo uniando las lecturas de secuencia como se describe en Faham y Willis, patente de los Estados Unidos No. 8.628.927 o usando etiquetas de secuencia como se describe en Faham et al., publicación de la patente de los Estados Unidos No. 2013/0236895A1. En otra realización, la etapa de secuenciación comprende secuenciar bidireccionalmente cada una de las moléculas individuales aisladas espacialmente para producir al menos una lectura de secuencia directa y al menos una lectura de secuencia inversa.

En algunas realizaciones, la etapa de secuenciación incluye coalescencia con al menos una pluralidad de lecturas de secuencia para formar cada clonotipo. Como se describe más detalladamente a continuación, dicha etapa de coalescencia es un proceso de combinación de lecturas de secuencia con tasas de error (por ejemplo, de errores de secuenciación y/o amplificación) para producir clonotipos correctos con un alto grado de probabilidad, tal como con un 99% de nivel de confianza.

En un aspecto, para cada muestra de un individuo, la técnica de secuenciación utilizada en los métodos de la invención genera secuencias de al menos 1000 lecturas de secuencia por proceso; en otro aspecto, dicha técnica genera secuencias de al menos 10.000 lecturas de secuencia por proceso; en otro aspecto, dicha técnica genera secuencias de al menos 100.000 lecturas de secuencia por proceso; en otro aspecto, dicha técnica genera secuencias de al menos 500.000 lecturas de secuencia por proceso; y en otro aspecto, dicha técnica genera secuencias de al menos 1.000.000 de lecturas de secuencia por proceso. A partir de dichas lecturas de secuencia, se determinan los clonotipos, por ejemplo, como se describe a continuación, o como se divulga en Faham y Willis (descrita anteriormente).

Las técnicas de secuenciación utilizadas en los métodos generan lecturas de secuencia que tienen longitudes de al menos 30 nucleótidos. En algunas realizaciones, una etapa de secuenciación genera lecturas de secuencia que tienen longitudes de al menos 50 nucleótidos; y en algunas realizaciones, una etapa de secuenciación genera lecturas de secuencia que tienen longitudes de al menos 100 nucleótidos.

Además de la última realización, al menos una de las lecturas de secuencia directa y al menos una de las lecturas de secuencia inversa tienen una región de solapamiento de tal manera que las bases de dicha región de solapamiento están determinadas por una relación complementaria inversa entre dichas lecturas de secuencia. En otra realización más, cada una de las regiones reorganizadas somáticamente comprende una región V y una región J y la etapa de secuenciación incluye además determinar una secuencia de cada una de las moléculas de ácido nucleico individuales a partir de una o más de sus lecturas de secuencia directa y al menos una lectura de secuencia inversa comenzando desde una posición en una región J y extendiéndose en la dirección de su región V asociada. En otra realización, las moléculas individuales comprenden ácidos nucleicos seleccionados del grupo que consiste en moléculas de IgH completas, moléculas de IgH incompletas, moléculas de IgK completas, moléculas inactivas de IgK, moléculas de TCR β , moléculas de TCR γ , moléculas de TCR δ completas y moléculas de TCR δ incompletas. En otra realización, la etapa de secuenciación comprende generar las lecturas de secuencia que tienen puntajes de calidad que disminuyen monotónicamente. En otro caso, el método anterior comprende las siguientes etapas: (a) obtener una muestra de ácido nucleico de células T y/o células B del individuo; (b) aislar espacialmente moléculas individuales derivadas de dicha muestra de ácido nucleico, comprendiendo las moléculas individuales conjuntos anidados de plantillas, cada una generada a partir de un ácido nucleico en la muestra y cada una de las cuales contiene una región reorganizada somáticamente o una porción de la misma, cada conjunto anidado puede producir un la pluralidad de lecturas de secuencia que se extienden cada una en la misma dirección y cada una comenzando desde una posición diferente en el ácido nucleico a partir del cual se generó el conjunto anidado; (c) secuenciar dichas moléculas individuales aisladas espacialmente; y (d) determinar abundancias de diferentes secuencias de las moléculas de ácido nucleico de la muestra de ácido nucleico para generar el perfil del clonotipo. En una realización, la etapa de secuenciación incluye producir una pluralidad de lecturas de secuencia para cada uno de los conjuntos anidados. En otra realización, cada una de las regiones reorganizadas somáticamente comprende una región V y una región J, y cada una de la pluralidad de lecturas de secuencia comienza desde una posición diferente en la región V y se extiende en la dirección de su región J asociada. La orientación para aplicar la última realización se describe en Faham et al., patente de los Estados Unidos No. 8.691.510.

Determinación del clonotipo a partir de los datos de secuencia

La construcción de clonotipos a partir de datos de lectura de secuencia depende en parte del método de secuenciación utilizado para generar dichos datos, ya que los diferentes métodos tienen diferentes longitudes de lectura y calidad de datos esperadas. Se pueden usar una variedad de enfoques para determinar los clonotipos a partir de datos de lectura de secuencia, por ejemplo, como se describe en Faham et al., publicación de la patente de los Estados Unidos No. 2013/0236895, Asbury et al., publicación internacional de patente WO/US2015/002908; y referencias similares. En un enfoque, se emplea un secuenciador Solexa para generar datos de lectura de secuencia para el análisis. En una realización, se obtiene una muestra que proporciona al menos $0,5 - 1,0 \times 10^6$ linfocitos para producir al menos 1 millón de moléculas de plantilla, que después de la amplificación opcional puede producir un millón o más de poblaciones clonales correspondientes de moléculas de plantilla (o grupos). Para la mayoría de los enfoques de secuenciación de alto rendimiento, incluido el enfoque Solexa, es conveniente realizar un muestreo excesivo a nivel de grupos para que cada secuencia de plantilla se determine con un alto grado de redundancia para aumentar la precisión de la determinación de la secuencia. Para implementaciones con base en Solexa, preferiblemente la secuencia de cada plantilla independiente se determina 10 veces o más. Para otros enfoques de secuenciación con diferentes longitudes de lectura esperadas y calidad de datos, se pueden usar diferentes niveles de redundancia para una precisión comparable de la determinación de la secuencia. Los expertos en la materia reconocen que los parámetros anteriores, por ejemplo, el tamaño de la muestra, la redundancia y similares son opciones de diseño relacionadas con aplicaciones particulares.

En un aspecto de la invención, las secuencias de clonotipos (que incluyen pero no se limitan a aquellas derivadas de TCR α , TCR β , TCR γ y/o TCR δ , pueden determinarse combinando información de una pluralidad de lecturas de

secuencia, por ejemplo, a lo largo de las regiones V(D)J de las cadenas seleccionadas. En otro aspecto, las secuencias de clonotipos se determinan combinando información de una pluralidad de lecturas de secuencia. Dichas pluralidades de lecturas de secuencia pueden incluir una o más lecturas de secuencia a lo largo de una cadena sentido (es decir, lecturas de secuencia "directa") y una o más lecturas de secuencia a lo largo de su cadena complementaria (es decir, lecturas de secuencia "inversa").

Las lecturas de secuencia de la invención pueden tener una amplia variedad de longitudes, dependiendo en parte de la técnica de secuenciación empleada. Por ejemplo, para algunas técnicas, pueden surgir varias compensaciones en su implementación, por ejemplo, (i) el número y la duración de las lecturas de secuencia por plantilla y (ii) el costo y la duración de una operación de secuenciación. En una realización, las lecturas de secuencia están en el intervalo de 20 a 400 nucleótidos; en otra realización, las lecturas de secuencia están en un intervalo de 30 a 200 nucleótidos; En otra realización más, las lecturas de secuencia están en el intervalo de 30 a 120 nucleótidos. En una realización, se generan de 2 a 1000 lecturas de secuencia para determinar la secuencia de cada clonotipo; en otra realización, se generan de 2 a 100 lecturas de secuencia para determinar la secuencia de cada clonotipo; y en otra realización más, se generan al menos 10 lecturas de secuencia para determinar la secuencia de cada clonotipo. En las realizaciones anteriores, los números dados son exclusivos de las lecturas de secuencia utilizadas para identificar muestras de diferentes individuos. Las longitudes de las diversas lecturas de secuencia utilizadas en las realizaciones descritas a continuación también pueden variar en función de la información que se pretende capturar mediante la lectura; por ejemplo, la ubicación inicial y la longitud de una lectura de secuencia pueden diseñarse para proporcionar la longitud de una región NDN así como su secuencia de nucleótidos; por lo tanto, se seleccionan las lecturas de secuencia que abarcan toda la región NDN. En otros aspectos, son suficientes una o más lecturas de secuencia que en combinación (pero no por separado) abarcan una región D y/o NDN.

En otro aspecto de la invención, las secuencias de clonotipos se determinan en parte alineando las lecturas de secuencia con una o más secuencias de referencia de la región V y una o más secuencias de referencia de la región J, y en parte mediante la determinación de bases sin alineación con las secuencias de referencia tal como en la región NDN altamente variable. Se puede aplicar una variedad de algoritmos de alineación a las lecturas de secuencia y secuencias de referencia. Por ejemplo, la guía para seleccionar métodos de alineación está disponible en Batzoglou, Briefings in Bioinformatics, 6: 6-22 (2005). En un aspecto, siempre que las lecturas de V o las lecturas de C (como se mencionó anteriormente) estén alineadas con las secuencias de referencia de las regiones V y J, se puede emplear un algoritmo de búsqueda de árbol, por ejemplo, como se describe generalmente en Gusfield (citado anteriormente) y Cormen et al., Introduction to Algorithms, Tercera edición (The MIT Press, 2009).

En otro aspecto, un extremo de al menos una lectura directa y un extremo de al menos una lectura inversa se superponen en una región de superposición (por ejemplo, 308 en la Figura 3B), de modo que las bases de las lecturas están en una relación complementaria inversa entre sí. Por lo tanto, por ejemplo, si una lectura directa en la región de superposición es "5'-acgttgc", entonces la lectura inversa en una relación complementaria inversa es "5'-gcaacgt" dentro de la misma región de superposición. En un aspecto, las bases dentro de dicha región de superposición se determinan, al menos en parte, a partir de una relación complementaria inversa de este tipo. Es decir, la probabilidad de una llamada de base (o un puntaje de calidad relacionado) en una región de superposición prospectiva aumenta si conserva, o es consistente con, una relación complementaria inversa entre las dos lecturas de secuencia. En un aspecto, los clonotipos de las cadenas de TCR β y de IgH (ilustrados en la Figura 3B) se determinan mediante al menos una lectura de secuencia que comienza en su región J y se extiende en la dirección de su región V asociada (denominada en el presente documento como una "lectura de C" (304)) y al menos una lectura de secuencia que comienza en su región V y se extiende en la dirección de su región J asociada (denominada en este documento como una "lectura de V" (306)). La región de superposición (308) puede o no abarcar la región NDN (315) como se muestra en la Figura 3B. La región de superposición (308) puede estar completamente en la región J, completamente en la región NDN, completamente en la región V, o puede abarcar un límite de la región J-región NDN o un límite de la región V-región NDN, o ambos límites (como se ilustra en la Figura 3B). Típicamente, tales lecturas de secuencia se generan extendiendo los cebadores de secuenciación, por ejemplo, (302) y (310) en la Figura 3B, con una polimerasa en una reacción de secuenciación por síntesis, por ejemplo, Metzger, Nature Reviews Genetics, 11: 31-46 (2010); Fuller et al., Nature Biotechnology, 27: 1013-1023 (2009). Los sitios de unión para los cebadores (302) y (310) están predeterminados, de modo que pueden proporcionar un punto de partida o un punto de anclaje para la alineación inicial y el análisis de las lecturas de secuencia. En una realización, se coloca una lectura de C de modo que abarque la región D y/o NDN de la cadena de TCR β o de IgH e incluye una porción de la región V adyacente, por ejemplo, como se ilustra en las Figuras 3B y 3C. En un aspecto, la superposición de la lectura de V y la lectura de C en la región V se usa para alinear las lecturas entre sí. En otras realizaciones, dicha alineación de lecturas de secuencia no es necesaria, por ejemplo, con cadenas de TCR β , de modo que una lectura de V solo puede ser lo suficientemente larga como para identificar la región V particular de un clonotipo. Este último aspecto se ilustra en la Figura 3C. La lectura de secuencia (330) se usa para identificar una región V, con o sin superposición de otra lectura de secuencia, y otra lectura de secuencia (332) atraviesa la región NDN y se usa para determinar la secuencia de la misma. La porción (334) de la lectura de secuencia (332) que se extiende dentro de la región V se usa para asociar la información de la secuencia de la lectura de secuencia (332) con aquella de la lectura de secuencia (330) para determinar un clonotipo. Para algunos métodos de secuenciación, tal como los enfoques de base por base como el método de secuenciación de Solexa, el tiempo del proceso de secuenciación y los costos de reactivos se reducen al minimizar el número de

ciclos de secuenciación en un análisis. Opcionalmente, como se ilustra en la Figura 3B, el amplicón (300) se produce con la etiqueta de muestra (312) para distinguir entre los clonotipos que se originan de diferentes muestras biológicas, por ejemplo, diferentes pacientes. La etiqueta de la muestra (312) se puede identificar mediante hibridación de un cebador con la región de unión del cebador (316) y extendiéndola (314) para producir una lectura de secuencia a través de la etiqueta (312), a partir de la cual se decodifica la etiqueta de muestra (312).

En las realizaciones en las que se unen múltiples etiquetas de secuencia a un ácido nucleico recombinado original o copias del mismo, se puede llevar a cabo una etapa de lecturas de secuencia coalescentes (o lecturas de secuencia de consenso de grupos) para determinar los clonotipos. Reducir un conjunto de lecturas de secuencia para una muestra dada a un conjunto de clonotipos distintos y registrar el número de lecturas para cada clonotipo sería trivial si la tecnología de secuenciación estuviera libre de errores. Sin embargo, en presencia de errores de secuenciación, cada clonotipo genuino está rodeado por una 'nube' de lecturas de secuencia con un número variable de errores con respecto a su secuencia. La "nube" de errores de secuenciación disminuye en densidad a medida que aumenta la distancia desde el clonotipo en el espacio de la secuencia. Hay una variedad de algoritmos disponibles para convertir lecturas de secuencia en clonotipos. En un aspecto, la coalescencia de las lecturas de secuencia (es decir, la coalescencia de clonotipos candidatos que se determina que tienen uno o más errores de secuenciación) depende de al menos tres factores: el número de secuencias obtenidas para cada uno de los clonotipos que se comparan; el número de bases en las que difieren; y el puntaje de calidad de secuenciación en las posiciones en las que son discordantes. En algunas realizaciones, se puede construir y evaluar una relación de probabilidad que se basa en las tasas de error esperadas y la distribución binomial de errores. Por ejemplo, dos clonotipos, uno con 150 lecturas y el otro con 2 lecturas con una diferencia entre ellas en un área de mala calidad de secuenciación, probablemente coalescerán, ya que es probable que se generen por error de secuenciación. Por otro lado, dos clonotipos, uno con 100 lecturas y el otro con 50 lecturas con dos diferencias entre ellos, no coalescen, ya que se considera que es poco probable que se generen por error de secuencia. En algunas realizaciones, el algoritmo descrito a continuación puede usarse para determinar clonotipos a partir de lecturas de secuencia. Algunos de estos conceptos se ilustran en la Figura 10A. En algunas realizaciones de una etapa de coalescencia, las lecturas de secuencia se convierten primero en clonotipos candidatos. Tal conversión depende de la plataforma de secuenciación empleada. Para plataformas que generan lecturas de secuencia larga de alta puntuación Q, la secuencia leída o una porción de la misma puede tomarse directamente como un clonotipo candidato. Para plataformas que generan lecturas de secuencia más cortas con menor puntuación Q, se pueden requerir algunas etapas de alineación y ensamblaje para convertir un conjunto de lecturas de secuencia relacionadas en un clonotipo candidato. Por ejemplo, para plataformas con base en Solexa, en algunas realizaciones, los clonotipos candidatos se generan a partir de colecciones de lecturas emparejadas de múltiples grupos, por ejemplo, 10 o más, como se mencionó anteriormente.

Las frecuencias de los clonotipos candidatos se pueden graficar en el espacio de la secuencia, como se ilustra en la Figura 10A, en la que dicho espacio se reduce a una dimensión (el eje horizontal) en aras de ilustración. El eje vertical da la magnitud de la frecuencia de cada clonotipo candidato, log (recuento de lectura) o alguna medida similar. En la figura, los clonotipos candidatos están representados por los diversos símbolos (1530). De acuerdo con una realización de la invención, si dos clonotipos candidatos están unidos depende de sus respectivas frecuencias o recuentos de lectura (como se indicó anteriormente), el número de diferencias de bases entre ellas (cuantas más diferencias, menos probable es la coalescencia) y los puntajes de calidad de las bases en los lugares en los que difieren las secuencias respectivas (los puntajes de mayor calidad hacen que la coalescencia sea menos probable). Los clonotipos candidatos pueden considerarse en el orden de sus respectivas frecuencias. La Figura 10A muestra el clonotipo 1 candidato (1532), el clonotipo 7 candidato (1534) y el clonotipo 11 candidato (1536) como los tres candidatos con las tres frecuencias más altas. Relacionado con cada uno de dichos clonotipos candidatos hay otros clonotipos candidatos que están próximos en secuencia, pero con frecuencias menores, tal como (i) para el clonotipo candidato 1 (1532), existen el clonotipo candidato 2 (1538) y los clonotipos candidatos 3, 4, 5 y 6 encerrados por un cono (1540); para el clonotipo candidato 7 (1534) existen los clonotipos candidatos 8, 9 y 10 encerrados por un cono (1542); y (iii) para el clonotipo candidato 11, existe el clonotipo candidato 12 encerrado por un cono (1544). Los conos representan límites de probabilidad dentro de los cuales un clonotipo candidato de menor frecuencia coalescería con uno de los clonotipos candidatos de mayor frecuencia 1, 7 u 11. Tales límites de probabilidad son funciones de la frecuencia de los clonotipos candidatos cercanos (3, 4, 5 y 6 para 1; 8, 9 y 10 para 7; y 12 para 11) y sus distancias en el espacio de la secuencia desde los respectivos clonotipos candidatos de mayor frecuencia. El clonotipo candidato 2 (1538) está fuera del cono (1540); por lo tanto, no coalescería con el clonotipo candidato 1 (1532). Nuevamente, los límites de probabilidad (de coalescencia) se muestran como conos debido a que los clones candidatos con frecuencias más altas tienen más probabilidades de ser clonotipos genuinamente diferentes que los de frecuencias más bajas y las diferencias múltiples en frecuencias más bajas son probablemente errores que múltiples diferencias en frecuencias más altas.

La nube de lecturas de secuencia que rodea a cada clonotipo candidato puede modelarse usando la distribución binomial y un modelo simple para la probabilidad de un error de base única. Este último modelo de error se puede inferir de la asignación de segmentos V y J o del algoritmo de búsqueda de clonotipos en sí, a través de la autoconsistencia y la convergencia. Se construye un modelo para la probabilidad de que una secuencia de 'nube' dada Y con errores de conteo de lectura C2 y E (con respecto a la secuencia X) sea parte de una secuencia del clonotipo verdadero X con un conteo de lectura perfecto C1 bajo el modelo nulo de que X es el único clonotipo verdadero en esta región del espacio de la secuencia. Se toma la decisión de coalescer o no la secuencia Y en el clonotipo X de acuerdo con los parámetros C1, C2 y E. Para cualquier C1 y E dado, se calcula previamente un valor máximo C2 para

decidir la coalescencia de la secuencia Y. Los valores máximos para C2 se eligen de modo que la probabilidad de que no coalesca Y bajo la hipótesis nula de que Y es parte del clonotipo X es menor que un valor P después de integrar todas las secuencias posibles Y con el error E en la vecindad de la secuencia X. El valor P controla el comportamiento del algoritmo y hace que la coalescencia sea más o menos permisiva.

5 Si una secuencia Y no coalesce en el clonotipo X porque su recuento de lectura está por encima del umbral C2 para que coalesca en el clonotipo X, entonces se convierte en un candidato para sembrar clonotipos separados (tal como con el clonotipo candidato 2 (1538) en la Figura 10A). Un algoritmo que implemente tales principios también se aseguraría de que cualquier otra secuencia Y2, Y3, etc., que esté 'más cerca' de esta secuencia Y (que se había considerado independiente de X) no esté agregada en X. Este concepto de 'cercanía' incluye tanto conteos de error con respecto a Y y X como el recuento absoluto de lectura de X e Y, es decir, se modela de la misma manera que el modelo anterior para la nube de secuencias de error alrededor del clonotipo X. De esta manera, las secuencias de 'nube' pueden atribuirse adecuadamente a su clonotipo correcto si ocurre que están más 'cerca' que un clonotipo. Por lo tanto, con respecto a la a la Figura 10A, si se considera que el clonotipo candidato 2 es genuinamente distinto del clonotipo candidato 1 (1532), entonces una rutina especial o subalgoritmo proporcionaría una regla para determinar cuál de los clonotipos candidatos 1 (1532) y 2 (1538), candidatos 4 y 5, entre 1 y 2, deben coalescer (si acaso).

En una realización, un algoritmo procede de arriba hacia abajo comenzando con la secuencia X con el recuento de lectura más alto. Esta secuencia siembra el primer clonotipo. Las secuencias vecinas se unen en este clonotipo si sus recuentos están por debajo de los umbrales calculados previamente (ver arriba), o se dejan solos si están por encima del umbral o 'más cerca' de otra secuencia que no se unió. Después de buscar todas las secuencias vecinas dentro de un recuento máximo de error, el proceso de lecturas de coalescencia en el clonotipo X se finaliza. Sus lecturas y todas las lecturas que han coalescido en ellas se contabilizan y eliminan de la lista de lecturas disponibles para hacer otros clonotipos. Luego se pasa a la siguiente secuencia con el recuento de lectura más alto. Las lecturas vecinas coalescen en este clonotipo como se indicó anteriormente y este proceso continúa hasta que no haya más secuencias con recuentos de lectura por encima de un umbral dado, por ejemplo, hasta que todas las secuencias con más de 1 recuento se hayan utilizado como semillas para clonotipos.

Como se mencionó anteriormente, en otra realización del algoritmo anterior, se puede agregar una prueba adicional para determinar si coalesce una secuencia candidata Y en un clonotipo X existente, que tiene en cuenta la puntuación de calidad de las lecturas de secuencia relevantes. El puntaje o puntajes promedio de calidad se determinan para la secuencia o secuencias Y (promediada en todas las lecturas con la secuencia Y) en as que las secuencias Y y X difieren. Si el puntaje promedio está por encima de un valor predeterminado, entonces es más probable que la diferencia indique un clonotipo verdaderamente diferente que no debería coalescer y si el puntaje promedio está por debajo de dicho valor predeterminado, entonces es más probable que la secuencia Y sea causada por errores de secuencia y, por lo tanto, debe coalescer en X.

La implementación exitosa del algoritmo anterior para la coalescencia de los clonotipos candidatos depende de tener una forma eficiente de encontrar todas las secuencias con menos errores E (es decir, menos de una medida de distancia de secuencia) de una secuencia de entrada X. Este problema puede resolverse usando un árbol de secuencia. La implementación de tales árboles tiene algunas características inusuales en que los nodos del árbol no se limitan a ser letras únicas de las secuencias de ADN de los clonotipos candidatos, como se ilustra en la Figura 10D. Los nodos pueden tener secuencias arbitrariamente largas, lo que permite un uso más eficiente de la memoria del ordenador.

Todas las lecturas de una muestra dada se colocan en el árbol de secuencia. Cada nodo de hoja contiene punteros para sus lecturas asociadas. Una secuencia única de un clonotipo candidato se recupera atravesando hacia atrás en el árbol desde la hoja hasta el nodo de la raíz. La primera secuencia se coloca en un árbol simple con un nodo de raíz y un nodo de hoja que contiene la secuencia completa de la lectura. Las secuencias se agregan a continuación una por una. Para cada secuencia agregada, se forma una nueva rama en el último punto de secuencia común entre la lectura y el árbol existente o agrega la lectura a un nodo de hoja existente si el árbol ya contiene la secuencia. Después de haber colocado todas las lecturas en el árbol, es fácil usar el árbol para los siguientes propósitos: 1) encontrar el recuento de lectura más alto: clasificar los nodos de hoja por recuento de lectura permite encontrar el nodo de hoja (es decir, la secuencia) con la mayor cantidad de lecturas y sucesivamente menores números de lecturas; 2) encontrar hojas vecinas: para cualquier secuencia, todas las rutas a través del árbol que tienen menos de X errores con respecto a esta secuencia se pueden buscar. Se inicia una ruta en la raíz y se ramifica en rutas separadas que avanzan a lo largo del árbol. Se anota el recuento de errores actual de cada ruta a medida que avanza el árbol. Cuando el recuento de errores excede los errores máximos permitidos, la ruta se da por finalizada. De esta forma, se podan grandes partes del árbol lo antes posible. Esta es una manera eficiente de encontrar todas las rutas (es decir, todas las hojas) dentro de los errores X de cualquier secuencia dada.

Las características de los conceptos anteriores se ilustran con más detalle en el diagrama de flujo de la Figura 10B. Se obtiene un conjunto de clonotipos candidatos a partir de datos de secuencia obtenidos secuenciando ácidos nucleicos recombinados extraídos de una muestra de células T o células B. En un aspecto, los clonotipos candidatos incluyen una región NDN y porciones de regiones V y J. Estas secuencias están organizadas en una estructura de datos (1550), que puede ser un árbol de secuencias. No se muestra en la Figura 10B, como parte de la generación de un conjunto de clonotipos candidatos, en una realización, los árboles de secuencia también pueden construirse para

regiones V conocidas y regiones J conocidas. Las lecturas de secuencia que forman un clonotipo candidato pueden luego mapearse, o alinearse, con estas secuencias conocidas a través de los árboles de secuencias para determinar eficientemente las secuencias V y J conocidas más probables de los clonotipos candidatos. Volviendo a la Figura 10B, una vez que se generan los clonotipos candidatos, se construye una estructura de datos, tal como un árbol de secuencia, para usar en un método para distinguir los clonotipos genuinos de los clonotipos candidatos que contienen errores experimentales o de medición, tal como errores de secuenciación. El clonotipo candidato que tiene la frecuencia más alta de ocurrencia entre los clonotipos candidatos actuales (HFCC_k) se selecciona (1552) de la estructura de datos, por ejemplo, un árbol de secuencia; en otras palabras, HFCC_k es el clonotipo candidato con el mayor número de copias o recuentos de lectura en el ciclo k. A continuación, se identifican los clonotipos candidatos vecinos de menor frecuencia (LFCC) (1554); es decir, se identifican los clonotipos candidatos dentro una distancia de D_k. En un aspecto de la invención, esta identificación se lleva a cabo usando un árbol de secuencia, que permite comparaciones de secuencia eficientes de secuencias relativamente cortas (<300 pb).

En una realización, las comparaciones, o alineamientos de secuencia, se llevan a cabo usando programación dinámica, por ejemplo, de acuerdo con lo revelado por Gusfield (citado anteriormente). En una realización adicional, dicha programación dinámica es programación dinámica en bandas en la que las secuencias que difieren del HFCC seleccionado en más de una distancia predeterminada no se consideran, lo que acelera el cálculo. Los candidatos HFCC_k y LFCC_j pueden compararse en función de muchos criterios o propiedades diferentes. En un aspecto, como se mencionó anteriormente, los clonotipos candidatos se comparan sobre la base de al menos dos propiedades: (i) frecuencia o recuentos de lectura y (ii) diferencias de secuencia. En otro aspecto, como se mencionó anteriormente, los clonotipos candidatos se comparan sobre la base de al menos tres propiedades: (i) frecuencia o recuentos de lectura, (ii) diferencias de secuencia y (iii) puntajes o medidas de calidad de las bases donde se producen diferencias. En una realización, las diferencias de secuencia incluyen sustituciones de bases; en otra realización, las diferencias de secuencia incluyen sustituciones, eliminaciones e inserciones de bases. La última realización es especialmente aplicable siempre que los datos de secuencia se generen mediante químicas de secuenciación por síntesis que no emplean terminadores, tales como secuenciadores 454 y secuenciadores Ion Torrent. Tales enfoques de secuenciación diferencian tramos de homopolímeros de diferentes tamaños por amplitud de señal; por lo tanto, las rutinas de llamada de bases en tales enfoques son propensas a errores de inserción y eliminación, porque la diferencia en el nivel de señal de los homopolímeros que difieren en un nucleótido cae precipitadamente al aumentar el tamaño del homopolímero (es decir, un 2 mer se distingue fácilmente de un 3 mer, pero un 8 mer es casi indistinguible de un 9 mer). En un aspecto, las comparaciones de HFCC y LFCC pueden implementarse utilizando una función (denominada en el presente documento "función de probabilidad de coalescencia"), tal como P(HFCC_k, LFCC_j, D, Q) que se muestra en el cuadro de decisión (1558), que depende de las cantidades (i) a (iii) descritas anteriormente. Tal función puede tomar muchas formas diferentes, pero generalmente el valor de P cambia con los cambios en (i), (ii) y (iii) de la siguiente manera: el valor de P preferiblemente aumenta monótonicamente con la frecuencia de HFCC y la relación de la frecuencia de HFCC a la de LFCC, de modo que cuanto mayor sea la proporción de la frecuencia de HFCC con respecto a la de LFCC, mayor será la probabilidad de que LFCC coalesca en HFCC. Del mismo modo, el valor de P preferiblemente disminuye monótonicamente con el grado en que difieren las secuencias de HFCC y LFCC, de modo que cuanto mayor sea la diferencia entre HFCC y LFCC (por ejemplo, medido por el número mínimo de sustituciones, inserciones o eliminaciones para cambiar una por otro) menor es la probabilidad de que LFCC coalesca con HFCC. Finalmente, el valor de P preferiblemente disminuye monótonicamente al aumentar la calidad de los puntajes de los lugares donde las secuencias de HFCC y LFCC difieren, de modo que para los puntajes de mayor calidad, menor será la probabilidad de que LFCC coalesca con HFCC.

Cuando las secuencias de HFCC y LFCC difieren en más de una ubicación, los puntajes de calidad en las diferentes ubicaciones pueden combinarse de varias maneras diferentes. En una realización, siempre que haya una pluralidad de tales diferencias, la pluralidad de puntajes de calidad se expresa como un valor promedio, que puede ser un promedio no ponderado o un promedio ponderado. La Figura 10C muestra un ejemplo de una función, P, calculada para diferentes valores de calidad (curvas de la a a la e) para una diferencia de secuencia dada. Como se ilustra en la Figura 10C, cada vez que el HFCC alcanza un nivel de aproximadamente 200 recuentos de lectura (1570), si los puntajes de calidad se determinan mediante la curva (a), cualquier LFCC con menos de aproximadamente 50 recuentos de lectura (1572) coalescen en HFCC. El argumento, D, de la función P es una medida de la distancia entre las secuencias HFCC_k y LFCC_j y su valor puede variar de un ciclo a otro a medida que avanza un análisis. (Los índices "k" indican que los valores de las constantes con un subíndice "k" pueden depender del ciclo computacional, k). En una realización, D = D_k, de modo que su valor es una función del número de ciclo. En otra realización, D = D (frecuencia de HFCC), de modo que su valor es una función de la frecuencia de HFCC, independiente del número de ciclo. Por ejemplo, a medida que disminuye la frecuencia de HFCC, disminuye la distancia, D, de los candidatos a comparar. En una realización, D es una distancia de Hamming entre HFCC_k y LFCC_j; sin embargo, se pueden usar otras medidas de distancia. En una realización, D_k es una función no creciente de k; y en otra realización, D_k es una función decreciente de k. Disminuir la magnitud de D con el aumento del número de ciclos, o con la frecuencia decreciente de HFCC, es ventajoso en algunas realizaciones porque a medida que el cálculo avanza a clonotipos candidatos de frecuencia cada vez más baja, la mayoría de dichos candidatos son singletones, de modo que la distancia de secuencia (en lugar de la diferencia de frecuencia) se convierte en la comparación predominante. Al reducir D a medida que avanza el cálculo, se reducen las comparaciones improductivas con clonotipos candidatos distantes de baja frecuencia, lo que acelera el cálculo. La función P puede ser una expresión complicada dependiendo del número de factores considerados. La Figura 10C ilustra valores calculados para una realización de P que relaciona umbrales de

recuento de lectura para coalescer un LFCC dado un recuento de lectura de un HFCC para diferentes puntajes de calidad, como se describió anteriormente. Las curvas "a" a "e" representan las relaciones para diferentes puntajes de calidad (con la curva "a" correspondiente al puntaje de mayor calidad).

5 Volviendo a la Figura 10B, si $P < P_k$, entonces LFCC_j no coalesce con HFCC_k y se selecciona otro LFCC (1560). Si $P >$
 P_k, entonces LFCC_j coalesce con HFCC_k (1562), en cuyo caso se selecciona otro LFCC (1566), a menos que no haya
 más LFCC para evaluar (1564). Si no hay más LFCC para evaluar (1564), entonces el HFCC_k actual (incluyendo todos
 los LFCC unidos en él) se elimina (1568) de la estructura de datos, como el árbol de secuencia. Dicha eliminación se
 10 ilustra en el árbol de secuencia simple (1590) de las Figuras 10D-10E. Allí, la ruta (1592) (indicada por una línea
 punteada) en el árbol de secuencia (1590) corresponde a HFCC (1596), que coalesce con LFCC (1598). Después de
 la coalescencia, el segmento de ruta (1592) en el área sombreada (1599) se elimina del árbol de secuencia (1590)
 para obtener un árbol de secuencia reducido (1597) que se muestra en la Figura 10E, que se utiliza en cálculos
 posteriores para encontrar LFCC vecino (1554). Después de dicha eliminación, la determinación del clonotipo finaliza
 si se cumple un criterio de detención (1570). En una realización, el criterio de detención (1570) es si se ha procesado
 el último clonotipo candidato no singletón (1552). En otra realización, el criterio de detención (1570) es si la frecuencia
 15 o los recuentos de lectura del HFCC seleccionado son inferiores a los correspondientes a un solo linfocito. En un
 aspecto del método de la invención, una etapa de amplificación puede dar como resultado que cada linfocito en una
 muestra esté representado por múltiples copias del mismo clonotipo; por lo tanto, en una realización, cualquiera que
 sea el HFCC tiene un número de recuentos de lectura por debajo del número correspondiente a un solo linfocito,
 entonces se detiene el cálculo. En algunas realizaciones, tal número de recuentos leídos (o copias de clonotipo
 20 candidato) es al menos 10; en otra realización, dicho número es al menos 20; en otra realización, dicho número es al
 menos 30; en otra realización, dicho número es al menos 40. Si no se cumple el criterio de detención, se selecciona
 el siguiente HFCC (1572). Las etapas analíticas resumidos en el diagrama de flujo de la Figura 10B puede
 implementarse en cualquier lenguaje de programación adecuado, tal como C, C++, Java, C #, Fortran, Pascal o
 similares.

25 De acuerdo con un aspecto de la invención, el método anterior para determinar clonotipos y/o perfiles de clonotipos
 comprende las etapas de (a) formación de una estructura de datos de moléculas inmunes recombinadas a partir de
 lecturas de secuencia obtenidas por secuenciación de ácido nucleico de alto rendimiento, (b) coalescencia con un
 clonotipo candidato de frecuencia más alta cualquier clonotipo candidato de frecuencia menor siempre que dicha
 frecuencia menor esté por debajo de un valor de frecuencia predeterminado y una diferencia de secuencia entre ellos
 30 esté por debajo de un valor de diferencia predeterminado para formar un clonotipo, (c) eliminar el clonotipo candidato
 unido de la estructura de datos y (d) repetir las etapas (b) y (c) hasta que se forme un perfil de clonotipo. En una
 realización, la estructura de datos es un árbol de secuencia.

De acuerdo con otro aspecto de la invención, el método anterior para determinar los clonotipos puede llevarse a cabo
 mediante etapas que comprenden: (a) proporcionar un conjunto de lecturas de secuencia de un repertorio de
 35 moléculas inmunes recombinadas que tienen cada una, una región V, una región NDN y una región J en la que para
 cada una de tales moléculas, al menos una lectura de secuencia abarca al menos una porción de la región NDN
 de dicha molécula; (b) formar a partir de las lecturas de secuencia que abarcan al menos una porción de una región NDN
 un árbol de secuencia que tiene hojas que representan clonotipos candidatos, cada hoja y su correspondiente clonotipo
 40 candidato tienen una frecuencia; (c) coalescencia con un clonotipo candidato de frecuencia más alta cualquier
 clonotipo candidato de frecuencia menor siempre que dicha frecuencia menor esté por debajo de un valor de frecuencia
 predeterminado y una diferencia de secuencia entre ellos esté por debajo de un valor de diferencia predeterminado
 para formar el clonotipo que tenga una secuencia del clonotipo candidato de frecuencia más alta; (d) eliminar las hojas
 correspondientes a los clonotipos candidatos unidos del árbol de secuencia; y (e) repetir las etapas (c) y (d) hasta que
 45 una frecuencia más alta de un clonotipo candidato de menor frecuencia esté por debajo de un valor de detención
 predeterminado. En una realización, la etapa de formación incluye además seleccionar un clonotipo candidato de
 frecuencia más alta e identificar todos dichos clonotipos candidatos de frecuencia menor que tienen una diferencia de
 secuencia con el mismo menor que un valor de diferencia predeterminado para formar un subconjunto de
 coalescencia. Por lo tanto, en tal realización, se puede limitar el número total de LFCC que se deben comparar para
 la operación de coalescencia (solo se consideran los que se encuentran dentro del valor de diferencia predeterminado).
 50 Dicho valor es una entrada de proceso que depende de la aplicación, por ejemplo, el tamaño del repertorio, la cantidad
 de tiempo de computación utilizado, etc. Como se mencionó anteriormente, la función utilizada para decidir si coalesce
 un HFCC con un LFCC puede tener una variedad de formas. En un aspecto general, para la etapa de coalescencia,
 dicha función puede tener las siguientes propiedades: depende de las frecuencias de HFCC, LFCC, la diferencia de
 secuencia entre ellas (que puede expresarse como una medida de diferencia de cuerda convencional, tal como una
 55 distancia de Hamming) y puntajes de calidad de una o más ubicaciones de nucleótidos en el que difieren el HFCC y
 el LFCC; tal que la función (i) aumenta monótonicamente al aumentar la proporción de frecuencia de HFCC y de
 frecuencia de LFCC, (ii) disminuye monótonicamente al aumentar la diferencia de secuencia entre HFCC y LFCC, y
 (iii) disminuye monótonicamente al aumentar los puntajes de calidad de una o más ubicaciones de nucleótidos. Es
 decir, con respecto a la propiedad (iii), lo más seguro es que HFCC y LFCC sean diferentes (por ejemplo, porque hay
 60 un alto nivel de confianza en las llamadas de base), entonces es menos probable que coalezcan.

En algunas realizaciones, se selecciona una función de probabilidad de coalescencia de modo que las lecturas de
 secuencia coalezcan en diferentes clonotipos (o polinucleótidos objetivo, tales como ácidos nucleicos recombinados)
 siempre que dichas lecturas de secuencia sean distintas con una probabilidad de al menos 95 por ciento; en otras

realizaciones, se selecciona una función de probabilidad de coalescencia de modo que las lecturas de secuencia coalezcan en diferentes clonotipos siempre que dichas lecturas de secuencia sean distintas con una probabilidad de al menos 99 por ciento; en otras realizaciones, se selecciona una función de probabilidad de coalescencia de modo que las lecturas de secuencia coalezcan en diferentes clonotipos siempre que dichas lecturas de secuencia sean distintas con una probabilidad de al menos 99,9 por ciento. Como se mencionó anteriormente, en algunas realizaciones, una función de probabilidad de coalescencia depende de una tasa de error de una química de secuenciación utilizada, el número de nucleótidos discrepantes en las lecturas de secuencia que se comparan, y las frecuencias relativas de las lecturas de secuencia que se comparan; en otra realización, una función de probabilidad de coalescencia depende de una tasa de error de una química de secuenciación utilizada, el número de nucleótidos discrepantes en las lecturas de secuencia que se comparan, las frecuencias relativas de las lecturas de secuencia que se comparan y los puntajes de calidad de los nucleótidos discrepantes. En lo anterior, la selección de un valor de frecuencia predeterminado y un valor de diferencia predeterminado es una elección de diseño que depende de aplicaciones particulares. Los factores que afectan tales elecciones pueden incluir detalles de la biología, la velocidad de implementación y similares.

15 Análisis del repertorio de TCRB

En este ejemplo, se analizan las cadenas de TCR β y se determinan los clonotipos. El análisis incluye amplificación, secuenciación y análisis de las secuencias de TCR β . Un cebador es complementario a una secuencia común en C β 1 y C β 2, y hay cebadores de 34 V capaces de amplificar todos los segmentos de 48 V. C β 1 o C β 2 difieren entre sí en las posiciones 10 y 14 de la unión J/C. El cebador para C β 1 y C β 2 termina en la posición 16 pb y no tiene preferencia por C β 1 o C β 2. Los cebadores de 34 V se modifican de un conjunto original de cebadores revelados en Van Dongen et al., publicación de la patente de Estados Unidos No. 2006/0234234. Los cebadores modificados se describen en Faham et al., publicación de la patente de los Estados Unidos No. 2010/0151471.

El analizador de genoma Illumina se usa para secuenciar el amplicón producido por los cebadores anteriores. Se realiza una amplificación de dos etapas en transcripciones de ARN mensajero (200), como se ilustra en las Figuras 2A-2B, la primera etapa que emplea los cebadores anteriores y una segunda etapa para agregar cebadores comunes para amplificación puente y secuenciación. Como se muestra en la Figura 2A, se realiza una PCR primaria usando en un lado un cebador de 20 pb (202) cuyo extremo 3' está a 16 bases de la unión J/C (204) y que es perfectamente complementario a C β 1 (203) y los dos alelos de C β 2. En la región V (206) de los transcritos de ARN (200), se proporciona un conjunto de cebadores (212) que contiene secuencias de cebadores complementarias a las diferentes secuencias de la región V (34 en una realización). Los cebadores del conjunto (212) también contienen una cola no complementaria (214) que produce amplicón (216) que tiene un sitio de unión del cebador (218) específico para cebadores P7 (220). Después de una PCR multiplex convencional, se forma un amplicón (216) que contiene la porción altamente diversa de la región J(D)V (206, 208 y 210) de los transcritos de ARNm y sitios de unión de cebadores comunes (203 y 218) para una amplificación secundaria para agregar una etiqueta de muestra (221) y cebadores (220 y 222) para la formación de grupos mediante PCR de puente. En la PCR secundaria, en el mismo lado de la plantilla, se usa un cebador (222 en la Figura 2B y al que se hace referencia en este documento como "C10-17-P5") que tiene en su extremo 3' la secuencia de las 10 bases más cercanas a la unión J/C, seguida de 17 pb con la secuencia de las posiciones 15-31 desde la unión J/C, seguida de la secuencia P5 (224), que desempeña un papel en la formación de grupos por PCR de puente en la secuenciación de Solexa. (Cuando el cebador C10-17-P5 (222) se hibrida a la plantilla generada a partir de la primera PCR, se crea un bucle de 4 pb (posición 11-14) en la plantilla, a medida que el cebador se hibrida con la secuencia de las 10 bases más cercanas a la unión J/C y las bases en las posiciones 15-31 desde la unión J/C. El bucle de las posiciones 11-14 elimina la amplificación diferencial de las plantillas que portan C β 1 o C β 2. La secuenciación se realiza con un cebador complementario a la secuencia de las 10 bases más cercanas a la unión J/C y las bases en las posiciones 15-31 de la unión J/C (este cebador se llama C'). El cebador C10-17-P5 puede purificarse por HPLC para garantizar que todo el material amplificado tiene extremos intactos que pueden ser utilizados eficientemente en la formación del grupo).

En la Figura 2A, la longitud del voladizo en los cebadores V (212) es preferiblemente de 14 pb. La PCR primaria se ayuda con un voladizo más corto (214). Alternativamente, por el bien de la PCR secundaria, el voladizo en el cebador V se usa en la PCR primaria el mayor tiempo posible porque la PCR secundaria se ceba a partir de esta secuencia. Se investigó un tamaño mínimo de voladizo (214) que soporta una PCR secundaria eficiente. Se realizaron dos series de cebadores V (para dos segmentos V diferentes) con tamaños de voladizo de 10 a 30 con etapas de 2 pb. Usando las secuencias sintéticas apropiadas, se realizó la primera PCR con cada uno de los cebadores de la serie y se realizó electroforesis en gel para mostrar que todos se amplificaron.

Como se ilustra en la Figura 2A, la PCR primaria usa 34 cebadores V diferentes (212) que se hibrida con la región V (206) de las plantillas de ARN (200) y contienen un voladizo común de 14 pb en la cola 5'. Los 14 pb es la secuencia parcial de uno de los cebadores de secuenciación Illumina (denominado el cebador Read 2). El cebador de amplificación secundario (220) en el mismo lado incluye la secuencia P7, una etiqueta (221) y la secuencia del cebador Read 2 (223) (este cebador se llama Read2_tagX_P7). La secuencia P7 se usa para la formación de grupos. El cebador Read 2 y su complemento se usan para secuenciar el segmento V y la etiqueta respectivamente. Se crea un conjunto de 96 de estos cebadores con etiquetas numeradas del 1 al 96 (véase más abajo). Estos cebadores se purifican por HPLC para garantizar que todo el material amplificado tenga extremos intactos que puedan utilizarse de

manera eficiente en la formación del grupo.

Como se mencionó anteriormente, el cebador de la segunda etapa, C-10-17-P5 (222, Figura 2B) ha interrumpido la homología con la plantilla generada en la primera etapa de la PCR. La eficacia de la amplificación con este cebador ha sido validada. Un cebador alternativo para C-10-17-P5, denominado CsegP5, tiene una homología perfecta con el cebador de la primera etapa C y la cola 5' que porta P5. La eficacia del uso de C-10-17-P5 y CsegP5 en la amplificación de plantillas de PCR de la primera etapa se comparó mediante la realización de PCR en tiempo real. En varias réplicas, se encontró que la PCR que usaba el cebador C-10-17-P5 tenía poca o ninguna diferencia en la eficiencia en comparación con la PCR que usaba el cebador CsegP5.

El amplicón (230) resultante de la amplificación de 2 etapas ilustrada en las Figuras 2A-2C tiene la estructura típicamente utilizada con el secuenciador Illumina como se muestra en la Figura 2C. Dos cebadores que se hibridan con la parte más externa de la molécula, los cebadores Illumina P5 y P7 se usan para la amplificación de la molécula en fase sólida (formación de grupos). Se realizan tres lecturas de secuencia por molécula. La primera lectura de 100 pb se realiza con el cebador C', que tiene una temperatura de coalescencia adecuada para el proceso de secuenciación de Illumina. La segunda lectura tiene solo 6 pb de longitud y es únicamente con el propósito de identificar la etiqueta de la muestra. Se genera utilizando un cebador de etiquetas proporcionado por el fabricante (Illumina). La lectura final es el cebador Read 2, también proporcionada por el fabricante (Illumina). Usando este cebador, se genera una lectura de 100 pb en el segmento V comenzando con la primera secuencia del cebador V de PCR.

Ejemplo

En este ejemplo, las etapas comunes a algunos casos, tal como el caso de la Figura 1C, se describen para aplicaciones particulares, que incluyen, pero no se limitan, la exposición de una muestra de tejido que comprende células T al antígeno, la activación de las células T en una muestra de tejido por antígeno, la obtención de ácidos nucleicos recombinados de las células T de una muestra de tejido, el aislamiento (o recuperación, o clasificación o separación) de células T activadas, la secuenciación de ácidos nucleicos recombinados, la formación de clonotipos y la determinación de clonotipos de células T específicas de antígeno.

Muestras de tejido. Las PBMC caracterizadas se adquirieron a través de Cellular Technology Limited. Las células se descongelaron, se lavaron y se lisaron con tampón RLT plus (Qiagen) para la purificación de ácido nucleico o se cultivaron durante la noche en presencia de péptidos (véase más adelante) para identificar células T específicas de antígeno.

Ensayos de células T específicas de antígeno, citometría de flujo y clasificación celular. Las células específicas de antígeno se identificaron usando una variedad de ensayos: unión de pentámero, sobreexpresión de marcadores de superficie celular (CD137, CD107) después de la incubación de péptidos a corto plazo y proliferación después de la incubación de péptidos a relativamente largo plazo. Las células T específicas de pentámero se identificaron incubando PBMC con pentámero Hpp65495_504 del HCMB (ProlImmune) de acuerdo con las instrucciones del fabricante. Los procedimientos para obtener células T específicas de antígeno viables basadas en la adquisición de los marcadores de superficie celular CD137/107 (para células T específicas de antígeno CD8) después de una breve incubación *in vitro* con péptidos son bien conocidos, por ejemplo, Chattopadhyay et al., Nature Medicine, 11: 1113-1117 (2005); Meier et al., Cytometry A, 73: 1035-1042 (2008); Wolf et al., Blood, 110: 201-210 (2007); Wolf et al., Cytometry A, 73: 1043-1049 (2008); y similares. En resumen, se usaron medios completos que contenían suero fetal bovino al 15% (FBS), aminoácidos no esenciales, glutamina y antibióticos para incubaciones de péptidos. Las PBMC descongeladas se lavaron y suspendieron a razón de ~400.000 células por pozo (placas de fondo en U de 96 pozos) en medio completo. Los anticuerpos no conjugados dirigidos contra CD28 y CD49d se añadieron luego a los pozos que contenían las células suspendidas. Los péptidos derivados de pp65 de CMV (HCMVA (pp65) (JPT Peptide Technologies) se agregaron directamente a la mezcla de células/anticuerpos, de acuerdo con las instrucciones del fabricante. Un péptido único derivado de pp65 de CMV (secuencia NLVPMVATV; en lo sucesivo denominado 'pp65₄₉₅') se usó a razón de 2 µg/mL. Se añadieron directamente pp65 "PepMix" y los conjuntos de péptidos CEF + (JPT Peptide Technologies) a la mezcla de células/anticuerpos, de acuerdo con las instrucciones del fabricante. Después de la adición de los péptidos, las células se incubaron a 37 °C durante ~18 horas. Las incubaciones de control negativo se prepararon como se describió anteriormente sin la adición de péptidos. Al final de la incubación, las células se recogieron del cultivo y se tñieron con anticuerpos para análisis y clasificación por citometría de flujo. Para cada ensayo específico de antígeno CD8 (CD137 y CD107), se usaron anticuerpos conjugados fluorescentemente a los siguientes marcadores de superficie celular para citometría de flujo: CD8, CD3 y CD137 o CD107a y CD107b. Las células se lavaron y suspendieron en PBS que contenía FBS (2%) y 4',6-diamidino-2-fenilindol (DAPI) para la exclusión de células no viables. Se incubaron PBMC marcadas con diacetato de carboxifluoresceína, éster succinimidílico (CFSE) como se describió anteriormente durante 6 días en presencia de péptidos y anticuerpos dirigidos contra CD28 y CD49d. Las células T CD8+ específicas de antígeno se identificaron y clasificaron de acuerdo con la dilución de CFSE en el día 6. Las células se adquirieron y clasificaron usando un instrumento FACSaria (BD Biosciences). Las células clasificadas específicas de antígeno (CD3+ CD8+ CMV pentámero+, CD3+ CD8+ CD137+, CD3+ CD8+ CD107a/b+ o CD8+ CFSE^{bajo}) y no específicas de antígeno (CD3+ CD8+ CD137-, CD3+ CD8+ CD107a/b-) se sedimentaron y se lisaron en tampón RLT Plus para aislamiento de ácido nucleico. El análisis de los archivos de datos de citometría de flujo se realizó utilizando FlowJo (Ashland, OR).

Preparación de ARN y ADNc. El ARN (y el ADN) se aisló utilizando mini y/o micro kits de ADN/ARN AllPrep, de acuerdo con las instrucciones del fabricante (Qiagen). El ARN se transcribió en forma inversa a ADNc usando kits Vilo (Life Technologies).

5 Amplificación y secuenciación de TCR β . El ADNc se amplificó usando conjuntos de cebadores específicos de locus para TCR β . Esta reacción de amplificación amplificó de manera reproducible todas las transcripciones de ARN posibles que se encuentran en la muestra que contiene el locus de TCR β reorganizado independientemente de qué segmento variable (V) y de qué alelo de la región constante (C) común cada molécula reorganizada poseía, al tiempo que agrega las secuencias necesarias para la formación de grupos e indexación de muestras.

10 Los cebadores de la primera etapa se diseñaron para permitir la amplificación de todos los alelos conocidos de las secuencias de la línea germinal, tal como se describió anteriormente y a continuación; Faham et al., Blood, 120: 5173-5180 (2012). En los extremos 5' de los cebadores del segmento V, se añadieron secuencias universales complementarias a los cebadores de PCR de la segunda etapa. Los cebadores se optimizaron de modo que cada posible segmento V y C se amplificara a una velocidad similar para sesgar mínimamente la distribución de frecuencia del repertorio durante el proceso de amplificación. La especificidad de los cebadores, por el contrario, no se optimizó ya que las secuencias del cebador podrían mapearse y eliminarse de la lectura eventual de la secuencia. Por lo tanto, una secuencia dada puede haber sido amplificada por múltiples cebadores.

20 En la segunda etapa de la PCR, los cebadores en el lado C se hibridaron con el segmento C con una cola 5' que contenía el cebador de secuencia y la secuencia P5 utilizada para la formación de grupos en el secuenciador del analizador de genoma Illumina. Los cebadores en el lado V hibridaron con el segmento V con una cola 5' que contenía el cebador de secuencia y la secuencia P7 utilizada para la formación de grupos. Para cada muestra, se utiliza un par de cebadores en la segunda etapa. En el lado C, siempre es el mismo cebador. En el lado V, es uno de un conjunto de cebadores que difiere en un índice de 6 bases. Específicamente, los cebadores en el lado V de la amplificación constituían uno de un conjunto de cebadores, cada uno de los cuales tenía una región 3' que se hibridaba con la secuencia del voladizo adjunta en la primera reacción pero que además contenía uno de los múltiples índices de 6 pares de bases que permitían la multiplexación de la muestra en el secuenciador. Cada uno de estos cebadores contenía además una cola 5' con la secuencia P7 utilizada para la formación de grupos en el secuenciador del analizador de genoma Illumina.

30 La primera etapa de la PCR se llevó a cabo utilizando una polimerasa de alta fidelidad (AccuPrime, Life Technologies) durante 16 ciclos. Se llevó a cabo una segunda etapa de PCR durante 22 ciclos en 1/100 de los productos de amplificación de la primera etapa de PCR. Cada muestra contenía una etiqueta de identificación única. Las muestras se agruparon y purificaron usando el kit de purificación de PCR QIAquick (Qiagen). La formación y secuenciación del grupo en ambas direcciones se realizó de acuerdo con el protocolo del fabricante (Illumina, Inc., La Jolla, CA). Específicamente, se realizaron tres reacciones de secuenciación. Primero se secuenciaron 115 pb del lado C suficiente para secuenciar a través de la secuencia de unión de C a V. En este punto, la cadena sintetizada se desnaturalizó y se lavó. Se hibridó un segundo cebador de secuenciación que permitió secuenciar el índice de muestra durante 6 ciclos para identificar la muestra. En este punto, la cadena del complemento inverso se generó de acuerdo con el protocolo de Illumina. Se obtuvo una lectura de secuenciación final de 95 pb de la dirección V a C proporcionando una secuencia amplia para mapear el segmento V con precisión. Los datos de secuenciación se analizaron luego para determinar las secuencias de clonotipo, tal como se describió anteriormente.

40 Determinación de clonotipo. Se definió un clonotipo cuando se obtuvieron al menos 2 lecturas de secuencia idénticas. En resumen, después de la exclusión de lecturas de baja calidad, los datos de la secuencia se analizaron para determinar las secuencias del clonotipo, incluido el mapeo de las secuencias consenso V y J de la línea germinal. Primero, las secuencias de índice de muestra se usaron para identificar cuál de las secuencias se origina a partir de cuál de las muestras agrupadas. Se excluyeron las secuencias cuyo índice no coincidía perfectamente con uno de los índices utilizados en un proceso específica. Luego, la lectura directa se usó para mapear el segmento J. Como todas las secuencias comenzaron desde la misma posición de los segmentos J, todos los segmentos J comenzaron en una posición de secuenciación definida previamente. Los primeros 25 pb de los segmentos J se usaron para mapear el segmento J. Cualquier lectura con más de 5 desajustes de alta calidad con los segmentos J conocidos se excluyó de un análisis posterior.

50 Después de la identificación del segmento J, los segmentos V se mapearon. La lectura inversa se utilizó para este propósito. Primero, el cebador V fue mapeado y excluido. Posteriormente, las siguientes 70 bases de la lectura inversa se mapearon para los segmentos V conocidos. Se excluyeron las lecturas que no se asignaron a segmentos J y V. La siguiente etapa en el mapeo involucró la identificación del marco que relacionaba las lecturas directas e inversas y esto permitió construir una secuencia continua de J a V. Esto se realizó utilizando las últimas 15 bases de la lectura directa que estaban dentro del segmento V independientemente de la longitud de NDN. Si bien estas bases podrían ser de una calidad de secuencia relativamente más baja, ya que estaban en el extremo terminal de una lectura larga, podrían usarse para mapear dentro de un solo segmento V identificado con el fin de identificar la posición en la que podrían unirse las dos lecturas. Finalmente, se usaron las secuencias conocidas de V y J a las que se asignaron el mapa de lecturas para identificar el punto en la lectura directa en la que las secuencias en las uniones divergían de estos segmentos mapeados.

60

Después de la determinación del clonotipo, se analizaron las frecuencias relativas de los clonotipos en las poblaciones no clasificadas, específicas de antígeno y no específicas de antígeno. Las comparaciones de frecuencia de clonotipo entre dos muestras se muestran en varias figuras. Los clonotipos que están presentes en la muestra A pero no en la muestra B (en los que se comparan las frecuencias en la muestra A y B) se representan tener la frecuencia de un clonotipo con una sola lectura en la muestra B. Por lo tanto, la frecuencia del clonotipo faltante en una muestra depende de la profundidad de secuenciación de una muestra particular. En estos casos donde falta un clonotipo en una muestra, debido a que la frecuencia de una sola lectura se asigna a estos clonotipos, la frecuencia observada se sobreestima. Por lo tanto, en la gran mayoría de estos casos, es probable que se sobreestime la frecuencia real del clonotipo. Los clonotipos ausentes en ambas muestras aparecen donde se cruzan los ejes. Los clonotipos presentes en una muestra, pero no en la otra, se encuentran a lo largo de los ejes x o y.

Los clonotipos de los análisis de células T específicas de antígeno se seleccionaron con base en los siguientes tres criterios. Primero, los clonotipos tenían frecuencias que aumentaban al menos 10 veces en células clasificadas específicas de antígeno frente a no específicas de antígeno o no clasificadas, por ejemplo, la Figura 6 paneles A y B. Segundo, estos clonotipos también estaban en una frecuencia más baja en células clasificadas "no específicas de antígeno" en comparación con las células no clasificadas si eran mayores de 1/100.000 para evitar el error de submuestreo (ruido de Poisson) asociado con clonotipos de muy baja frecuencia en muestras clasificadas. Tercero, debido al número limitado de células específicas de antígeno de entrada después de la clasificación (generalmente menos de <30.000 células), se aplicó un umbral equivalente mayor que '20 células' con base en el número de entrada relativamente bajo de células en estas muestras. Este umbral mínimo permitió la exclusión de clonotipos que parecían enriquecidos en muestras clasificadas específicas de antígeno pero que se debían solo a la presencia de una o unas pocas células en la muestra. Por ejemplo, considere una población clasificada de 10.000 células pentámero+ de una muestra con un millón de células T. Si una sola célula con una frecuencia de 1 por millón en la muestra sin clasificar se clasifica de manera incidental en la muestra pentámero+, su frecuencia en la muestra clasificada será de 1/10.000 y parecería estar 100 veces enriquecida en la muestra pentámero+ en comparación con la muestra sin clasificar. Para mejorar este problema, se requirió un clonotipo para representar 20 células en la muestra clasificada de pentámero+. Específicamente, el umbral de frecuencia \log_{10} requerido en la muestra pentámero+ se calculó como $\log_{10}(1/(n/20))$, en la que n es el número de células clasificadas de entrada para esa muestra de acuerdo con lo determinado por citometría de flujo (por ejemplo, en la Figura 7 panel A, 16.281 es el número de células clasificadas de entrada y la frecuencia de umbral calculada es $10^{-2.9}$). Esas secuencias que reúnen los tres criterios descritos anteriormente se clasificaron como clonotipos de células T específicas de antígeno.

Resultados. La combinación de clasificación y secuenciación se usó para identificar clonotipos específicos de antígeno en un individuo con una respuesta positiva conocida a un antígeno de citomegalovirus (CMV). Primero, la secuenciación de TCR β se combinó con un ensayo inmune basado en multímeros para validar este método para la identificación de clonotipos de TCR β CD8 específicos de antígeno. Un péptido derivado de pp65 (495-404) de CMV (denominado en este documento como el "péptido pp65₄₉₅") es un epítipo inmunodominante restringido a HLA-A*0201 que induce respuestas en individuos positivos para CMV. Para identificar directamente las células T específicas de este antígeno, se utilizó un reactivo pentámero disponible comercialmente que contiene el péptido pp65₄₉₅ unido a una molécula de MHC. En principio, todas las células T que portan las secuencias que se unen al pentámero deben detectarse independientemente de su potencial funcional. Las células T CD8 específicas de pp65₄₉₅ se identificaron secuenciando el repertorio de TCR β de células que fueron clasificadas con base en la unión del pentámero (pentámero+).

Se obtuvieron PBMC congeladas de un individuo con una respuesta caracterizada para pp65₄₉₅ mediante el ensayo ELISPOT. Se clasificaron dos poblaciones de este individuo: células T CD8 pentámero+ y pentámero-. Los ácidos nucleicos que codifican los clonotipos de TCR β se secuenciaron en estas dos poblaciones junto con la muestra de PBMC sin clasificar, de modo que se pudieran determinar las frecuencias relativas de los clonotipos en cada población. Se utilizaron tres criterios para identificar los clonotipos de TCR β específicos de pp65₄₉₅: 1) Los clonotipos que se enriquecieron (es decir, tienen una frecuencia sustancialmente más alta) en la población de pentámero+ en comparación con la población de pentámero-; 2) Los clonotipos que se enriquecieron en la población de pentámero+ en comparación con la muestra sin clasificar; y 3) Los clonotipos que se desenriquecieron (es decir, tienen una frecuencia más baja) en la población de pentámero- en comparación con la muestra sin clasificar.

Se identificaron ocho clonotipos que se enriquecieron sustancialmente (~1.000 veces) en la población de pentámero+ en comparación con la población de pentámero- (Figura 6 panel A). Las frecuencias de estos clonotipos se compararon en las poblaciones de pentámero+ y las no clasificadas (Figura 6 panel B). El más alto de estos clonotipos tiene una frecuencia de 0,81% en la muestra sin clasificar, lo que es consistente con la respuesta elevada esperada para pp65₄₉₅ en este individuo. Sin embargo, varios de los otros clonotipos estaban presentes en un nivel inferior a 10^{-4} . Los 8 clonotipos se enriquecieron en la población de pentámero+ por un factor de ~100 veces en comparación con su frecuencia en PBMC sin clasificar.

Se usaron PBMC del mismo individuo para evaluar si los ensayos inmunes que se basan en cambios indirectos o funcionales en las células T después de la estimulación con antígeno son eficaces para la identificación de clonotipos de TCR β CD8 específicos de pp65₄₉₅. Las PBMC se estimularon con pp65₄₉₅ seguido de citometría de flujo 18 horas después de la estimulación para capturar células basándose en la expresión del marcador de activación CD137. El

repertorio de TCR β se amplificó y secuenció a partir de células CD137+ y CD137- clasificadas. Los criterios para la identificación de clonotipos de TCR β específicos de pp65₄₉₅ con este ensayo fueron similares a los utilizados en el ensayo de pentámero. Específicamente, se esperaba que los clonotipos de TCR β específicos de pp65₄₉₅ estuvieran presentes a frecuencias mucho más altas en la población CD137+ en comparación con la población CD137-.

- 5 Se identificaron nueve clonotipos que se enriquecieron sustancialmente (~1.000 veces) en la población CD137+ en comparación con la población CD137- (Figura 7 panel A). La frecuencia de estos clonotipos en la muestra sin clasificar osciló desde tan alto como 0,81% hasta tan bajo como 0,004% (Figura 7 panel B). Estos clonotipos se enriquecieron en la población CD137+ en comparación con la muestra de PBMC sin clasificar por ~100 veces. Para garantizar que estas células se activaran debido a la estimulación con el péptido, se realizó un experimento de control sin péptido.
10 Ninguno de los 9 clonotipos que se enriquecieron con el péptido en la población CD137+ enriquecida después de la incubación sin péptido en las células CD137+ (Figura 7 panel C).

- Se compararon los clonotipos específicos identificados por los ensayos de pentámero y CD137 y se encontró que se solapaban sustancialmente. Los 8 clonotipos que se identificaron con el ensayo de pentámero también se identificaron por el ensayo CD137 (Figura 8 panel A), aunque se identificó un clonotipo adicional por el ensayo CD137 que no se identificó en el ensayo de pentámero.
15

- Se realizó un tercer ensayo funcional para la identificación de clonotipos específicos de antígeno combinando la captura de células proliferantes después de la incubación con péptido pp65₄₉₅ y secuenciación del repertorio. Las células se marcaron con CFSE y se incubaron con pp65₄₉₅ o sin péptido durante 6 días. Las células CD8 proliferantes se clasificaron luego con base en la dilución de CFSE. Los clonotipos específicos de pp65₄₉₅ se identificaron en función de su frecuencia relativa en la población de CFSE^{bajo} en comparación con la de las PBMC frescas sin clasificar.
20

- Se identificaron dieciséis clonotipos que aumentaron sustancialmente en la población de CFSE^{bajo}, y la frecuencia de algunos de los clonotipos identificados fue inferior a 10⁻⁵ (Figura 9 panel A). Se usó un ensayo de proliferación idéntico que carecía del péptido como control. Ninguno de los 16 clonotipos identificados por el ensayo de proliferación se enriquecieron en la población CFSE^{bajo} cuando no se usó péptido (Figura 9 panel B).
25

- Una ventaja de usar ensayos de monitoreo inmunitario indirecto en comparación con reactivos de pentámero es la capacidad de evaluar respuestas a más de un antígeno peptídico al mismo tiempo. Se usó un grupo de 138 péptidos que abarcaban la proteína pp65 completa (denominada en este documento pp65grupo) para estimular las PBMC en el ensayo de proliferación para identificar células T específicas de pp65grupo. El análisis del repertorio de células en proliferación después de la incubación de pp65grupo permitió la identificación de 25 clonotipos. De estos 25 clonotipos identificados usando el pp65grupo, 12 de estos también se consideraron específicos de antígeno con el péptido pp65₄₉₅ único, lo que demuestra la repetibilidad del enfoque.
30

- Se identificaron siete de los ocho clonotipos identificados por el ensayo de pentámero en el ensayo de proliferación de pp65grupo, lo que demuestra que el uso del grupo de péptidos no disminuye sustancialmente la sensibilidad. Además, el ensayo de proliferación con el pp65grupo permitió la identificación de clonotipos adicionales que presumiblemente son específicos de otros péptidos dentro del grupo. La mayoría de los clonotipos adicionales identificados con el pp65grupo no estaban enriquecidos en la población de pentámero+ (Figura 9 panel D), lo que concuerda con que no son específicos para el péptido pp65₄₉₅.
35

Definiciones

- A menos que se defina específicamente lo contrario en este documento, los términos y símbolos de la química, bioquímica, genética y biología molecular de los ácidos nucleicos utilizados en este documento siguen los de los tratados y textos estándar en el campo, por ejemplo, Komberg y Baker, DNA Replication, Segunda edición (W.H. Freeman, Nueva York, 1992); Lehninger, Biochemistry, Second Edition (Worth Publishers, Nueva York, 1975); Strachan y Read, Human Molecular Genetics, segunda edición (Wiley-Liss, Nueva York, 1999); Abbas et al., Cellular and Molecular Immunology, 6ª edición (Saunders, 2007).
40

- "Activación" o "activación inmune" o "activadas", especialmente en referencia a las células T, significa una fase de una respuesta inmune adaptativa que sigue a la fase de reconocimiento del antígeno (durante la cual los linfocitos específicos de antígeno se unen a los antígenos) y está caracterizada por la proliferación de linfocitos y su diferenciación en células efectoras, por ejemplo, Abbas et al., Cellular and Molecular Immunology, cuarta edición, (W.B. Saunders Company, 2000). La activación de las células T puede estar asociada con la secreción de ciertas citocinas que son detectables mediante ensayos convencionales, tal como un ensayo ELISPOT, y puede estar asociada con la expresión de marcadores característicos de la superficie celular, tales como CD25, CD134, CD69, CD137, CD154, o similares, por ejemplo, Gratama et al., Cytometry A, 73A: 971-974 (2008).
45
50

- "Alinear" significa un método para comparar una secuencia de prueba, tal como una lectura de secuencia, con una o más secuencias de referencia para determinar qué secuencia de referencia o qué porción de una secuencia de referencia es la más cercana con base en una medida de distancia de secuencia. Un ejemplo de método para alinear secuencias de nucleótidos es el algoritmo de Smith Waterman. Las medidas de distancia pueden incluir la distancia de Hamming, la distancia de Levenshtein o similares. Las medidas de distancia pueden incluir un componente
55

relacionado con los valores de calidad de los nucleótidos de las secuencias que se comparan.

"Amplicón" significa el producto de una reacción de amplificación de polinucleótidos; es decir, una población clonal de polinucleótidos, que pueden ser monocatenarios o bicatenarios, que se replican a partir de una o más secuencias de partida. Una o más secuencias iniciales pueden ser una o más copias de la misma secuencia, o pueden ser una mezcla de diferentes secuencias. Preferiblemente, los amplicones se forman mediante la amplificación de una única secuencia de partida. Los amplicones pueden producirse mediante una variedad de reacciones de amplificación cuyos productos comprenden réplicas de uno o más ácidos nucleicos iniciales u objetivo. En un aspecto, las reacciones de amplificación que producen amplicones están "impulsadas por una plantilla" en ese emparejamiento de base de reactivos, ya sea nucleótidos u oligonucleótidos, tienen complementos en un polinucleótido plantilla que se requieren para la creación de productos de reacción. En un aspecto, las reacciones impulsadas por una plantilla son extensiones de cebador con una polimerasa de ácido nucleico o ligamientos de oligonucleótidos con una ligasa de ácido nucleico. Dichas reacciones incluyen, entre otras, reacciones en cadena de polimerasa (PCR), reacciones de polimerasa lineal, amplificación con base en una secuencia de ácido nucleico (NASBA), amplificaciones de círculo rodante y similares, descritas en las siguientes referencias: Mullis et al., patentes de los Estados Unidos Nos 4.683.195; 4.965.188; 4.683.202; 4.800.159 (PCR); Gelfand et al., patente de los Estados Unidos No. 5.210.015 (PCR en tiempo real con sondas "taqman"); Wittwer et al., patente de los Estados Unidos No. 6.174.670; Kacian et al., patente de Estados Unidos No. 5.399.491 ("NASBA"); Lizardi, patente de los Estados Unidos No. 5.854.033; Aono et al., publicación de la patente japonesa No. JP 4-262799 (amplificación de círculo rodante); y similares. En un aspecto, los amplicones de la invención son producidos por PCR. Una reacción de amplificación puede ser una amplificación en "tiempo real" si hay disponible una química de detección que permita medir un producto de reacción a medida que avanza la reacción de amplificación, por ejemplo, "PCR en tiempo real" descrita a continuación, o "NASBA en tiempo real" como se describe en Leone et al., *Nucleic Acids Research*, 26: 2150-2155 (1998), y referencias similares. Como se usa en el presente documento, el término "amplificar" significa realizar una reacción de amplificación. Una "mezcla de reacción" significa una solución que contiene todos los reactivos necesarios para realizar una reacción, que puede incluir, entre otros, agentes tamponantes para mantener el pH a un nivel seleccionado durante una reacción, sales, cofactores, depuradores y similares.

"Clonotipo" significa una secuencia de nucleótidos recombinada de un linfocito que codifica un receptor inmune o una porción del mismo. Más particularmente, clonotipo significa una secuencia de nucleótidos recombinada de una célula T o célula B que codifica un receptor de células T (TCR) o un receptor de células B (BCR), o una porción del mismo. En diversas realizaciones, los clonotipos pueden codificar todo o una porción de un reordenamiento VDJ de IgH, un reordenamiento DJ de IgH, un reordenamiento VJ de IgK, un reordenamiento VJ de IgL, un reordenamiento VDJ de TCR β , un reordenamiento DJ de TCR β , un VJ reordenamiento de TCR α , un reordenamiento VJ de TCR γ y, un reordenamiento VDJ de TCR δ , un reordenamiento VD de TCR δ , un reordenamiento Kde-V, o similares. Los clonotipos también pueden codificar regiones de punto de ruptura de translocación que involucran genes de receptores inmunes, tales como Bcl-IgH o Bcl-IgH. En un aspecto, los clonotipos tienen secuencias que son suficientemente largas para representar o reflejar la diversidad de las moléculas inmunes de las que derivan; en consecuencia, los clonotipos pueden variar ampliamente en longitud. En algunas realizaciones, los clonotipos tienen longitudes en el intervalo de 25 a 400 nucleótidos; en otras realizaciones, los clonotipos tienen longitudes en el intervalo de 25 a 200 nucleótidos.

"Perfil de clonotipo" significa una lista de clonotipos distintos y sus abundancias relativas que se derivan de una población de linfocitos. Típicamente, la población de linfocitos se obtiene de una muestra de tejido. El término "perfil de clonotipo" está relacionado con, pero más general que, el concepto de inmunología del "repertorio" inmune como se describe en referencias, tales como las siguientes: Arstila et al., *Science*, 286: 958-961 (1999); Yassai et al., *Immunogenetics*, 61: 493-502 (2009); Kedzierska et al., *Mol. Immunol.*, 45 (3): 607-618 (2008); y similares. El término "perfil de clonotipo" incluye una amplia variedad de listas y abundancias de ácidos nucleicos que codifican receptores inmunes reorganizados, que pueden derivarse de subconjuntos seleccionados de linfocitos (por ejemplo, linfocitos infiltrantes de tejidos, subconjuntos inmunofenotípicos o similares), o que pueden codificar porciones de receptores inmunes que tienen diversidad reducida en comparación con los receptores inmunes completos. En algunas realizaciones, los perfiles de clonotipo pueden comprender al menos 10^3 clonotipos distintos; en otras realizaciones, los perfiles de clonotipo pueden comprender al menos 10^4 clonotipos distintos; en otras realizaciones, los perfiles de clonotipo pueden comprender al menos 10^5 clonotipos distintos. En tales realizaciones, tales perfiles de clonotipo pueden comprender además abundancias o frecuencias relativas de cada uno de los clonotipos distintos. En un aspecto, un perfil de clonotipo es un conjunto de secuencias de nucleótidos recombinados distintos (con sus abundancias) que codifican receptores de células T (TCR) o receptores de células B (BCR), o fragmentos de los mismos, respectivamente, en una población de linfocitos de un individuo, en la que las secuencias de nucleótidos del conjunto tienen una correspondencia uno a uno con distintos linfocitos o sus subpoblaciones clonales para sustancialmente todos los linfocitos de la población. En un aspecto, los segmentos de ácido nucleico que definen los clonotipos se seleccionan de modo que su diversidad (es decir, el número de secuencias de ácido nucleico distintas en el conjunto) sea lo suficientemente grande como para que sustancialmente cada célula T o célula B o clon de las mismas en un individuo porte una secuencia de ácido nucleico única de dicho repertorio. Es decir, preferiblemente cada clon diferente de una muestra tiene un clonotipo diferente. En otros aspectos de la invención, la población de linfocitos correspondiente a un repertorio puede ser células B circulantes, o células T circulantes, o pueden ser subpoblaciones de cualquiera de

las poblaciones anteriores, incluidas, entre otras, células T CD4+ o células T CD8+ u otras subpoblaciones definidas por marcadores de superficie celular o similares. Dichas subpoblaciones pueden adquirirse tomando muestras de tejidos particulares, por ejemplo, médula ósea, o ganglios linfáticos, o similares, o clasificando o enriqueciendo células de una muestra (tal como sangre periférica) en función de uno o más marcadores de superficie celular, tamaño, morfología o similares. En otros aspectos, la población de linfocitos correspondiente a un repertorio puede derivarse de tejidos de enfermedades, tales como un tejido tumoral, un tejido infectado o similares. En una realización, un perfil de clonotipo que comprende cadenas de TCR β humano o fragmentos de las mismas comprende varias secuencias de nucleótidos distintas en el intervalo de $0,1 \times 10^6$ a $1,8 \times 10^6$, o en el intervalo de $0,5 \times 10^6$ a $1,5 \times 10^6$, o en el intervalo de $0,8 \times 10^6$ a $1,2 \times 10^6$. En otra realización, un perfil de clonotipo que comprende cadenas de IgH humanas o fragmentos de las mismas comprende varias secuencias de nucleótidos distintas en el intervalo de $0,1 \times 10^6$ a $1,8 \times 10^6$, o en el intervalo de $0,5 \times 10^6$ a $1,5 \times 10^6$, o en el intervalo de $0,8 \times 10^6$ a $1,2 \times 10^6$. En una realización particular, un perfil de clonotipo de la invención comprende un conjunto de secuencias de nucleótidos que codifican sustancialmente todos los segmentos de la región V(D)J de una cadena de IgH. En un aspecto, "sustancialmente todo", como se usa en este documento, significa cada segmento que tiene una abundancia relativa de 0,001 por ciento o más; o en otro aspecto, "sustancialmente todo" como se usa en el presente documento significa cada segmento que tiene una abundancia relativa de 0,0001 por ciento o más. En otra realización particular, un perfil de clonotipo de la invención comprende un conjunto de secuencias de nucleótidos que codifica sustancialmente todos los segmentos de la región V(D)J de una cadena de TCR β . En otra realización, un perfil de clonotipo de la invención comprende un conjunto de secuencias de nucleótidos que tienen longitudes en el intervalo de 25-200 nucleótidos e incluyen segmentos de las regiones V, D y J de una cadena de TCR β . En otra realización, un perfil de clonotipo de la invención comprende un conjunto de secuencias de nucleótidos que tienen longitudes en el intervalo de 25-200 nucleótidos e incluyen segmentos de las regiones V, D y J de una cadena de IgH. En otra realización, un perfil de clonotipo de la invención comprende varias secuencias de nucleótidos distintas que es sustancialmente equivalente al número de linfocitos que expresan una cadena de IgH distinta. En otra realización, un perfil de clonotipo de la invención comprende varias secuencias de nucleótidos distintas que son sustancialmente equivalentes a la cantidad de linfocitos que expresan una cadena de TCR β distinta. En otra realización más, "sustancialmente equivalente" significa que, con una probabilidad del noventa y nueve por ciento, un perfil de clonotipo incluirá una secuencia de nucleótidos que codifica una IgH o TCR β o una porción de la misma transportada o expresada por cada linfocito de una población de un individuo con una frecuencia de 0,001 por ciento o mayor. En otra realización más, "sustancialmente equivalente" significa que, con una probabilidad del noventa y nueve por ciento, un repertorio de secuencias de nucleótidos incluirá una secuencia de nucleótidos que codifica una IgH o TCR β o una porción de la misma transportada o expresada por cada linfocito presente con una frecuencia de 0,0001 por ciento o más. En algunas realizaciones, los perfiles de clonotipo se derivan de muestras que comprenden de 10^5 a 10^7 linfocitos. Dichas cantidades de linfocitos pueden obtenerse de muestras de sangre periférica de 1-10 mL.

"Coalescencia" significa tratar dos clonotipos candidatos con diferencias de secuencia iguales determinando que tales diferencias se deben a un error experimental o de medición y no a diferencias biológicas genuinas. En un aspecto, una secuencia de un clonotipo candidato de frecuencia más alta se compara con la de un clonotipo candidato de frecuencia más baja y si se satisfacen los criterios predeterminados, entonces el número de clonotipos candidatos de frecuencia más baja se agrega al del clonotipo candidato de frecuencia más alta y el clonotipo candidato de frecuencia más baja es por lo tanto ignorado posteriormente. Es decir, los recuentos de lectura asociados con el clonotipo candidato de frecuencia más baja se agregan a los del clonotipo candidato de frecuencia más alta.

"Regiones determinantes de complementariedad" (CDR) significan regiones de una inmunoglobulina (es decir, anticuerpo) o receptor de células T en las que la molécula complementa una conformación de antígeno, determinando por lo tanto, la especificidad de la molécula y el contacto con un antígeno específico. Los receptores de células T y las inmunoglobulinas tienen cada uno tres CDR: CDR1 y CDR2 se encuentran en el dominio variable (V), y CDR3 incluye algunos de los dominios V, todos los diversos (D) (solo cadenas pesadas) y de unión (J), y algunos de los constantes (C).

"Estructura de datos" significa una organización de información, usualmente en un ordenador o dispositivo de memoria, para una mejor eficiencia del algoritmo. Los ejemplos de estructuras de datos incluyen colas, apilamientos, listas vinculadas, cúmulos, tablas hash, matrices, árboles y similares. Las estructuras de datos pueden tener subestructuras que corresponden a unidades de información o subconjuntos de información relacionada. Por ejemplo, las matrices tienen filas y columnas de entradas; los árboles tienen nodos, ramas, árboles y hojas; o similar. En un aspecto, una estructura de datos utilizada en este documento es un árbol de secuencia, una matriz o una tabla hash.

"Dispositivo de microfluidos" significa un sistema integrado de una o más cámaras, puertos y canales que están interconectados y en comunicación fluida y diseñados para llevar a cabo una reacción o proceso analítico, ya sea solo o en cooperación con un aparato o instrumento que proporciona funciones de soporte, tales como introducción de muestras, medios de conducción de fluidos y/o reactivos, control de temperatura, sistemas de detección, recolección de datos y/o sistemas de integración, y similares. Los dispositivos de microfluidos pueden incluir además válvulas, bombas y recubrimientos funcionales especializados en paredes interiores, por ejemplo, para evitar la adsorción de los componentes o reactivos de la muestra, facilitar el movimiento de los reactivos por electroosmosis o similares. Tales dispositivos generalmente se fabrican en o como un sustrato sólido, que puede ser vidrio, plástico u otros materiales poliméricos sólidos, y generalmente tienen un formato plano para facilitar la detección y monitoreo del

movimiento de la muestra y el reactivo, especialmente a través de métodos ópticos o electroquímicos. Las características de un dispositivo de microfluidos generalmente tienen dimensiones de sección transversal de menos de unos pocos cientos de micrómetros cuadrados y los pasajes generalmente tienen dimensiones capilares, por ejemplo, que tiene dimensiones de sección transversal máximas de aproximadamente 500 μm a aproximadamente 0,1 μm . Los dispositivos de microfluidos generalmente tienen capacidades de volumen en el intervalo de 1 μL a unos pocos nL, por ejemplo, 10-100 nL. La fabricación y el funcionamiento de los dispositivos de microfluidos son bien conocidos en la técnica, como se ejemplifica en las siguientes referencias: Ramsey, patentes de los Estados Unidos Nos. 6.001.229; 5.858.195; 6.010.607; y 6.033.546; Soane et al., patentes de los Estados Unidos Nos. 5.126.022 y 6.054.034; Nelson et al., patente de los Estados Unidos No. 6.613.525; Maher et al., patente de los Estados Unidos No. 6.399.952; Ricco et al., publicación internacional de patente WO 02/24322; Bjomson et al., publicación internacional de patente WO 99/19717; Wilding et al., patentes de los Estados Unidos Nos. 5.587.128; 5,498,392; Sia et al., *Electrophoresis*, 24: 3563-3576 (2003).

"Porcentaje de homología", "porcentaje de identidad" o términos similares utilizados en referencia con la comparación de una secuencia de referencia y otra secuencia ("secuencia de comparación") significan que en una alineación óptima entre las dos secuencias, la secuencia de comparación es idéntica a la secuencia de referencia en varias posiciones de subunidades equivalentes al porcentaje indicado, siendo las subunidades nucleótidos para comparaciones de polinucleótidos o aminoácidos para comparaciones de polipéptidos. Como se usa en el presente documento, una "alineación óptima" de las secuencias que se comparan es aquella que maximiza las coincidencias entre subunidades y minimiza el número de espacios empleados en la construcción de una alineación. Las identidades porcentuales pueden determinarse con implementaciones de algoritmos disponibles comercialmente, tales como los descritos por Needleman y Wunsch, *J. Mol. Biol.*, 48: 443-453 (1970) (programa "GAP" del Paquete de Análisis de Secuencia de Wisconsin, Genetics Computer Group, Madison, WI), o similares. Otros paquetes de software en la técnica para construir alineaciones y calcular el porcentaje de identidad u otras medidas de similitud incluyen el programa "BestFit", basado en el algoritmo de Smith y Waterman, *Advances in Applied Mathematics*, 2: 482-489 (1981) (Paquete de Análisis de Secuencia de Wisconsin, Genetics Computer Group, Madison, WI). En otras palabras, por ejemplo, para obtener un polinucleótido que tiene una secuencia de nucleótidos al menos 95 por ciento idéntica a una secuencia de nucleótidos de referencia, hasta el cinco por ciento de los nucleótidos en la secuencia de referencia pueden eliminarse o sustituirse con otro nucleótido, o un número de los nucleótidos hasta el cinco por ciento del número total de nucleótidos en la secuencia de referencia pueden insertarse en la secuencia de referencia.

"Reacción en cadena de la polimerasa" o "PCR" significa una reacción para la amplificación *in vitro* de secuencias específicas de ADN mediante la extensión simultánea de cebadores de cadenas complementarias de ADN. En otras palabras, la PCR es una reacción para hacer múltiples copias o réplicas de un ácido nucleico objetivo flanqueado por sitios de unión del cebador, dicha reacción comprende una o más repeticiones de las siguientes etapas: (i) desnaturar el ácido nucleico objetivo, (ii) hibridación de cebadores con los sitios de unión del cebador, y (iii) extensión de los cebadores por medio de una polimerasa de ácido nucleico en presencia de trifosfatos de nucleósidos. Por lo general, la reacción se cicla a través de diferentes temperaturas optimizadas para cada etapa en un instrumento termociclador. Las temperaturas particulares, las duraciones en cada etapa y las velocidades de cambio entre las etapas dependen de muchos factores bien conocidos por los expertos en la materia, por ejemplo, ejemplificados por las referencias: McPherson et al., editores, *PCR: A Practical Approach and PCR2: A Practical Approach* (IRL Press, Oxford, 1991 y 1995, respectivamente). Por ejemplo, en una PCR convencional que usa ADN polimerasa Taq, un ácido nucleico objetivo bicatenario puede desnaturarse a una temperatura $> 90\text{ }^{\circ}\text{C}$, los cebadores se hibridan a una temperatura en el intervalo de 50-75 $^{\circ}\text{C}$, y los cebadores se extienden a una temperatura en el intervalo de 72-78 $^{\circ}\text{C}$. El término "PCR" abarca formas derivadas de la reacción, que incluyen, pero no se limitan a, RT-PCR, PCR en tiempo real, PCR anidada, PCR cuantitativa, PCR multiplexada y similares. Los volúmenes de reacción varían desde unos pocos cientos de nanolitros, por ejemplo, 200 nL, a unos pocos cientos de μL , por ejemplo, 200 μL . "PCR de transcripción inversa" o "RT-PCR" significa una PCR precedida por una reacción de transcripción inversa que convierte un ARN objetivo en un ADN monocatenario complementario, que luego se amplifica, por ejemplo, Tecott et al., patente de los Estados Unidos No. 5.168.038. "PCR en tiempo real" significa una PCR para la cual se controla la cantidad de producto de reacción, es decir, el amplicón, se controla a medida que avanza la reacción. Existen muchas formas de PCR en tiempo real que difieren principalmente en las químicas de detección utilizadas para controlar el producto de reacción, por ejemplo, Gelfand et al., patente de los Estados Unidos No. 5.210.015 ("taqman"); Wittwer et al., patentes de los Estados Unidos Nos 6.174.670 y 6.569.627 (colorantes de intercalación); Tyagi et al., patente de los Estados Unidos No. 5.925.517 (balizas moleculares). Las químicas de detección para PCR en tiempo real se revisan en Mackay et al., *Nucleic Acids Research*, 30: 1292-1305 (2002). "PCR anidada" significa una PCR de dos etapas en la que el amplicón de una primera PCR se convierte en la muestra para una segunda PCR utilizando un nuevo conjunto de cebadores, al menos uno de los cuales se une a una ubicación interior del primer amplicón. Como se usa en el presente documento, "cebadores iniciales" en referencia a una reacción de amplificación anidada significan los cebadores utilizados para generar un primer amplicón, y "cebadores secundarios" significan uno o más cebadores utilizados para generar un segundo amplicón anidado. "PCR multiplexada" significa una PCR en la que múltiples secuencias objetivo (o una secuencia objetivo única y una o más secuencias de referencia) se llevan a cabo simultáneamente en la misma mezcla de reacción, por ejemplo, Bernard et al., *Anal. Biochem.*, 273: 221-228 (1999) (PCR en tiempo real de dos colores). Por lo general, se emplean distintos conjuntos de cebadores para cada secuencia que se amplifica. Típicamente, el número de secuencias objetivo en una PCR multiplex está en el intervalo de 2 a 50, o de 2 a 40, o de 2 a 30. "PCR cuantitativa" significa una PCR diseñada para medir la abundancia de una o más secuencias específicas

objetivo en una muestra o espécimen. La PCR cuantitativa incluye tanto la cuantificación absoluta como la cuantificación relativa de tales secuencias objetivo. Las mediciones cuantitativas se realizan utilizando una o más secuencias de referencia o estándares internos que se pueden analizar por separado o junto con una secuencia objetivo. La secuencia de referencia puede ser endógena o exógena a una muestra o espécimen y, en este último caso, puede comprender una o más plantillas competidoras. Las secuencias de referencia endógenas típicas incluyen segmentos de transcripciones de los siguientes genes: β -actina, GAPDH, β_2 -microglobulina, ARN ribosómico y similares. Las técnicas para PCR cuantitativa son bien conocidas por aquellos expertos en la técnica, como se ejemplifica en las siguientes referencias: Freeman et al., *Biotechniques*, 26: 112-126 (1999); Becker-Andre et al., *Nucleic Acids Research*, 17: 9437-9447 (1989); Zimmerman et al., *Biotechniques*, 21: 268-279 (1996); Diviacco et al., *Gene*, 122: 3013-3020 (1992); Becker-Andre et al., *Nucleic Acids Research*, 17: 9437-9446 (1989); y similares.

"Ensamblaje del ciclo de polimerasa" o reacción "PCA" (también denominada en este documento "PCR enlazada") significa una PCR que comprende al menos un par de cebadores externos y al menos un par de cebadores internos. Un cebador interno tiene una porción 3' que es complementario a un ácido nucleico objetivo (o su complemento) y una porción 5' que es complementario a la porción 5' de otro cebador interno correspondiente a un ácido nucleico objetivo diferente.

"Cebador" significa un oligonucleótido, ya sea natural o sintético que es capaz, al formar un dúplex con una plantilla polinucleotídica, de actuar como un punto de inicio de la síntesis de ácido nucleico y extenderse desde su extremo 3' a lo largo de la plantilla para que se forma un dúplex extendido. La extensión de un cebador generalmente se lleva a cabo con una polimerasa de ácido nucleico, tal como ADN o ARN polimerasa. La secuencia de nucleótidos añadidos en el proceso de extensión está determinada por la secuencia del polinucleótido plantilla. Por lo general, los cebadores se extienden mediante una ADN polimerasa. Los cebadores generalmente tienen una longitud en el intervalo de 14 a 40 nucleótidos, o en el intervalo de 18 a 36 nucleótidos. Los cebadores se emplean en una variedad de reacciones de amplificación nucleica, por ejemplo, reacciones de amplificación lineal usando un cebador único, o reacciones en cadena de la polimerasa, que emplean dos o más cebadores. La guía para seleccionar las longitudes y secuencias de cebadores para aplicaciones particulares es bien conocida por los expertos en la materia, tal como lo demuestran las siguientes referencias: Dieffenbach, editor, *PCR Primer: A Laboratory Manual*, 2nd Edition (Cold Spring Harbor Press, Nueva York, 2003).

"Puntaje de calidad" significa una medida de la probabilidad de que una asignación de base en una ubicación de secuencia particular sea correcta. Los expertos en la materia conocen una variedad de métodos para calcular puntajes de calidad para circunstancias particulares, tales como llamado de bases como resultado de diferentes químicas de secuenciación, sistemas de detección, algoritmos de llamadas de bases, etc. En general, los valores de puntajes de calidad están monótonicamente relacionados con las probabilidades de una correcta llamada de bases. Por ejemplo, un puntaje de calidad, o Q, de 10 puede significar que hay un 90 por ciento de posibilidades de que una base sea llamada correctamente, una Q de 20 puede significar que hay un 99 por ciento de posibilidades de que una base sea llamada correctamente, etc. Para algunas plataformas de secuenciación, particularmente aquellas que usan química de secuenciación por síntesis, los puntajes de calidad promedio disminuyen en función de la longitud de lectura de secuencia, de modo que los puntajes de calidad al comienzo de una lectura de secuencia son más altos que aquellos al final de una lectura de secuencia, tales disminuciones se deben a fenómenos tales como extensiones incompletas, extensiones de arrastre, pérdida de plantilla, pérdida de polimerasa, fallas de protección, fallas de desprotección, y similares.

"Lectura de secuencia" significa una secuencia de nucleótidos determinada a partir de una secuencia o flujo de datos generados por una técnica de secuenciación, cuya determinación se realiza, por ejemplo, mediante un software de llamado de bases asociado con la técnica, por ejemplo, software de llamado de bases de un proveedor comercial de una plataforma de secuenciación de ADN. Una lectura de secuencia generalmente incluye puntajes de calidad para cada nucleótido en la secuencia. Típicamente, las lecturas de secuencia se realizan extendiendo un cebador a lo largo de un ácido nucleico plantilla, por ejemplo, con una ADN polimerasa o una ADN ligasa. Los datos se generan al registrar señales, tales como señales ópticas, químicas (por ejemplo, cambio de pH) o eléctricas, asociadas con dicha extensión. Dichos datos iniciales se convierten en una lectura de secuencia.

"Etiqueta de secuencia" (o "etiqueta") o "código de barras" significa un oligonucleótido que está unido, generalmente a través de un enlace covalente, a otra molécula o complejo molecular y que se utiliza para identificar y/o rastrear la otra molécula en una reacción o una serie de reacciones. Las etiquetas de secuencia pueden variar ampliamente en tamaño y composición; las siguientes referencias proporcionan una guía para seleccionar conjuntos de etiquetas de secuencia apropiadas para realizaciones particulares: Brenner, patente de los Estados Unidos No. 5.635.400; Brenner y Macevicz, patente de los Estados Unidos No. 7.537.897; Brenner et al., *Proc. Natl. Acad. Sci.*, 97: 1665-1670 (2000); Church et al., publicación de la patente europea 0303 459; Shoemaker et al., *Nature Genetics*, 14: 450-456 (1996); Morris et al., publicación de la patente europea 0799897A1; Wallace, patente de los Estados Unidos NO. 5.981.179; y similares. Las longitudes y composiciones de etiquetas de secuencia pueden variar ampliamente, y la selección de longitudes y/o composiciones particulares depende de varios factores que incluyen, entre otros, cómo se usan las etiquetas para generar una lectura, por ejemplo, mediante una reacción de hibridación o mediante una reacción enzimática, tal como amplificación y secuenciación; si están etiquetados, por ejemplo, con un tinte fluorescente o similar; el número de etiquetas de secuencia distinguibles requeridas para identificar inequívocamente un conjunto de moléculas de interés y similares, y qué tan diferentes deben ser las etiquetas de un conjunto para garantizar una

identificación confiable, por ejemplo, libertad de hibridación cruzada, identificación errónea de errores de secuenciación o similares. En algunas realizaciones, las etiquetas de secuencia pueden tener cada una, una longitud dentro de un intervalo de 6 a 36 nucleótidos, o de 4 a 30 nucleótidos, o de 8 a 40 nucleótidos, o de 6 a 50 nucleótidos, respectivamente; siempre, sin embargo, que el término "etiqueta de secuencia" también pueda usarse en referencia a una etiqueta de secuencia de las longitudes anteriores intercaladas entre un par de cebadores que pueden usarse para amplificar o bien manipular la etiqueta de secuencia, por ejemplo, para identificarla mediante secuenciación de ADN. En un aspecto, se usan conjuntos de etiquetas de secuencia en las que cada etiqueta de secuencia de un conjunto tiene una secuencia de nucleótidos única que difiere de aquella de cualquier otra etiqueta del mismo conjunto por una plural de bases; en algunas realizaciones, dicha pluralidad es al menos de tres bases; en otro aspecto, se usan conjuntos de etiquetas de secuencia en las que la secuencia de cada etiqueta de un conjunto difiere de aquella de cualquier otra etiqueta del mismo conjunto en al menos cuatro bases.

"Árbol de secuencia" significa una estructura de datos de árbol para representar secuencias de nucleótidos. En un aspecto, una estructura de datos de árbol de la invención es un árbol dirigido enraizado que comprende nodos y bordes que no incluyen ciclos o rutas cíclicas. Los bordes de los nodos de las estructuras de datos de árbol de la invención generalmente se ordenan. Los nodos y/o bordes son estructuras que pueden contener o estar asociadas con un valor. Cada nodo en un árbol tiene cero o más nodos secundarios, que por convención se muestran debajo en el árbol. Un nodo que tiene un hijo se llama nodo padre del hijo. Un nodo tiene como máximo un padre. Los nodos que no tienen hijos se denominan nodos hoja. El nodo superior de un árbol se llama nodo raíz. Al ser el nodo superior, el nodo raíz no tendrá padres. Es el nodo en el que comúnmente comienzan las operaciones en el árbol (aunque algunos algoritmos comienzan con los nodos de hoja y terminan en la raíz). Se puede llegar a todos los demás nodos siguiendo bordes o enlaces.

REIVINDICACIONES

1. Un método para identificar clonotipos de células T específicas de antígeno en una muestra que comprende células T, comprendiendo el método las etapas de:
- 5 (a) exponer la muestra a una pluralidad de antígenos en una pluralidad de mezclas de reacción diferentes, en las que cada antígeno de dicha pluralidad de antígenos está presente en una subpluralidad única predeterminada de la pluralidad de mezclas de reacción;
- (b) seleccionar células T específicas de antígeno de cada una de la pluralidad de mezclas de reacción;
- (c) secuenciar ácidos nucleicos recombinados que codifican clonotipos del mismo segmento del receptor de células T a partir de:
- 10 (i) las células T específicas de antígeno seleccionadas en la etapa (b) de cada una de la pluralidad de mezclas de reacción, en la que la secuenciación produce una pluralidad de lecturas de secuencia para cada mezcla de reacción a partir de la cual se determinan los clonotipos y las frecuencias de clonotipos, y
- (ii) células T obtenidas de la muestra antes de la etapa (a), en la que la secuenciación produce una pluralidad de lecturas de secuencia a partir de las cuales se determinan los clonotipos y las frecuencias de clonotipos;
- 15 (d) para cada mezcla de reacción, comparar las frecuencias de clonotipos determinadas en la etapa (c) (i) con las frecuencias de clonotipos determinadas en la etapa (c) (ii); y
- (e) para cada uno de la pluralidad de antígenos, identificar un clonotipo de una célula T específica de antígeno que tiene una mayor frecuencia aumentada en cada una de las mezclas de reacción en las que estaba presente el antígeno y no tiene una mayor frecuencia en cada una de las mezclas de reacción en las que el antígeno estaba ausente.
- 20 2. El método de la reivindicación 1, en el que dicha etapa de secuenciación se realiza mediante secuenciación de alto rendimiento que incluye aislar espacialmente moléculas individuales de dichos ácidos nucleicos recombinados en una superficie sólida y secuenciar las moléculas individuales aisladas en paralelo para proporcionar dichas lecturas de secuencia.
- 25 3. El método de la reivindicación 1 o la reivindicación 2, en el que cada clonotipo se determina mediante coalescencia de una pluralidad de dichas lecturas de secuencia.
4. El método de una cualquiera de las reivindicaciones 1-3, en el que las células T específicas de antígeno se seleccionan con base en la interacción con un antígeno en la mezcla de reacción.
5. El método de una cualquiera de las reivindicaciones 1-4, en el que dicha muestra es una muestra de sangre o una muestra de células mononucleares de sangre periférica derivadas de sangre periférica.
- 30 6. El método de una cualquiera de las reivindicaciones 1-5, en el que la muestra es una muestra de tejido escasa.
7. El método de la reivindicación 6, en el que la muestra de tejido escasa es de un sujeto con cáncer.
8. El método de una cualquiera de las reivindicaciones 1-7, en el que dicha pluralidad de antígenos es de 2 a 100 antígenos proteicos.

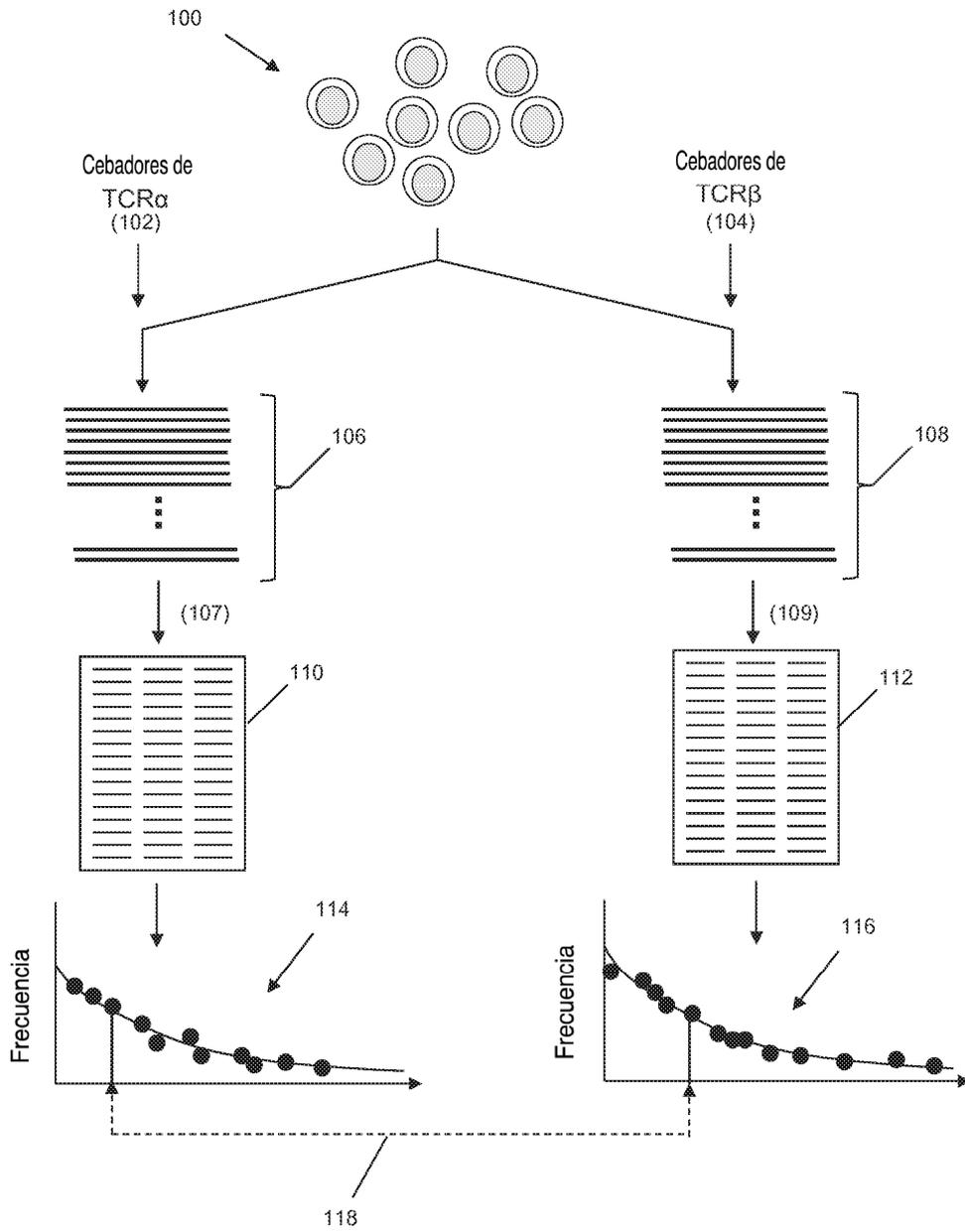


Fig. 1A

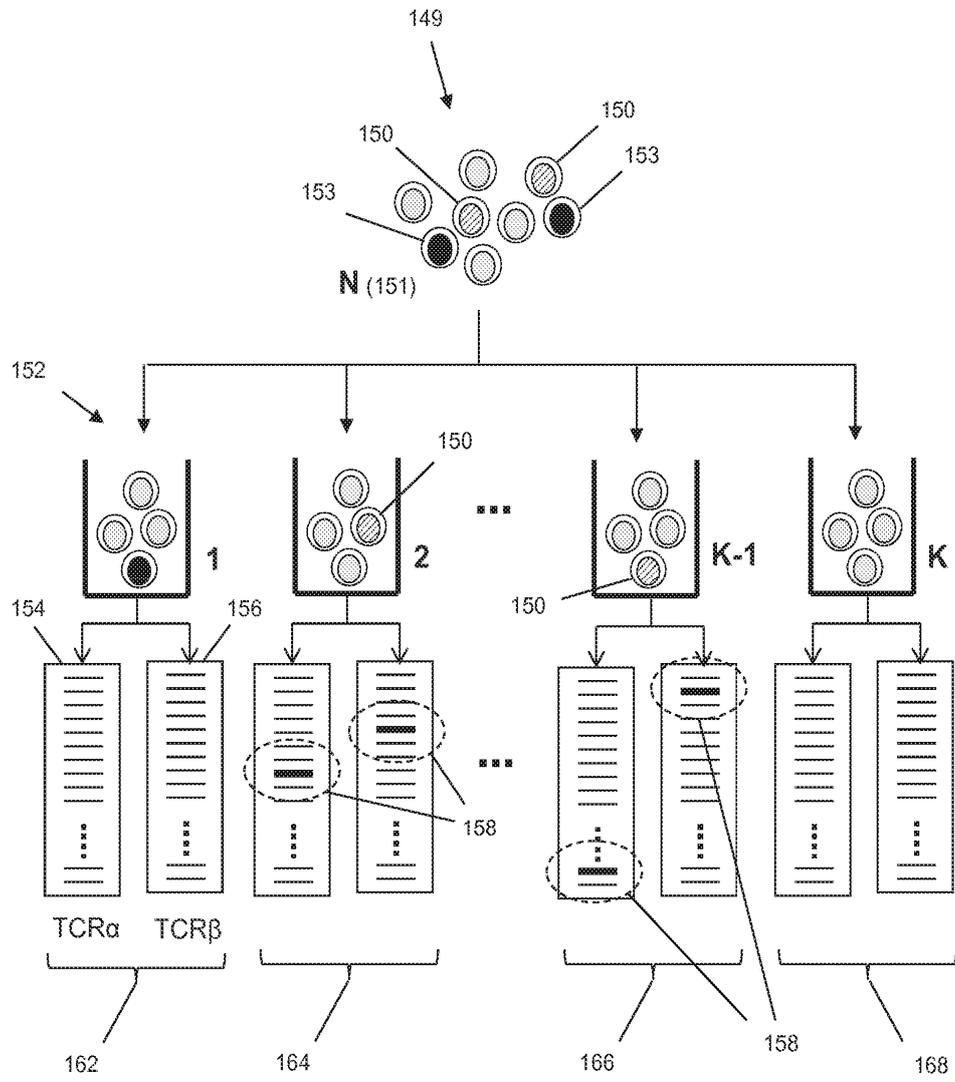


Fig. 1B

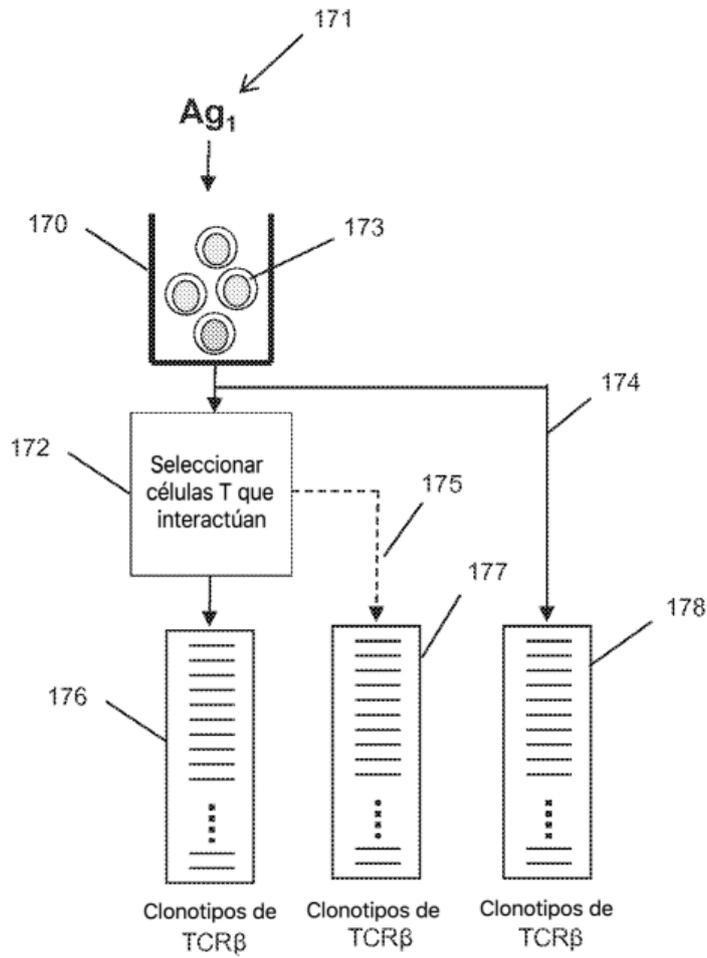


Fig. 1C

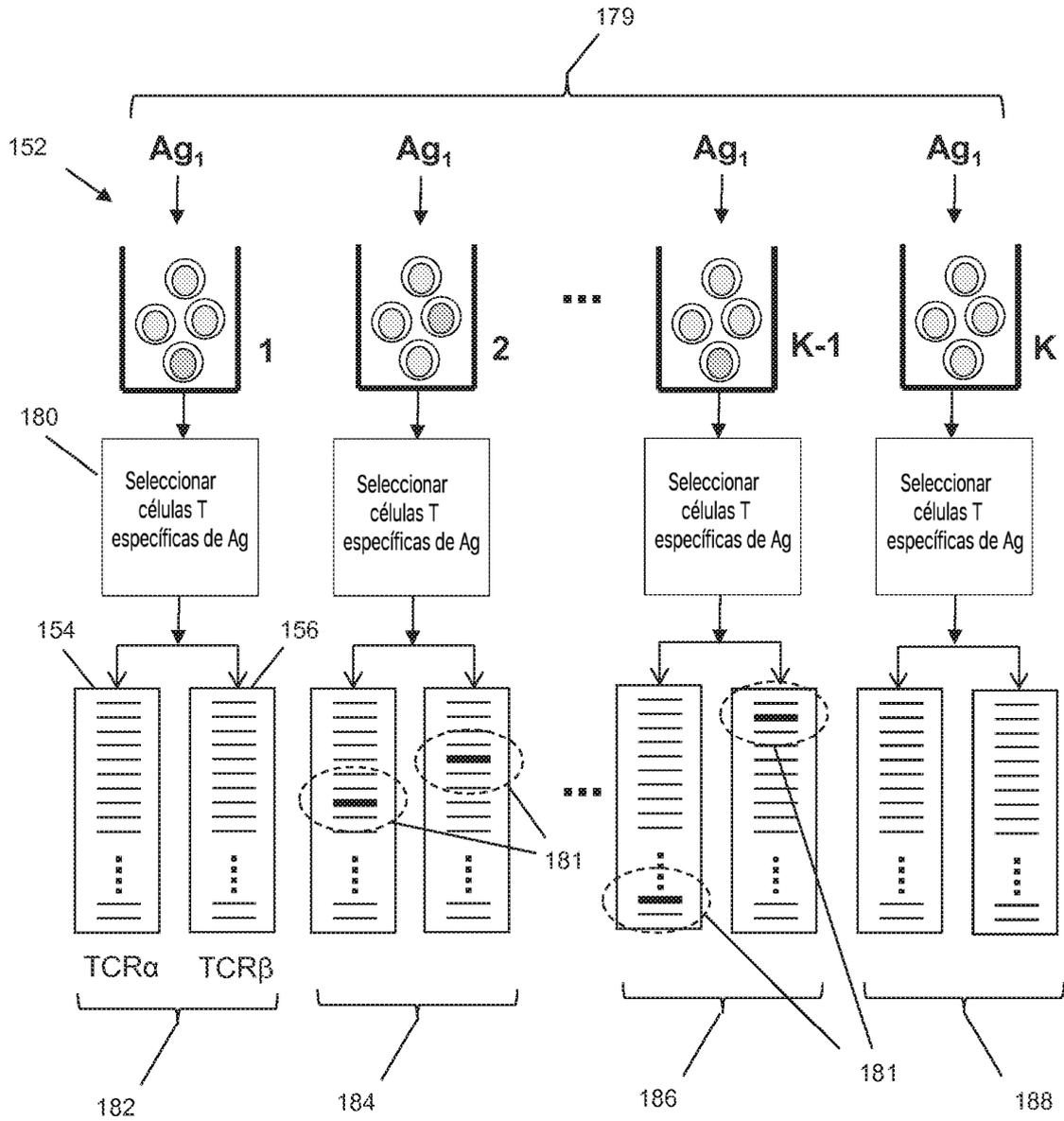


Fig. 1D

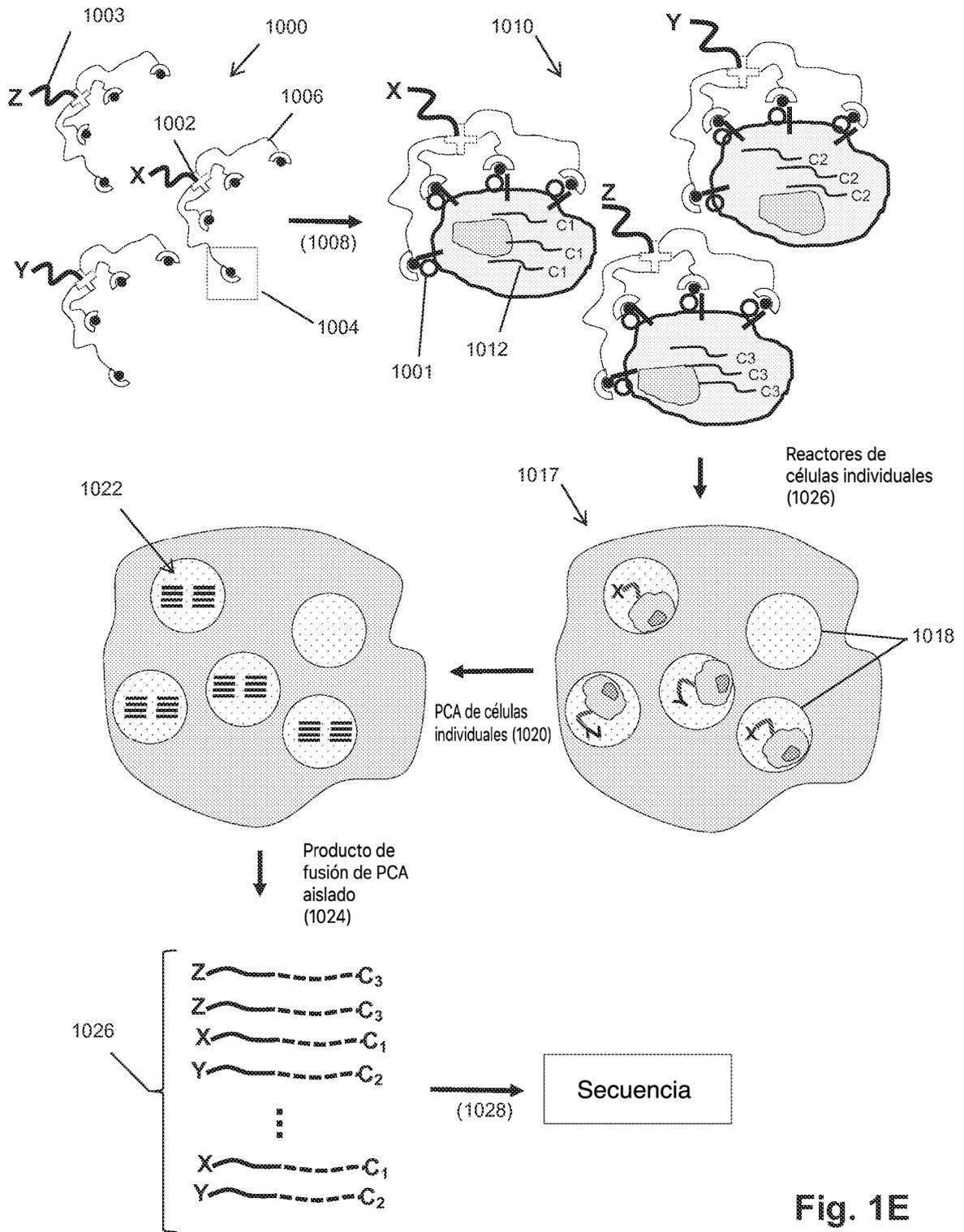


Fig. 1E

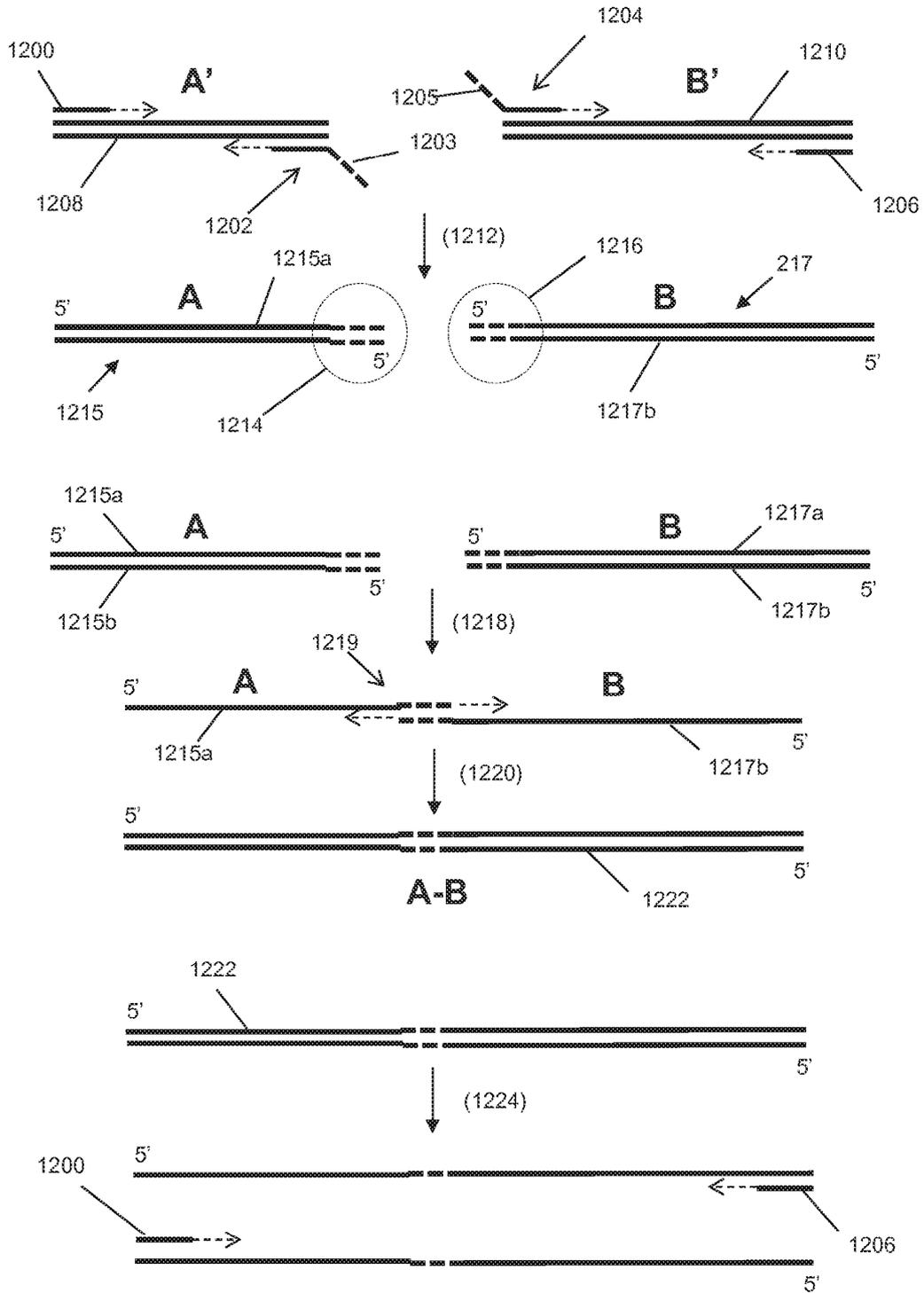


Fig. 1F

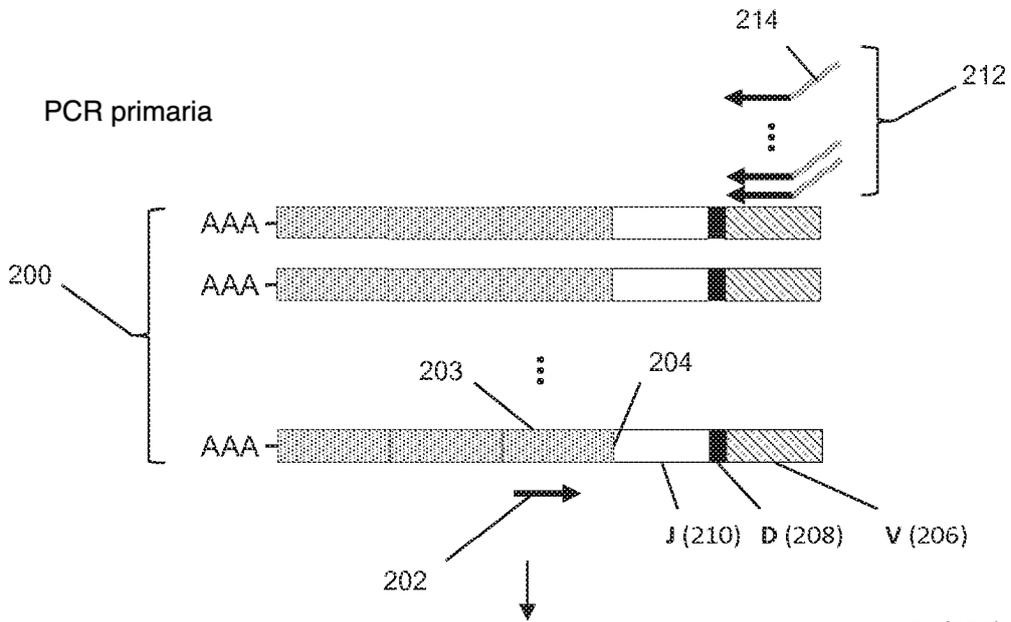


Fig. 2A

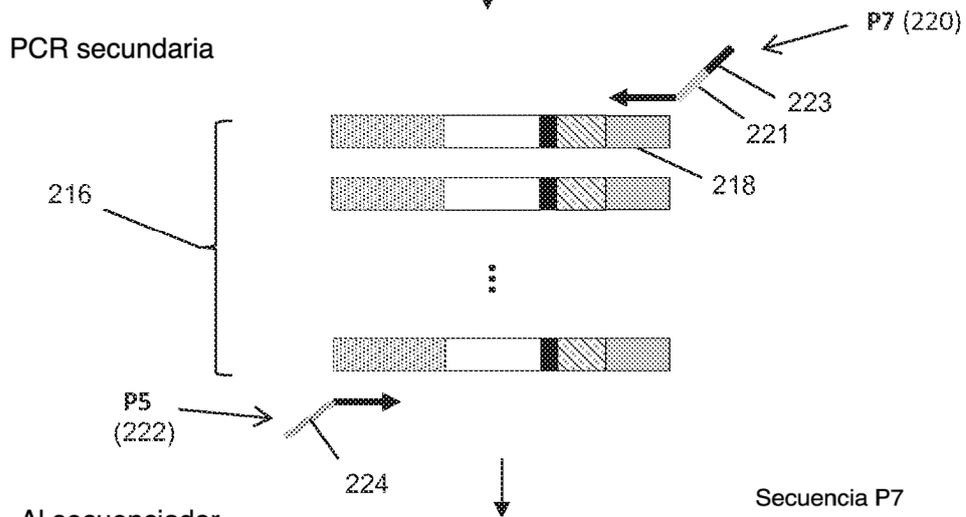


Fig. 2B

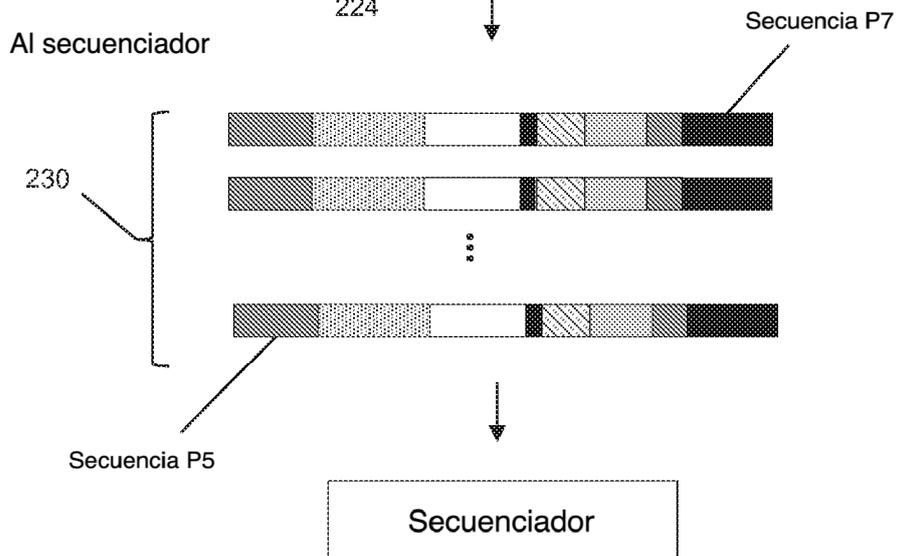


Fig. 2C

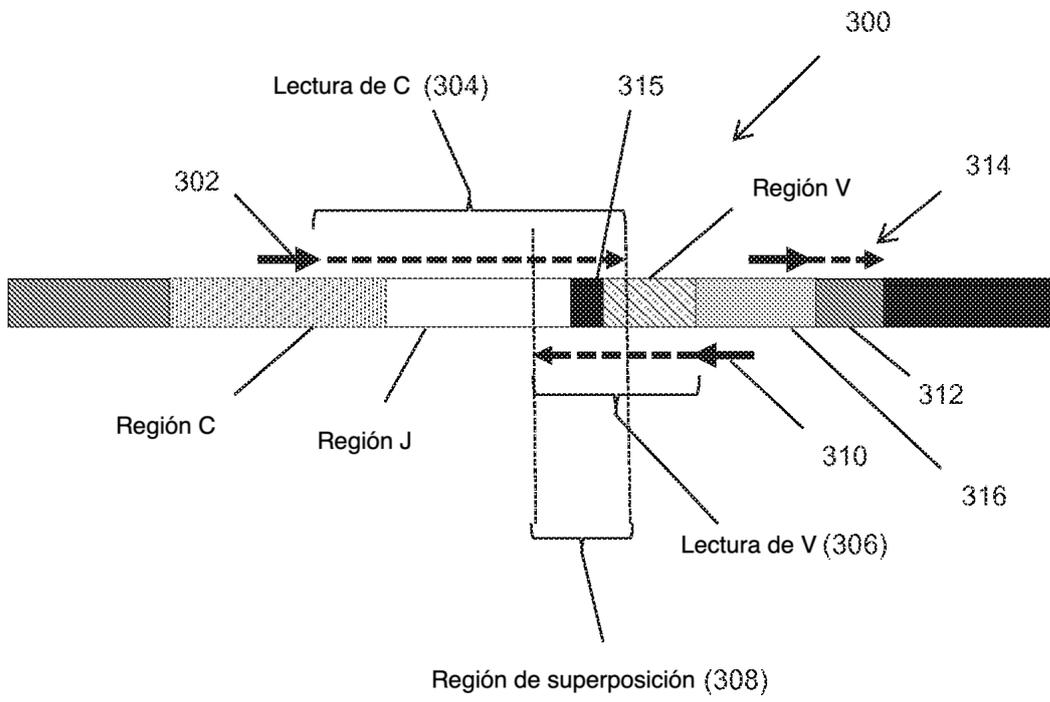


Fig. 3A

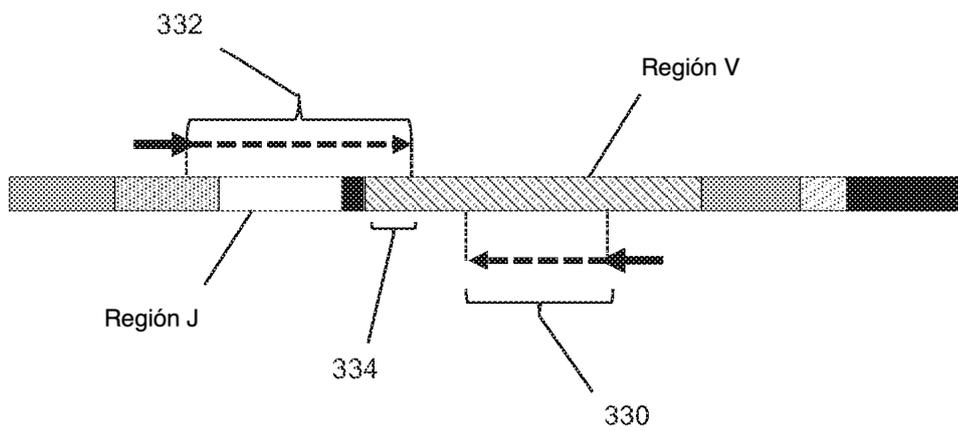


Fig. 3B

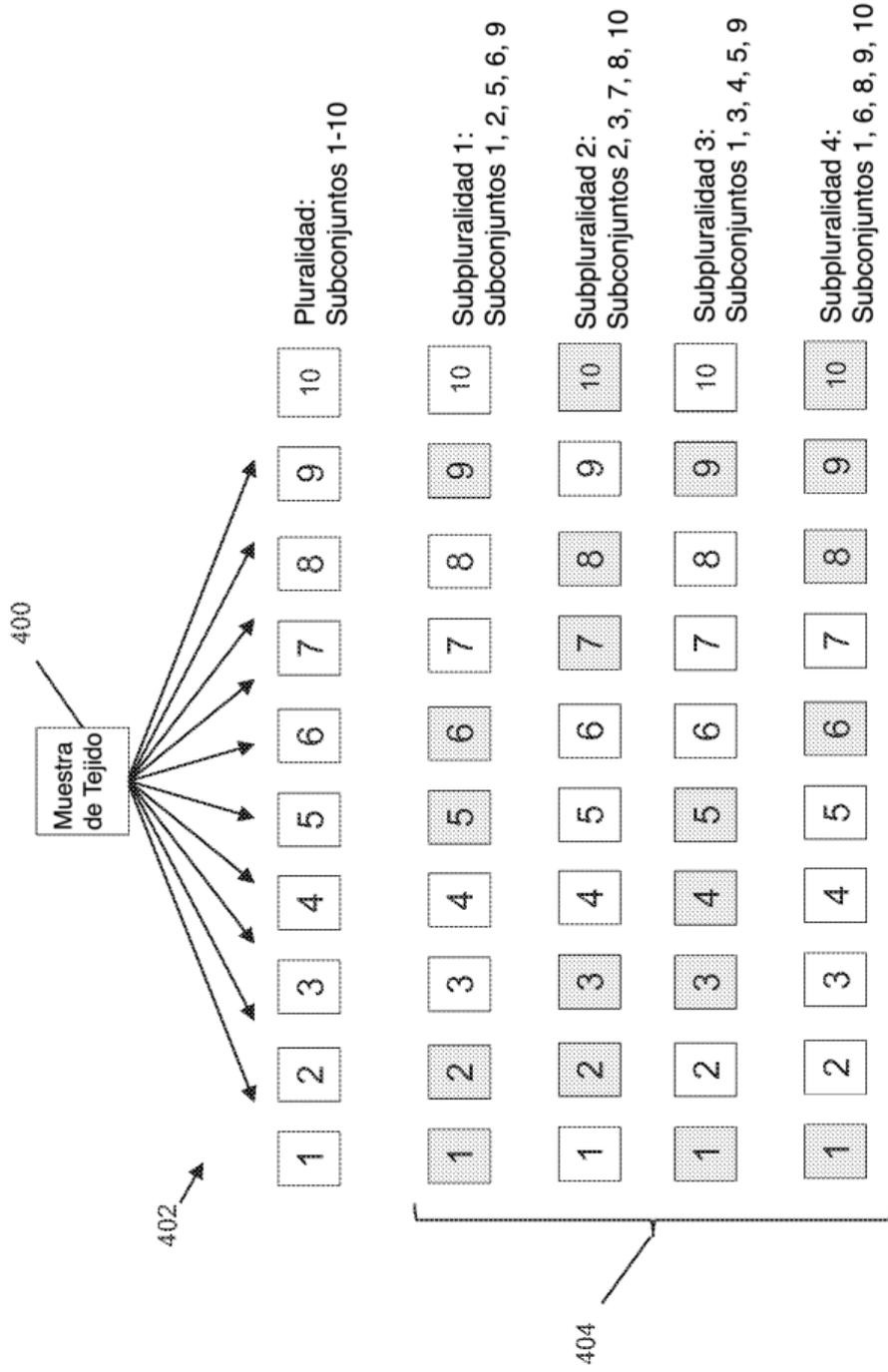


Fig. 4

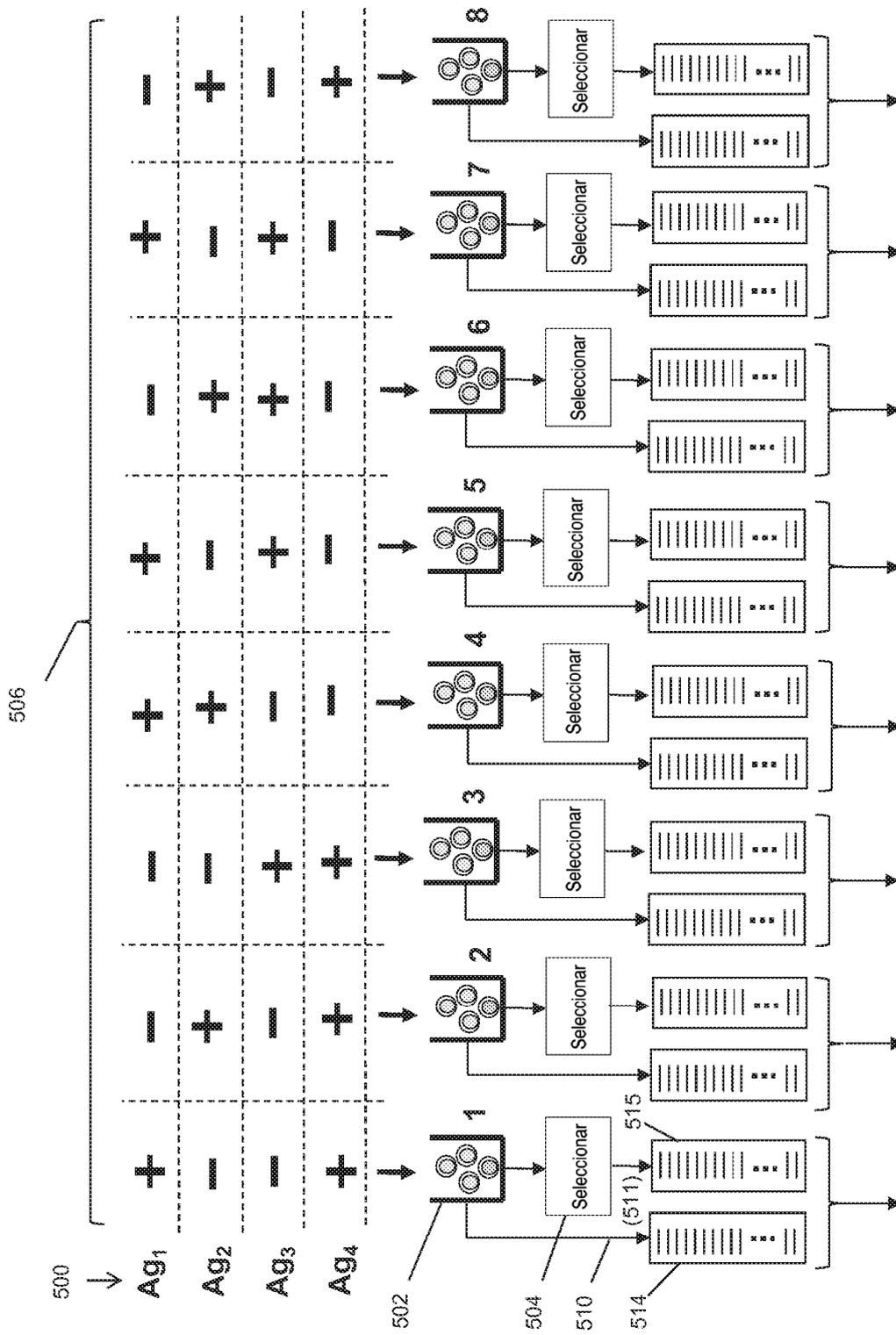


Fig. 5A

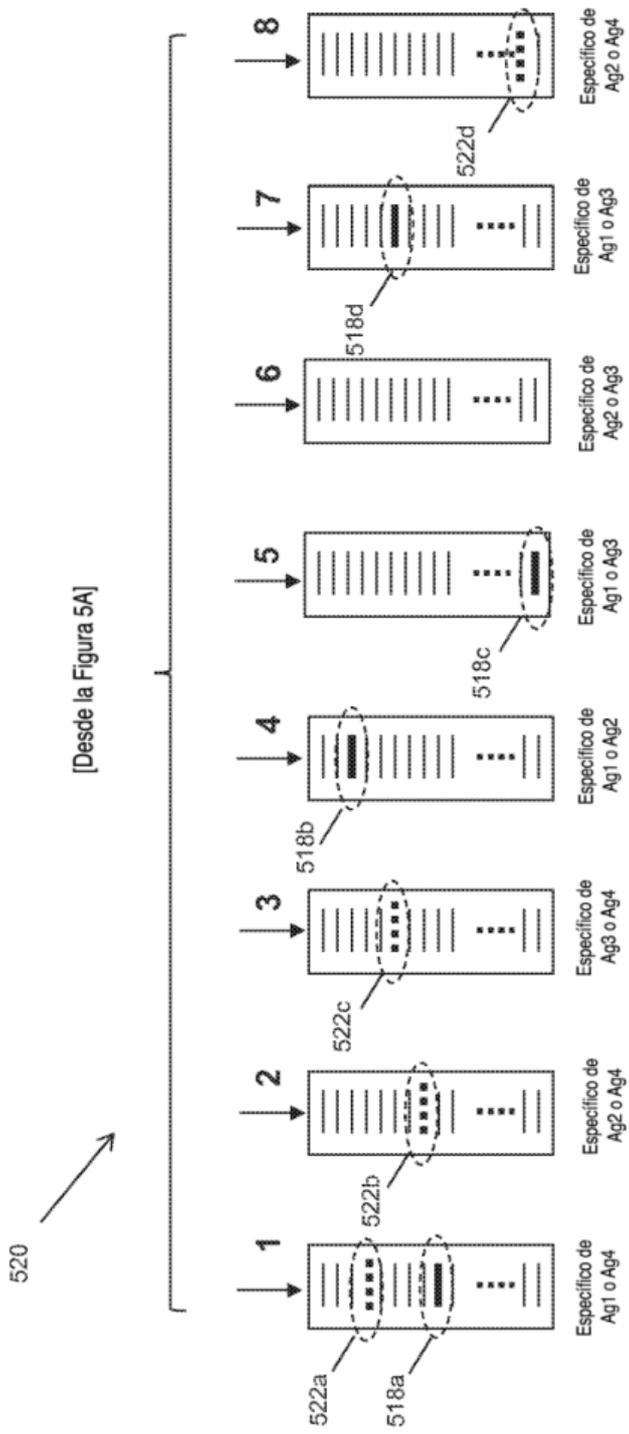


Fig. 5B

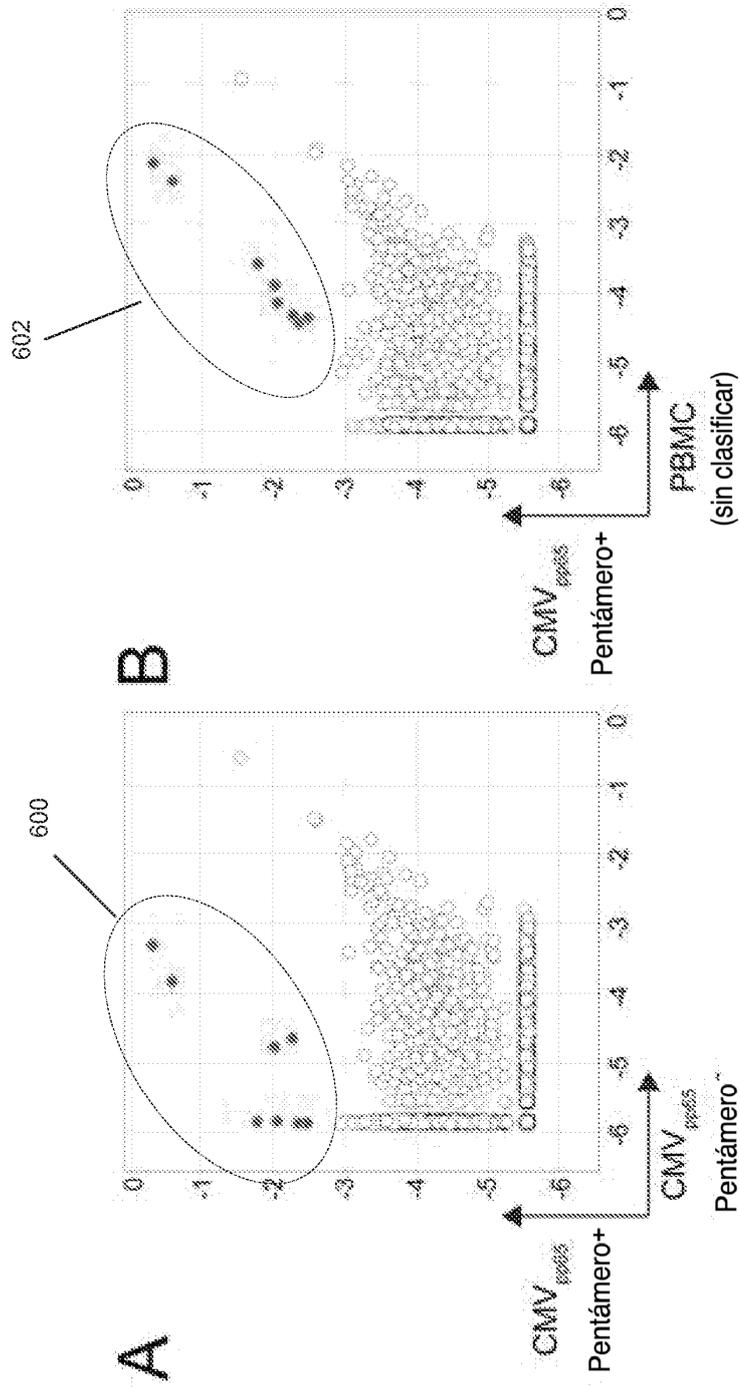


Fig. 6

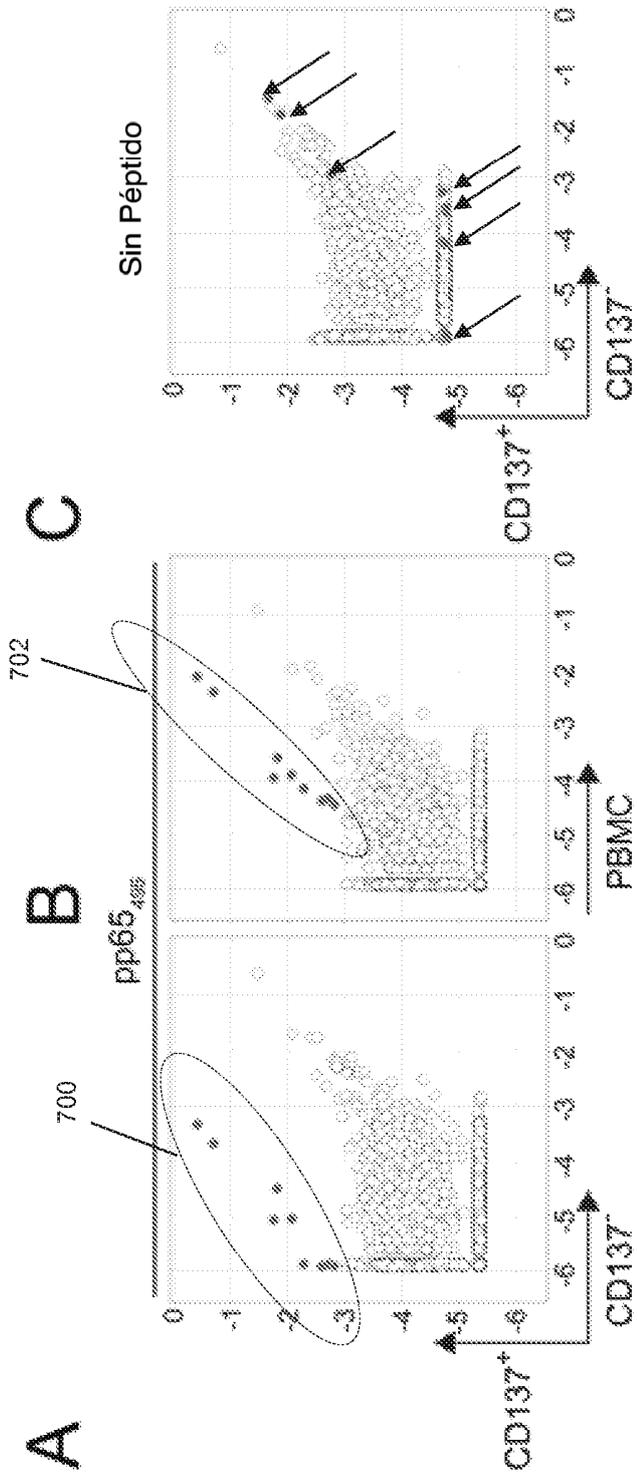


Fig. 7

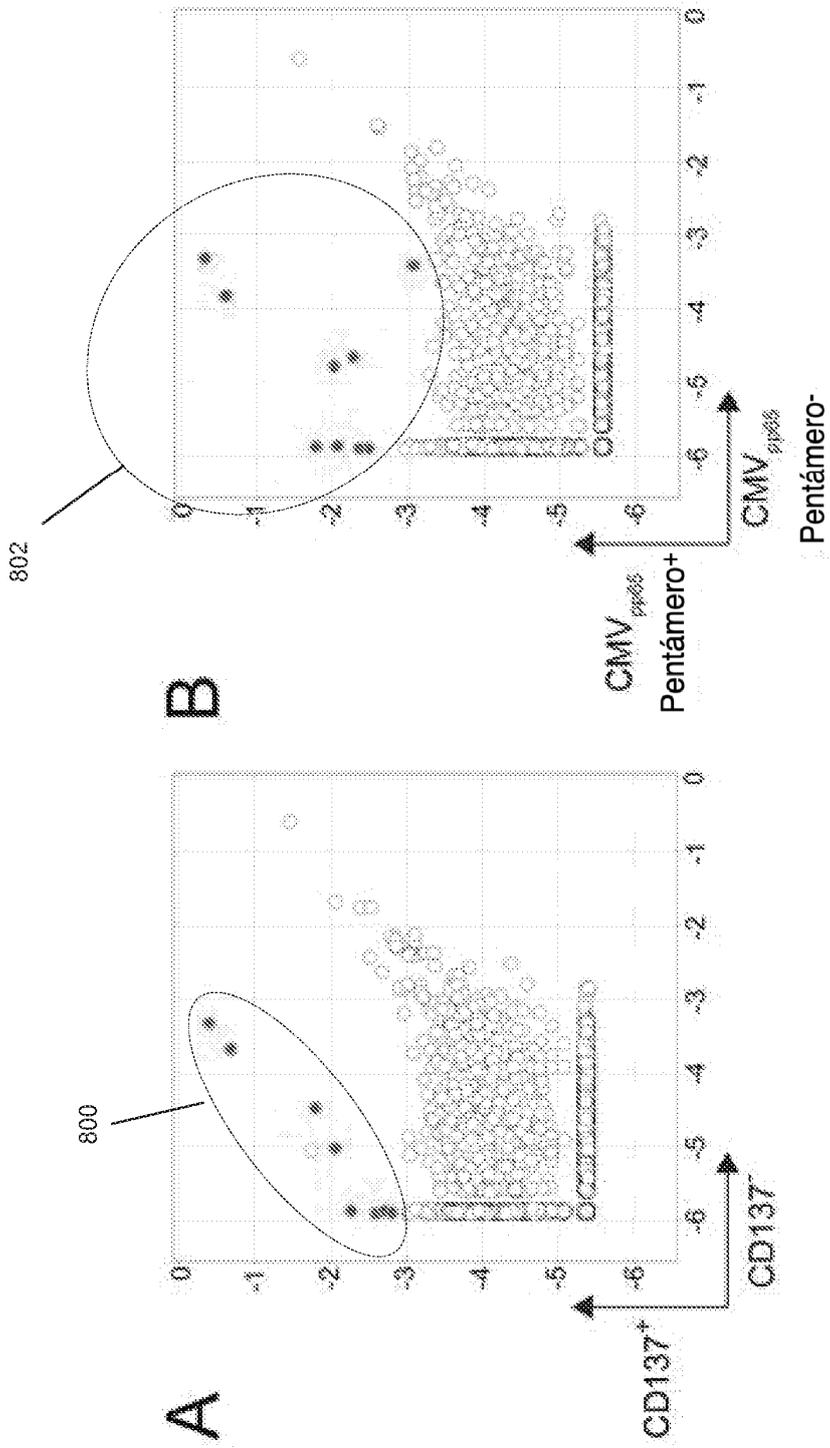


Fig. 8

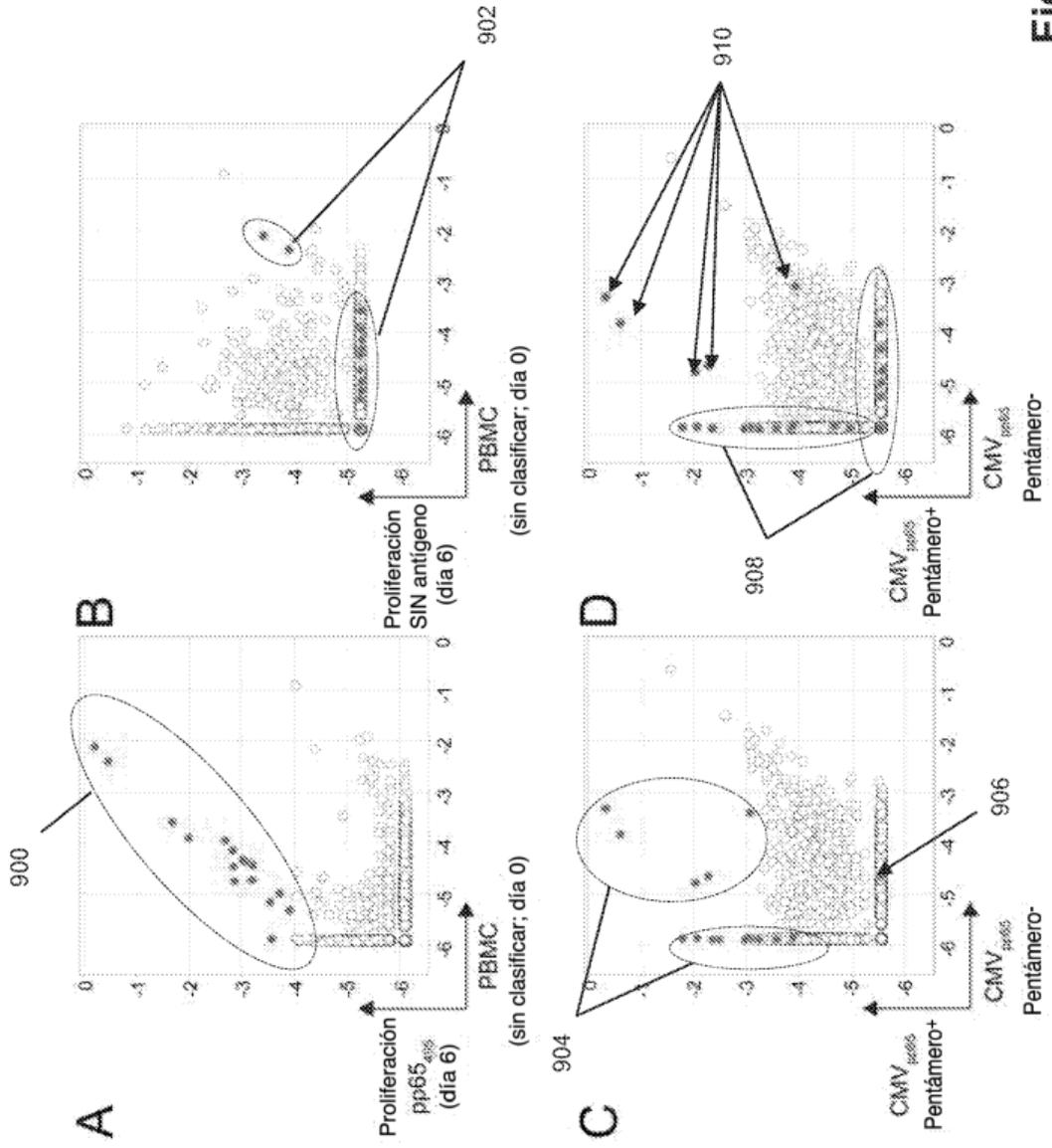


Fig. 9

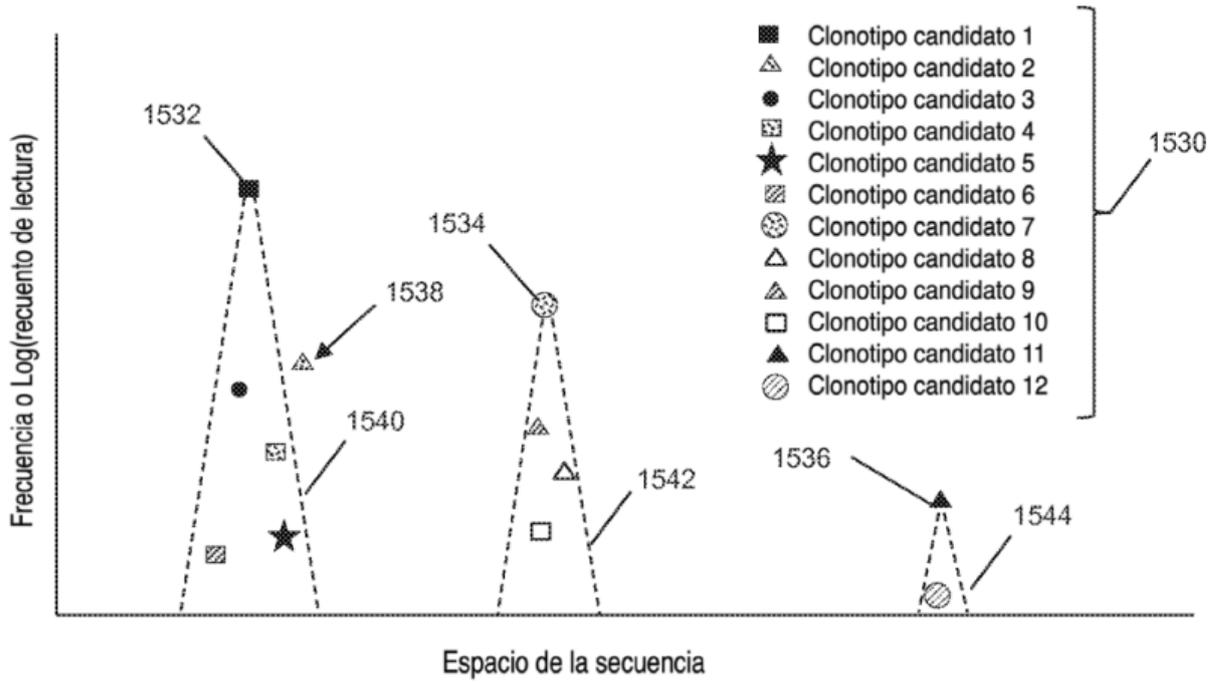


Fig. 10A

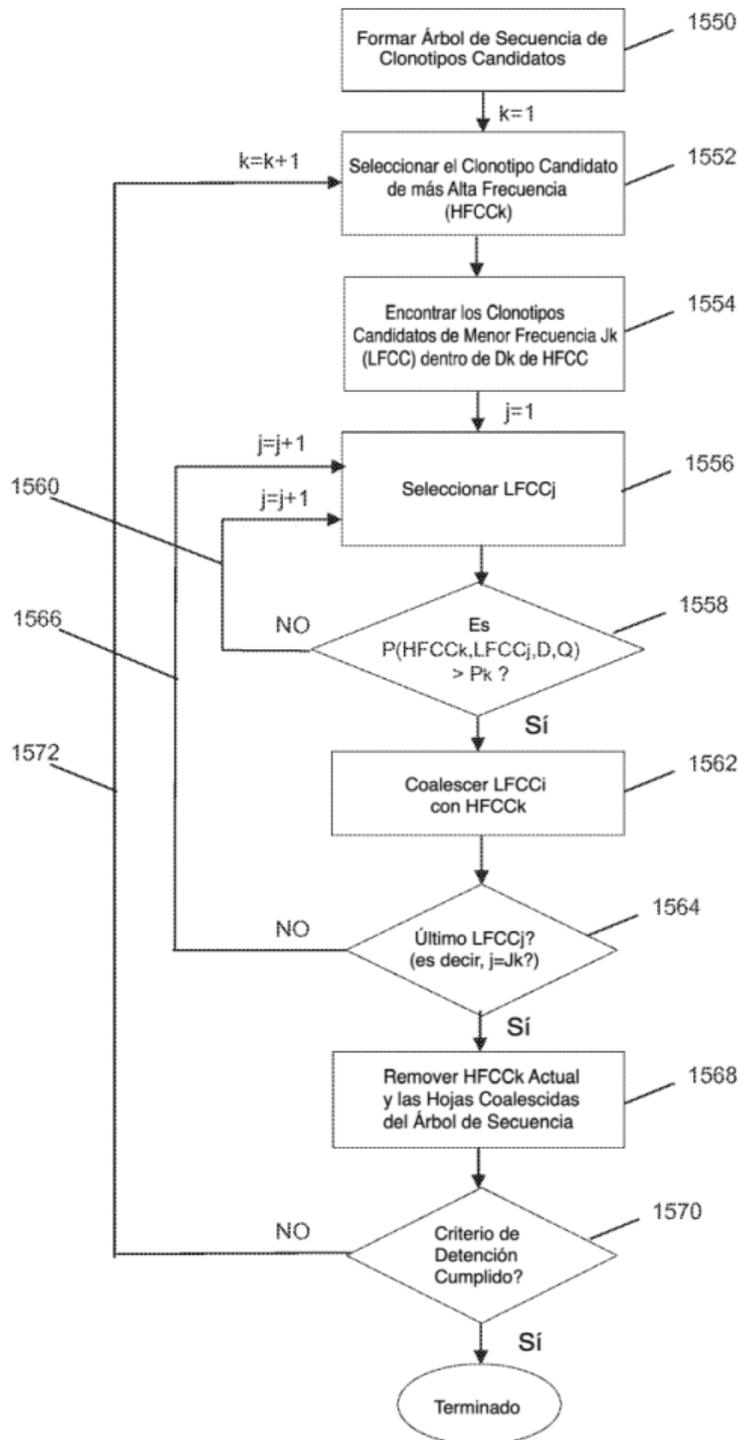


Fig. 10B

Ejemplo de Función para Determinar Si Coalescen los Clonotipos Candidatos Dependiendo de los Recuentos de Lecturas, Diferencias de Base y Puntajes Q

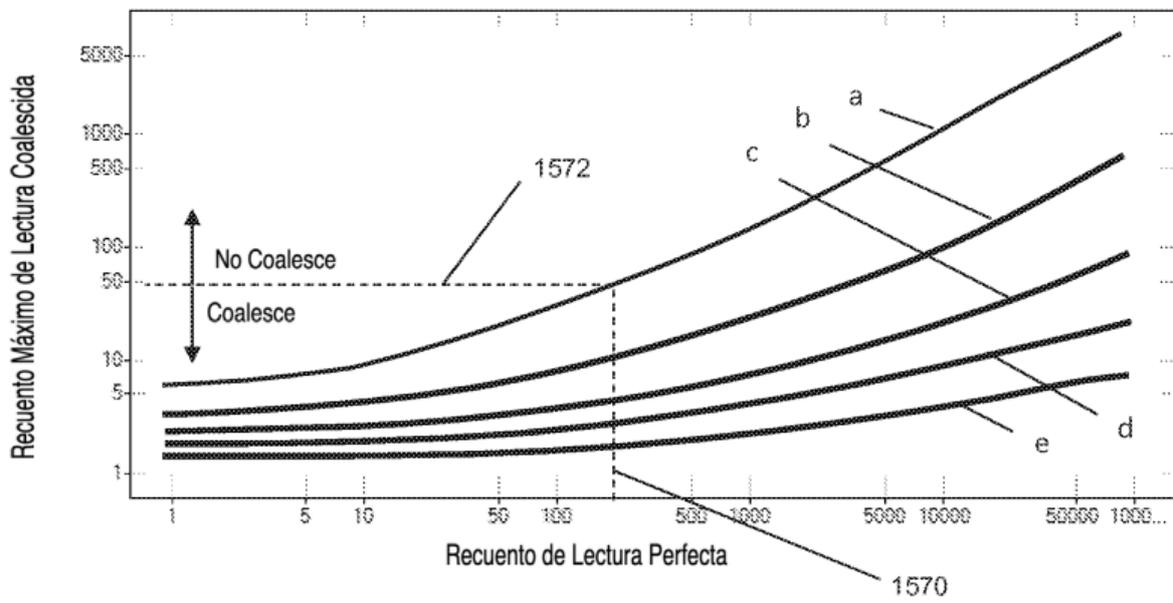


Fig. 10C

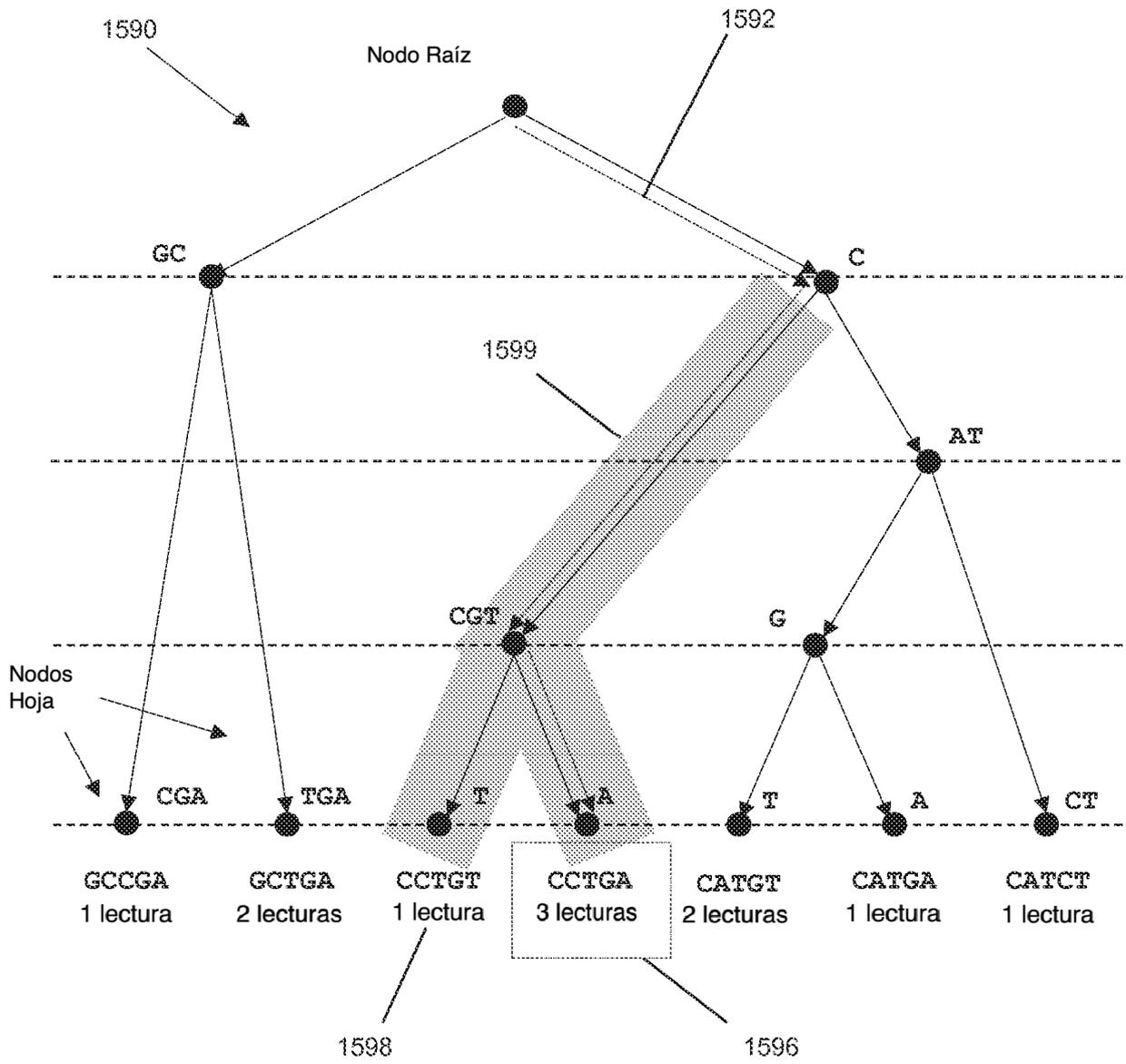


Fig. 10D

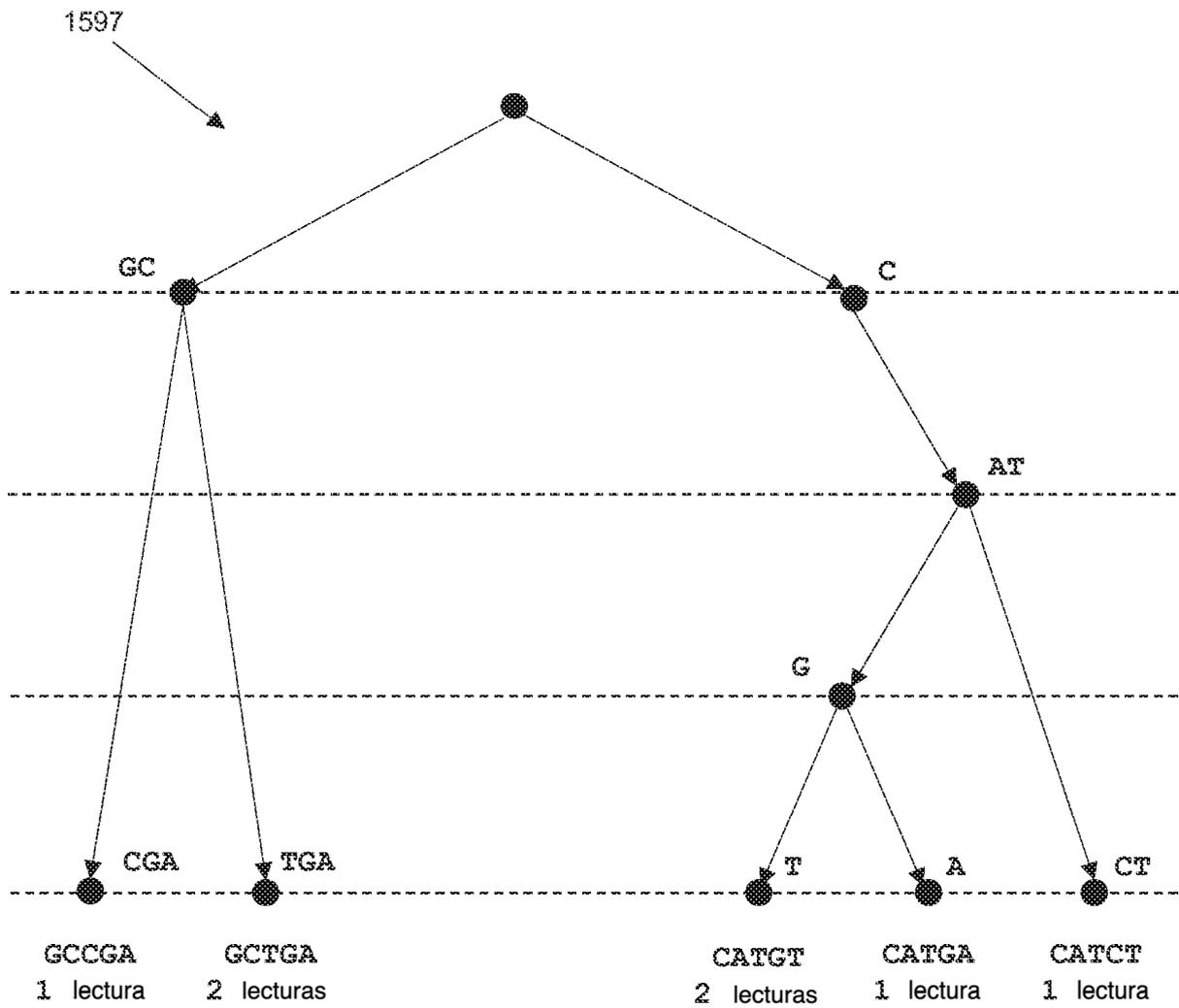


Fig. 10E