

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 593 614**

51 Int. Cl.:

**C12Q 1/68** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **04.05.2011 PCT/US2011/000791**

87 Fecha y número de publicación internacional: **10.11.2011 WO11139371**

96 Fecha de presentación y número de la solicitud europea: **04.05.2011 E 11777704 (5)**

97 Fecha y número de publicación de la concesión europea: **10.08.2016 EP 2567226**

54 Título: **Monitorización de estados de salud y enfermedad usando perfiles de clonotipos**

30 Prioridad:

**06.05.2010 US 332175 P**  
**25.10.2010 US 455743 P**  
**25.02.2011 US 446822 P**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:  
**12.12.2016**

73 Titular/es:

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

72 Inventor/es:

**FAHAM, MALEK y**  
**WILLIS, THOMAS**

74 Agente/Representante:

**UNGRÍA LÓPEZ, Javier**

**ES 2 593 614 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**

Monitorización de estados de salud y enfermedad usando perfiles de clonotipos

**Campo técnico**

5 La invención se refiere en general a la monitorización de afecciones de salud y enfermedad de un individuo por medidas moleculares, y más particularmente, a la monitorización de afecciones de salud y enfermedad de un individuo mediante la medición de perfiles de moléculas del sistema inmune usando secuenciación de ADN de alto rendimiento.

**Antecedentes de la invención**

10 El sistema inmune adaptativo, que comprende las respuestas humoral (o mediada por células B) y citotóxica (o mediada por células T), ha evolucionado para atacar características moleculares específicas en sus dianas respectivas. La aparición de una respuesta frente a una diana específica proporciona a un huésped "memoria" de ella, proporcionándole una capacidad de organizar una respuesta más fuerte si la misma diana apareciera otra vez. Habitualmente, cualquier proteína o polisacárido puede servir como la diana para algún subconjunto de las células de la respuesta inmune adaptativa o sus productos que reconocen características moleculares específicas, o epitopos, en la diana.

15 Como la enfermedad autoinmune implica el reconocimiento por algún componente del sistema inmune adaptativo de dianas propias, se han examinado aspectos del sistema inmune adaptativo para ayudar en el diagnóstico y pronóstico de dichas enfermedades. Usando tecnologías inmunológicas estándar, se ha investigado el sistema inmune humoral mediante la búsqueda de autoanticuerpos circulantes. Se han identificado autoanticuerpos, como antinuclear anti-ADNs, y el factor reumatoide, para varias enfermedades. Estos anticuerpos pueden no ser patológicos en sí mismos, ni tampoco la diana que reconocen en el cuerpo necesariamente la misma que la ensayada in vitro; sin embargo, la medida de sus niveles ayuda en el diagnóstico y, en algunos casos, tiene algunas implicaciones de pronóstico y tratamiento.

20 Otra metodología para estudiar el sistema inmune adaptativo en enfermedades autoinmunes y linfoides se basa en el análisis de la diversidad de las células inmunes adaptativas. La activación de las células inmunes adaptativas da lugar a su expansión clonal. La evidencia de esta expansión clonal se obtiene habitualmente por amplificación del ARN o ADN de la sangre de parte de la secuencia de ácido nucleico que codifica la región de reconocimiento de antígeno. Por ejemplo, se usan cebadores de PCR para amplificar secuencias que tienen un segmento V específico de la cadena  $\beta$  en el receptor de las células T (análogo a la cadena pesada de anticuerpo) para amplificar los segmentos J o segmentos J y D conectados al segmento V específico. Cuando está presente una población celular diversa, se espera amplificar fragmentos con una distribución de amplicones con un tamaño ligeramente diferente, pero la expansión clonal causa el enriquecimiento en tamaños específicos y así más intensos según se visualizan como bandas en un gel. En la técnica denominada "espectrotipificación", cada uno de los segmentos V se amplifica con los segmentos J y D para evaluar si alguno de estos amplicones muestra una expansión clonal.

25 Un problema de la estrategia de espectrotipificación es que muchas secuencias distintas pueden tener la misma longitud y por lo tanto son indistinguibles. Por lo tanto, por espectrotipificación sólo puede discernirse la expansión clonal dramática. Existe una necesidad de mejorar los métodos de diagnóstico y ayuda en el pronóstico de enfermedad autoinmune y estados de enfermedad autoinmune así otras enfermedades para las que el sistema inmune juega un papel central.

30 Aunque una especificidad adicional en el perfilado del sistema inmune sería de gran utilidad para permitir una mejor predicción de su impacto en la salud humana, se obtendría una utilidad incluso mayor si se desarrollaran métodos que permitieran identificar las células T y B específicas implicadas en los procesos de enfermedad incluso si esas secuencias particulares no se hubieran observado nunca anteriormente. La gran diversidad del sistema inmune le proporciona una reserva inmensa de células potencialmente útiles pero también presenta un reto para el investigador que intenta usar este repertorio para propósitos predictivos. Cualquier secuencia simple dirigida a un antígeno es una de un gran número que podría estar implicada con y/o correlacionarse con el proceso de la enfermedad en un individuo dado. Los métodos que identificarán cuáles de las muchas células en un individuo dado están implicadas con los procesos de la enfermedad serían muy valiosos para la salud humana.

35 El perfilado de las células inmunes también tiene utilidad en el diagnóstico y gestión de los cánceres. El tratamiento de los cánceres implica frecuentemente la evaluación de la respuesta al tratamiento así como la monitorización para la recurrencia de la enfermedad. Las metodologías más comunes para monitorizar la respuesta y la recurrencia del cáncer son evaluaciones radiográficas así como biomarcadores sanguíneos. Por ejemplo, los escaneos CT se usan frecuentemente para monitorizar la recurrencia del cáncer en múltiples enfermedades incluyendo cáncer de colon. De manera similar, los biomarcadores proteicos, como PSA y CEA, son biomarcadores sanguíneos usados para seguir los cánceres de próstata y colon. Las reorganizaciones genómicas específicas generan otra diana atractiva para uso para seguir a las células del cáncer. Por ejemplo, la translocación BCR-ABL presente en la gran mayoría de pacientes con leucemia mielógena crónica (CML) ha surgido como un analito para evaluar el estado de la enfermedad. La especificidad de la translocación para las células leucémicas y su susceptibilidad de ser ensayada por tecnología PCR permitió la generación de un ensayo altamente específico y sensible que se usa ahora rutinariamente para monitorizar a los pacientes con CML.

40 El perfilado de células inmunes (o clonotipo) puede usarse para generar marcadores para neoplasmas linfoides. El cáncer en el linaje de células linfoides es un conjunto heterogéneo de enfermedades clínicas

que refleja frecuentemente el estadio de desarrollo de la célula que ha experimentado la transformación a una célula cancerosa. La leucemia linfoblástica aguda (ALL) surge lo más frecuentemente de linfocitos inmaduros. Por otra parte, el mieloma múltiple (MM) ocurre en células plasmáticas que se han diferenciado para producir anticuerpos. De manera similar, los diferentes tipos de linfomas reflejan frecuentemente diferentes estadios de desarrollo celular. Estas enfermedades ocurren en diferentes grupos de edad, tienen diferentes pronósticos y mortalidad, y pueden tratarse con regímenes distintos. Estas enfermedades se tratan frecuentemente con quimioterapia, radioterapia, y/o trasplante de médula ósea. La recurrencia de la enfermedad se monitoriza entonces por diferentes métodos dependiendo de la situación clínica particular. Estos métodos incluyen la evaluación de la sangre y/o médula ósea usando recuentos y morfología de la sangre estándar, citometría de flujo (FCM) usando marcadores de la superficie celular, electroforesis de proteínas, así como técnicas moleculares como PCR y FISH. Además, frecuentemente se utilizan estudios radiográficos como escaneo CT y PET para monitorizar la recurrencia de algunos de los cánceres linfoides. Estos métodos implican invasión (médula ósea), coste y riesgo de radiación, y/o carecen de sensibilidad.

Algunos marcadores moleculares específicos para una célula de cáncer detectables por PCR de una manera sensible están presentes en una fracción de los neoplasmas linfoides. Por ejemplo, BCR-ABL está presente en una fracción de los pacientes con ALL y puede servir como un marcador para monitorizar la recidiva del tumor. Desafortunadamente, para la mayoría de los pacientes no existen dichos marcadores que pueden usarse para la detección sensible y específica de recidiva. FCM puede usarse para detectar enfermedad residual mínima (MRD) lo que es útil para propósitos de pronóstico. En esta técnica que usa separación de células activada por flujo multicolor (FACS), una célula de cáncer puede identificarse gracias a los marcadores particulares de superficie celular que tiene. La sensibilidad de esta técnica en las manos de expertos está limitada a  $<10^{-4}$  (1 célula de cáncer en 10.000 células normales) y los marcadores presentes en un punto de tiempo pueden desaparecer posteriormente. Por lo tanto, FCM generalmente no es útil para detectar la recidiva temprana en muestras de sangre.

La PCR proporciona una metodología sensible para la detección de secuencias específicas y se ha usado para detectar la reorganización particular en el receptor de células B (BCR) o receptores de células T (TCR) de la célula de cáncer (por ejemplo, véanse Greenberg et al., 1995, Journal of Leukocyte Biology, 57, 856-864; US 6.416.948 B1; y Campana 2009, Hematol Oncol Clin North Am. 23(5), 1083-1098). Esta técnica saca provecho del hecho de que los receptores de las células B o T en un linfocito se crean después de eventos de recombinación imperfectos que generan secuencias únicas para los diferentes linfocitos. Por ejemplo, un TCR está comprendido por cadenas TCR $\alpha$  y TCR $\beta$ . TCR $\alpha$  se crea mediante la recombinación que une una de las varias regiones V diferentes con una de las varias regiones J. De manera similar, TCR $\beta$  se crea mediante la recombinación que crea un segmento V, D, y J en tándem. En ambos casos, la recombinación frecuentemente no es perfecta y algunas bases pueden deleccionarse de las secuencias del segmento de la línea germinal y otras bases (denominadas bases N y P) pueden añadirse. La secuencia entre los segmentos V y J se refiere como la región NDN.

Estas secuencias pueden servir entonces como una etiqueta para estos linfocitos y su progenie. Como estos eventos de recombinación también ocurren en las células que finalmente se vuelven malignas, las secuencias únicas de los receptores de células B y T pueden servir como etiquetas para detectar las células de cáncer. La secuencia etiqueta es específica del paciente, y de hecho puede cambiar en el mismo paciente debido a evolución clonal. Para definir la secuencia del receptor de células T o B de las células leucémicas para un paciente, se usa la muestra de leucemia de diagnóstico que está habitualmente altamente enriquecida para el clon leucémico. Por ejemplo, el ADN del receptor de células T y/o B se amplifica de una muestra de diagnóstico, y el producto se corre en un gel que puede separar el ADN sobre la base del tamaño (algunas veces referido como "espectrotipificación"); o alternativamente, puede hacerse análisis de heterodúplex. Un alto grado de sesgo de la distribución de tamaño observada indica expansión monoclonal, que puede confirmarse secuenciando una muestra del pico de separación sesgado. Sin dicha secuenciación posterior, frecuentemente es difícil determinar si dicho sesgo tiene orígenes monoclonal o policlonal, por ejemplo, Van Dongen et al, publicación de patente US 2006/0234234.

Una vez se ha identificado la etiqueta de secuencia, puede usarse PCR en tiempo real usando sondas Taqman para monitorizar el nivel de esa secuencia. La región NDN no es habitualmente lo suficientemente larga como para englobar los cebadores de PCR y el oligonucleótido de detección. Por lo tanto, típicamente, se usan cebadores de PCR complementarios a las regiones V y J y una sonda Taqman que incluye algunas de las bases de NDN del clon leucémico. Los cebadores proporcionan parte de la especificidad, ya que sólo amplifican una fracción del repertorio completo. La especificidad frente al clonotipo particular se proporciona por la hibridación de la sonda Taqman. Por lo tanto, la sensibilidad del ensayo no es habitualmente tan buena como en una PCR típica (por ejemplo, BCR-ABL), en la que la pareja cebadora (con o sin la sonda Taqman) proporciona la especificidad. Se ha mostrado que la sensibilidad puede ser tan alta como  $10^{-5}$  para algunas secuencias pero puede ser significativamente peor dependiendo de la especificidad de hibridación proporcionada por la sonda Taqman cuya secuencia es complementaria al menos a parte de la región NDN. Dada la baja sensibilidad para algunas sondas, el ensayo puede no funcionar para ninguna de las reorganizaciones en un paciente particular. El asunto de la evolución clonal también ha surgido previamente reduciendo más la probabilidad de detectar un bajo nivel de leucemia. Además, esta técnica es engorrosa ya que requiere la generación de sondas Taqman específicas de paciente así como molde para ser usados como estándares. Es necesario usar estos

estándares específicos de paciente cada vez que se va a ensayar una muestra del paciente. La inconsistencia de la sensibilidad entre pacientes, la naturaleza engorrosa, y los problemas logísticos de obtener controles apropiados para el ensayo han limitado en gran medida su uso. Por lo tanto, existe una necesidad de generar marcadores que puedan usarse para monitorizar la recidiva en pacientes con neoplasmas linfoides. En algunas realizaciones, la invención descrita en la presente memoria facilita que se desarrolle un conjunto de marcadores muy general, sensible, y específico para la gestión de pacientes con cánceres linfoides usando secuenciación de células inmunes.

Sería ventajoso para muchos campos, incluyendo particularmente los campos autoinmunes y de cáncer linfoide, si hubiera ensayos disponibles para evaluar los perfiles de clonotipos de individuos que fueran más sensibles y exhaustivos que las técnicas actuales y que fueran aplicables generalmente sin la necesidad de fabricar reactivos individualizados.

#### Resumen de la invención

La invención proporciona un método para detectar la recurrencia de un neoplasma linfoide en un paciente con clonotipos de cáncer definidos que comprende las etapas de (a) determinar un perfil de secuencias de ADN recombinadas en células T y/o células B que comprende las etapas de (i) amplificar moléculas de ácido nucleico de células T y/o células B de una muestra obtenida del paciente; (ii) aislar espacialmente las moléculas individuales del ácido nucleico amplificado; y (iii) secuenciar las moléculas individuales de ácido nucleico por secuenciación de nueva generación para formar el perfil de secuencias de ADN recombinadas; (b) evaluar el perfil de secuencias de ADN recombinadas para determinar los niveles de los clonotipos de cáncer definidos y los clones que evolucionan de los mismos; y (c) usar los niveles para detectar la recurrencia del neoplasma linfoide.

En la presente memoria se describe un método para monitorizar una enfermedad que comprende las etapas de (a) identificar uno o más clonotipos específicos de paciente correlacionados con una enfermedad mediante la determinación de un perfil de clonotipos de una muestra de linfocitos en un tejido relacionado con la enfermedad, comprendiendo la muestra un repertorio de clonotipos del tejido relacionado con la enfermedad; (b) determinar un perfil de clonotipos de una muestra de células de sangre periférica para identificar la presencia, ausencia y/o nivel de uno o más clonotipos específicos de paciente correlacionados con la enfermedad, comprendiendo dicha muestra de sangre periférica un repertorio de clonotipos; y (c) repetir la etapa (b) para monitorizar la enfermedad o afección en el paciente. En un método, la etapa de identificación incluye además determinar un perfil de clonotipos de una muestra de linfocitos en un tejido no relacionado con la enfermedad en el mismo paciente y comparar dicho perfil de clonotipos con el de dicho tejido relacionado con la enfermedad para identificar dicho uno o más clonotipos específicos de paciente. Las enfermedades que pueden monitorizarse incluyen, pero no están limitadas a, trastornos proliferativos linfoides, tumores sólidos, enfermedades infecciosas, y enfermedades autoinmunes. El tamaño del repertorio puede variar ampliamente dependiendo de las aplicaciones particulares; pero, en un método, un repertorio incluye cada clonotipo en una muestra de un individuo presente a una frecuencia de 0,01 por ciento o mayor con una probabilidad de noventa y nueve por ciento. En otros métodos, un repertorio incluye cada clonotipo en una muestra de un individuo presente a una frecuencia de 0,001 por ciento o mayor con una probabilidad de noventa y nueve por ciento.

En la presente memoria también se describe un método para monitorizar una enfermedad que comprende las etapas del método A de monitorizar una enfermedad en un paciente, comprendiendo el método las etapas de: (a) determinar un perfil de clonotipos de una muestra de linfocitos de un individuo afligido con una enfermedad y un perfil de clonotipos del mismo individuo de una muestra de linfocitos enriquecida sobre la base de marcadores de la superficie celular asociados con la enfermedad para identificar uno o más clonotipos específicos de paciente correlacionados con la enfermedad, en el que cada una de las muestras comprende un repertorio de clonotipos de las poblaciones de linfocitos enriquecida y no enriquecida; (b) determinar el nivel de cada uno de los uno o más clonotipos específicos de paciente en un perfil de clonotipos de una muestra de células de sangre periférica, teniendo dicha muestra un volumen definido y comprendiendo un repertorio de clonotipos de éstos; y (c) repetir la etapa (b) para monitorizar la enfermedad o afección en el paciente.

En casos adicionales de los métodos anteriores, las etapas respectivas de determinación de un repertorio de una muestra de células de sangre periférica comprende además incluir como uno o más clonotipos específicos de paciente cualesquiera de los clonotipos no registrados previamente que son clonotipos filogénicos del uno o más clonotipos específicos de paciente. Siempre que la enfermedad es un trastorno proliferativo linfoide, el uno o más clonotipos específicos de paciente de dichos métodos puede incluir además mutaciones y reorganizaciones genéticas relacionadas con el cáncer adicionales, tales como reemplazos en la región V (descrito con más detalle más adelante) identificados fácilmente por los métodos descritos en la presente memoria.

En la presente memoria también se describe un método para medir simultáneamente el número de linfocitos y los niveles de expresión de clonotipos en una muestra que comprende las etapas de: (i) obtener de un individuo una muestra que comprende células T y/o células B; (ii) secuenciar moléculas individuales aisladas espacialmente derivadas de ADN genómico de dichas células, comprendiendo dichas moléculas individuales aisladas espacialmente un número de clonotipos correspondiente a un número de linfocitos en la muestra; (iii) secuenciar moléculas individuales aisladas espacialmente derivadas de ARN de dichas células, comprendiendo dichas moléculas individuales aisladas espacialmente números de clonotipos correspondientes a niveles de expresión de éstos en los linfocitos de la muestra; y (iv) determinar los niveles de expresión de clonotipos en los linfocitos de la muestra

comparando para cada clonotipo el número determinado de moléculas individuales aisladas derivadas de ADN genómico de dichas células y el número determinado de moléculas individuales aisladas derivadas de ARN de dichas células.

5 La invención supera varias deficiencias en la técnica anterior proporcionando, entre otras ventajas, métodos basados en secuencia para medir con mucha mayor sensibilidad los clonotipos correlacionados con afecciones de enfermedad o salud. La invención proporciona además dichos ensayos en un formato general aplicable a cualquier paciente sin la necesidad de fabricar reactivos individualizados o específicos de paciente. Dichos avances tienen aplicaciones particularmente útiles en las áreas de autoinmunidad y cánceres linfoides. En la última área, la invención proporciona además métodos de ensayo y  
10 monitorización que son capaces de detectar y seguir no sólo niveles muy bajos de clonotipos correlacionados con enfermedad sino también clonotipos que han experimentado modificaciones que escaparían a la detección por metodologías anteriores. Esta última característica tiene un valor tremendo, por ejemplo, en la monitorización de enfermedad residual mínima en cánceres linfoides.

#### Descripción breve de los dibujos

15 Las características nuevas de la invención se muestran con particularidad en las reivindicaciones adjuntas. Una mejor comprensión de las características y ventajas de la presente invención se obtiene por referencia a la descripción detallada siguiente que muestra realizaciones ilustrativas, en las que se utilizan los principios de la invención, y los dibujos adjuntos de los cuales:

La FIG. 1A es un diagrama de flujo de un ejemplo de un método para determinar perfiles de clonotipos.

20 La FIG. 1B ilustra la distribución relativa de mutaciones somáticas que ocurren en inmunoglobulinas producidas por células B.

Las FIGS. 2A-2B muestran un esquema de PCR de dos etapas para amplificar los genes TCR $\beta$ .

25 La FIG. 3A ilustra un producto de PCR que se amplificó usando el esquema de las Figs 2A-2B, que va a experimentar una PCR secundaria para añadir sitios de unión de cebadores para amplificación en puente y secuenciación para secuenciación basada en Solexa. La FIG. 3B ilustra detalles de un ejemplo de determinación de una secuencia de nucleótidos del producto de PCR de la Fig. 3A. La FIG. 3C ilustra detalles de otro ejemplo de determinación de una secuencia de nucleótidos del producto de PCR de la Fig. 3A.

30 La FIG. 4A ilustra un esquema de PCR para generar tres moldes de secuenciación de una cadena de IgH en una única reacción. Las FIGS. 4B-4C ilustran un esquema de PCR para generar tres moldes de secuenciación de una cadena de IgH en tres reacciones separadas después de lo cual los amplicones resultantes se combinan para una PCR secundaria para añadir sitios de unión de cebador P5 y P7. La Fig. 4D ilustra las localizaciones de lecturas de secuencia generadas para una cadena de IgH. La Fig. 4E ilustra el uso de la estructura de codones de las regiones V y J para mejorar las llamadas de base en la región NDN.

La FIG. 5 muestra datos que evidencian la reproducibilidad de las PCR múltiples de la invención.

La FIG. 6 muestra datos que demuestran que las PCR múltiples de la invención introducen un sesgo de amplificación mínimo.

40 La FIG. 7A muestra el log<sub>10</sub> de la frecuencia de cada clonotipo en las dos muestras duplicadas usando Accuprime y ADNc correspondiente a 500 ng de ARN como molde de entrada.

La FIG. 7B representa el log<sub>10</sub> de la frecuencia de cada clonotipo usando ADNc correspondiente a 500 ng de ARN como molde de entrada y Accuprime (eje de las X) o Taq de alta fidelidad (eje de las Y).

La FIG. 7C muestra el log<sub>10</sub> de la frecuencia de cada clonotipo usando ADNc correspondiente a 50 ng de ARN como molde de entrada y Accuprime (eje de las X) o Taq de alta fidelidad (eje de las Y).

45 La FIG. 8 es el número de moléculas de TCR $\beta$  de muestras. 8A y 8B muestran datos de la amplificación de IgH de ADN genómico.

La FIG. 9 muestra datos que indican que las amplificaciones múltiples según la invención tienen un sesgo de amplificación mínimo.

La FIG. 10 muestra datos que comparan clonotipos de dos individuos.

#### 50 Descripción detallada de la invención

Un aspecto de esta invención utiliza tecnologías de secuenciación de nueva generación para evaluar los niveles de reorganizaciones de TCR o BCR en una población de linfocitos. Estas tecnologías de secuenciación pueden obtener 1 millón o más de lecturas de una muestra a un coste razonable. Un clonotipo presente a una frecuencia de 1/1.000.000 o menor aún puede detectarse de manera específica  
55 usando estas tecnologías. La amplificación múltiple para amplificar todos los tipos diferentes de secuencias de una porción específica de gen o transcrito puede conseguirse a partir de ADN de una muestra sangre o médula ósea. Por ejemplo, para amplificar secuencias de IgH, pueden usarse varios cebadores complementarios a todos los segmentos V conocidos y alelos junto con varios cebadores complementarios a todos los segmentos J y alelos, La Fig. 1A ilustra las etapas de dicho método para una realización que emplea una clase de secuenciadores de ADN (por ejemplo, secuenciación por síntesis Solexa, como se describe más adelante) para el perfilado de clonotipos de una muestra. Se obtiene una muestra que contiene células B o células T (100) después de lo cual se extrae el ADN o ARN y amplifica (102) en una reacción que amplifica preferentemente clonotipos y una secuencias terminales para amplificación y secuenciación posterior. Las moléculas individuales de los clonotipos amplificados se  
60 distribuyen aleatoriamente en una superficie sólida (104), tal como, una superficie de vidrio, que se ha configurado para permitir una segunda amplificación in situ para producir poblaciones clonales (o colonias) de cada molécula individual (106). Las moléculas de cada colonia se secuencian (108), por

ejemplo, usando una técnica de secuenciación por síntesis, después de lo cual los tipos y abundancias de las secuencias se tabulan para formar un perfil de clonotipos (110), o de manera equivalente un perfil de repertorios. El método puede realizarse con un sesgo de amplificación pequeño entre las diferentes secuencias. El ARN de los genes de TCR $\beta$  e IgH puede amplificarse sólo con pequeñas diferencias en las eficiencias de los diferentes cebadores de V, validando de esta manera la posibilidad de hacer lo mismo a partir de ADN. Este esquema puede mejorar los problemas para la lectura en tiempo real para la detección de reorganizaciones de bajo nivel de TCR y/o BCR.

La sensibilidad se determina mediante estadísticas de conteo (esto es, la sensibilidad se incrementa incrementando los tamaños de la muestra de células y secuenciación) y amplificación equivalente (esto es, los clonotipos que tienen secuencias variadas pueden amplificarse sin sesgo significativo en una reacción de amplificación múltiple, tal como una PCR, como se ilustra más adelante). Como la sensibilidad está limitada finalmente por estadísticas de conteo, para obtener más sensibilidad simplemente se pueden obtener más células (es decir, muestras mayores) y más lecturas de secuenciación. Con lecturas de secuenciación suficientes, la sensibilidad está limitada por el número de linfocitos en la muestra. Por el contrario, la sensibilidad para el ensayo de PCR en tiempo real está limitada por el fondo. Además, los clones específicos de un paciente pueden determinarse secuenciando una muestra de leucemia o linfoma de diagnóstico. Una vez se ha determinado el clonotipo, su nivel puede determinarse en muestras a puntos de tiempo posteriores. En algunas realizaciones preferidas, no existe ningún requerimiento para sondas o cebadores específicos de paciente o la utilización de moldes específicos de paciente para utilizarse como estándares. En su lugar, los clones específicos de paciente se siguen almacenando los datos respecto a las secuencias relevantes para cada paciente, y el mismo ensayo funciona para todos los pacientes.

En general, en la presente memoria se describen métodos para aplicar técnicas de secuenciación de ácidos nucleicos a la tarea de monitorizar el repertorio de células de inmunidad adaptativa para hacer un perfil del sistema inmune. Los perfiles del sistema inmune generados pueden usarse para el diagnóstico de enfermedades y trastornos, y para el diagnóstico de estados de enfermedades y trastornos. Los métodos de perfilado inmune de la invención proporcionada pueden usarse para monitorizar enfermedades y trastornos y evaluar el tratamiento de enfermedades y trastornos. Las enfermedades y trastornos a los que se pueden aplicar los métodos de la invención proporcionada incluyen enfermedad autoinmune, incluyendo lupus eritematoso sistémico (SLE), esclerosis múltiple (MS), artritis reumatoide (RA), y espondilitis anquilosante (AS). Los métodos descritos en la presente memoria pueden aplicarse al diagnóstico, monitorización, y tratamiento de rechazo de trasplantes y envejecimiento inmune. Además, los métodos de perfilado inmune de la invención proporcionada pueden usarse para diagnosticar, monitorizar, y tratar otras enfermedades relacionadas con el sistema inmune, incluyendo cáncer y enfermedad infecciosa.

La secuenciación de moléculas amplificadas individuales puede distinguir diferentes secuencias y por lo tanto tiene la sensibilidad de detectar cambios cuantitativos en la expansión clonal. En general, en un caso, se proporciona un método para determinar un perfil de secuencias de ADN recombinadas en células T y/o células B. El método puede comprender etapas que incluyen aislar muestras de un sujeto, una o más rondas de amplificación de ácidos nucleicos, aislar espacialmente ácidos nucleicos individuales, y secuenciar los ácidos nucleicos. Los ácidos nucleicos pueden ser ADN o ARN. Las secuencias de ADN recombinadas en células T y/o células B pueden denominarse clonotipos.

En un caso, se proporciona un método para determinar uno o más clonotipos que se correlacionan en un sujeto o individuo. En otro caso, se proporciona un método para desarrollar un algoritmo que puede predecir uno o más clonotipos que se correlacionan en cualquier muestra de un sujeto con una enfermedad. En otro caso, se proporciona un método para descubrir uno o más clonotipos que se correlacionan para un individuo usando un algoritmo que puede predecir uno o más clonotipos que se correlacionan en cualquier muestra de un sujeto. En otro caso, se proporciona un método para generar un algoritmo que calcula una actividad de enfermedad. En otro caso, se proporciona un método para monitorizar el estado de una enfermedad de un individuo.

El perfilado del repertorio de células T y B puede ser valioso para enfermedades con aspectos inflamatorios. Esta inflamación se debe frecuentemente a una reacción autoinmune y/o de hipersensibilidad. Estas enfermedades incluyen enfermedad cardiovascular, enfermedad de Alzheimer, y pre-eclampsia. La inflamación también se ha asociado con estados metabólicos anormales incluyendo obesidad y diabetes. Existen otras enfermedades relacionadas con inflamación. En un aspecto de la invención, un segmento de ácido nucleico de células B recombinado se amplifica por una PCR con una pluralidad de cebadores directos o una pluralidad de cebadores inversos para generar un conjunto anidado de moldes (véase la Fig. 4A y 4B y sus descripciones más adelante). Los moldes de dicho conjunto pueden amplificarse adicionalmente en una superficie para formar amplicones separados (por ejemplo, por PCR en puente usando un instrumento cBot, Illumina, San Diego, CA). Los moldes del mismo conjunto anidado pueden estar asociados entre sí por lecturas de secuencia generadas en sus extremos comunes. Los conjuntos anidados de moldes permiten usar una química de secuenciación con proporciones de error relativamente altas para analizar secuencias más largas de lo que de otro modo sería posible, mientras al mismo tiempo se mantiene un alto promedio de puntuaciones de calidad sobre la longitud completa de la secuencia. Los conjuntos anidados también aseguran que se obtiene al menos una lectura de secuencia de una región V incluso si se ha sometido a hipermutación somática. En una realización, pueden usarse químicas de secuenciación para analizar ácidos nucleicos altamente variables,

tales como moléculas de IgH, que tienen proporciones de error no mejores de las siguientes: 0,2 por ciento de lecturas de secuencia contienen al menos un error en las posiciones 1-50; 0,2-1,0 por ciento de lecturas de secuencia contienen al menos un error en las posiciones 51-75; 0,5-1,5 por ciento de lecturas de secuencia contienen al menos un error en las posiciones 76-100; y 1-5 por ciento de lecturas de secuencia contienen al menos un error en las posiciones 101-125. En otra realización, se posicionan sitios de unión del cebador de secuenciación de manera que cuando se extienden producen una serie de lecturas de secuencia en las que cada lectura de secuencia excepto la última se superpone con su sitio de unión a cebador aguas abajo inmediatamente adyacente y/o lectura de secuencia, proporcionando de esta manera una cobertura de secuencia continua con puntuaciones de mayor calidad de que lo que sería posible si se usara un único molde largo para generar una única lectura de secuencia larga.

#### I. Descripción adicional

Se describe un método para determinar un perfil de secuencias de ADN recombinadas en células T y/o células B que comprende: obtener una muestra de un sujeto que comprende células T y/o células B, aislar espacialmente moléculas individuales del ADN genómico de dichas células; secuenciar dichas moléculas individuales aisladas espacialmente del ADN genómico, y determinar los niveles de diferentes secuencias de dicha muestra para generar dicho perfil de secuencias de ADN recombinadas. Se proporciona un método para determinar un perfil de secuencias de ADN recombinadas en células T y/o células B que comprende: obtener una muestra de un sujeto que comprende células T y/o células B, aislar espacialmente moléculas individuales del ADN genómico de dichas células, amplificar dichas moléculas individuales del ADN genómico, secuenciar dicho ADN amplificado, y determinar los niveles de diferentes secuencias de dicha muestra para generar dicho perfil de secuencias de ADN recombinadas. Se proporciona un método para determinar un perfil de secuencias de ADN recombinadas en células T y/o células B que comprende: obtener una muestra de un sujeto que comprende células T y/o células B, amplificar el ADN genómico de dichas células, aislar espacialmente moléculas individuales de dicho ADN amplificado, secuenciar dichas moléculas individuales aisladas espacialmente de ADN amplificado; y determinar los niveles de diferentes secuencias de dicha muestra para generar dicho perfil de secuencias de ADN recombinadas. Se proporciona un método para determinar un perfil de secuencias de ADN recombinadas en células T y/o células B que comprende: obtener una muestra de un sujeto que comprende células T y/o células B, amplificar el ADN genómico de dichas células, aislar espacialmente moléculas individuales de dicho ADN amplificado, re-amplificar dichas moléculas de ADN amplificado, secuenciar dichas moléculas de ADN re-amplificado, y determinar los niveles de diferentes secuencias de dicha muestra para generar dicho perfil de secuencias de ADN recombinadas. Se proporciona un método para determinar un perfil de secuencias de ADN recombinadas en células T y/o células B que comprende: obtener una muestra de un sujeto que comprende células T y/o células B, transcribir de forma inversa el ARN de dichas células para formar ADNc, aislar espacialmente moléculas individuales de dicho ADNc, opcionalmente re-amplificar dichas moléculas individuales aisladas espacialmente de ADNc, secuenciar dicho ADNc y/o ADNc re-amplificado; y determinar los niveles de diferentes secuencias de dicha muestra para generar dicho perfil de secuencias de ADN recombinadas. Se proporciona un método para determinar un perfil de secuencias de ADN recombinadas en células T y/o células B que comprende: obtener una muestra de un sujeto que comprende células T y/o células B; aislar espacialmente células individuales en dicha muestra, secuenciar moléculas individuales de ácido nucleico de dichas células; y determinar los niveles de diferentes secuencias de dicha muestra para generar dicho perfil de secuencias de ADN recombinadas. En un caso, dicha amplificación y/o re-amplificación comprende PCR, PCR múltiple, TMA, NASBA, o LAMP. En otro caso, dicho aislamiento espacial comprende separar dicho ADN o ADNc en dos dimensiones en un soporte sólido, separar dicho ADN o ADNc en tres dimensiones en una disolución con micelas, o separar moléculas usando cámaras de micro-reacción. En otro caso, dicha amplificación y/o re-amplificación es por crecimiento de bacterias que portan ADN o ADNc subclonado, amplificación de ADN o ADNc en un portaobjetos, o amplificación de ADN o ADNc en un lecho. En otro caso, dicha secuenciación comprende secuenciación didesoxi. En otro caso, dicha secuenciación comprende secuenciación por síntesis usando nucleótidos marcados en extremo reversiblemente. En otro caso, dicha secuenciación comprende detección de la liberación de pirofosfato en la incorporación de nucleótidos. En otro caso, dicha secuenciación comprende hibridación específica de alelo a una biblioteca de sondas oligonucleotídicas marcadas. En otro caso, dicha secuenciación comprende secuenciación por síntesis usando hibridación específica de alelo a una biblioteca de sondas oligonucleotídicas marcadas seguido de ligación de dichas sondas. En otro caso, dicha secuenciación comprende monitorización en tiempo real de la incorporación de nucleótidos marcados durante una etapa de polimerización. En otro caso, dichas secuencias de ADN recombinadas comprenden genes del receptor de células T y/o genes de inmunoglobulina. En otro caso, dicha secuenciación comprende secuenciación de un subconjunto de las secuencias clonales completas de genes de inmunoglobulina y/o del receptor de células T. En otro caso, dicho subconjunto de secuencia clonal completa comprende la unión V-D, unión D-J del gen de una inmunoglobulina o receptor de células T, la región variable completa de un gen de inmunoglobulina o receptor de células T, la región de reconocimiento de antígeno, o la región determinante de la complementariedad 3 (CDR3). En otro caso, dichos genes del receptor de células T comprenden genes del receptor  $\beta$  de células T. En otro caso, dichos genes de inmunoglobulina comprenden genes de inmunoglobulina pesados. En otro caso, dicha amplificación o re-amplificación comprende una pluralidad de cebadores complementarios a los segmentos V y un cebador complementario a un segmento C. En otro caso, dicha amplificación o re-amplificación comprende una pluralidad de cebadores complementarios

a los segmentos V y una pluralidad de cebadores complementarios a los segmentos C. En otro caso, dicha pluralidad de cebadores complementarios a los segmentos V comprende al menos tres cebadores diferentes para cada segmento V y la pluralidad de cebadores complementarios a los segmentos C comprenden al menos 1, al menos 2, al menos 3, al menos 4, al menos 5, o al menos 6 cebadores. En otro caso, dichas células T o B son subconjuntos de las células T y B totales. En otro caso, dicho subconjunto de células T son células CD4+, CD8+, o células con CD27 alto. En otro caso, dicha muestra comprende al menos 100.000, al menos 500.000, al menos 750.000, de al menos 1.000.000 de células T. En otro caso, dicha secuenciación comprende al menos 1.000 de lecturas por operación, al menos 10.000 de lecturas por operación, al menos 100.000 de lecturas por operación, o al menos 1.000.000 de lecturas por operación. En otro caso, dicha secuenciación comprende generar aproximadamente 30 pb, aproximadamente 40 pb, aproximadamente 50 pb, aproximadamente 60 pb, aproximadamente 70 pb, aproximadamente 80 pb, aproximadamente 90 pb, aproximadamente 100 pb, aproximadamente 110, o aproximadamente 120 pb por lectura. En otro caso, dicha muestra se toma cuando el sujeto está en un estado de crisis de una enfermedad autoinmune. En otro caso, dicha muestra se toma de un sujeto que tiene o se sospecha que tiene lupus eritematoso sistémico. Se proporciona un método para determinar uno o más clonotipos que se correlacionan en un sujeto que comprende: generar uno o más perfiles de clonotipos por secuenciación de ácido nucleico de moléculas individuales aisladas espacialmente de al menos una muestra del sujeto, en el que la al menos una muestra está relacionada con un primer estado de la enfermedad, y determinar uno o más clonotipos que se correlacionan en el sujeto sobre la base de uno o más perfiles de clonotipos. En un caso, dicha al menos una muestra es de un tejido afectado por la enfermedad. En otro caso, dicha determinación de uno o más clonotipos que se correlacionan comprende comparar los perfiles de clonotipos de al menos dos muestras. En otro caso, el primer estado de la enfermedad es un estado de pico de la enfermedad. En otro caso, dichos uno o más clonotipos que se correlacionan están presentes en el estado de pico de la enfermedad, En otro caso, dichos uno o más clonotipos que se correlacionan están ausentes en el estado de pico de la enfermedad. En otro caso, dichos uno o más clonotipos que se correlacionan están altos en el estado de pico de la enfermedad. En otro caso, dichos uno o más clonotipos que se correlacionan están bajos en el estado de pico de la enfermedad. En otro caso, dicha muestra comprende células T y/o células B. En otro caso, dichas células T y/o células B comprenden un subconjunto de células T y/o células B. En otro caso, dicho subconjunto de células T y/o células B está enriquecido por interacción con un marcador. En otro caso, dicho marcador es un marcador de superficie celular en el subconjunto de células T y/o células B. En otro caso, dicho subconjunto de células T y/o células B interacciona con un antígeno presente específicamente en la enfermedad. En otro caso, la enfermedad es lupus eritematoso sistémico o esclerosis múltiple. Se describe un método para desarrollar un algoritmo que puede predecir uno o más clonotipos que se correlacionan en cualquier muestra de un sujeto con una enfermedad, que comprende: a) generar una pluralidad de perfiles de clonotipos de un conjunto de muestras, en el que las muestras son relevantes para la enfermedad, b) identificar uno o más clonotipos que se correlacionan del conjunto de muestras, c) usar parámetros de secuencia y/o datos funcionales de uno o más clonotipos que se correlacionan identificados en b) para desarrollar el algoritmo que puede predecir clonotipos que se correlacionan en cualquier muestra de un sujeto con la enfermedad. En un caso, el conjunto de muestras se toma de uno o más tejidos afectados por la enfermedad. En otro caso, dicha identificación de uno o más clonotipos que se correlacionan comprende comparar perfiles de clonotipos de al menos dos muestras. En otro caso, dichos datos funcionales incluyen la capacidad de unión de marcadores en la superficie de células T y/o células B o interacción con antígeno por una célula T o célula B. En otro caso, dichos parámetros de secuencia comprenden secuencia de ácido nucleico y secuencia de aminoácidos predicha. En otro caso, las muestras son de uno o más individuos en un estado de pico de la enfermedad. En otro caso, dichos uno o más clonotipos que se correlacionan están presentes en el estado de pico de la enfermedad. En otro caso, dichos uno o más clonotipos que se correlacionan están a un alto nivel en el estado de pico de la enfermedad. En otro caso, dichos uno o más clonotipos que se correlacionan están a un bajo nivel en el estado de pico de la enfermedad. En otro caso, el uno o más clonotipos que se correlacionan están ausentes en el estado de pico de la enfermedad. En otro caso, la enfermedad es lupus eritematoso sistémico o esclerosis múltiple. En otro caso, se proporciona un método para descubrir uno o más clonotipos que se correlacionan para un individuo, que comprende introducir un perfil de clonotipos de una muestra del individuo en un algoritmo, y usar el algoritmo para determinar uno o más clonotipos que se correlacionan para el individuo. En un caso, se proporciona el algoritmo que es un algoritmo que puede predecir uno o más clonotipos que se correlacionan en cualquier muestra de un sujeto con una enfermedad, que comprende, desarrollar dicho algoritmo por: a) generación de una pluralidad de perfiles de clonotipos de un conjunto de muestras, en el que las muestras son relevantes para la enfermedad, b) identificación de uno o más clonotipos que se correlacionan del conjunto de muestras, c) uso de parámetros de secuencia y/o datos funcionales de uno o más clonotipos que se correlacionan identificados en b) para desarrollar un algoritmo que puede predecir clonotipos que se correlacionan en cualquier muestra de un sujeto con la enfermedad. En un caso, dicha muestra se toma en un estado de pico de la enfermedad. En otro caso, la muestra se toma de tejido afectado por la enfermedad. Se describe un método para generar un algoritmo que calcula una puntuación de la actividad de una enfermedad, que comprende: desarrollar un algoritmo que usa un conjunto de factores para combinar niveles de clonotipos que se correlacionan en una puntuación de actividad de la enfermedad, comparar la puntuación de actividad de la enfermedad con datos clínicos respecto al estado de la enfermedad, y



optimizar los factores con el fin de maximizar la correlación entre los datos clínicos y la puntuación de actividad de la enfermedad. En un caso, se proporciona un método para monitorizar el estado de la enfermedad de un individuo, que comprende: a) determinar un perfil de clonotipos de una muestra del individuo, b) introducir la información del perfil de clonotipos de a) en un algoritmo que calcula una puntuación de actividad de la enfermedad, en el que el algoritmo se genera desarrollando un algoritmo que usa un conjunto de factores para combinar niveles de clonotipos que se correlacionan en una puntuación de actividad de la enfermedad, comparar la puntuación de actividad de la enfermedad con datos clínicos respecto al estado de la enfermedad, y optimizar los factores con el fin de maximizar la correlación entre los datos clínicos y la puntuación de actividad de la enfermedad, y c) usar el algoritmo que calcula una puntuación de actividad de la enfermedad para generar una puntuación predictiva del estado de la enfermedad del individuo. En otro caso, el método para monitorizar el estado de la enfermedad de un individuo comprende además determinar uno o más clonotipos que se correlacionan en el individuo, e introducir la información del uno o más clonotipos que se correlacionan en el algoritmo. En otro caso, dicha determinación de uno o más clonotipos que se correlacionan en el individuo comprende a) generar uno o más perfiles de clonotipos por secuenciación de ácido nucleico de moléculas individuales aisladas espacialmente de al menos una muestra del sujeto, en el que la al menos una muestra está relacionada con un primer estado de la enfermedad, y b) determinar uno o más clonotipos que se correlacionan en el sujeto sobre la base del uno o más perfiles de clonotipos. En otro caso, dicha determinación de uno o más clonotipos que se correlacionan en el individuo comprende a) introducir un perfil de clonotipos de una muestra del individuo en un algoritmo que puede predecir uno o más clonotipos que se correlacionan, en el que dicho algoritmo que puede predecir uno o más clonotipos que se correlacionan se desarrolla por i) generación de una pluralidad de perfiles de clonotipos de un conjunto de muestras, en el que las muestras son relevantes para la enfermedad, ii) identificación de uno o más clonotipos que se correlacionan del conjunto de muestras, iii) uso de parámetros de secuencia y/o datos funcionales de uno o más clonotipos que se correlacionan identificados en ii) para desarrollar el algoritmo que puede predecir clonotipos que se correlacionan en cualquier muestra de un sujeto con la enfermedad, y c) uso del algoritmo que puede predecir uno o más clonotipos que se correlacionan para determinar uno o más clonotipos que se correlacionan para el individuo. En otro caso, la enfermedad es lupus eritematoso sistémico o esclerosis múltiple. Se describe un método para determinar uno o más clonotipos de células T o B que se correlacionan, que comprende: a) dividir una muestra de células de un sujeto en al menos dos muestras, b) generar uno o más perfiles de clonotipos por secuenciación de ácido nucleico de moléculas aisladas espacialmente individualmente de una de las muestras de células del sujeto, c) enriquecer otra muestra de células de ese sujeto sobre la base de al menos un parámetro molecular de las células, d) generar uno o más perfiles de clonotipos por secuenciación de ácido nucleico de moléculas aisladas espacialmente individualmente de la muestra enriquecida del sujeto, y e) identificar al menos un clonotipo sobre la base de clonotipos cuya abundancia en la muestra se ha alterado entre la muestra enriquecida y la muestra no enriquecida. En otro caso, el parámetro molecular es un marcador de la superficie celular. En otro caso, el enriquecimiento se hace capturando células usando un marcador de afinidad inmovilizado en una fase sólida. En otro caso, la superficie sólida es un conjunto de lechos. En otro caso, la superficie sólida es una columna. En otro caso, el marcador se marca usando un resto fluorescente. En otro caso, el enriquecimiento se consigue por citometría de flujo usando el marcador fluorescente. En otro caso, las células son linfocitos B y el enriquecimiento se hace usando antígenos que se unen al receptor de células B. En otro caso, el enriquecimiento se hace mediante captura en una superficie sólida en la que se inmovilizan los antígenos. En otro caso, el antígeno se usa para marcar los linfocitos B y el enriquecimiento se consigue usando citometría de flujo usando este marcador. En otro caso, las células son linfocitos T y el enriquecimiento se hace usando un método que permite marcar las células T que reaccionan con un antígeno específico y enriquecer usando citometría de flujo. En otro caso, las células T se marcan usando el método de tinción intracelular de citoquinas. En otro caso, las células T se marcan usando el método de captura de citoquinas. En otro caso, el parámetro molecular es el receptor de células B que es capaz de unirse al menos a un antígeno que es específico para un patógeno. En otro caso, el parámetro molecular es el receptor de células T que es capaz de unirse al menos a un antígeno que es específico para un patógeno. En otro caso, la muestra se toma de un paciente que se ha expuesto a un patógeno en un primer punto de tiempo. Se describe un método para determinar un conjunto de clonotipos en un individuo que se correlacionan con una reacción inmune a un patógeno, que comprende: a) dividir una muestra de células de un sujeto en al menos dos muestras, b) generar uno o más perfiles de clonotipos por secuenciación de ácido nucleico de moléculas aisladas espacialmente individualmente de una muestra de células del sujeto, c) enriquecer otra muestra de células de ese sujeto sobre la base de la capacidad de las células de unirse al menos a un antígeno del patógeno, d) generar uno o más perfiles de clonotipos por secuenciación de ácido nucleico de moléculas aisladas espacialmente individualmente de la muestra enriquecida del sujeto y, e) identificar al menos un clonotipo que se correlaciona sobre la base de clonotipos cuya abundancia en la muestra se ha alterado entre la muestra enriquecida y la muestra no enriquecida. Se describe un método para determinar un conjunto de clonotipos en un individuo que se correlaciona con una reacción inmune a un tumor, que comprende: a) dividir una muestra de células de un sujeto en al menos dos muestras, b) generar uno o más perfiles de clonotipos por secuenciación de ácido nucleico de moléculas aisladas espacialmente individualmente de una muestra de células del sujeto, c) enriquecer otra muestra de células de ese sujeto sobre la base de la capacidad de las células de unirse al menos a un autoantígeno presente en el tumor, d) generar uno o más perfiles de clonotipos por

secuenciación de ácido nucleico de moléculas aisladas espacialmente individualmente de la muestra enriquecida del sujeto, e) identificar al menos un clonotipo que se correlaciona sobre la base de clonotipos cuya abundancia en la muestra se ha alterado entre la muestra enriquecida y la muestra no enriquecida. En otro caso, se usan los niveles de clonotipos que se correlacionan para evaluar el riesgo de que el individuo tenga un tumor. En otro caso, se sabe que los antígenos están presentes en un tumor que ya ha ocurrido en ese individuo y los clonotipos que se correlacionan se usan para evaluar el riesgo de recurrencia del tumor. En otro caso, se sabe que los antígenos están presentes en tumores en otros individuos y los clonotipos se usan para evaluar el riesgo de cáncer en un paciente que no ha tenido un tumor detectado previamente. Se describe un método para determinar un conjunto de clonotipos en un individuo que se correlaciona con una reacción inmune a un material liberado en la corriente sanguínea por daño a un órgano, que comprende: a) dividir una muestra de células de un sujeto en al menos dos muestras, b) generar uno o más perfiles de clonotipos por secuenciación de ácido nucleico de moléculas aisladas espacialmente individualmente de una muestra de células del sujeto, c) enriquecer otra muestra de células de ese sujeto sobre la base de la capacidad de las células de unirse al menos a un autoantígeno presente en los órganos dañados, d) generar uno o más perfiles de clonotipos por secuenciación de ácido nucleico de moléculas aisladas espacialmente individualmente de la muestra enriquecida del sujeto y, e) identificar al menos un clonotipo que se correlaciona sobre la base de clonotipos cuya abundancia en la muestra se ha alterado entre la muestra enriquecida y la muestra no enriquecida. En otro caso, se usan los niveles de clonotipos que se correlacionan para evaluar el riesgo de que el individuo tenga daño orgánico. Se describe un método para determinar un conjunto de clonotipos en un individuo que se correlaciona con una reacción inmune a un agente terapéutico, que comprende: a) dividir una muestra de células de un sujeto en al menos dos muestras, b) generar uno o más perfiles de clonotipos por secuenciación de ácido nucleico de moléculas aisladas espacialmente individualmente de una muestra de células del sujeto, c) enriquecer otra muestra de células de ese sujeto sobre la base de la capacidad de las células de unirse al menos a un antígeno contenido en el agente terapéutico, d) generar uno o más perfiles de clonotipos por secuenciación de ácido nucleico de moléculas aisladas espacialmente individualmente de la muestra enriquecida del sujeto y, e) identificar al menos un clonotipo que se correlaciona sobre la base de clonotipos cuya abundancia en la muestra se ha alterado entre la muestra enriquecida y la muestra no enriquecida. En otro caso, se usan los niveles de clonotipos que se correlacionan para evaluar el riesgo de que el individuo esté presentando hipersensibilidad a un agente terapéutico. Se describe un método para determinar un conjunto de clonotipos en un individuo que se correlaciona con una reacción inmune frente a una placa arterial, que comprende: a) dividir una muestra de células de un sujeto en al menos dos muestras, b) generar uno o más perfiles de clonotipos por secuenciación de ácido nucleico de moléculas aisladas espacialmente individualmente de una muestra de células del sujeto, c) enriquecer otra muestra de células de ese sujeto sobre la base de la capacidad de las células de unirse al menos a un antígeno presente en la placa arterial, d) generar uno o más perfiles de clonotipos por secuenciación de ácido nucleico de moléculas aisladas espacialmente individualmente de la muestra enriquecida del sujeto y, e) identificar al menos un clonotipo que se correlaciona sobre la base de clonotipos cuya abundancia en la muestra se ha alterado entre la muestra enriquecida y la muestra no enriquecida. En otro caso, se usan los niveles de clonotipos que se correlacionan para evaluar el riesgo de que el individuo tenga una enfermedad cardiovascular. En otro caso, se usan los niveles de clonotipos que se correlacionan para evaluar el riesgo de que la placa arterial sea inestable. Se describe un método para determinar un identificador de secuencia encontrado en células implicadas en un neoplasma linfóide, que comprende: a) obtener una muestra de células del individuo afectado en la que se sabe que existen las células cancerosas, b) generar uno o más perfiles de clonotipos relacionados con al menos una reorganización genómica de células inmunes por secuenciación de ácido nucleico de moléculas aisladas espacialmente individualmente de las células en la muestra, c) identificar el identificador de secuencia como la secuencia del clonotipo asociado con el tumor. En otro caso, la muestra es de la médula ósea del paciente. En otro caso, la muestra es de la sangre del paciente. En otro caso, la muestra es de una biopsia de un tumor linfóide sólido. En otro caso, la reorganización genómica de células inmunes es una reorganización VDJ de IgH en una célula B. En otro caso, la reorganización genómica de células inmunes es una reorganización DJ de IgH en una célula B. En otro caso, la reorganización genómica de células inmunes es una reorganización VJ de IgK en una célula B. En otro caso, la reorganización genómica de células inmunes es una reorganización VJ de IgL en una célula B. En otro caso, la reorganización genómica de células inmunes es una reorganización VDJ de TCR  $\beta$  en una célula T. En otro caso, la reorganización genómica de células inmunes es una reorganización DJ de TCR  $\beta$  en una célula T. En otro caso, la reorganización genómica de células inmunes es una reorganización VJ de TCR  $\alpha$  en una célula T. En otro caso, la reorganización genómica de células inmunes es una reorganización VJ de TCR  $\lambda$  en una célula T. En otro caso, la reorganización genómica de células inmunes es una reorganización VDJ de TCR  $\delta$  en una célula T. En otro caso, la reorganización genómica de células inmunes es una reorganización VD de TCR  $\delta$  en una célula T. En otro caso, la reorganización genómica de células inmunes es una translocación de un segmento J de IgH a otra región del genoma. En otro caso, la reorganización genómica de células inmunes es una translocación de cualquier segmento J a otra región del genoma. En otro caso, la identificación del clonotipo asociado a tumor se hace por frecuencia de clonotipo. En otro caso, la identificación del clonotipo asociado a tumor se hace por frecuencia de clonotipo. En otro caso, la identificación del clonotipo asociado a tumor se hace por la detección de reorganización cruzada de linaje. En otro caso, la

identificación del clonotipo asociado a tumor se hace por la identificación de reorganizaciones no funcionales. En otro caso, la identificación del clonotipo asociado a tumor se hace por la asociación de clonotipos celulares con al menos un marcador molecular asociado con el tumor. Se describe un método para determinar los niveles de células tumorales linfoides circulantes en un individuo cuyo tumor se ha asociado con un único identificador de secuencia en un primer punto de tiempo, que comprende: a) obtener una muestra de células del paciente, b) generar uno o más perfiles de clonotipos relacionados con al menos una reorganización genómica de células inmunes por secuenciación de ácido nucleico de moléculas aisladas espacialmente individualmente de las células en la muestra, c) determinar los niveles de células tumorales a partir del nivel de los clonotipos asociados con el identificador de secuencia. Se describe un método para determinar los niveles de células tumorales linfoides circulantes en un individuo cuyo tumor se ha asociado con un único identificador de secuencia en un primer punto de tiempo, que comprende: a) obtener una muestra de células del paciente, b) enriquecer las células sobre la base de al menos un marcador molecular, c) generar uno o más perfiles de clonotipos relacionados con al menos una reorganización genómica de células inmunes por secuenciación de ácido nucleico de moléculas aisladas espacialmente individualmente de las células en la muestra, d) determinar los niveles de células tumorales a partir del nivel de clonotipos asociados con el identificador de secuencia. En otro caso, la muestra es una muestra de sangre. En otro caso, la muestra es una muestra de médula ósea. En otro caso, la muestra es una muestra de linfa. En otro caso, la muestra es una muestra de tejido. En otro caso, las células se marcan fluorescentemente y se enriquecen usando citometría de flujo. En otro caso, las células se enriquecen mediante la unión a un soporte sólido. En otro caso, el clonotipo se define como el clonotipo que contiene el único identificador de secuencia. En otro caso, se determina que los clonotipos son aquellos clonotipos que probablemente hayan resultado de mutaciones y reorganizaciones en el identificador de secuencia. En otro caso, se usan los niveles de células tumorales circulantes y/o el cambio en los niveles de células tumorales circulantes en un algoritmo para producir una puntuación que se correlaciona con el riesgo de tener una recurrencia de tumor clínica. En otro caso, se usan los niveles de células tumorales circulantes y/o el cambio en los niveles de las células tumorales circulantes para tomar una decisión sobre el tratamiento.

## II. Métodos para determinar perfiles de clonotipos

Los métodos descritos en la presente memoria pueden usarse para generar perfiles de secuencias de ADN recombinadas, o clonotipos, en una muestra de un sujeto. En un caso, se describe un método para determinar un perfil de secuencias de ADN recombinadas en células T y/o células B que incluye obtener una muestra de un sujeto que comprende células T y/o células B, aislar moléculas individuales de ADN genómico de dichas células, secuenciar las moléculas individuales aisladas de ADN genómico, y determinar los niveles de diferentes secuencias de la muestra para generar dicho perfil de secuencias de ADN recombinadas.

En otro caso, se describe un método para determinar un perfil de secuencias de ADN recombinadas en células T y/o células B que incluye obtener una muestra de un sujeto que comprende células T y/o células B, aislar moléculas individuales de ADN genómico de las células, amplificar las moléculas individuales de ADN genómico, secuenciar el ADN amplificado, y determinar los niveles de diferentes secuencias de la muestra para generar dicho perfil de secuencias de ADN recombinadas.

En otro caso, se describe un método para determinar un perfil de secuencias de ADN recombinadas en células T y/o células B que incluye obtener una muestra de un sujeto que comprende células T y/o células B, amplificar el ADN genómico de las células, aislar moléculas individuales del ADN amplificado, secuenciar las moléculas individuales aisladas de ADN amplificado, y determinar los niveles de diferentes secuencias de la muestra para generar el perfil de secuencias de ADN recombinadas.

En otro caso, se describe un método para determinar un perfil de secuencias de ADN recombinadas en células T y/o células B que incluye obtener una muestra de un sujeto que incluye células T y/o células B, amplificar el ADN genómico de las células, aislar moléculas individuales del ADN amplificado, re-amplificar las moléculas de ADN amplificado, secuenciar las moléculas de ADN re-amplificado, y determinar los niveles de diferentes secuencias de la muestra para generar el perfil de secuencias de ADN recombinadas.

En otro caso, se describe un método para determinar un perfil de secuencias de ADN recombinadas en células T y/o células B que incluye obtener una muestra de un sujeto que comprende células T y/o células B, aislar el ARN de dicha muestra, transcribir de forma inversa el ARN de dichas células para formar ADNc, aislar moléculas individuales de dicho ADNc, opcionalmente re-amplificar dicho ADNc, secuenciar dichas moléculas individuales aisladas de dicho ADNc o ADN re-amplificado, y determinar los niveles de diferentes secuencias de dicha muestra para generar dicho perfil de secuencias de ADN recombinadas.

En otro caso, se describe un método para determinar un perfil de secuencias de ADN recombinadas en células T y/o células B que incluye obtener una muestra de un sujeto que incluye células T y/o células B, aislar moléculas individuales de ARN de dicha muestra, secuenciar las moléculas individuales de ARN, y determinar los niveles de diferentes secuencias de dicha muestra para generar el perfil de secuencias de ADN recombinadas.

## Sujetos y muestras

Los métodos de la invención proporcionada pueden usar muestras de sujetos o individuos (por ejemplo, pacientes). El sujeto puede ser un paciente, por ejemplo, un paciente con una enfermedad autoinmune. El sujeto puede ser un paciente con una enfermedad infecciosa o cáncer, tal como una leucemia o un linfoma. El sujeto puede ser un mamífero, por ejemplo, un ser humano. El sujeto puede ser macho o

hembra. El sujeto puede ser un bebé, un niño, o un adulto. En algunas realizaciones, el sujeto ya no vive. En algunas realizaciones, el sujeto vive. El sujeto puede ser un individuo que se ha expuesto a un arma biológica.

El sujeto también podría ser un animal no humano. El animal no humano podría ser una mascota doméstica o un animal de granja. El animal no humano podría ser un perro, gato, vaca, caballo, cabra, o cerdo. El animal no humano podría ser un animal clonado. El animal no humano podría estar implicado en la producción de productos farmacéuticos.

Las muestras usadas en los métodos de la invención proporcionada pueden incluir, por ejemplo, un fluido corporal de un sujeto, incluyendo fluido amniótico que rodea a un feto, humor acuoso, bilis, sangre y plasma sanguíneo, cerumen (cera de los oídos), fluido de Cowper o fluido pre-eyaculatorio, quilo, quimo, eyaculado femenino, fluido intersticial, linfa, menstruación, leche de mama, mucus (incluyendo mocos y flema), fluido pleural, pus, saliva, sebo (aceite de la piel), semen, suero, sudor, lágrimas, orina, lubricación vaginal, vómito, agua, heces, fluidos corporales internos, incluyendo fluido cerebroespinal que rodea el cerebro y la médula espinal, fluido sinovial que rodea las articulaciones de los huesos, fluido intracelular que es el fluido que está en el interior de las células, y humor vítreo que es el fluido en el globo ocular. 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, ó 5,0 mL. La muestra puede ser fluido cerebro espinal (CSF) cuando el sujeto tiene esclerosis múltiple, fluido sinovial cuando el sujeto tiene artritis reumatoide, y biopsia de la piel (u otro órgano) cuando el sujeto tiene lupus sistémico. En una realización, el clonotipo puede identificarse del fluido/tejido corporal disponible lo más probablemente para reflejar la patología seguido de monitorización posterior de los niveles de los clonotipos de un fluido corporal diferente, por ejemplo, sangre. Las muestras también pueden incluir disolventes en los que se ha disuelto el material biológico. Las muestras pueden analizarse en un momento en el que la enfermedad es inactiva. Las muestras pueden analizarse en un momento en el que la enfermedad es activa. Las muestras pueden obtenerse en un momento en el que la enfermedad es inactiva. Las muestras pueden obtenerse en un momento en el que la enfermedad es activa. La muestra la puede obtener un responsable sanitario, por ejemplo, un médico, asistente de médico, enfermera, veterinario, dermatólogo, reumatólogo, dentista, paramédico, o cirujano. La muestra la puede obtener un técnico de investigación. La muestra la puede proporcionar el sujeto. La muestra puede ser proporcionada de manera anónima. La muestra puede ser proporcionada por correo. La muestra puede ser proporcionada por una agencia de orden público o por un investigador. Puede obtenerse más de una muestra de un sujeto.

La muestra puede ser una biopsia, por ejemplo, una biopsia de la piel. La biopsia puede ser, por ejemplo, del cerebro, hígado, pulmón, corazón, colon, riñón, o médula ósea. Puede usarse cualquier técnica de biopsia usada por los expertos en la técnica 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 hace un corte más pequeño que en una biopsia abierta. La biopsia puede ser una biopsia con aguja gruesa o incisional, en la que se retira parte del tejido. La biopsia puede ser una biopsia excisional, en la que se intenta retirar una lesión entera. La biopsia puede ser una biopsia de aspiración con aguja fina, en la que se retira una muestra de tejido o fluido con una aguja.

La muestra puede obtenerse de material corporal que es dejado a un lado por un sujeto. Dicho material desechado puede incluir desperdicios humanos. El material desechado también puede incluir células de la piel de muda, sangre, dientes o pelo.

La muestra puede incluir células inmunes, por ejemplo, las células inmunes pueden incluir células T y/o células B. Las células T (linfocitos T) incluyen, por ejemplo, células que expresan receptores de células T. Las células T incluyen células T auxiliares (células T efectoras o células Th), células T citotóxicas (CTL), células T de memoria, y células T reguladoras. La muestra puede incluir una única célula en algunas aplicaciones (por ejemplo, un ensayo de calibración para definir células T relevantes) o más generalmente al menos 1.000, al menos 10.000, al menos 100.000, al menos 250.000, al menos 500.000, al menos 750.000, o al menos 1.000.000 células T.

Las células B incluyen, por ejemplo, células B plasmáticas, células B de memoria, células B1, células B2, células B de la zona marginal, y células B foliculares. Las células B pueden expresar inmunoglobulinas (anticuerpos, receptor de células B). La muestra puede incluir una única célula en algunas aplicaciones (por ejemplo, un ensayo de calibración para definir células B relevantes) o más generalmente al menos 1.000, al menos 10.000, al menos 100.000, al menos 250.000, al menos 500.000, al menos 750.000, o al menos 1.000.000 células B.

La muestra puede incluir ácido nucleico, por ejemplo, ADN (por ejemplo, ADN genómico o ADN mitocondrial) o ARN (por ejemplo, ARN mensajero o microARN). El ácido nucleico puede ser ADN o ARN sin células, por ejemplo, extraído del sistema circulatorio, Vlassov et al, Curr. Mol. Med., 10: 142-165 (2010); Swarup et al, FEBS Lett., 581: 795-799 (2007). En los métodos de la invención proporcionada, la cantidad de ARN o ADN de un sujeto que puede analizarse incluye, por ejemplo, tan poco como una única célula en algunas aplicaciones (por ejemplo, un ensayo de calibración) y tantas como 10 millones de células o más que se traduce en un intervalo de ADN de 6pg-60µg, y ARN de aproximadamente 1pg-10µg.

Como se discute más completamente más adelante (Definiciones), una muestra de linfocitos es lo suficientemente grande como para que sustancialmente cada célula T o célula B con un clonotipo distinto esté representada en ella, formando de esta manera un repertorio (como se usa el término en la presente

memoria). En un caso, se toma una muestra que contiene con una probabilidad de noventa y nueve por ciento cada clonotipo de una población presente a una frecuencia de 0,001 por ciento o mayor. En otro caso, se toma una muestra que contiene con una probabilidad de noventa y nueve por ciento cada clonotipo de una población presente a una frecuencia de 0,0001 por ciento o mayor. En una realización, una muestra de células B o células T incluye al menos medio millón de células, y en otra realización dicha muestra incluye al menos un millón de células.

Siempre que una fuente de material de la que se toma una muestra es escasa, tal como, muestras de estudios clínicos, o semejantes, el ADN del material puede amplificarse por una técnica no sesgada, tal como amplificación de genoma completo (WGA), amplificación por desplazamiento múltiple (MDA); o técnica semejante, por ejemplo, Hawkins et al, *Curr. Opin. Biotech.*, 13: 65-67 (2002); Dean et al, *Genome Research*, 11: 1095-1099 (2001); Wang et al, *Nucleic Acids Research*, 32: e76 (2004); Hosono et al, *Genome Research*, 13: 954-964 (2003); y semejantes.

Las muestras de sangre tienen un interés particular, especialmente en la monitorización de neoplasmas linfoides, tales como linfomas, leucemias, o semejantes, y pueden obtenerse usando técnicas convencionales, por ejemplo, Innis et al, editores, *PCR Protocols* (Academic Press, 1990); o semejantes. Por ejemplo, las células sanguíneas blancas pueden separarse de las muestras de sangre usando técnicas convencionales, por ejemplo, kit RosetteSep (Stem Cell Technologies, Vancouver, Canadá). Las muestras de sangre pueden variar en volumen de 100  $\mu$ L a 10 mL; en un aspecto, los volúmenes de la muestra de sangre están en el intervalo de 200, 100  $\mu$ L a 2 mL. El ADN y/o ARN puede extraerse de dicha muestra de sangre usando técnicas convencionales para uso en métodos de la invención, por ejemplo, el kit de sangre y tejido DNeasy (Qiagen, Valencia, CA). Opcionalmente, pueden aislarse adicionalmente subconjuntos de células sanguíneas blancas, por ejemplo, linfocitos, usando técnicas convencionales, por ejemplo, separación de células activada por fluorescencia (FACS) (Becton Dickinson, San Jose, CA), separación de células activada magnéticamente (MACS) (Miltenyi Biotec, Auburn, CA), o semejantes.

En otras realizaciones, los ácidos nucleicos se analizan de una muestra de un subconjunto de células. Puede emplearse un método para separar células, por ejemplo, usando un marcador de la superficie celular. Por ejemplo, las células pueden aislarse por citometría de flujo para separar células, separación por flujo, separación de células activada por fluorescencia (FACS), separación basada en lechos tal como separación de células magnética (MACS; por ejemplo, usando partículas magnéticas recubiertas de anticuerpo), separación basada en el tamaño (por ejemplo, un tamiz, una matriz de obstáculos, o un filtro), separación en un dispositivo de microfluído, separación basada en anticuerpo, sedimentación, adsorción por afinidad, extracción por afinidad, o centrifugación en gradiente de densidad. Las células pueden purificarse por microdissección con captura de láser. La separación puede basarse en el tamaño, morfología celular o en marcadores intracelulares o extracelulares. Los métodos para aislar o separar células tumorales se describen, por ejemplo, en Nagrath S. et al. (2007) *Nature* 450: 1235-1239; Patentes US nos. 6008002, 7232653, y 7332288; Publicación PCT No. WO2008157220A1; y Solicitudes de Patente US Nos. US20080138805A1 y US20090186065; y Rosenberg R. et al. (2002) *Cytometry* 49 150-158.

El subconjunto de células puede ser un subconjunto de células T y/o células B. El subconjunto de células T puede ser células CD4+, CD8+, o altas en CD27. Las mezclas de anticuerpos para marcar y/o separar una gran variedad de subconjuntos de células T y células B están disponibles comercialmente de proveedores tales como Quest Diagnostic (San Juan Capistrano, CA); Dako (Dinamarca); y semejantes. Lo siguiente son ejemplos de kits disponibles para subconjuntos relacionados con enfermedad (en los que se lista la especificidad de antígeno de los anticuerpos: leucemia/linfoma linfoblástico de precursor de B (CD19, CD79a (citoplásmico), CD20, CD10, Tdt, HLADR, CD34, IgM (citoplásmico)); linfoma de células B grandes difuso (CD20, CD19, CD22, CD79a, CD30); linfoma folicular (CD20, CD10, CD10, BCL2, BCL6); leucemia de células del manto (CD19, CD20, CD5, CD23-, BCL1); y semejantes.

La separación de células activada por fluorescencia (FACS) usa características de dispersión de la luz y fluorescentes para separar células. Puede conferirse una propiedad fluorescente en una célula usando, por ejemplo, sondas de ácidos nucleicos o anticuerpos conjugados con un colorante fluorescente. Una suspensión celular puede formar una corriente de líquido que fluye. La corriente de células forma gotas que contienen aproximadamente una célula por gota. Antes de que la corriente forme gotas, se mide una característica fluorescente de cada célula. Se pone una carga en un anillo de carga eléctrica antes de la medición de la intensidad de fluorescencia y la carga opuesta es portada en la gota al desprenderse de la corriente. Las gotas cargadas pasan a través de dos placas de deflexión de alto voltaje que desvían las gotas en diferentes contenedores sobre la base de su carga. La carga puede aplicarse directamente a la corriente y la gota que se desprende retiene la carga del mismo signo que la corriente. La corriente se vuelve a hacer neutra después de que la gota se desprenda.

En FACS puede usarse inmunofluorescencia directa o indirecta. En la inmunofluorescencia directa, un anticuerpo se conjuga directamente con un colorante fluorescente, En la inmunofluorescencia indirecta, el anticuerpo primario no está marcado, y un anticuerpo secundario se conjuga con un colorante fluorescente.

Como las recombinaciones de identificación están presentes en el ADN de las células de inmunidad adaptativa de cada individuo así como en sus transcritos de ARN asociados, bien el ARN o ADN puede secuenciarse en los métodos de la invención proporcionada. Una secuencia recombinada de una célula T o célula B que codifica un receptor de células T o molécula de inmunoglobulina, o una parte de ésta, se refiere como un clonotipo. El ADN o ARN puede corresponder a secuencias de genes del receptor de células T (TCR) o genes de inmunoglobulina (Ig) que codifican anticuerpos. Por ejemplo, el ADN y ARN

puede corresponder a secuencias que codifican las cadenas  $\alpha$ ,  $\beta$ ,  $\gamma$ , o  $\delta$  de un TCR. En la mayoría de las células T, el TCR es un heterodímero que consiste en una cadena  $\alpha$  y una cadena  $\beta$ . La cadena TCR $\alpha$  se genera por recombinación VJ, y el receptor de cadena  $\beta$  se genera por recombinación V(D)J. Para la cadena TCR $\beta$ , en los seres humanos, hay 48 segmentos V, 2 segmentos D, y 13 segmentos J. Varias bases pueden deletionarse y otras añadirse (denominados nucleótidos N y P) en cada una de las dos uniones. En una minoría de células T, los TCR consisten en cadenas  $\gamma$  y delta  $\delta$ . La cadena TCR  $\gamma$  se genera por recombinación VJ, y la cadena TCR  $\delta$  se genera por recombinación V(D)J (Kenneth Murphy, Paul Travers, y Mark Walport, *Janeway's Immunology* 7ª edición, Garland Science, 2007).

El ADN y ARN analizados en los métodos de la invención pueden corresponder a secuencias que codifican inmunoglobulinas de cadena pesada (IgH) con regiones constantes ( $\alpha$ ,  $\delta$ ,  $\epsilon$ ,  $\gamma$ , o  $\mu$ ) o inmunoglobulinas de cadena ligera (IgK o IgL) con regiones constantes  $\lambda$  o  $\kappa$ . Cada anticuerpo tiene dos cadenas ligeras idénticas y dos cadenas pesadas idénticas. Cada cadena está compuesta por una región constante (C) y una región variable. Para la cadena pesada, la región variable está compuesta por segmentos variables (V), de diversidad (D), y de unión (J). Las varias secuencias distintas que codifican cada tipo de estos segmentos están presentes en el genoma. Un evento específico de recombinación VDJ ocurre durante el desarrollo de una célula B, marcando a esa célula para generar una cadena pesada específica. La diversidad en la cadena ligera se genera de una manera similar excepto en que no hay región D de manera que sólo hay recombinación VJ. La mutación somática ocurre frecuentemente cerca del sitio de la recombinación, causando la adición o delección de varios nucleótidos, incrementando adicionalmente la diversidad de las cadenas pesada y ligera generadas por las células B. La posible diversidad de los anticuerpos generados por una célula B es entonces el producto de las diferentes cadenas pesadas y ligeras. Las regiones variables de las cadenas pesadas y ligeras contribuyen para formar la región o sitio de reconocimiento (o unión) al antígeno. Además de esta diversidad está un proceso de hipermutación somática que puede ocurrir después de que se organice una respuesta específica frente a algún epítipo.

Como se ha mencionado anteriormente, según la invención, pueden seleccionarse cebadores para generar amplicones de subconjuntos de ácidos nucleicos recombinados extraídos de linfocitos. Dichos subconjuntos pueden referirse en la presente memoria como "regiones reorganizadas somáticamente". Las regiones reorganizadas somáticamente pueden comprender ácidos nucleicos de linfocitos en desarrollo o completamente desarrollados, en los que los linfocitos en desarrollo son células en las que la reorganización de genes inmunes no se ha completado para formar moléculas que tienen regiones V(D)J completas. Las regiones reorganizadas somáticamente incompletas ejemplares incluyen moléculas de IgH incompletas (tales como, moléculas que contienen sólo regiones D-J), moléculas de TCR $\delta$  incompletas (tales como, moléculas que contienen sólo regiones D-J), e IgK inactiva (por ejemplo, que comprende regiones Kde-V). Las reorganizaciones incompletas también se encuentran en células inmunes completamente desarrolladas asociadas con el segundo cromosoma en una célula en la que el primer cromosoma formó una reorganización productiva.

#### **Control de las cantidades de muestra y estimación del número de células**

El muestreo adecuado de las células es un aspecto importante para interpretar los datos del repertorio, como se describe adicionalmente más adelante en las definiciones de "clonotipo" y "repertorio". Por ejemplo, empezando a partir de 1.000 células crea una frecuencia mínima de que el ensayo es sensible independientemente de cuántas lecturas de secuencia se obtienen. Por lo tanto, en la presente memoria se describen métodos para cuantificar el número de moléculas de receptor inmune introducido. Esto se ha implementado para secuencias de TCR $\beta$  e IgH. En cada caso se usa un conjunto de cebadores que es capaz de amplificar todas las secuencias diferentes. Con el fin de obtener un número absoluto de copias, se realiza una PCR en tiempo real con los múltiples cebadores junto con un estándar con un número conocido de copias de receptor inmune. Un ejemplo de datos de PCR en tiempo real asociados con el ejemplo de vacunación de ratón se muestra en la Fig. 9. Esta medida de PCR en tiempo real puede hacerse a partir de la reacción de amplificación que posteriormente se secuenciará o puede hacerse en una alícuota separada de la misma muestra. En el caso de ADN, el número absoluto de moléculas de receptor inmune reorganizadas puede convertirse fácilmente en número de células (en 2 veces ya que algunas células tendrán 2 copias reorganizadas del receptor inmune específico evaluado y otras tendrán una). En el caso de ADNc, el número total medido de moléculas reorganizadas en la muestra en tiempo real puede extrapolarse para definir el número total de estas moléculas usado en otra reacción de amplificación de la misma muestra. Además, este método puede combinarse con un método para determinar la cantidad total de ARN para definir el número de moléculas de receptor inmune reorganizadas en una cantidad unitaria (por ejemplo, 1  $\mu$ g) de ARN asumiendo una eficiencia específica de síntesis de ADNc. Si la cantidad total de ADNc se mide entonces la eficiencia de la síntesis de ADNc no necesita considerarse. Si también se conoce el número de células, puede computarse las copias de receptor inmune reorganizadas por célula. Si no se conoce el número de células, se puede estimar a partir del ARN total ya que las células de un tipo específico generan habitualmente una cantidad de ARN comparable. Por lo tanto, a partir de las copias de moléculas de receptor inmune reorganizadas por 1  $\mu$ g se puede estimar el número de estas moléculas por célula.

Una desventaja de hacer una PCR en tiempo real separada de la reacción que se procesaría para secuenciación es que podría haber efectos inhibidores que son diferentes en la PCR en tiempo real de la otra reacción ya que pueden utilizarse enzimas diferentes, ADN de entrada, y otras condiciones. El procesamiento de los productos de la PCR en tiempo real para secuenciación mejoraría este problema.

Sin embargo, un bajo número de copias usando PCR en tiempo real puede deberse bien a bajo número de copias o a efectos inhibidores, u otras condiciones subóptimas en la reacción.

Otra estrategia que puede utilizarse es añadir una cantidad conocida de moléculas reorganizadas de receptor inmune únicas con una secuencia conocida, es decir, cantidades conocidas de uno o más estándares internos, al ADNc o ADN genómico de una muestra con una cantidad desconocida. Mediante el recuento del número relativo de moléculas que se obtiene de la secuencia conocida añadida comparado con el resto de las secuencias de la misma muestra, se puede estimar el número de moléculas de receptor inmune reorganizadas en la muestra de ADNc inicial. (Dichas técnicas para recuento molecular son muy conocidas, por ejemplo, Brenner et al, patente US 7.537.897. Los datos de la secuenciación de la secuencia única añadida pueden usarse para distinguir las diferentes posibilidades si también se está usando una calibración de PCR en tiempo real. Un número bajo de copias de receptor inmune reorganizado en el ADN (o ADNc) crearía una alta proporción entre el número de moléculas para la secuencia adicionada comparado con el resto de las secuencias de la muestra. Por otra parte, si el bajo número de copias medido por PCR en tiempo real se debe a ineficiencia en la reacción, la proporción no sería alta.

En la presente memoria se describen métodos para medir la expresión de clonotipos a nivel celular. Esto es, como se ha indicado anteriormente, los clonotipos pueden usarse para contar linfocitos; por lo tanto, midiendo los clonotipos derivados de ADN genómico y los mismos clonotipos derivados de ARN, puede determinarse la expresión de clonotipos basada en células. Un método para medir simultáneamente el número de linfocitos y los niveles de expresión de clonotipos en una muestra puede comprender las etapas de: (a) obtener de un individuo una muestra que comprende células T y/o células B; (b) secuenciar moléculas individuales aisladas espacialmente derivadas del ADN genómico de dichas células, comprendiendo dichas moléculas individuales aisladas espacialmente un número de clonotipos correspondiente a un número de linfocitos en la muestra; (c) secuenciar las moléculas individuales aisladas espacialmente derivadas de ARN de dichas células, comprendiendo dichas moléculas individuales aisladas espacialmente números de clonotipos correspondientes a los niveles de expresión de éstos en los linfocitos de la muestra; y (d) determinar los niveles de expresión de los clonotipos en los linfocitos de la muestra comparando para cada clonotipo el número determinado a partir de moléculas individuales aisladas derivadas del ADN genómico de dichas células y el número determinado a partir de moléculas individuales aisladas derivadas del ARN de dichas células. El ADN genómico y ARN se extraen fácilmente de la muestra usando kits disponibles comercialmente, tales como el mini kit AIIPrep DNA/RNA (Qiagen GmbH, Alemania). Como se ha mencionado anteriormente, en un caso, la etapa de determinación incluye además determinar dicho número de linfocitos en dicha muestra añadiendo una cantidad conocida de un estándar interno a dicho ADN genómico. En otro caso, cuando por ejemplo la muestra es sangre periférica, la muestra tiene un volumen definido que permite determinar una concentración de dichos linfocitos en dicha muestra. Típicamente, dicho volumen definido está en el intervalo de 1 mL a 50 mL, y más habitualmente, en el intervalo de 1 mL a 10 mL. En otro caso, los números del mismo clonotipo derivado de ADN genómico y ARN se comparan simplemente dividiendo el número de clonotipos determinado a partir de las moléculas individuales aisladas derivadas del ARN por el número de clonotipos determinado a partir de las moléculas individuales aisladas derivadas de dicho ADN genómico. Estos dos conjuntos de clonotipos se distinguen fácilmente en la misma ronda de secuenciación por el uso de marcadores, particularmente etiquetas de oligonucleótidos que se unen durante el proceso de preparación de las muestras. Para la secuenciación basada en Solexa, dichos marcadores pueden incorporarse con las etiquetas usadas para identificar diferentes muestras (por ejemplo) añadiendo un único nucleótido a la etiqueta para indicar ADN o ARN, o simplemente usando una etiqueta adicional de manera que la muestra de cada paciente se marca con dos etiquetas, una para la fracción de ADN genómico y una para la fracción de ARN. Así, dicha etapa de secuenciación de dichas moléculas individuales aisladas espacialmente derivadas de dicho ARN puede incluir el marcaje de cada una de dichas moléculas individuales aisladas espacialmente con un primer marcador que indica su origen de ARN y dicha etapa de secuenciación de dichas moléculas individuales aisladas espacialmente derivadas de dicho ADN genómico puede incluir el marcaje de cada una de dichas moléculas individuales aisladas espacialmente con un segundo marcador que indica su origen de ADN genómico de manera que el primer marcador es distinguible del segundo marcador. En un caso, dichos marcadores son etiquetas de oligonucleótido distintas que se identifican por secuenciación.

Asimismo, los métodos descritos en la presente memoria pueden usarse para proporcionar simultáneamente (esto es, sobre la base de medidas en una única muestra), el número de linfocitos y la clonalidad. Dicho método puede implementarse con las etapas siguientes: (a) obtener de un individuo una muestra que comprende células T y/o células B; (b) secuenciar moléculas individuales aisladas espacialmente derivadas del ácido nucleico de dichas células, comprendiendo dichas moléculas individuales aisladas espacialmente un número de clonotipos correspondiente a un número de linfocitos en la muestra; (c) determinar el número de linfocitos a partir del número de moléculas individuales aisladas espacialmente; (d) determinar las abundancias de las diferentes secuencias de las moléculas individuales aisladas espacialmente para generar un perfil de clonotipos y una medida de la clonalidad basada en éste. El ácido nucleico de los linfocitos puede ser ADN genómico y/o ARN; sin embargo, preferiblemente el ácido nucleico es ADN genómico. De manera similar a anteriormente, en un caso, la etapa de determinar dicho número incluye además determinar dicho número de linfocitos en dicha muestra añadiendo una cantidad conocida de un estándar interno a dicho ADN genómico. Y, de manera

similar, cuando la muestra es una muestra de sangre periférica tiene un volumen definido de manera que se determina una concentración de dichos linfocitos en dicha muestra. En algunos casos de lo anterior, sólo se emplean células B y, en otras realizaciones, sólo se emplean células T.

#### **Amplificación de poblaciones de ácido nucleico**

5 Como se ha indicado anteriormente, los amplicones de poblaciones de ácidos nucleicos diana pueden generarse por una variedad de técnicas de amplificación. En un aspecto de la invención, se usa PCR múltiple para amplificar miembros de una mezcla de ácidos nucleicos, particularmente mezclas que comprenden moléculas inmunes recombinadas tales como receptores de células T, receptores de células B, o partes de éstos. La guía para llevar a cabo PCR múltiples de dichas moléculas inmunes se encuentra  
10 en las referencias siguientes: Morley, patente U.S. 5.296.351; Gorski, patente U.S. 5.837.447; Dau, patente U.S. 6.087.096; Von Dongen et al, publicación de patente U.S. 2006/0234234; publicación de patente europea EP 1544308B1; y semejantes. Las referencias anteriores describen la técnica referida como "espectrotipificación", en la que una población de moléculas inmunes se amplifica por PCR múltiple después de lo cual las secuencias del amplicón resultante se separan físicamente, por ejemplo, por  
15 electroforesis, con el fin de determinar si hay una clase con un tamaño predominante. Dicha clase indicaría una población clonal predominante de linfocitos que, a su vez, sería indicativa de estado de enfermedad. En la espectrotipificación, es importante seleccionar cebadores que presentan una pequeña o ninguna reactividad cruzada (es decir, que no hibridan con sitios de unión de otros cebadores); de otra forma, puede haber una falsa representación de clases de tamaño en el amplicón. En la presente invención, siempre que los ácidos nucleicos de una población se amplifiquen uniformemente, la reactividad cruzada de los cebadores es permisible porque en la presente invención se analizan las secuencias de los ácidos nucleicos amplificados, no meramente sus tamaños. Como se describe más completamente más adelante, en un aspecto, la etapa de aislar espacialmente las moléculas de ácido nucleico individuales se consigue llevando a cabo una amplificación múltiple primaria de una región reorganizada somáticamente preseleccionada o parte de ésta (es decir, secuencias diana) usando  
20 cebadores directos e inversos que tienen cada uno colas no complementarias con las secuencias diana para producir un primer amplicón cuyas secuencias miembro tienen secuencias comunes en cada extremo para permitir manipulación adicional. Por ejemplo, dichos extremos comunes pueden incluir sitios de unión de cebador para amplificación continua usando sólo un único cebador directo y un único cebador inverso en lugar de varios de cada uno, o para amplificación en puente de moléculas individuales en una superficie sólida, o semejantes. Dichos extremos comunes pueden añadirse en una única amplificación como se ha descrito anteriormente, o pueden añadirse en un procedimiento de dos etapas para evitar las dificultades asociadas con la fabricación y ejercicio de control de calidad sobre las mezclas de cebadores largos (por ejemplo, 50-70 bases o más). En dicho proceso de dos etapas (descrito más completamente más adelante e ilustrado en las Figs. 4A-4B), la amplificación primaria se lleva a cabo como se ha descrito anteriormente, excepto que las colas del cebador están limitadas en longitud para proporcionar sólo sitios de unión de cebador directo e inverso en los extremos de las secuencias del primer amplicón. Entonces se lleva a cabo una amplificación secundaria usando cebadores de amplificación secundaria específicos para estos sitios de unión de cebador para añadir secuencias adicionales a los extremos de un segundo amplicón. Los cebadores de la amplificación secundaria tienen colas no complementarias con las secuencias diana, que forman los extremos del segundo amplicón y que pueden usarse en conexión con la secuenciación de los clonotipos del segundo amplicón. En una realización, dichas secuencias añadidas pueden incluir sitios de unión de cebador para generar lecturas de secuencia y sitios de unión de cebador para llevar a cabo la PCR en puente en una superficie sólida para generar poblaciones clonales de moléculas individuales aisladas espacialmente, por ejemplo, cuando se usa secuenciación basada en  
45 Solexa. En esta última estrategia, una muestra de secuencias del segundo amplicón se dispone en una superficie sólida que tiene unidos oligonucleótidos complementarios capaces de hibridar con secuencias de la muestra, después de lo cual se implementan ciclos de extensión de cebador, desnaturalización, hibridación hasta que se forman las poblaciones clonales de moldes. Preferiblemente, el tamaño de la muestra se selecciona de manera que (i) incluye una representación efectiva de clonotipos en la muestra original, y (ii) la densidad de las poblaciones clonales en la superficie sólida está en un intervalo que permite la determinación inequívoca de la secuencia de los clonotipos.

Además de asegurar que la muestra contiene células suficientes para ser representativa de la muestra original, es importante que los amplicones generados por la reacción de PCR múltiple sea representativa de las células en la reacción. Con el fin de conseguir esto, las condiciones de cebador deben seleccionarse de manera que se produzca la amplificación de cada célula en la reacción.

Las secuencias de TCR o BCR o partes de éstas pueden amplificarse a partir de ácido nucleico en una reacción múltiple usando al menos un cebador que hibrida con la región C y uno o más cebadores que puedan hibridar con uno o más segmentos V (como se ilustra en las Figs. 2A-2B y Figs. 4A-4B y se discute más completamente más adelante). El número de cebadores que hibridan con segmentos V en una reacción múltiple puede ser, por ejemplo, al menos 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, u 80. El número de cebadores que hibridan con segmentos V en una reacción múltiple puede ser, por ejemplo, 10-60, 20-50, 30-50, 40-50, 20-40, 30-40, ó 35-40. Los cebadores pueden hibridar con diferentes segmentos V. Para los genes de IgH, debido a la posibilidad de mutaciones somáticas en los segmentos V, pueden usarse múltiples cebadores que hibridan con cada segmento V;



por ejemplo, 1, 2, 3, 4, ó 5 cebadores por segmento V. El número de cebadores que hibridan con segmentos C en una reacción múltiple puede incluir, por ejemplo, al menos 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, ó 15. El número de cebadores que hibridan con segmentos C en una reacción múltiple puede ser 1-10, 2-9, 3-8, 4-7, 3-8, ó 3-6. La amplificación de genes de TCR o inmunoglobulina puede ocurrir como se describe en el Ejemplo 3 y/o Ejemplo 4.

La región que se va a amplificar puede incluir la secuencia clonal completa o un subconjunto de la secuencia clonal, incluyendo la unión V-D, la unión D-J de un gen de inmunoglobulina o receptor de células T, la región variable completa de un gen de inmunoglobulina o receptor de células T, la región de reconocimiento de antígeno, o una CDR, por ejemplo, región determinante de la complementariedad 3 (CDR3).

La secuencia de TCR o inmunoglobulina puede amplificarse usando una etapa de amplificación primaria y una secundaria. Cada una de las diferentes etapas de amplificación puede comprender diferentes cebadores. Los diferentes cebadores pueden introducir secuencia que no está originalmente presente en la secuencia del gen inmune. Por ejemplo, el procedimiento de amplificación puede añadir nuevos sitios de unión a cebador a los extremos de las secuencias diana para convertir una amplificación múltiple en una amplificación única o el procedimiento de amplificación puede añadir una o más etiquetas al extremo 5' y/o 3' de la secuencia amplificada de TCR o inmunoglobulina (como se ilustra en las Figs. 3A-3C). La etiqueta puede ser una secuencia que facilita la secuenciación posterior del ADN amplificado. La etiqueta puede ser una secuencia que facilita la unión de la secuencia amplificada a un soporte sólido.

Otros métodos para la amplificación pueden no emplear ningún cebador en la región V. En lugar de esto, puede usarse un cebador específico del segmento C y un cebador genérico puede ponerse en el otro lado (5'). El cebador genérico puede adjuntarse en la síntesis de ADNc a través de diferentes métodos incluyendo los métodos bien descritos de intercambio de cadenas. De manera similar, el cebador genérico puede adjuntarse después de la preparación del ADNc a través de diferentes métodos incluyendo ligación.

Otros medios para amplificar ácido nucleico que pueden usarse en los métodos de la invención proporcionada incluyen, por ejemplo, PCR de transcripción inversa, PCR en tiempo real, PCR en tiempo real cuantitativa, PCR digital (dPCR), PCR en emulsión digital (dePCR), PCR clonal, PCR de polimorfismo en la longitud del fragmento amplificado (AFLP PCR), PCR específica de alelo, PCR asimétrica (en la que se usa un gran exceso de cebadores para una cadena elegida), PCR de colonia, amplificación dependiente de helicasa (HDA), PCR de inicio caliente, PCR inversa (IPCR), PCR in situ, PCR larga (extensión de ADN mayor de aproximadamente 5 kilobases), PCR múltiple, PCR anidada (usa más de una pareja de cebadores), PCR de célula única, PCR "touchdown", PCR isothermal mediada por bucle (LAMP), y amplificación basada en la secuencia de ácido nucleico (NASBA). Otros esquemas de amplificación incluyen: reacción en cadena de la ligasa, amplificación de ADN ramificado, amplificación por círculo rodante, amplificación círculo a círculo, amplificación SPIA, amplificación de diana por amplificación de captura y ligación (TACL), y amplificación RACE.

La información en ARN en una muestra puede convertirse en ADNc usando transcripción inversa. Pueden usarse cebadores poliA, cebadores aleatorios, y/o cebadores específicos de gen en las reacciones de transcripción inversa según protocolos convencionales.

Después de la amplificación del ADN del genoma (o amplificación de ácido nucleico en la forma de ADNc por transcripción inversa del ARN), las moléculas de ácido nucleico individuales pueden aislarse, opcionalmente re-amplificarse, y después secuenciarse individualmente. Los protocolos de amplificación ejemplares pueden encontrarse en van Dongen et al, *Leukemia*, 17: 2257-2317 (2003) o van Dongen et al, publicación de patente U.S. 2006/0234234. Brevemente, un protocolo ejemplar es como sigue: tampón de reacción: tampón ABI II o tampón ABI Gold (Life Technologies, San Diego, CA); volumen final de reacción 50  $\mu$ L; 100 ng de ADN de muestra; 10 pmoles de cada cebador (sujeto a ajustes para equilibrar la amplificación como se describe más adelante); dNTP a concentración final 200  $\mu$ M;  $MgCl_2$  a concentración final 1,5 mM (sujeto a optimización dependiendo de las secuencias diana y la polimerasa); polimerasa Taq (1-2 U/tubo); condiciones de ciclado: preactivación 7 min a 95°C; hibridación a 60°C; tiempos de ciclado: 30s desnaturalización; 30s hibridación; 30s extensión.

Las polimerasas que pueden usarse para la amplificación en los métodos de la invención están disponibles comercialmente e incluyen, por ejemplo, polimerasa Taq, polimerasa AccuPrime, o Pfu. La elección de polimerasa para usarse puede basarse en si se prefiere la fidelidad o eficiencia.

Los métodos para el aislamiento de ácidos nucleicos de un conjunto incluyen separación espacial de las moléculas en dos dimensiones en un sustrato sólido (por ejemplo, portaobjetos de vidrio), separación espacial de las moléculas en tres dimensiones en una disolución en micelas (tal como se puede conseguir usando emulsiones de aceite con o sin la inmovilización de las moléculas en una superficie sólida tal como lechos), o usando cámaras de microreacción, por ejemplo, en chips microfluidos o nanofluidos. La dilución puede usarse para asegurar que de media una única molécula está presente en un volumen dado, región espacial, lecho, o cámara de reacción. La guía para dichos métodos de aislar moléculas de ácido nucleico individuales se encuentra en las referencias siguientes: Sambrook, *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, 2001s); Shendure et al, *Science*, 309: 1728-1732 (incluyendo material suplementario) (2005); patente U.S. 6.300.070; Bentley et al, *Nature*, 456: 53-59 (incluyendo material suplementario) (2008); patente U.S. 7.323.305; Matsubara et al, *Biosensors & Bioelectronics*, 20: 1482-1490 (2005); patente U.S. 6.753.147; y semejantes.

Las mediciones de PCR en tiempo real, tinción de picogreen, electroforesis en nanofluido (por ejemplo, LabChip) o absorción UV pueden usarse en una etapa inicial para juzgar la cantidad funcional de material

amplificable.

Los métodos para la re-amplificación de ácidos nucleicos incluyen crecimiento bacteriano de colonias aisladas transformadas con ácido nucleico, amplificación en un portaobjetos (por ejemplo, colonias de PCR (polonias)), y amplificación en un lecho (por ejemplo, en una PCR de emulsión). El mismo método puede usarse para amplificar y re-amplificar el ácido nucleico o puede usarse un método diferente para amplificar y reamplificar el ácido nucleico.

En determinadas realizaciones, las etapas de subclonación incluyen una etapa en la que un cebador común se une al ADN o ARN a través de una etapa de amplificación o ligación. Este cebador se usa entonces para amplificar los clones y como una secuencia de reconocimiento para la hibridación de un cebador para secuenciación (por ejemplo, como se ilustra en las Figs. 2A-2B y 4A-4B, y se discute más completamente más adelante).

En un aspecto, las amplificaciones múltiples se llevan a cabo de manera que cantidades relativas de secuencias en una población de partida son sustancialmente las mismas que aquellas en la población amplificada, o amplicón. Esto es, las amplificaciones múltiples se llevan a cabo con sesgo mínimo de amplificación entre las secuencias miembro de una población de muestra. En una realización, dichas cantidades relativas son sustancialmente las mismas si cada cantidad relativa en un amplicón está en cinco veces su valor en la muestra de partida. En otra realización, dichas cantidades relativas son sustancialmente las mismas si cada cantidad relativa en un amplicón está en dos veces su valor en la muestra de partida. Como se discute más completamente más adelante, el sesgo de la amplificación en PCR puede detectarse y corregirse usando técnicas convencionales de manera que puede seleccionarse un conjunto de cebadores de PCR para un repertorio predeterminado que proporciona la amplificación no sesgada de cualquier muestra.

Respecto a muchos repertorios basados en secuencias de TCR o BCR, una amplificación múltiple usa opcionalmente todos los segmentos V. La reacción se optimiza para intentar obtener una amplificación que mantiene la abundancia relativa de las secuencias amplificadas por diferentes cebadores de segmentos V. Algunos de los cebadores están relacionados, y por lo tanto muchos de los cebadores pueden "intercomunicarse", amplificando moldes que no concuerdan perfectamente con él. Las condiciones se optimizan de manera que cada molde puede amplificarse de una manera similar independientemente de qué cebador lo amplifique. En otras palabras, si hay dos moldes, entonces después de una amplificación de 1.000 veces, ambos moldes pueden amplificarse aproximadamente 1.000 veces, y no importa que para uno de los moldes la mitad de los productos amplificados porten un cebador diferente debido a la intercomunicación. En análisis posteriores de los datos de secuenciación, la secuencia de cebador se elimina del análisis, y por lo tanto no importa qué cebador se usa en la amplificación siempre que los moldes se amplifiquen igualmente.

Como la cantidad de cada molde es desconocida en una población de ADNc generada a partir de ARNm, puede generarse un conjunto de estándares usando PCR únicas de una población de ADNc de clonotipos. Esto se llevó a cabo para un repertorio de clonotipos de TCR $\beta$ . El producto en cada una de 34 de dichas PCR (usando en las reacciones separadas cebadores del Ejemplo 3) comprendió una pluralidad de secuencias con un cebador de V. Los diferentes productos se cuantificaron cuidadosamente para crear un conjunto de estándares a la misma concentración. Se usó un combinado de todos los 34 cebadores y se realizaron 34 PCR en tiempo real usando el combinado de cebadores y cada una de las secuencias estándar como un molde. Idealmente, sin sesgo, todos los 34 estándares mostrarán una eficiencia de amplificación igual por PCR en tiempo real. Esto sugiere que cada secuencia se amplifica igualmente aunque la presencia de intercomunicación hace que no esté claro qué cebadores están llevando a cabo la amplificación. Esta optimización es consistente con el objetivo de tener una amplificación igual independientemente de los cebadores reales que se incorporan en el producto de amplificación. El incremento de la concentración total del combinado de cebadores redujo significativamente el rango dinámico como se esperaba a partir de incrementar la eficiencia de la amplificación. Además, para los moldes que parecían amplificar más eficientemente que la media, disminuyó la concentración de su cebador con concordancia perfecta en el combinado. A la inversa, para los moldes que se amplificaron de manera ineficiente, se incrementó la concentración de su cebador con concordancia perfecta. Esta optimización demostró que todos los moldes se amplifican en dos veces de la amplificación promedio.

El sesgo de la amplificación también puede evitarse llevando a cabo una amplificación de dos etapas (como se ilustra en las Figs. 2A-2B) 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 diana. Las colas incluyen sitios de unión de cebador que se añaden a los extremos de las secuencias del amplicón primario de manera que dichos sitios se usan en una segunda etapa de amplificación usando sólo un único cebador directo y un único cebador inverso, eliminando de esta manera una causa primaria de sesgo de la amplificación. Preferiblemente, la PCR primaria tendrá un número de ciclos lo suficientemente pequeño (por ejemplo, 5-10) como para minimizar la amplificación diferencial por los diferentes cebadores. La amplificación secundaria se hace con una pareja de cebadores y por lo tanto el problema de amplificación diferencial es mínimo. Un uno por ciento de la PCR primaria se toma directamente para la PCR secundaria. Treinta y cinco ciclos (equivalente a ~28 ciclos sin la etapa de dilución de 100 veces) usados entre las dos amplificaciones fueron suficientes para mostrar una amplificación robusta independientemente de si la división de ciclos fue: un ciclo primario y 34 secundarios ó 25 primarios y 10 secundarios. Aunque idealmente haciendo sólo 1 ciclo en la PCR primaria puede disminuir el sesgo de la amplificación, hay otras consideraciones. Un aspecto de esto es

representación. Esto juega un papel cuando la cantidad de partida introducida no está en exceso respecto al número de lecturas obtenidas finalmente. Por ejemplo, si se obtienen 1.000.000 de lecturas y empezando con 1.000.000 de moléculas introducidas entonces el tomar sólo 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 no ser que la amplificación de PCR primaria generara significativamente más de 100 moléculas. Esto indica que puede usarse un mínimo de 8 ciclos (256 veces), pero de manera más confortable 10 ciclos (~1.000 veces). La alternativa a esto es tomar más de 1% de la PCR primaria en la secundaria pero debido a la alta concentración de cebador usada en la PCR primaria, puede usarse un gran factor de dilución para asegurar que estos cebadores no interfieren en la amplificación y empeoran el sesgo de la amplificación entre secuencias. Otra alternativa es añadir una etapa de purificación o enzimática para eliminar los cebadores de la PCR primaria para permitir una dilución menor de ésta. En este ejemplo, la PCR primaria fue 10 ciclos y la segunda 25 ciclos.

La reproducibilidad de una PCR múltiple puede evaluarse como sigue, como se ejemplifica con el conjunto de ensayos de cebadores del Ejemplo 2. Se realizan dos reacciones de PCR primarias usando el conjunto de ensayo de cebadores, por ejemplo, los cebadores combinados de TCR $\beta$  y el cebador de C (del Ejemplo 2) y una muestra de ADNc como un molde. La abundancia relativa en cada molde amplificado se evalúa usando PCR en tiempo real. Usando cada uno de los dos productos amplificados como un molde, se realizaron treinta y cuatro reacciones de PCR en tiempo real diferentes usando el cebador de C y uno de los cebadores de V en cada reacción. Los datos mostrados en la Fig. 5 demuestran que la abundancia relativa determinada por PCR en tiempo real fue altamente reproducible usando todos los cebadores de V para las dos muestras, indicando que la amplificación múltiple es altamente reproducible. El número de ciclos (valor Ct) para cada una de las amplificaciones por PCR en tiempo real usando el producto de amplificación múltiple como un molde se muestra en el eje de las X y usando el segundo producto de amplificación múltiple como un molde se representa en el eje de las Y.

La cantidad de sesgo en la amplificación de un conjunto de cebadores puede evaluarse usando el procedimiento siguiente, que se ejemplifica usando el conjunto de cebadores del Ejemplo 2. El conjunto de cebadores de ensayo (como anteriormente) se usa para amplificar ADNc como un molde (por ejemplo, obtenido de ARNm extraído de linfocitos). La cantidad de molde amplificado por cada uno de los 34 cebadores diferentes (junto con el cebador del segmento C) se determina usando PCR en tiempo real y esta cantidad se compara con la cantidad amplificada usando los mismos cebadores con el ADNc. Como puede haber intercomunicación incluso si la abundancia relativa entre las secuencias internas en el producto amplificado y el ADNc fueran las mismas, sólo pueden detectarse diferencias significativas en la amplificación usando este resultado. Esta posibilidad puede ensayarse sintetizando una colección de cebadores para amplificar segmentos internos de un número de las secuencias de ADNc de partida. Por ejemplo, se diseñaron 12 oligos que pueden, cuando se usan con cebador del segmento C, amplificar secuencias internas a los cebadores del segmento V anteriores. Si hay un sesgo de la amplificación mínimo, entonces la concentración de estas secuencias internas debería cambiar poco entre el ADNc de partida y los productos amplificados. Los datos de este ejemplo se muestran en la FIG. 6. Aquí, se usó una muestra de ADNc como un molde para una amplificación múltiple usando los cebadores de TCR $\beta$  y cebador de C combinados (del Ejemplo 2). El cebador de C y los cebadores internos aguas abajo se usaron para la amplificación inicial de material molde de la amplificación múltiple. De manera similar, se usó PCR en tiempo real para evaluar la abundancia relativa de las mismas secuencias en el ADNc. Si la amplificación múltiple tuviera algún sesgo significativo, la abundancia relativa en el material amplificado sería muy diferente de la de en el ADNc. Como puede observarse en la FIG. 6, se observó una alta correlación lo que demuestra un sesgo mínimo en la amplificación en la amplificación múltiple. El número de ciclos (valor Ct) para cada una de las amplificaciones por PCR en tiempo real usando cebadores internos, y ADNc y el producto de la amplificación múltiple como molde, se muestra en los ejes de las X y las Y, respectivamente.

La amplificación inicial puede hacerse a partir de ADN o ARN (por ejemplo, después de conversión a ADNc).

#### **Secuenciación de poblaciones de ácidos nucleicos**

Las técnicas de secuenciación de ADN para uso en los métodos de la invención incluyen secuenciación por síntesis usando nucleótidos marcados en los extremos reversiblemente, pirosecuenciación, secuenciación 454, hibridación específica de alelo a una biblioteca de sondas de oligonucleótidos marcadas, secuenciación por síntesis usando hibridación específica de alelo a una biblioteca de clones marcados que está seguida de ligación, monitorización en tiempo real de la incorporación de nucleótidos marcados durante una etapa de polimerización, secuenciación de colonias, y secuenciación SOLiD. La secuenciación de las moléculas separadas se ha demostrado más recientemente por reacciones de extensión secuencial o única usando polimerasas o ligasas así como por hibridaciones diferenciales únicas o secuenciales con bibliotecas de sondas. Estas reacciones se han realizado en muchas secuencias clonales en paralelo incluyendo demostraciones en aplicaciones comerciales actuales de más de 100 millones de secuencias en paralelo. Estas estrategias de secuenciación pueden usarse así para estudiar el repertorio de receptor de células T (TCR) y/o receptor 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 en la que se secuencian en paralelo. Dichas superficies sólidas pueden incluir superficies no porosas (tal como en secuenciación Solexa, por

ejemplo, Bentley et al, Nature, 456: 53-59 (2008) o secuenciación Complete Genomics, por ejemplo, Drmanac et al, Science, 327: 78-81 (2010)), matrices de pocillos, que pueden incluir moldes unidos a lechos o partículas (tales como con 454, por ejemplo, Margulies et al, Nature, 437: 376-380 (2005) o secuenciación Ion Torrent, publicación de patente U.S. 2010/0137143 ó 2010/0304982), membranas micromecanizadas (tales como con secuenciación SMRT, por ejemplo, Eid et al, Science, 323: 133-138 (2009)), o matrices de lechos (como con secuenciación SOLiD o secuenciación de colonias, por ejemplo, Kim et al, Science, 316: 1481-1414 (2007)). En otro aspecto, dichos métodos comprenden amplificar las moléculas aisladas bien antes o después de que sean aisladas espacialmente en una superficie sólida. La amplificación anterior puede comprender amplificación basada en emulsión, tal como PCR de emulsión, o amplificación por círculos rodantes. Es particularmente interesante la secuenciación basada en Solexa en la que moléculas molde individuales se aíslan espacialmente en una superficie sólida, después de lo cual se amplifican en paralelo por PCR en puente para formar poblaciones clonales separadas, o agrupaciones, y después 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™, Illumina, Inc., San Diego, CA, 2010); y adicionalmente en las referencias siguientes: patentes U.S. 6.090.592; 6.300.070; 7.115.400; y EP0972081B1. En una realización, las moléculas individuales dispuestas y amplificadas en una superficie sólida forman agrupaciones en una densidad de al menos  $10^5$  agrupaciones por  $\text{cm}^2$ ; o en una densidad de al menos  $5 \times 10^5$  por  $\text{cm}^2$ ; o en una densidad de al menos  $10^6$  agrupaciones por  $\text{cm}^2$ . En una realización, se emplean químicas de secuenciación que tienen proporciones de error relativamente altas. En dichas realizaciones, las puntuaciones de calidad media producidas por dichas químicas son funciones que declinan monotónicamente de longitudes de lectura de secuencia. En una realización, dicha declinación corresponde a que 0,5 por ciento de las lecturas de secuencia tienen al menos un error en las posiciones 1-75; 1 por ciento de las lecturas de secuencia tienen al menos un error en las posiciones 76-100; y 2 por ciento de las lecturas de secuencia tienen al menos un error en las posiciones 101-125.

En un aspecto, para cada muestra de un individuo, la técnica de secuenciación usada en los métodos de la invención genera secuencias de al menos 1.000 clonotipos por operación; en otro aspecto, dicha técnica genera secuencias de al menos 10.000 clonotipos por operación; en otro aspecto, dicha técnica genera secuencias de al menos 100.000 clonotipos por operación; en otro aspecto, dicha técnica genera secuencias de al menos 500.000 clonotipos por operación; y en otro aspecto, dicha técnica genera secuencias de al menos 1.000.000 clonotipos por operación. En otro aspecto más, dicha técnica genera secuencias de entre 100.000 a 1.000.000 clonotipos por operación por muestra individual.

La técnica de secuenciación usada en los métodos de la invención proporcionada puede generar aproximadamente 30 pb, aproximadamente 40 pb, aproximadamente 50 pb, aproximadamente 60 pb, aproximadamente 70 pb, aproximadamente 80 pb, aproximadamente 90 pb, aproximadamente 100 pb, aproximadamente 110 pb, aproximadamente 120 pb por lectura, aproximadamente 150 pb, aproximadamente 200 pb, aproximadamente 250 pb, aproximadamente 300 pb, aproximadamente 350 pb, aproximadamente 400 pb, aproximadamente 450 pb, aproximadamente 500 pb, aproximadamente 550 pb, o aproximadamente 600 pb por lectura.

La técnica de secuenciación usada en los métodos de la invención proporcionada puede generar al menos 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 150, 200, 250, 300, 350, 400, 450, 500, 550, ó 600 pb por lectura. Un perfil de clonotipos basado en secuencia de un individuo puede obtenerse usando las etapas siguientes: (a) obtener una muestra de ácido nucleico de células T y/o células B del individuo; (b) aislar espacialmente las moléculas individuales derivadas de dicha muestra de ácido nucleico, comprendiendo las moléculas individuales conjuntos anidados de moldes generado cada uno de un ácido nucleico en la muestra y conteniendo cada uno una región reorganizada somáticamente o una parte de ésta, siendo capaz cada conjunto anidado de producir una pluralidad de lecturas de secuencia cada una extendiéndose en la misma dirección y cada una empezando a partir de una posición diferente en el ácido nucleico del que se generó el conjunto anidado; (c) secuenciar dichas moléculas individuales aisladas espacialmente; y (d) determinar las abundancias de diferentes secuencias de las moléculas de ácido nucleico de la muestra de ácido nucleico para generar el perfil de clonotipos. 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 empieza a partir de una posición diferente en la región V y se extiende en la dirección de su región J asociada. 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. Adicionalmente a la última realización, al menos una de las lecturas de secuencia directa y al menos una de las lecturas de secuencia inversa tiene una región de superposición de manera que las bases de dicha región de superposición se determinan por una relación de complementariedad 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 empezando a partir de 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, IgK

completa, moléculas de IgK inactivas, moléculas de TCR $\beta$ , moléculas de TCR $\gamma$ , moléculas de TCR $\delta$  completas, y moléculas de TCR $\delta$  incompletas. En otra realización, la etapa de secuenciación comprende generar las lecturas de secuencia que tienen puntuaciones de calidad que disminuyen monótonicamente. Adicionalmente a la última realización, las puntuaciones de calidad que disminuyen monótonicamente son tales que las lecturas de secuencia tienen proporciones de errores no mejores de las siguientes: 0,2 por ciento de lecturas de secuencia contienen al menos un error en las posiciones de base 1 a 50, 0,2 a 1,0 por ciento de lecturas de secuencia contienen al menos un error en las posiciones 51-75, 0,5 a 1,5 por ciento de lecturas de secuencia contienen al menos un error en las posiciones 76-100.

Como se indica más adelante en la definición de repertorio, pueden secuenciarse diferentes regiones predeterminadas de los genes de inmunoglobulina o receptor de células T. En algunas realizaciones, puede secuenciarse la secuencia completa de las regiones variables para identificar y cuantificar un clonotipo.

Puede secuenciarse un único subconjunto de las secuencias clonales completas. En algunas realizaciones, se secuencian nucleótidos que comprenden las uniones VD y DJ para identificar y cuantificar de manera única un clonotipo. En otras realizaciones, el fragmento que puede secuenciarse es la región variable completa. En otra realización más, se secuencia la región de reconocimiento de antígeno o la región determinante de la complementariedad 3 (CDR3). Un fragmento que contiene la CD3 completa o la región variable completa puede amplificarse para permitir la secuenciación de la CD3 que comprende partes de los segmentos V, D, y J.

En una realización, sólo se amplifica y secuencia CDR3. La amplificación y secuenciación de CDR3 pueden conseguirse usando cebadores específicos de una o más secuencias de segmento V (así como uno o más cebadores en el otro lado del amplicón en el segmento C). Pueden utilizarse cebadores para cada uno de los segmentos V en una o más reacciones de amplificación dando lugar a la amplificación del repertorio completo de secuencias. Este repertorio de secuencias puede mezclarse entonces y someterse a separación, con o sin amplificación, y secuenciarse usando cualquiera de las técnicas de secuenciación descritas. Cuando la amplificación con los varios cebadores de V se hace en tubos separados, el número de moléculas que portan los diferentes segmentos V puede "normalizarse" debido a la saturación de la PCR. Por ejemplo, si un segmento V particular tuviera una o varias expansiones clonales dando lugar a su representación más que otros segmentos, esta información puede borrarse o disminuirse ya que la reacción de PCR para cada segmento puede dirigirse a saturación o cerca de ésta. La PCR en tiempo real puede usarse para cuantificar cuánto de cada segmento V está presente. Puede secuenciarse la CDR3 completa, o puede secuenciarse un subconjunto de la secuencia de CDR3.

En una realización, sólo se analiza un subconjunto de clonotipos. Esto puede conseguirse amplificando con un cebador específico para el subconjunto de clonotipos, por ejemplo, un cebador que es específico para el segmento V. Los clonotipos únicos pueden identificarse por secuenciación con lecturas contiguas largas que proporcionan conectividad completa. En algunas realizaciones, cuando están presentes varias secuencias de interés, una longitud de lectura corta a lo largo de sólo una de las uniones puede generar etiquetas degeneradas que no son únicas para un clonotipo específico sino que están compartidas entre clonotipos múltiples. Por ejemplo, la secuenciación a lo largo de la unión V/J puede agrupar todas las secuencias con la misma V/J independientemente del segmento D como un clonotipo. La información sobre la conectividad completa de todos los segmentos permite distinguir secuencias que pueden compartir los mismos segmentos V y J pero que están conectadas a diferentes segmentos, por ejemplo.

#### **Determinación de clonotipos a partir de datos de secuencia**

En un aspecto de la invención, las secuencias de los clonotipos (incluyendo pero no limitado a aquellas derivadas de IgH, TCR $\alpha$ , TCR $\beta$ , TCR $\gamma$ , TCR $\delta$ , y/o IgL $\kappa$  (IgK)) pueden determinarse combinando la información de una o más lecturas de secuencia, por ejemplo, junto con las regiones V(D)J de las cadenas seleccionadas. En otro aspecto, las secuencias de los clonotipos se determinan combinando la información de una pluralidad de lecturas de secuencia. (Tal y como se usa en la presente memoria, una "lectura de secuencia" es una secuencia de datos generada por una técnica de secuenciación a partir de la cual se determina una secuencia de nucleótidos. Típicamente, las lecturas de secuencia se hacen extendiendo un cebador a lo largo de un ácido nucleico molde, por ejemplo, con una ADN polimerasa o una ADN ligasa. Los datos se generan registrando las señales, tales como señales ópticas, químicas (por ejemplo, cambio en el pH), o eléctricas, asociadas con dicha extensión). Dichas pluralidades de lecturas de secuencia pueden incluir una o más lecturas de secuencia a lo largo de una cadena con sentido (es decir, lecturas de secuencia "directas") y una o más lecturas de secuencia a lo largo de su cadena complementaria (es decir, lecturas de secuencia "inversas"). Cuando se generan múltiples lecturas de secuencia a lo largo de la misma cadena, se generan en primer lugar moldes separados mediante la amplificación de moléculas de muestra con cebadores seleccionados para las diferentes posiciones de las lecturas de secuencia. Este concepto se ilustra en la Fig. 4A, en la que los cebadores (404, 406 y 408) se emplean para generar amplicones (410, 412, y 414, respectivamente) en una única reacción. Dichas amplificaciones pueden llevarse a cabo en la misma reacción o en reacciones separadas. En un aspecto, siempre que se emplee la PCR, se usan reacciones de amplificación separadas para generar los moldes separados que, a su vez, se combinan y usan para generar múltiples lecturas de secuencia a lo largo de la misma cadena. Esta última estrategia es preferible para evitar la necesidad de equilibrar las concentraciones de cebadores (y/o otros parámetros de la reacción) para asegurar una amplificación igual de los múltiples moldes (algunas veces referido en la presente memoria como "amplificación equilibrada" o "amplificación sin sesgo"). La generación de moldes en reacciones separadas se ilustra en las Figs. 4B-

4C. Aquí, una muestra que contiene IgH (400) se divide en tres partes (472, 474, y 476) que se añaden a PCR separadas usando cebadores de la región J (401) y cebadores de la región V (404, 406, y 408, respectivamente) para producir amplicones (420, 422 y 424, respectivamente). Los últimos amplicones se combinan (478) en PCR secundaria (480) usando los cebadores P5 y P7 para preparar los moldes (482) para PCR en puente y secuenciación en un secuenciador Illumina GA, o instrumento semejante.

Las lecturas de secuencia pueden tener una amplia variedad de longitudes, dependiendo en parte de la técnica de secuenciación que se está empleando. Por ejemplo, para algunas técnicas, pueden surgir varias soluciones intermedias en su implementación, por ejemplo, (i) el número y longitudes de las lecturas de secuencia por molde y (ii) el coste y 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 1 a 4 lecturas de secuencia para determinar la secuencia de cada clonotipo; en otra realización, se generan 2 a 4 lecturas de secuencia para determinar la secuencia de cada clonotipo; y en otra realización, se generan 2 a 3 lecturas de secuencia para determinar la secuencia de cada clonotipo. En las realizaciones anteriores, los números proporcionados son exclusivos de lecturas de secuencia usadas para identificar muestras de diferentes individuos. Las longitudes de las varias lecturas de secuencia usadas en las realizaciones descritas más adelante también varían sobre la base de la información que se busca que sea capturada por la lectura; por ejemplo, la localización de partida y la longitud de una lectura de secuencia puede diseñarse para proporcionar la longitud de una región NDN así como su secuencia de nucleótidos; así, se seleccionan las lecturas de secuencia que abarcan la región NDN completa. En otros aspectos, una o más lecturas de secuencia engloban las regiones D y/o NDN.

Las secuencias de los clonotipos pueden determinarse en parte alineando 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 por la determinación de bases sin alineamiento con las secuencias de referencia, tal como en la región NDN altamente variable. Pueden aplicarse una variedad de algoritmos de alineamiento a las lecturas de secuencia y las secuencias de referencia. Por ejemplo, la guía para seleccionar métodos de alineamiento está disponible en Batzoglu, *Briefings in Bioinformatics*, 6: 6-22 (2005). En un aspecto, siempre que las lecturas de V o las lecturas de C (descrito más completamente más adelante) se alinean con las secuencias de referencia de la región V y J, se emplea un algoritmo de búsqueda en árbol, por ejemplo, Cormen et al, *Introduction to Algorithms*, Tercera Edición (The MIT Press, 2009). Las estructuras de codones de las secuencias de referencia de V y J pueden usarse en un proceso de alineamiento para eliminar errores de la secuenciación y/o para determinar un nivel de confianza en el alineamiento resultante, como se describe más completamente más adelante. 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 Fig. 3B), de manera que las bases de las lecturas están en una relación complementaria inversa entre sí. Así, por ejemplo, si una lectura directa en la región de superposición es "5'-acgttc", entonces una lectura inversa en una relación complementaria inversa es "5'-gcaacgt" en la misma región de superposición. En un aspecto, se determinan las bases en dicha región de superposición, al menos en parte, a partir de dicha relación complementaria inversa. Esto es, una probabilidad de una llamada de base (o puntuación de calidad relacionada) en una región de superposición potencial se incrementa si conserva, o es consistente con, una relación complementaria inversa entre las dos lecturas de secuencia. En un aspecto, se determinan los clonotipos de las cadenas TCR  $\beta$  e IgH (ilustrados en la Fig. 3B) por al menos una lectura de secuencia empezando en su región J y extendiendo en la dirección de su región V asociada (referido en la presente memoria como una "lectura C" (304)) y al menos una lectura de secuencia empezando en su región V y extendiendo en la dirección de su región J asociada (referido en la presente memoria como una "lectura V" (306)). La región de superposición (308) puede englobar o no la región NDN (315) como se muestra en la Fig. 3B. La región de superposición (308) puede estar enteramente en la región J, enteramente en la región NDN, enteramente en la región V, o puede englobar un límite región J-región NDN o un límite región V-región NDN, o ambos de dichos límites (como se ilustra en la Fig. 3B). Típicamente, dichas lecturas de secuencia se generan extendiendo cebadores de secuenciación, por ejemplo (302) y (310) en la Fig. 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 manera que pueden proporcionar un punto de partida o punto de anclaje para el alineamiento inicial y análisis de las lecturas de secuencia. En una realización, una lectura C se posiciona de manera que engloba la región D y/o NDN de la cadena de TCR  $\beta$  o IgH e incluye una parte de la región V adyacente, por ejemplo, como se ilustra en las Figs. 3B y 3C. En un aspecto, la superposición de la lectura V y la lectura C en la región V se usa para alinear las lecturas entre sí. En otra realización, dicho alineamiento de lecturas de secuencia no es necesario, por ejemplo, con las cadenas de TCR $\beta$ , de manera que una lectura V puede ser sólo lo suficientemente larga para identificar la región V particular de un clonotipo. Este último aspecto se ilustra en la Fig. 3C. La lectura de secuencia (330) se usa para identificar una región V, con o sin superposición con otra lectura de secuencia, y otra lectura de secuencia (332) atraviesa la región NDN y se usa para determinar la secuencia de ésta. La parte (334) de la lectura de secuencia (332) que se extiende en la región V se usa para asociar la información de la secuencia de la lectura de secuencia (332) con la de la lectura de secuencia (330) para determinar un clonotipo. Para algunos métodos de secuenciación, tal como estrategias base por base como el método

de secuenciación Solexa, el tiempo de operación de la secuenciación y los costes de los reactivos se reducen minimizando el número de ciclos de secuenciación en un análisis. Opcionalmente, como se ilustra en la Fig. 3B, se produce el amplicón (300) con etiqueta de muestra (312) para distinguir entre clonotipos que se originan de diferentes muestras biológicas, por ejemplo, diferentes pacientes. La etiqueta de muestra (312) puede identificarse hibridando una región de unión cebador a cebador (316) y extendiéndola (314) para producir una lectura de secuencia a lo largo de la etiqueta (312), a partir de la cual se descodifica la etiqueta de la muestra (312).

La cadena de IgH presenta un mayor reto en su análisis que la cadena de TCR $\beta$  debido al menos a dos factores: i) la presencia de mutaciones somáticas hace que el mapeo o alineamiento sean más difíciles, y ii) la región NDN es mayor por lo que frecuentemente no es posible mapear una parte del segmento V para la lectura C. En un aspecto de la invención, este problema se supera usando una pluralidad de conjuntos de cebadores para generar lecturas V, que se localizan en diferentes localizaciones a lo largo de la región V, preferiblemente de manera que los sitios de unión del cebador no se superponen y están espaciados entre sí, y con al menos un sitio de unión de cebador adyacente a la región NDN, por ejemplo, en una realización, de 5 a 50 bases de la unión V-NDN, o en otra realización de 10 a 50 bases de la unión V-NDN. La redundancia de una pluralidad de conjuntos de cebadores minimiza el riesgo de no detectar un clonotipo debido a un fallo de uno o dos cebadores que tienen sitios de unión afectados por mutaciones somáticas. Además, la presencia de al menos un sitio de unión de cebador adyacente a la región NDN hace más probable que una lectura V se superponga con la lectura C y por lo tanto extiende de manera efectiva la longitud de la lectura C. Esto permite la generación de una secuencia continua que abarca todos los tamaños de regiones NDN y que también puede mapear sustancialmente las regiones V y J enteras en ambos lados de la región NDN. Las realizaciones para llevar a cabo dicho esquema se ilustran en las Figs. 4A y 4D. En la Fig. 4A, una muestra que comprende cadenas de IgH (400) se secuencian generando una pluralidad de amplicones para cada cadena mediante la amplificación de las cadenas con un único conjunto de cebadores de la región J (401) y una pluralidad (tres mostrados) de conjuntos de cebadores de la región V (402) (402, 406, 408) para producir una pluralidad de amplicones anidados (por ejemplo, 410, 412, 416) todos comprendiendo la misma región NDN y teniendo diferentes longitudes que engloban sucesivamente partes mayores (411, 413, 415) de la región V (402). Los miembros de un conjunto anidado pueden agruparse conjuntamente después de secuenciar indicando la identidad (o identidad sustancial) de sus regiones NDN, J y/o C respectivas, permitiendo de esta manera la reconstrucción de un segmento V(D)J más largo de lo que sería el caso de otra manera para una plataforma de secuenciación con longitud de lectura limitada y/o calidad de secuencia. En una realización, la pluralidad de conjuntos de cebadores puede ser un número en el intervalo de 2 a 5. En otra realización, la pluralidad es 2-3; y otra realización más, la pluralidad es 3. Las concentraciones y posiciones de los cebadores en una pluralidad pueden variar ampliamente. Las concentraciones de los cebadores de la región V pueden ser o no las mismas. En una realización, el cebador más cercano a la región NDN tiene una mayor concentración que los demás cebadores de la pluralidad, por ejemplo, para asegurar que los amplicones que contienen la región NDN están representados en el amplicón resultante. Uno o más cebadores (por ejemplo, 435 y 437 en la Fig. 4B) adyacentes a la región NDN (444) pueden usarse para generar una o más lecturas de secuencia (por ejemplo, 434 y 436) que se superponen a la lectura de secuencia (442) generada por el cebador de la región J (432), mejorando de esta manera la calidad de las llamadas de base en la región de superposición (440). Las lecturas de secuencia de la pluralidad de cebadores pueden superponerse o no al sitio de unión del cebador aguas abajo adyacente y/o lectura de secuencia aguas abajo adyacente. En una realización, las lecturas de secuencia próximas a la región NDN (por ejemplo, 436 y 438) pueden usarse para identificar la región V particular asociada con el clonotipo. Dicha pluralidad de cebadores reduce la probabilidad de amplificación incompleta o fallida en el caso de que uno de los sitios de unión del cebador esté hipermutado durante el desarrollo de la inmunoglobulina. También incrementa la probabilidad de que la diversidad introducida por la hipermutación de la región V se capture en una secuencia de clonotipo. Puede realizarse una PCR secundaria para preparar los amplicones anidados para secuenciación, por ejemplo, amplificando con los cebadores P5 (401) y P7 (404, 406, 408) como se ilustra para producir amplicones (420, 422, y 424), que pueden distribuirse como moléculas únicas en una superficie sólida, en la que se amplifican adicionalmente por PCR en puente, o técnica semejante.

La llamada de base en las regiones NDN (particularmente de las cadenas de IgH) puede mejorarse usando la estructura de codones de las regiones J y V flanqueantes, como se ilustra en la Fig. 4C. (Tal y como se usa en la presente memoria, "estructura de codones" significa los codones del marco de lectura natural de los segmentos de transcritos de TCR o BCR o genes fuera de las regiones NDN, por ejemplo, la región V, la región J, o semejantes). Aquí, el amplicón (450), que es una visión ampliada del amplicón de la Fig. 4B, se muestra junto con las posiciones relativas de la lectura C (442) y lectura V adyacente (434) encima y las estructuras de codones (452 y 454) de la región V (430) y la región J (446), respectivamente, debajo. Según este aspecto de la invención, después de que las estructuras de codones (452 y 454) se identifiquen por alineamiento convencional con las secuencias de referencia de V y J, las bases en la región NDN (456) se llaman (o identifican) base por base moviendo desde la región J (446) hacia la región V (430) y en la dirección opuesta desde la región V (430) hacia la región J (446) usando lecturas de secuencia (434) y (442). En condiciones biológicas normales, sólo las secuencias de TCR o IgH recombinadas que tienen codones en marco de la región V a través de la región NDN y hasta la región J se expresan como proteínas. Esto es, de las variantes generadas somáticamente las únicas que

se expresan son aquellas cuyos marcos de codones de la región J y la región V están en marco entre sí y permanecen en marco a través de la región NDN. (Aquí, los marcos correctos de las regiones V y J se determinan a partir de las secuencias de referencia). Si se identifica una secuencia fuera de marco sobre la base de una o más llamadas de base de baja calidad, el clonotipo correspondiente está señalado para re-evaluación o como una anomalía relacionada con enfermedad potencial. Si la secuencia identificada está en marco y se basa en llamadas de base de alta calidad, entonces hay una mayor confianza de que el clonotipo correspondiente se haya llamado correctamente. De acuerdo con esto, en la presente memoria se describe un método para determinar clonotipos basados en V(D)J a partir de lecturas de secuencia bidireccionales que comprende las etapas de: (a) generar al menos una lectura de secuencia de la región J que empieza en una región J y se extiende en una región NDN y al menos una lectura de secuencia de la región V que empieza en las regiones V y se extiende hacia la región NDN tal que la lectura de secuencia de la región J y la lectura de secuencia de la región V se superponen en una región de superposición, y la región J y la región V tienen cada una una estructura de codones; (b) determinar si la estructura de codones de la región J extendida en la región NDN está en marco con la estructura de codones de la región V extendida hacia la región NDN. En un caso adicional, la etapa de generación incluye generar al menos una lectura de secuencia de la región V que empieza en la región V y se extiende a través de la región NDN hacia la región J, tal que la lectura de secuencia de la región J y la lectura de secuencia de la región V se superponen en una región de superposición.

Análisis de las lecturas de secuencia. Coalescencia de lecturas de secuencia en clonotipos. La construcción de clonotipos a partir de datos de lecturas de secuencia depende en parte del método de secuenciación usado para generar dichos datos, ya que los diferentes métodos tienen diferentes longitudes de lectura y calidad de datos esperadas. En una estrategia, se emplea un secuenciador Solexa para generar datos de lecturas de secuencia para 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 molde, lo que después de amplificación opcional puede producir poblaciones clonales correspondientes de un millón o más de moléculas molde (o agrupaciones). Para la mayor parte de las estrategias de secuenciación de alto rendimiento, incluyendo la estrategia Solexa, dicho sobre muestreo a nivel de agrupación es deseable de manera que cada secuencia de molde se determina con un gran grado de redundancia para incrementar la exactitud de la determinación de la secuencia. Para las implementaciones basadas en Solexa, preferiblemente la secuencia de cada molde independiente se determina 10 veces o más. Para otras estrategias de secuenciación con diferentes longitudes de secuencia y calidad de datos esperadas, pueden usarse diferentes niveles de redundancia para una exactitud comparable en la determinación de la secuencia. Los expertos en la técnica reconocen que los parámetros anteriores, por ejemplo, tamaño de la muestra, redundancia, y semejantes, son elecciones de diseño relacionadas con aplicaciones particulares.

La reducción de un conjunto de lecturas para una muestra dada en sus distintos clonotipos y el registro del número de lecturas para cada clonotipo sería un problema computacional trivial si la tecnología de secuenciación no tuviera errores. Sin embargo, en presencia de errores en la secuenciación, cada clonotipo está rodeado de una "nube" de lecturas con números variados de errores respecto a la secuencia verdadera del clonotipo. Cuanto mayor es el número de dichos errores menor es la densidad si la nube circundante, es decir, la densidad de la nube disminuye al movernos fuera del clonotipo en el espacio de la secuencia. Está disponible una variedad de algoritmos para convertir lecturas de secuencia en clonotipos. En un aspecto, la coalescencia de las lecturas de secuencia depende de tres factores: el número de secuencias obtenido para cada uno de los dos clonotipos de interés; el número de bases en el que se diferencian; y la calidad de la secuenciación en las posiciones en las que son discordantes. Se evalúa una proporción de probabilidad que se basa en las proporciones 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 ellos en un área de baja calidad de secuenciación coalescerán probablemente ya que probablemente se generen por error de secuenciación. Por otra parte, 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 improbable que se generen por error de secuenciación. En una realización de los métodos de la invención, el algoritmo descrito más adelante puede usarse para determinar clonotipos a partir de lecturas de secuencia.

Esta nube de lecturas que rodea a cada clonotipo puede modelarse usando distribución binomial y un modelo simple para la probabilidad de un único error de base. Este último modelo de error puede inferirse del mapeo de los segmentos V y J o del algoritmo de descubrimiento de clonotipo en sí mismo, mediante auto-consistencia y convergencia. Se construye un modelo para la probabilidad de una secuencia Y "nube" dada con cuenta de lectura C2 y errores E (respecto a la secuencia X) sea parte de una secuencia verdadera de clonotipo X con cuenta de lectura perfecta C1 bajo el modelo nulo de que X es el único clonotipo verdadero en esta región de espacio de secuencia. Se toma una decisión de si coalescer o no la secuencia Y en el clonotipo X según los parámetros C1, C2, y E. Para cualquier C1 y E dados, se precalcula un valor C2 máximo para decidir coalescer la secuencia Y. Los valores máximos para C2 se eligen de manera que la probabilidad de fracaso en la coalescencia de Y bajo la hipótesis nula que Y es parte del clonotipo X es menor de algún valor P después de integrar sobre todas las posibles secuencias Y con 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.

Si una secuencia Y no coalesce en el clonotipo X porque su cuenta de lectura está por encima del umbral C2 para coalescer en el clonotipo X entonces se convierte en un candidato para sembrar clonotipos



deparados. El algoritmo también se asegura de que cualesquiera otras secuencias Y2, Y3, etc. que están "más cerca" de esta secuencia Y (que se ha considerado independiente de X) no se agreguen en X. Este concepto de "ceranía" incluye tanto cuentas de error respecto a Y y X como la cuenta de lectura absoluta 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 "nube" pueden atribuirse apropiadamente a su clonotipo correcto si aparecen "cerca" de más de un clonotipo.

El algoritmo procede de manera arriba a abajo empezando con la secuencia X con la cuenta de lectura más alta. Esta secuencia siembra el primer clonotipo. Las secuencias vecinas se colaescen en este clonotipo si sus cuentas están por debajo de los umbrales precalculados (véase anteriormente), o se dejan solas si están por encima del umbral o "más cerca" de otra secuencia que no ha coalescido. Después de buscar todas las secuencias vecinas en una cuenta de error máxima, termina el proceso de coalescer lecturas en el clonotipo X. Sus lecturas y todas las lecturas que se han coalescido en él se tienen en cuenta y retiran de la lista de lecturas disponible para hacer otros clonotipos. La siguiente secuencia se mueve entonces con la cuenta de lectura más alta. Las lecturas vecinas se coalescen en este clonotipo como anteriormente y este proceso se continúa hasta que no hay más secuencias con cuentas de lectura por encima de un umbral dado, por ejemplo, hasta que todas las secuencias con más de 1 cuenta se han usado como semillas para clonotipos.

En otra realización del algoritmo anterior, puede añadirse un ensayo adicional para determinar si coalescer una secuencia candidata Y en un clonotipo X existente, que tiene en cuenta la puntuación de calidad de las lecturas de secuencia relevantes. La o las puntuaciones de calidad medias se determinan para la o las secuencias Y (promediada a lo largo de todas las lecturas con secuencia Y) en la que las secuencias Y y X son diferentes. Si la puntuación media 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 la puntuación media está por debajo de dicho valor predeterminado entonces es más probable que la secuencia Y esté causada por errores de secuenciación y por lo tanto debería coalescer en X.

Árbol de secuencia. El algoritmo anterior de lecturas coalescentes en clonotipos depende de tener una manera eficiente de encontrar todas las secuencias con menos de errores E a partir de la misma secuencia de entrada X. Este problema se resuelve usando un árbol de secuencia. La implementación de este árbol tiene algunas características no habituales ya que los nodos del árbol no están restringidos a ser simples letras de ADN. Los nodos pueden tener secuencias arbitrariamente largas. Esto permite un uso más eficiente de la memoria del ordenador.

Todas las lecturas de una muestra dada se ponen en el árbol de secuencia. Cada nodo de la hoja contiene punteros para sus lecturas asociadas. Corresponde a una única secuencia proporcionada atravesando hacia atrás el árbol desde la hoja hasta el nodo de la raíz. La primera secuencia se pone en un árbol simple con un nodo de raíz y un nodo de hoja que contiene la secuencia completa de la lectura. A continuación, se añaden secuencias una en una. Para cada secuencia añadida, se forma bien una nueva rama en el último punto de la secuencia común entre la lectura y el árbol existente o se añade la lectura a un nodo de hoja existente su el árbol ya contiene la secuencia.

Habiendo puesto todas las lecturas en el árbol, es fácil usar el árbol para los propósitos siguientes: 1. Cuenta de lectura más alta: la separación de los nodos de hoja por cuenta de lectura nos permite encontrar el nodo de hoja (es decir, secuencia) con más lecturas. 2. Descubrimiento de hojas vecinas: para cualquier secuencia se pueden buscar todos los caminos a lo largo del árbol que tienen menos de errores X respecto a esta secuencia. Un camino empieza en la raíz y este camino se ramifica en caminos separados a lo largo del árbol. Se anota la cuenta de error actual de cada camino al proceder a lo largo del árbol. Cuando la cuenta de error supera los errores máximos permitidos el camino dado se termina. De esta manera, se recortan grandes partes del árbol tan pronto como es posible. Ésta es una manera eficiente de encontrar todos los caminos (es decir, todas las hojas) en errores X de cualquier secuencia dada.

Hipermutaciones somáticas. En una realización, los clonotipos basados en IgH que han experimentado hipermutación somática se determinan como sigue. Una mutación somática se define como una base secuenciada que es diferente de la base correspondiente de una secuencia de referencia (del segmento relevante, habitualmente V, J o C) y que está presente en un número estadísticamente significativo de lecturas. En una realización, las lecturas de C pueden usarse para encontrar mutaciones somáticas respecto al segmento J mapeado y asimismo lecturas de V para el segmento V. Sólo se usan partes de las lecturas C y V que bien se mapearon directamente a segmentos J o V o que estaban dentro de la extensión de clonotipo hasta el límite de NDN. De esta manera, la región NDN se evita y la misma "información de secuencia" no se usa para el descubrimiento de mutaciones que se usó previamente para la determinación de clonotipos (para evitar clasificar erróneamente como mutaciones nucleótidos que son realmente sólo diferentes regiones NDN recombinadas). Para cada tipo de segmento, el segmento mapeado (alelo principal) se usa como un soporte y se considera que todas las lecturas que se han mapeado en este alelo durante la fase de mapeo de lectura. Cada posición de las secuencias de referencia en la que al menos una lectura se ha mapeado se analiza para mutaciones somáticas. En una realización, los criterios para aceptar una base no de referencia como mutación válida incluyen los siguientes: 1) al menos N lecturas con la base de mutación dada, 2) al menos una fracción dada de lecturas N/M (en la que M es el número total de lecturas mapeadas en esta posición de base) y 3) un corte estadístico basado en la distribución binomial, la puntuación Q media de las lecturas N en la base de

mutación así como el número (M-N) de lecturas con una base no de mutación. Preferiblemente, los parámetros anteriores se seleccionan de manera que la proporción de descubrimientos falsos de mutaciones por clonotipo es menos de 1 en 1.000, y más preferiblemente, menos de 1 en 10.000.

5 Clonotipos filogénicos (clanes). En algunas enfermedades, tales como cánceres, incluyendo trastornos proliferativos linfocitos, un único progenitor de linfocito puede dar lugar a mucha progenie de linfocitos relacionada, cada una poseyendo y/o expresando un TCR o BCR ligeramente diferente, y por lo tanto un clonotipo diferente, debido a hipermutación somática en curso o a mutación o mutaciones somáticas relacionadas con la enfermedad, tales como sustituciones de bases, reorganizaciones aberrantes, o semejantes. Las células que producen dichos clonotipos se refieren en la presente memoria como clones filogénicos, y un conjunto de dichos clones relacionados se refieren en la presente memoria como un "clan". Asimismo, los clonotipos de clones filogénicos se refieren como clonotipos filogénicos y un conjunto de clonotipos filogénicos puede referirse como un clan de clonotipos. En un aspecto, los métodos de la invención comprenden monitorizar la frecuencia de un clan de clonotipos (es decir, la suma de frecuencias de los clonotipos filogénicos constituyentes del clan), en lugar de una frecuencia de un clonotipo individual. (La expresión "uno o más clonotipos específicos de paciente" engloba el concepto de clanes). Los clonotipos filogénicos pueden identificarse por una o más medidas de existencia de relación con un clonotipo parental. En una realización, los clonotipos filogénicos pueden agruparse en el mismo clan por porcentaje de homología, como se describe más completamente más adelante. En otra realización, los clonotipos filogénicos se identifican por el uso común de regiones V, regiones J, y/o regiones NDN. Por ejemplo, un clan puede definirse por clonotipos que tienen regiones J y ND comunes pero diferentes regiones V (algunas veces referido como "reemplazo VH"); o pueden definirse por clonotipos que tienen las mismas regiones V y J (mutadas de forma idéntica por sustituciones de bases de sus secuencias de referencia respectivas) pero con diferentes regiones NDN; o pueden definirse por un clonotipo que ha experimentado una o más inserciones y/o deleciones de 1-10 bases, o de 1-5 bases, o de 1-3 bases, para generar miembros del clan. En otra realización, los clonotipos se asignan a algún clan si satisfacen los criterios siguientes: i) se mapean en los mismos segmentos V y J de referencia, con los mapeos ocurriendo en las mismas posiciones relativas en la secuencia del clonotipo, y ii) sus regiones NDN son sustancialmente idénticas. "Sustancial" en referencia a los miembros de un clan significa que se permiten algunas pequeñas diferencias en la región NDN porque se pueden haber producido mutaciones somáticas en esta región. Preferiblemente, en una realización, para evitar llamar de manera falsa a una mutación en la región NDN, si una sustitución de base se acepta como una mutación relacionada con el cáncer depende directamente del tamaño de la región NDN del clan. Por ejemplo, un método puede aceptar un clonotipo como un miembro del clan si tiene una diferencia de una base de la o las secuencias de NDN del clan como una mutación relacionada con cáncer si la longitud de la o las secuencias de NDN del clan es m nucleótidos o mayor, por ejemplo, 9 nucleótidos o mayor, de otra manera no se acepta, o si tiene una diferencia de dos bases de la o las secuencias de NDN del clan como mutaciones relacionadas con cáncer si la longitud de la o las secuencias de NDN del clan es n nucleótidos o mayor, por ejemplo, 20 nucleótidos o mayor, de otra manera no se acepta. En otra realización, los miembros de un clan se determinan usando los criterios siguientes: (a) la lectura V se mapea en la misma región V, (b) la lectura C se mapea en la misma región J, (c) la región NDN es sustancialmente idéntica (como se ha descrito anteriormente), y (d) la posición de la región NDN entre el límite V-NDN y el límite J-NDN es la misma (o de forma equivalente, el número de adiciones de bases aguas abajo a D y el número de adiciones de bases aguas arriba a D son las mismas). Tal y como se usa en la presente memoria, el término "lectura de C" puede referirse a una lectura generada a partir de un cebador de secuenciación que hibrida bien con una región C (en el caso de usar una muestra de ARN) o con una región J (en el caso de usar una muestra de ADN). Como se explica en otro lugar, esto es porque una región C se une con una región J en un proceso de corte y empalme posterior a la transcripción.

Los clonotipos filogénicos de una única muestra pueden agruparse en clanes y los clanes de muestras sucesivas adquiridas a tiempos diferentes pueden compararse entre sí. En particular, en un aspecto de la invención, los clanes que contienen clonotipos correlacionados con un neoplasma linfocito se identifican entre los clonotipos determinados de cada muestra en cada punto de tiempo. El conjunto (o clan) de clonotipos que se correlacionan de cada punto de tiempo se compara con el de la muestra inmediatamente anterior para determinar el estado de la enfermedad, por ejemplo, determinando en clanes sucesivos si una frecuencia de un clonotipo particular se incrementa o disminuye, si aparece un nuevo clonotipo que se correlaciona que se sabe que se correlaciona a partir de estudios poblacionales o bases de datos, o semejantes. Un estado determinado podría ser remisión continua, recidiva incipiente, evidencia de evolución clonal adicional, o semejante.

Uso de isotipos. Los perfiles de clonotipos pueden incluir información sobre el uso de isotipos. Siempre que se determinan los clonotipos basados en IgH o TCR $\beta$  a partir de ARN, el corte y empalme posterior a la transcripción une las regiones C a las regiones J, como se ilustra en la Fig. 3B. Los cebadores de secuenciación pueden usarse para generar lecturas de C (por ejemplo, 304) que hibridan con un sitio de unión de cebador predeterminado (302) en la región C (307) en la unión con la región J (309). Si el sitio de unión de cebador (302) se selecciona de manera que la lectura de C (304) incluye una parte (305) de la región C (307), entonces puede determinarse la identidad de la región C (307), que, a su vez, permite determinar el isotipo del BCR sintetizado. En una realización, el sitio de unión del cebador (302) se selecciona de manera que la lectura de C (304) incluye al menos seis nucleótidos de la región C (307); en otra realización, el sitio de unión del cebador (302) se selecciona de manera que la lectura de C (304)

incluye al menos 8 nucleótidos de la región C (307). Cada clonotipo determinado según esta realización incluye información de secuencia de parte (305) de su región C correspondiente y a partir de dicha información de secuencia se determina su isotipo correspondiente. Los clonotipos que se correlacionan pueden tener un primer isotipo en el momento en el que se determinan inicialmente, pero pueden cambiar a otro tipo de isotipo durante el tiempo en el que están siendo monitorizados. Esta realización es capaz de detectar dichos cambios percibiendo clonotipos previamente no registrados que tienen secuencias idénticas a los clonotipos que se correlacionan, excepto para la secuencia de parte (305) que corresponde a un isotipo diferente.

Se espera que el error de la PCR se concentre en algunas bases que estaban mutadas en los ciclos tempranos de PCR. Se espera que el error de secuenciación esté distribuido en muchas bases aunque es totalmente aleatorio ya que el error probablemente tenga algún sesgo sistemático. Se asume que algunas bases tendrán error de secuenciación en una proporción mayor, por ejemplo 5% (5 veces la media). Dadas estas asunciones, el error de secuenciación se vuelve el tipo dominante de error. El distinguir los errores de PCR de la aparición de clonotipos altamente relacionados jugará un papel en el análisis. Dada la significancia biológica para determinar que hay dos o más clonotipos altamente relacionados, se toma una estrategia conservadora para hacer dichas llamadas. Se considera la detección de bastantes de los clonotipos menores de manera que se asegura con una confianza alta (por ejemplo, 99,9%) que hay más de un clonotipo. Por ejemplo, de los clonotipos que están presentes a 100 copias/1.000.000, la variante menor se detecta 14 o más veces para designarla como un clonotipo independiente. De forma similar, para los clonotipos presentes a 1.000 copias/1.000.000 la variante menor puede detectarse 74 o más veces para designarla como un clonotipo independiente. Este algoritmo puede potenciarse usando la puntuación de calidad de base que se obtiene con cada base secuenciada. Si la relación entre la puntuación de calidad y la proporción de error se valida por encima, entonces en lugar de emplear la proporción de error conservativa del 5% para todas las bases, la puntuación de calidad puede usarse para decidir el número de lecturas que es necesario que estén presentes para llamar un clonotipo independiente. La puntuación de calidad mediana de la base específica en todas las lecturas puede usarse, o de forma más rigurosa, la probabilidad de siendo un error pueda computarse dada la puntuación de calidad de la base específica en cada lectura, y entonces las probabilidades pueden combinarse (asumiendo independencia) para estimar el número probable de error de secuenciación para esa base. Como resultado, hay diferentes umbrales para rechazar la hipótesis del error de secuenciación para diferentes bases con diferentes puntuaciones de calidad. Por ejemplo, para un clonotipo presente a 1.000 copias por 1.000.000, la variante menor se designa independiente cuando se detecta 22 y 74 veces si la probabilidad de error fuera 0,01 y 0,05, respectivamente.

### III. Clonotipos que se correlacionan y algoritmos médicos

En la presente memoria se describen métodos para identificar clonotipos cuya presencia, ausencia y/o nivel está correlacionado con un estado de enfermedad y para usar dicha información para tomar decisiones de diagnóstico o pronóstico. En un método, la información de perfiles de clonotipos, que puede acoplarse con otra información médica, tal como niveles de expresión de genes no TCR o no BCR, condición fisiológica, o semejantes, se presenta a los pacientes o responsables sanitarios en el contexto de un algoritmo; esto es, un conjunto de una o más etapas en el que se evalúan los resultados de ensayos y/o exámenes y (i) bien se determina un curso de acción o se toma una decisión respecto al estado de salud o enfermedad o (ii) se toma una serie de decisiones según un diagrama de flujo, o estructura de toma de decisión semejante, que da lugar a un curso de acción, o una decisión respecto al estado de salud o enfermedad. Los algoritmos descritos en la presente memoria pueden variar ampliamente en formato. Por ejemplo, un algoritmo puede sugerir simplemente que un paciente debe tratarse con un fármaco, si un determinado clonotipo, o subconjunto de clonotipos, supera una proporción predeterminada en un perfil de clonotipos, o se incrementa en proporción en más de una proporción predeterminada entre las medidas de monitorización. Incluso de forma más simple, un algoritmo puede indicar meramente que existe una correlación positiva entre un estado de enfermedad y un nivel de uno o más clonotipos y/o una función de TCR o BCR codificada por uno o más clonotipos. Los algoritmos más complejos pueden incluir información fisiológica del paciente además de información a partir de uno o más perfiles de clonotipos. Por ejemplo, en trastornos complejos, tales como algunos trastornos autoinmunes, la información del perfil de clonotipos puede combinarse en un algoritmo con otros datos del paciente tales como curso de tratamiento anterior, presencia, ausencia o intensidad de los síntomas, por ejemplo, erupción, inflamación articular, expresión de genes particulares, o semejantes. En un caso, un algoritmo para uso con la monitorización de trastornos linfoides proporciona un valor fraccionado predeterminado por encima del cual la proporción de un clonotipo (y/o clonotipos relacionados evolutivamente) en un perfil de clonotipos de una muestra (tal como una muestra de sangre) indica una recidiva de la enfermedad o una resistencia a un tratamiento. Dichos algoritmos pueden consistir en o incluir medidas convencionales de clonalidad de TCR o BCR. En otro caso, un algoritmo para uso con la monitorización de trastornos autoinmunes proporciona uno o más valores fraccionados predeterminados por encima de los cuales una proporción de clonotipos en un perfil de clonotipos que codifica TCR o BCR específicos para uno o más antígenos predeterminados, respectivamente, indica el inicio de una crisis autoinmune.

#### A. Clonotipos que se correlacionan frente a los que no se correlacionan

Los métodos descritos en la presente memoria proporcionan medios para distinguir a) clonotipos que se correlacionan (que pueden ser aquellos clonotipos cuyo nivel se correlaciona con enfermedad) de b) clonotipos que no se correlacionan (que pueden ser aquellos clonotipos cuyos niveles no se correlacionan

con enfermedad). En un caso, un clonotipo que se correlaciona puede presentar una correlación bien positiva o negativa con la enfermedad. En otro caso, un clonotipo presente en un estado de pico de una enfermedad pero no presente en un estado no de pico de una enfermedad puede ser un clonotipo que se correlaciona (correlación positiva con enfermedad). En otro caso, un clonotipo que es más abundante (es decir, está presente a un nivel mayor de moléculas) en un estado de pico (o estadio) de una enfermedad que en un estado no de pico de la enfermedad puede ser un clonotipo que se correlaciona (correlación positiva con la enfermedad). En otro caso, un clonotipo ausente en el estado de pico de una enfermedad pero presente durante un estado no de pico de la enfermedad puede ser un clonotipo que se correlaciona (correlación negativa con la enfermedad). En otro caso, un clonotipo que es menos abundante en un estado de pico de una enfermedad que en un estado no de pico de una enfermedad puede ser un clonotipo que se correlaciona (correlación negativa con la enfermedad). En otro caso, un clonotipo que se correlaciona para un individuo se determina por un algoritmo.

B. Descubrimiento de clonotipos que se correlacionan y no se correlacionan usando un ensayo de calibración sin un estudio poblacional.

Los clonotipos que se correlacionan pueden identificarse observando los clonotipos presentes en alguna muestra que tiene relevancia para un estado de enfermedad. Esta muestra podría ser sangre de una muestra en un estado de pico de una enfermedad (por ejemplo, una muestra de sangre de un paciente con MS o lupus durante una crisis aguda), o podría ser de un tejido afectado por la enfermedad, o relacionado con la enfermedad, que está enriquecido para células T y B implicadas en la enfermedad para ese individuo, tal como una inflamación o tumor. Los ejemplos de estos tejidos podrían ser biopsias de riñón de pacientes con lupus con inflamación de riñón, fluido cerebroespinal (CSF) en pacientes con MS durante una crisis, fluido sinovial para pacientes con artritis reumatoide, muestras de tumor de pacientes con cáncer. En todos estos ejemplos, es probable que los tejidos contendrán células T y B relevantes que están relacionadas con la enfermedad (aunque no necesariamente los agentes causales). Es importante que si este método se usa para identificar los clonotipos que son relevantes para la enfermedad, éstos sólo serán relevantes para el individuo en cuya muestra se detectaron. Como resultado, se necesita un ensayo de calibración específico con el fin de usar este método para identificar clonotipos que se correlacionan en cualquier individuo dado con una enfermedad. Esto es, los clonotipos que se correlacionan pueden descubrirse o determinarse generando un perfil de clonotipos de una muestra tomada de un tejido afectado directamente por, o relevante para, una enfermedad (algunas veces referido en la presente memoria como "tejido relacionado con enfermedad"). Dicha determinación puede incluir además generar un perfil de clonotipos de una muestra tomada de un tejido no afectado por, o no relevante para, una enfermedad (algunas veces referido en la presente memoria como "tejido no relacionado con enfermedad"), después comparar los primeros y últimos perfiles de clonotipos para identificar clonotipos que se correlacionan como aquellos que están a un nivel alto, nivel bajo o que son funcionalmente distintos, por ejemplo, codifican TCR o BCR específicos para un antígeno particular. Dicha determinación puede hacerse identificando clonotipos presentes en un perfil de clonotipos de un tejido afectado, o relacionado con enfermedad, a una frecuencia mayor que los mismos clonotipos en un perfil de clonotipos de tejido no afectado, o no relacionado con enfermedad.

Se describe un método para determinar uno o más clonotipos que se correlacionan en un sujeto. El método puede incluir etapas para a) generar uno o más perfiles de clonotipos por secuenciación de ácido nucleico de moléculas individuales aisladas espacialmente de al menos una muestra del sujeto, en el que la al menos una muestra está relacionada con un primer estado de la enfermedad, y b) determinar uno o más clonotipos que se correlacionan en el sujeto sobre la base del uno o más perfiles de clonotipos.

En un caso, al menos una muestra es de un tejido afectado por la enfermedad. En otro caso, dicha determinación de uno o más clonotipos que se correlacionan comprende comparar los perfiles de clonotipos de al menos dos muestras. En otro caso, el primer estado de la enfermedad es un estado de pico de la enfermedad. En otro caso, uno más clonotipos que se correlacionan están presentes en el estado de pico de la enfermedad. En otro caso, el uno más clonotipos que se correlacionan están ausentes en el estado de pico de la enfermedad. En otro caso, uno más clonotipos que se correlacionan están altos en el estado de pico de la enfermedad. En otro caso, uno más clonotipos que se correlacionan están bajos en el estado de pico de la enfermedad. En otro caso, la muestra comprende células T y/o células B. En otro caso, las células T y/o células B comprenden un subconjunto de células T y/o células B. En otro caso, el subconjunto de células T y/o células B está enriquecido por interacción con un marcador. En otro caso, el marcador es un marcador de la superficie celular en el subconjunto de células T y/o células B. En otro caso, el subconjunto de células T y/o células B interactúa con un antígeno presente específicamente en la enfermedad. Por ejemplo, en el caso de trastornos linfoproliferativos, tales como linfomas, puede obtenerse una muestra de calibración de tejidos linfoides, de lesiones causadas por el trastorno, por ejemplo, lesiones metastásicas, o de tejidos afectados indirectamente por el trastorno por enriquecimiento como se ha sugerido anteriormente. Para los neoplasmas linfoides, existe una guía ampliamente disponible y kits disponibles comercialmente para el inmunofenotipado y enriquecimiento de linfocitos relacionados con la enfermedad, por ejemplo, "U.S.-Canadian consensus recommendations on the immunophenotypic analysis of haematologic neoplasia by flow cytometry", *Cytometry*, 30: 214-263 (1997); paneles de anticuerpos Multmix™ para el inmunofenotipado de leucemia y linfoma por citometría de flujo (Dako, Dinamarca); y semejantes. Los tejidos linfoides incluyen nódulos linfoides, bazo, amígdalas, adenoides, timo, y semejantes.

En un caso, la enfermedad es una enfermedad autoinmune. En otro caso, la enfermedad autoinmune es

lupus eritematoso sistémico, esclerosis múltiple, artritis reumatoide, o espondilitis anquilosante.

En algunos casos, los clonotipos que se correlacionan se identifican observando los clonotipos presentes en alguna muestra que tiene relevancia para un estado distinto de un estado de enfermedad. Estos estados podrían incluir exposición a antígenos no causantes de enfermedad, tales como reacciones alérgicas sub-sintomáticas a pólenes locales. Esto podría usarse para identificar si un individuo ha vuelto recientemente a una localización geográfica que contenía el antígeno. Los estados podrían incluir exposición a un antígeno relacionado con un proceso industrial o la fabricación o producción de agentes de bioterrorismo.

C. Descubrimiento de clonotipos que se correlacionan y no se correlacionan usando un estudio poblacional.

Se describe un método para identificar clonotipos que se correlacionan usando un estudio poblacional. La utilidad del estudio poblacional es que permite generalizar la información específica acerca de clonotipos que se correlacionan que se han verificado en individuos con resultados de estado de enfermedad conocido para permitir que dichos clonotipos que se correlacionan se identifiquen en todos los sujetos futuros sin la necesidad de un ensayo de calibración. El conocimiento de un conjunto específico de clonotipos que se correlacionan puede usarse para extraer reglas acerca de los atributos probables (parámetros) de clonotipos que se correlacionarán en sujetos futuros. Dicho método se implementa con las etapas siguientes: (a) generar perfiles de clonotipos para cada uno de un conjunto de muestras de tejidos afectados por, o relevantes para, una enfermedad; (b) determinar los clonotipos que están a un nivel alto o nivel bajo respecto a los mismos clonotipos en muestras de tejidos no afectados o que son funcionalmente distintos de los clonotipos en muestras de tejidos no afectados. Tal y como se usa en la presente memoria, en un aspecto, "funcionalmente distinto" en referencia a clonotipos significa que TCR o BCR codificados por uno son específicos para un antígeno diferente, proteína o complejo que el otro. Opcionalmente, el método anterior puede incluir además una etapa de desarrollo de un algoritmo para predecir clonotipos que se correlacionan en cualquier muestra a partir de la información de secuencia de los clonotipos determinada en las etapas anteriores (a) y/o (b) o a partir de los datos funcionales, es decir, una determinación de que los clonotipos recién medidos codifican TCR o BCR específicos para un antígeno, proteína o complejo específico para la enfermedad en observación.

En conexión con lo anterior, uno o más clonotipos específicos de paciente pueden identificarse concordando los clonotipos determinados en una o más medidas iniciales ("clonotipos determinados") con clonotipos que se sabe que se correlacionan con dicha enfermedad, que pueden estar disponibles a través de un estudio poblacional, base de datos, o semejante. En un caso, el concordar dichos clonotipos comprende encontrar identidad entre una secuencia de aminoácidos codificada por el clonotipo determinado y la de una secuencia de aminoácidos codificada por un clonotipo que se sabe que está correlacionado con la enfermedad, o una variante sustancialmente idéntica del último clonotipo. Tal y como se usa en la presente memoria, "variante sustancialmente idéntica", en un caso, significa que las secuencias que se están comparando o concordando son al menos 80 por ciento idénticas, o al menos 90 por ciento idénticas, ya sea una secuencia de ácido nucleico o una secuencia de aminoácidos. En otro caso, variante sustancialmente idéntica significa que se diferencia por 5 o menos adiciones, deleciones y/o sustituciones de bases o aminoácidos. En otro caso, la concordancia de dichos clonotipos comprende encontrar identidad entre el clonotipo determinado y una secuencia de ácido nucleico de un clonotipo que se sabe que está correlacionado con la enfermedad, o una variante sustancialmente idéntica del último clonotipo. En otro caso más, la concordancia de dichos clonotipos comprende encontrar identidad entre el clonotipo determinado y una secuencia de ácido nucleico de un clonotipo que se sabe que está correlacionado con la enfermedad, o una variante sustancialmente idéntica del último clonotipo.

En la presente memoria se describen métodos que incluyen identificar clonotipos que se correlacionan y no se correlacionan secuenciando el repertorio de células inmunes en un estudio de muestras de pacientes con enfermedad o enfermedades y opcionalmente controles sanos a diferentes tiempos y, en el caso de los pacientes con una enfermedad, a diferentes (y conocidos) estados del curso de la enfermedad caracterizados por datos clínicos. La enfermedad puede ser, por ejemplo, una enfermedad autoinmune. Los clonotipos cuyo nivel se correlaciona con medidas de la enfermedad en estos diferentes estados pueden usarse para desarrollar un algoritmo que prediga la identidad de un conjunto mayor de secuencias que se correlacionarán con la enfermedad como distintas de aquellas que no se correlacionarán con la enfermedad en todos los individuos. A diferencia del caso del ensayo de calibración, no es necesario que las secuencias que se correlacionan hayan estado presentes en el estudio de descubrimiento pero pueden predecirse sobre la base de estas secuencias. Por ejemplo, una secuencia que se correlaciona puede ser una secuencia de ADN del gen de TCR que codifica la misma secuencia de aminoácidos que la secuencia de ADN de un clonotipo identificado en el estudio de descubrimiento. Además, el algoritmo que puede predecir uno o más clonotipos que se correlacionan puede usarse para identificar clonotipos en una muestra de cualquier individuo y no es de ninguna manera único para un individuo dado, permitiendo así predecir los clonotipos que se correlacionan en una muestra nueva sin conocimiento previo de los clonotipos presentes en ese individuo.

Se describe un método para desarrollar un algoritmo que predice uno o más clonotipos que se correlacionan en cualquier muestra de un sujeto con una enfermedad que comprende: a) generar una pluralidad de perfiles de clonotipos a partir de un conjunto de muestras, en el que las muestras son relevantes para la enfermedad, b) identificar uno o más clonotipos que se correlacionan del conjunto de muestras, c) usar parámetros de secuencia y/o datos funcionales de uno o más clonotipos que se

correlacionan identificados en b) para desarrollar un algoritmo que pueda predecir clonotipos que se correlacionan en cualquier muestra de un sujeto con la enfermedad.

En un caso, el conjunto de muestras se toma de uno o más tejidos afectados por la enfermedad.

En otro caso, la identificación de uno o más clonotipos que se correlacionan comprende comparar los perfiles de clonotipos de al menos dos muestras. En otro caso, los datos funcionales incluyen la capacidad de unión de los marcadores en las células T y/o células B o interacción con antígeno por una célula T o célula B. En otro caso, dichos parámetros de secuencia comprenden secuencia de ácido nucleico y secuencia de aminoácidos predicha. En otro caso, las muestras son de uno o más individuos en un estado de pico de la enfermedad. En otro caso, dichos uno o más clonotipos que se correlacionan están presentes en el estado de pico de la enfermedad. En otro caso, dichos uno o más clonotipos que se correlacionan están a un nivel alto en el estado de pico de la enfermedad. En otro caso, uno o más clonotipos que se correlacionan están a un nivel bajo en el estado de pico de la enfermedad. En otro caso, uno o más clonotipos que se correlacionan están ausentes en el estado de pico de la enfermedad.

En un caso, la enfermedad es una enfermedad autoinmune. En otro caso, la enfermedad autoinmune es lupus eritematoso sistémico, esclerosis múltiple, artritis reumatoide, o espondilitis anquilosante.

Se describe un método para descubrir uno o más clonotipos que se correlacionan para un individuo, que comprende a) introducir un perfil de clonotipos de una muestra del individuo en un algoritmo, y b) usar el algoritmo para determinar uno o más clonotipos que se correlacionan para el individuo. El algoritmo puede ser un algoritmo desarrollado por: a) generación de una pluralidad de perfiles de clonotipos de un conjunto de muestras, en el que las muestras son relevantes para la enfermedad, b) identificación de uno o más clonotipos que se correlacionan del conjunto de muestras, y c) uso de parámetros de secuencia y/o datos funcionales de uno o más clonotipos que se correlacionan identificados en b) para desarrollar el algoritmo que puede predecir clonotipos que se correlacionan en cualquier muestra de un sujeto con la enfermedad.

En algunos casos, los clonotipos que se correlacionan son clonotipos identificados presentes en poblaciones que se han expuesto a un antígeno que tiene relevancia para un estado distinto de un estado de enfermedad. Este estado podría incluir exposición a antígenos no causantes de enfermedad, tales como reacciones alérgicas sub-sintomáticas a pólenes locales. Esto podría usarse para identificar si un individuo ha viajado recientemente a una localización geográfica que contenía el antígeno. Los estados podrían incluir exposición a un antígeno relacionado con un proceso industrial o la fabricación o producción de agentes de bioterrorismo.

D. Descubrimiento de clonotipos que se correlacionan y no se correlacionan usando un ensayo de calibración combinado con un estudio poblacional.

Los clonotipos que se correlacionan pueden identificarse usando un ensayo de calibración combinado con un estudio poblacional. El estudio poblacional no resulta en un algoritmo que permita predecir los clonotipos en ninguna muestra sino que permite desarrollar un algoritmo para predecir clonotipos que se correlacionan en cualquier muestra de un sujeto para el que se ha generado un perfil de clonotipos de calibración particular. Un ejemplo de esto podría ser el desarrollo de un algoritmo que predeciría los clonotipos que se correlacionan en un paciente con lupus sobre la base del perfil de clonotipos medido de una muestra de sangre en cualquier estadio de la enfermedad después de haber sido sometido en primer lugar a un ensayo de sangre durante un estado de crisis clínica que se usó para calibrar el algoritmo. Así, los clonotipos que se correlacionan pueden identificarse en etapas: (a) generar perfiles de clonotipos de un conjunto de muestras de tejidos relevantes para o afectados por una enfermedad para identificar un conjunto de clonotipos asociados con la enfermedad bien por nivel y/o por función y para identificar una relación entre dicho nivel y/o función y el estado de la enfermedad; (b) medir un perfil de clonotipos de una muestra de un tejido de un primer estado de la enfermedad; (c) determinar un clonotipo que se correlaciona a partir de la relación de la etapa (a). Los clonotipos que se correlacionan pueden identificarse en etapas: (a) generar perfiles de clonotipos de un conjunto de muestras de tejidos relevantes para o afectados por una enfermedad para identificar un conjunto de clonotipos asociados con la enfermedad bien por nivel y/o por función y para identificar una relación entre dicho nivel y/o función y el estado de la enfermedad; (b) medir un perfil de clonotipos de calibración en un nuevo sujeto en un estadio relevante de la enfermedad en un estadio de pico o de tejido afectado por la enfermedad o en un estado caracterizado funcionalmente; (c) determinar un clonotipo que se correlaciona a partir de la relación de la etapa (a).

En la presente memoria se describen métodos para identificar clonotipos que se correlacionan y no se correlacionan secuenciando el repertorio de células inmunes en un estudio de muestras de pacientes de enfermedad o enfermedades y opcionalmente controles sanos a diferentes tiempos y, en el caso de los pacientes con una enfermedad, a diferentes (y conocidos) estados del curso de la enfermedad caracterizado por datos clínicos. Los clonotipos que se encuentran a diferente frecuencia (o nivel) en el primer estado que en el segundo estado se usan entonces para desarrollar un algoritmo que predice cuáles de las secuencias encontradas en los repertorios de cada individuo en el primer estado de la enfermedad se correlacionarán con la enfermedad en el estado posterior en cada individuo que es distinto de aquellos que no se correlacionarán con la enfermedad en ese individuo. A diferencia del caso del ensayo de calibración solo, las secuencias que se correlacionan pueden ser un subconjunto de todas las secuencias que se ha encontrado que son diferentes entre los estados de la enfermedad. También es posible que los clonotipos que se correlacionan no se encuentren en la muestra de calibración pero se predice sobre la base del algoritmo que se correlacionan si aparecen en una muestra futura. Como un ejemplo, un clonotipo que codifica la misma secuencia de aminoácidos que un clonotipo encontrado en

una muestra de calibración puede predecirse que es un clonotipo que se correlaciona sobre la base del algoritmo que resulta del estudio poblacional. A diferencia de las realizaciones previas, el algoritmo se desarrolla para predecir los clonotipos que se correlacionan sobre la base de un perfil de clonotipos de calibración que es el perfil de clonotipos generado en el individuo para el que se van a predecir los clonotipos que se correlacionan en un estado específico de la enfermedad. En este caso, el algoritmo no puede usarse para generar clonotipos que se correlacionan en un individuo particular hasta que se ha medido un perfil de clonotipos de calibración específico. Después de haber medido este perfil de calibración en un sujeto particular, todos los clonotipos que se correlacionan posteriores pueden predecirse sobre la base de la medida de los perfiles de clonotipos en ese individuo.

Se describe un método para descubrir uno o más clonotipos que se correlacionan para un individuo, que comprende a) introducir un perfil de clonotipos de una muestra del individuo en un algoritmo, y b) usar el algoritmo para determinar uno o más clonotipos que se correlacionan para el individuo. El algoritmo puede ser un algoritmo desarrollado por: a) generación de una pluralidad de perfiles de clonotipos de un conjunto de muestras, en el que las muestras son relevantes para la enfermedad, b) identificación de uno o más clonotipos que se correlacionan del conjunto de muestras, y c) uso de parámetros de secuencia y/o datos funcionales de uno o más clonotipos que se correlacionan identificados en b) para desarrollar un algoritmo que puede predecir clonotipos que se correlacionan en cualquier muestra de un sujeto con la enfermedad. En un caso, la muestra se toma en un estado de pico de la enfermedad. En otro caso, la muestra se toma de tejido afectado por la enfermedad.

En algunos casos, los clonotipos que se correlacionan y no se correlacionan usando un ensayo de calibración combinado con un estudio poblacional se realiza para clonotipos presentes en poblaciones que se han expuesto a un antígeno que tiene relevancia para un estado distinto de un estado de enfermedad. Este estado podría incluir la exposición a antígenos no causantes de enfermedad, tales como reacciones alérgicas sub-sintomáticas a pólenes locales. Esto podría usarse para identificar si un individuo ha viajado recientemente a una localización geográfica que contenía el antígeno. Los estados podrían incluir exposición a un antígeno relacionado con un proceso industrial o la fabricación o producción de agentes de bioterrorismo.

E1. Parámetros relacionados con la secuencia que pueden usarse para predecir clonotipos que se correlacionan

Con el fin de llevar a cabo un estudio poblacional puede usarse un conjunto de entrenamiento para entender las características de clonotipos que se correlacionan ensayando varios parámetros que pueden distinguir aquellos clonotipos que se correlacionan de aquellos que no lo hacen. Estos parámetros incluyen la secuencia o los segmentos V, D, y J específicos. En un caso, se muestra que los segmentos V específicos tienen mayor probabilidad de correlacionar con algunas enfermedades como es el caso si los clonotipos para una enfermedad específica probablemente reconozcan epítomos relacionados y por lo tanto pueden tener similitud de secuencia. Otros parámetros incluidos en casos adicionales incluyen el grado de hipermutación somática identificada y el nivel de un clonotipo en el pico de un episodio y su nivel cuando la enfermedad es relativamente inactiva. Otros parámetros que pueden predecir los clonotipos que se correlacionan incluyen sin limitación: 1) restos de la secuencia incluyendo región V o J, una combinación VJ, secuencias cortas en la región DJ; 2) longitud de la secuencia del clonotipo; 3) nivel del clonotipo incluyendo nivel absoluto (número de clones por millón de moléculas) o nivel de rango; 4) similitud en la secuencia de aminoácidos y ácido nucleico con otros clonotipos; la frecuencia de otros clonotipos altamente relacionados, incluyendo aquellos con cambios silenciosos (diferencias de nucleótidos que codifican los mismos aminoácidos) o aquellos con cambios conservativos de aminoácidos; 5) para los BCR el nivel de mutaciones somáticas en el clonotipo y/o el número de clonotipos distintos que se diferencian por mutaciones somáticas de algunos clonotipos de la línea germinal; 6) clonotipos cuyas proteínas asociadas tienen estructuras tridimensionales similares.

E2. Bases de datos de clonotipos que codifican anticuerpos específicos para un antígeno

Estos clonotipos que se correlacionan pueden codificar inmunoglobulinas o TCR que son específicos para uno o más epítomos de uno o más antígenos. Así, los clonotipos que se correlacionan pueden determinarse comparando los clonotipos medidos con entradas de una base de datos que comprende sustancialmente todos los clonotipos posibles para uno o más antígenos seleccionados (es decir, una "base de datos de clonotipos específicos de antígeno"). Dichas bases de datos pueden construirse secuenciando regiones seleccionadas de secuencias que codifican anticuerpos de linfocitos que producen anticuerpos con especificidad para los antígenos o epítomos de interés, o dichas bases de datos pueden poblarse llevando a cabo experimentos de unión con fagos que expresan y presentan anticuerpos o fragmentos de éstos en sus superficies. El último proceso se lleva a cabo fácilmente como se describe en Niro et al, *Nucleic Acids Research*, 38(9): e110 (2010). Brevemente, el método puede comprender las etapas siguientes: (a) un antígeno de interés, por ejemplo la proteína central de HCV, se une a un soporte sólido, (b) una biblioteca de anticuerpos codificados por fagos se expone al antígeno en condiciones para la unión del anticuerpo de manera que una fracción de los anticuerpos codificados por fagos se une al antígeno unido y otra fracción permanece libre, y (c) recoger y secuenciar los anticuerpos codificados por fagos que se unen para crear entradas de una base de datos de clonotipos que se correlacionan. Los anticuerpos codificados por fagos unidos se secuencian convenientemente usando una técnica de secuenciación de ADN de alto rendimiento como se ha descrito anteriormente. En un caso, los clonotipos del método codifican compuestos que se unen a fragmentos variables de cadena única (scFv). Pueden usarse condiciones para la unión de anticuerpos de diferentes astringencias. Las secuencias de ácido

nucleico determinadas del fago unido pueden tabularse e introducirse en la base de datos de clonotipos específicos de antígeno apropiada.

F. Datos funcionales para refinar la determinación de clonotipos que se correlacionan

Los datos funcionales pueden usarse para ayudar en la identificación de clonotipos que se correlacionan.

5 Por ejemplo, las células T y/o células B que contienen determinados marcadores que están enriquecidos en células que contienen clonotipos que se correlacionan pueden capturarse mediante métodos estándar como FACS o MACS. En otro caso, el marcador es un marcador de la superficie celular. En otro caso, la reactividad de las células T y/o células B frente a un antígeno relevante para la patología o para el tejido afectado sería una buena evidencia de la relevancia patológica de un clonotipo.

10 En otro caso, la secuencia de los clonotipos candidatos puede sintetizarse y ponerse en el contexto del TCR o BCR completo y evaluarse para la reactividad relevante. Alternativamente, los fragmentos amplificados de las diferentes secuencias pueden usarse como una entrada para técnicas de exposición en fago, ribosoma, o ARN. Estas técnicas pueden seleccionar secuencias con la reactividad relevante. La comparación de los resultados de la secuenciación para aquellos antes y después de la selección puede  
15 identificar aquellos clones que tienen la reactividad y por lo tanto probablemente sean patológicos. En otro caso, las técnicas de exposición específicas (por ejemplo, exposición en fago, ribosoma, o ARN) pueden usarse en un formato de matriz. Las moléculas individuales (o amplificaciones de estas moléculas individuales) que portan secuencias individuales del TCR o BCR (por ejemplo, secuencias CDR3) pueden organizarse bien como fagos, ribosomas, o ARN. Los antígenos específicos pueden estudiarse entonces  
20 para identificar la o las secuencias que codifican péptidos que se unen a ellas. Los péptidos que se unen a antígenos relevantes para la enfermedad probablemente sean patológicos.

G. Generación de un algoritmo de carga inmune

Puede usarse un algoritmo para computar una carga inmune, un valor o puntuación proporcionada por una función de los niveles de clonotipos que se correlacionan y no se correlacionan. La carga inmune  
25 puede usarse para tomar una decisión clínica. Usando los datos de un experimento (por ejemplo, un experimento que comprende muestras de sujetos en un primer estado de una enfermedad y muestras de sujetos en un segundo estado de la enfermedad), puede desarrollarse un algoritmo que combina la información acerca de los niveles de los clonotipos que se correlacionan y no se correlacionan en una única puntuación (carga inmune). Los parámetros de este algoritmo pueden ajustarse para maximizar la  
30 correlación entre carga inmune y los datos clínicos. Por ejemplo, los datos clínicos pueden ser una medida clínica de la gravedad de la enfermedad (por ejemplo, el grado de lesiones en un MRI para un paciente con esclerosis múltiple). Así, en un caso, una carga inmune puede calcularse por las etapas: (a) desarrollar un algoritmo que usa un conjunto de factores para combinar los niveles de los clonotipos que se correlacionan en una única puntuación de la actividad de la enfermedad; (b) comparar la puntuación  
35 generada en la etapa (a) con los datos clínicos acerca del estado de la enfermedad; y (c) optimizar los factores con el fin de maximizar la correlación entre los datos clínicos y la puntuación de la actividad de la enfermedad.

Los clonotipos que se correlacionan usados para generar un algoritmo de carga inmune pueden generarse usando un ensayo de calibración, un estudio poblacional, o un ensayo de calibración y un  
40 estudio poblacional como se ha descrito anteriormente.

Algunos de los factores que pueden considerarse en la combinación de los clonotipos que se correlacionan son el número de clonotipos que se correlacionan, su nivel, su proporción de cambio (velocidad), y la proporción de cambio en la velocidad (aceleración). Otros factores para evaluar incluyen el nivel de los clonotipos en el pico de episodio y en el estado inactivo de la enfermedad.

45 En un caso, la carga inmune generada se refiere a una enfermedad autoinmune. Dicha carga puede referirse como una carga autoinmune.

Se describe un método para generar un algoritmo que calcula una puntuación de la actividad de la enfermedad, que comprende: a) desarrollar un algoritmo que usa un conjunto de factores para combinar los niveles de clonotipos que se correlacionan en una puntuación de la actividad de la enfermedad, b)  
50 comparar la puntuación de la actividad de la enfermedad con datos clínicos respecto al estado de la enfermedad, y c) optimizar los factores con el fin de maximizar la correlación entre los datos clínicos y la puntuación de la actividad de la enfermedad.

H. Monitorización de la enfermedad usando el algoritmo de carga

1. Monitorización de la enfermedad sin un ensayo de calibración

55 Los clonotipos y el algoritmo de carga inmune pueden determinarse usando un estudio poblacional. La carga inmune puede usarse directamente sin tener que calibrar en primer lugar el paciente individual. Este ensayo puede hacerse cuando el paciente está en cualquier estado de la enfermedad. Este ensayo puede usarse para generar clonotipos que se correlacionan y no se correlacionan específicos sobre la base del algoritmo desarrollado anteriormente. La carga inmune puede calcularse entonces usando el segundo  
60 algoritmo generado en un estudio poblacional. Esta puntuación puede usarse clínicamente. Puede llevarse a cabo un ensayo de monitorización sin usar un ensayo de calibración por las etapas siguientes: (a) medir los clonotipos de un paciente en el momento en el que el paciente se va a monitorizar; y b) usar los clonotipos que se correlacionan predichos por el ensayo de algoritmo de descubrimiento y los datos del ensayo de monitorización para generar una puntuación que refleje el estado de la enfermedad del  
65 paciente usando el algoritmo de monitorización.

Se describe un método para monitorizar el estado de la enfermedad de un individuo, que comprende: a) determinar un perfil de clonotipos de una muestra de un sujeto, b) introducir la información del perfil de



clonotipos de a) en un algoritmo, y c) usar el algoritmo para generar una puntuación predictiva del estado de enfermedad del individuo. El algoritmo puede ser un algoritmo generado por a) desarrollo de un algoritmo que usa un conjunto de factores para combinar los niveles de los clonotipos que se correlacionan en una puntuación de la actividad de la enfermedad, b) comparación de la puntuación de la actividad de la enfermedad con datos clínicos respecto al estado de la enfermedad, y c) optimización de los factores con el fin de maximizar la correlación entre los datos clínicos y la puntuación de la actividad de la enfermedad.

## 2. Monitorización de la enfermedad usando un ensayo de calibración

Los clonotipos que se correlacionan y el algoritmo de carga inmune pueden determinarse usando un ensayo de calibración o un ensayo de calibración y un estudio poblacional. La carga inmune puede usarse en clínica realizando en primer lugar un ensayo de calibración. Este ensayo puede hacerse cuando el paciente está en un estado que es similar al primer estado usado en el estudio que generó los clonotipos que se correlacionan y que no se correlacionan que se usan en el algoritmo de carga inmune. Por ejemplo, este estado puede ser un estado de crisis de una enfermedad autoinmune si así es cómo se derivó el algoritmo de carga inmune. Este ensayo de calibración puede usarse entonces para generar los clonotipos que se correlacionan y no se correlacionan específicos para usarse en los ensayos de monitorización de la enfermedad posteriores. En un punto posterior en el tratamiento de este paciente, se hace otro ensayo en el paciente y la carga inmune puede calcularse usando el algoritmo generado en el estudio de descubrimiento, y la lista de los niveles de clonotipos generada en este ensayo de calibración específico del paciente. Esta puntuación de carga inmune puede usarse clínicamente. En un caso, un ensayo de monitorización usando un ensayo de calibración comprende las etapas siguientes: (a) ensayar un paciente en el estado de enfermedad 1 con el fin de determinar un perfil de clonotipos; (b) medir los clonotipos de un paciente en un momento posterior (el momento en el que el paciente se va a monitorizar); (c) usar el algoritmo de monitorización para generar una puntuación de la enfermedad que refleje un estado de la enfermedad a partir del perfil de clonotipos del estado de la enfermedad 1 del ensayo de calibración e información de un ensayo en un momento posterior.

Se describe un método para monitorizar el estado de la enfermedad de un individuo, que comprende: a) determinar un perfil de clonotipos de una muestra de un sujeto, b) introducir la información del perfil de clonotipos de a) en un algoritmo, y c) usar el algoritmo para generar una puntuación predictiva del estado de la enfermedad del individuo. El algoritmo puede ser un algoritmo generado por a) desarrollo de un algoritmo que usa un conjunto de factores para combinar los niveles de los clonotipos que se correlacionan en una puntuación de la actividad de la enfermedad, b) comparación de la puntuación de la actividad de la enfermedad con datos clínicos respecto al estado de la enfermedad, y c) optimización de los factores con el fin de maximizar la correlación entre los datos clínicos y la puntuación de la actividad de la enfermedad. En otro caso, el método puede comprender además determinar uno o más clonotipos que se correlacionan en el individuo por cualquiera de los métodos descritos en la presente memoria, e introducir la información del uno o más clonotipos que se correlacionan en el algoritmo.

En un caso, la enfermedad es una enfermedad autoinmune. En otro caso, la enfermedad autoinmune es lupus eritematoso sistémico, esclerosis múltiple, artritis reumatoide, o espondilitis anquilosante.

## 3. Otros factores relacionados con el uso de carga inmune

La misma carga inmune puede significar diferentes cosas para diferentes pacientes. Para uno, es necesario considerar la imagen clínica completa de un paciente. Desde una perspectiva de ensayo, se puede considerar la velocidad (proporción de cambio de la carga inmune en el tiempo) y la aceleración (proporción de cambio de la velocidad en el tiempo) además del nivel de carga inmune para tomar decisiones clínicas. Por ejemplo, si la puntuación de carga autoinmune se incrementa (alta velocidad) puede ser predictiva de una crisis incipiente en una enfermedad autoinmune.

Los ensayos adicionales que pueden integrarse en la puntuación de carga, por ejemplo, una puntuación de carga autoinmune, incluyen, por ejemplo, velocidad de sedimentación de eritrocitos (ESR), niveles de proteína C reactiva (CRP), anti-ADNs, otras titulaciones de autoanticuerpos, niveles de complemento, niveles de proteína en orina, proporción proteína en orina/creatinina, niveles de creatinina, niveles de nitrógeno de urea en la sangre (BUN), niveles de plaquetas, recuentos WBC, hematocrito (Hct), Hb, resultados de análisis de orina. Otros ensayos que están relacionados con SLE que pueden integrarse incluyen, por ejemplo, nivel de CD27, nivel celular de CD27++, genes que responden a INF (Baechler, EC et al. (2003) *Proc. Natl. Acad. Sci.* 100: 2610-2615), y puntuación de quimioquinas (Bauer JW et al. (2009) *Arthritis Rheum.* 60: 3098-3107). Otros ensayos no relacionados con lupus incluyen, por ejemplo, ensayo de la hormona estimulante de tiroides (TSH), ensayo de triyodotironina (T3), ensayo de tiroxina (T4), ensayos de función hepática (LFT), otros autoanticuerpos, ensayo de calprotectina, ensayo de lactoferrina, y análisis del fluido sinovial. Los ensayos adicionales pueden incluir ensayo de formación de imágenes, incluyendo, por ejemplo, MRI, escaneo CT, rayos X, y ultrasonidos.

I. El uso de tecnologías de secuenciación en combinación con enriquecimientos celulares parciales como parte de una etapa de calibración para encontrar clonotipos que se correlacionan.

Existen varias tecnologías que pueden usarse para separar células en la sangre o tejido sobre la base de marcadores celulares. Éstas incluyen separaciones en fase sólida tales como lechos o columnas en las que se inmovilizan reactivos de afinidad específicos tales como anticuerpos. La separación en fase líquida puede conseguirse usando técnicas tales como citometría de flujo en la que reactivos marcados que se unen específicamente a marcadores celulares seleccionados se usan para dirigir un dispositivo de flujo seleccionado en el que pueden usarse marcadores fluorescentes específicos para separar las células así

marcadas. La leucofóresis es otra técnica de separación en fase líquida que puede usarse para enriquecer poblaciones de leucocitos de la sangre, por ejemplo, Shelat, Am. J. Med. 123: 777-784 (2010); patente U.S. 5.846.928; o semejantes, después de lo cual puede llevarse a cabo un enriquecimiento adicional mediante marcadores de la superficie celular.

5 Algunas veces será ventajoso hacer un perfil de un subconjunto de células T y/o B cuando se calibra un algoritmo para predecir clonotipos que se correlacionan y/o para medir los niveles de estos clonotipos con el fin de medir una carga de enfermedad. Esto puede hacerse usando los métodos descritos anteriormente para una variedad de marcadores de superficie o internos. Un reto que se presenta al usar dichos métodos se basa en el hecho de que la selectividad para un marcador dado nunca es perfecta para  
10 un marcador dado. Como resultado, es improbable que el enriquecimiento de lugar a una población pura de las células seleccionadas. Otro método para conseguir enriquecimiento de subconjuntos de células sin la separación de las células es crecer selectivamente el subconjunto de células de interés. Por ejemplo, las células T pueden activarse con antígeno in vitro, y se puede permitir a las células activadas dividirse e incrementar su número lo que da lugar a su enriquecimiento.

15 Los perfiles de clonotipos de poblaciones de células T y/o B pueden generarse secuenciando moléculas individuales, aisladas espacialmente, tanto antes como después del enriquecimiento para al menos un marcador celular. Los dos perfiles se comparan entonces para determinar qué clonotipos tienen frecuencias significativamente alteradas entre las poblaciones de células enriquecidas y no enriquecidas con el fin de identificar clonotipos que están asociados con células que portan los marcadores  
20 seleccionados para el enriquecimiento. La ventaja de usar el método de secuenciación para conseguir esta identificación es que los clonotipos pueden identificarse incluso si el enriquecimiento es relativamente bajo ya que las frecuencias de clonotipos pueden medirse bien con suficiente profundidad de secuenciación. Esto, a su vez, permitirá usar una variedad de metodologías de enriquecimiento rutinariamente y con bajo coste ya que no serán necesarios los enriquecimiento múltiples costosos diseñados para conseguir pureza.

25 En un caso, dicha secuenciación de receptores de células T y B antes y después del enriquecimiento celular podría usarse como parte de un algoritmo de calibración con el fin de determinar qué clonotipos se correlacionan. En este caso, las células se secuencian antes y después del enriquecimiento para marcadores que son relevantes para la enfermedad en un primer estado de enfermedad. Los ejemplos de este primer estado de enfermedad podrían ser: muestras de sangre en un estado de pico de la enfermedad, muestras de tejido afectado, muestras de tejido afectado en un estado de pico de la enfermedad, etc. Se obtienen así clonotipos de la fracción celular que contiene y no contiene el marcador celular y pueden usarse como entradas en un algoritmo que se usa entonces para determinar clonotipos que se correlacionan en ese individuo.

35 En otro caso, el enriquecimiento no sólo se hace en una muestra recogida durante un primer estado de la enfermedad sino en muestras posteriores del mismo individuo. De esta manera, los clonotipos que se correlacionan se determinan en sub fracciones de las células en un momento cualquiera.

40 En otro caso, se usan marcadores celulares en conjugación con secuenciación para evaluar no sólo la frecuencia de clonotipos relevantes específicos sino también su estado funcional. En este caso, los marcadores celulares proporcionan información más allá de la identidad de los clonotipos que se correlacionan ya que refinan la predicción del estado de la enfermedad obtenido por la frecuencia de los clonotipos que se correlacionan. Mediante la secuenciación antes y después del enriquecimiento con marcadores específicos, se determina la frecuencia de un clonotipo específico. Además, se determina la frecuencia y la fracción de las células que contiene este clonotipo específico junto con otro marcador  
45 celular. Considérense dos pacientes con la misma frecuencia de clonotipos relevantes para algún estado clínico pero diferente frecuencia de marcadores celulares particulares (por ejemplo, marcadores de activación) en las células que contienen estos clonotipos. Estos pacientes pueden tener diferente actividad de la enfermedad a pesar de tener la misma frecuencia de los clonotipos relevantes.

50 Los marcadores celulares pueden ser marcadores de la activación celular. En general, los marcadores pueden determinarse midiendo la expresión de genes usando una población de células T y/o B que se sabe que se correlacionan con la enfermedad. Estos marcadores podrían ser marcadores de la superficie celular o células expresadas en la célula.

En un caso de esta invención, se enriquecen células que se muestra que tienen afinidad para un antígeno que se sabe que es relevante para una enfermedad particular. Hay varios métodos para hacer esto.

55 En otro caso, las células de interés son células B que interaccionan con antígeno específico. En este caso, existirán células B que tienen secuencias del receptor de células B que se unen a este antígeno específico. Estos receptores de células B pueden usarse así como marcadores de la superficie celular que pueden enriquecerse usando reactivos específicos de antígeno. En un caso, pueden usarse lechos o columnas en los que se inmoviliza el antígeno para enriquecer para células que expresan receptores de células B específicos para este antígeno. En otro caso, el antígeno se convierte en multimérico, por ejemplo tetramérico, con el fin de incrementar la afinidad de las células que expresan los receptores de células B apropiados. En otro caso, estas células pueden marcarse usando un reactivo antígeno marcado fluorescentemente. Estas células pueden enriquecerse entonces usando un método de citometría de flujo que separa sobre la base del marcador fluorescente. Este proceso puede hacerse en combinación con  
60 otros marcadores de células B en la metodología de citometría de flujo. En otra realización, el reactivo antígeno fluorescente se convierte en multimérico con el fin de incrementar la afinidad de las células que expresan el receptor de células B apropiado.

Puede definirse la fuerza de la interacción de los diferentes clonotipos con el antígeno específico. El grado de enriquecimiento de un clonotipo por la interacción con el antígeno proporciona una medida de la fuerza de la interacción. Como resultado, en lugar del nivel de "titulación" tradicional de un anticuerpo, puede obtenerse información más detallada. Específicamente, puede determinarse la frecuencia de diferentes clonotipos con diferentes avidedces. En un caso, el antígeno así usado en el enriquecimiento es una especie molecular única. En otro caso, el antígeno es una mezcla compleja de antígenos que son relevantes para una enfermedad. El antígeno puede ser un tipo celular o una mezcla de tipos celulares.

En otro caso, las células de interés son células T que interactúan con un antígeno específico en el contexto de una molécula MHC. En un caso, el péptido que forma complejo con las moléculas MHC se usa para capturar las células relevantes. Los tetrámeros de complejo MHC-péptido se han usado previamente con éxito para este propósito. En otro caso, la sangre o tejido relevante que contiene células capaces de presentar el antígeno y las células T se incuban con el antígeno para permitir que los péptidos sean presentados a las células T. Las células que se activan a través de la unión a estos antígenos pueden enriquecerse entonces por alguna característica de la activación. Potencialmente, puede utilizarse cualquier característica de la activación como proliferación celular, migración de leucocitos, citotoxicidad, y/o expresión de marcadores de activación. Por ejemplo, las células activadas proliferan y se puede permitir que se dividan y de esta manera enriquecerse. De forma similar, las células activadas pueden expresar algunos marcadores que pueden usarse para capturarlas. Estos marcadores pueden ser marcadores de la superficie o algún marcador interno como citoquinas, tales como INF $\gamma$ , IL-2, o IL-7. Las células que expresan marcadores de superficie pueden capturarse fácilmente usando diferentes técnicas tales como FACS o lechos recubiertos con anticuerpos frente al marcador de superficie. También se han desarrollado técnicas para capturar células que expresan marcadores intracelulares, particularmente citoquinas. Una técnica se denomina tinción de intracelular de citoquinas. En este método, las citoquinas específicas del proceso inmune en cuestión se atrapan en el interior de las células T que se permeabilizan posteriormente permitiendo que estas citoquinas específicas se marquen usando anticuerpos fluorescentes. Estas células marcadas pueden enriquecerse entonces usando un método de citometría de flujo. Otro método, captura de citoquinas, usa anticuerpos híbridos que tienen especificidad dual. Una de las especificidades es para algún marcador genérico en todas las células T (como molécula MHC) y el otro es para la citoquina de interés, por ejemplo, INF $\gamma$ , IL-2, o IL-7. La especificidad genérica une el anticuerpo en la superficie de todas las células T, y las citoquinas liberadas de las células T se capturan entonces por anticuerpos unidos en las mismas células. Pueden usarse entonces anticuerpos fluorescentes frente a la citoquina relevante permitiendo la captura de las células relevantes usando FACS.

Puede definirse la fuerza de la interacción de los diferentes clonotipos con el antígeno específico. El grado de enriquecimiento de un clonotipo por la interacción con el antígeno proporciona una medida de la fuerza de la interacción. Por lo tanto, puede determinarse la frecuencia de diferentes clonotipos con diferentes avidedces.

En un caso, el antígeno así usado en el enriquecimiento es una especie molecular única. En otro caso, el antígeno es una mezcla compleja de antígenos que son relevantes para una enfermedad. La mezcla de antígenos puede ser un tipo celular o una mezcla de tipos celulares.

J. Enriquecimiento de antígeno para la detección de recurrencia de una infección latente

Se usa frecuentemente en enfermedades infecciosas no sólo para medir la presencia o ausencia de un patógeno sino también para medir y monitorizar la respuesta inmune a este patógeno. Como resultado, la medida de anticuerpos producidos por el sistema inmune frente a antígenos de patógenos específicos es una metodología rutinaria en la práctica clínica. Dichas respuestas inmunes frente a antígeno de patógeno específico según se miden por anticuerpos no proporcionan, sin embargo, una visión completa de la respuesta inmune al antígeno. Los anticuerpos medidos, pueden ser el producto de muchos clones diferentes de células B cada uno de los cuales expresa un anticuerpo ligeramente diferente cada uno de los cuales puede portar información ligeramente diferente acerca del estado de la enfermedad. Además, las respuestas de las células T a estos antígenos no se están midiendo en absoluto.

La respuesta inmune de un paciente a una infección patogénica podría perfilarse de forma exhaustiva usando los métodos descritos en la presente memoria. En un caso, la respuesta de las células B puede medirse de forma exhaustiva en un individuo infectado con un patógeno usando un enriquecimiento de células B en un punto del curso de la enfermedad para averiguar los clonotipos de células B que son relevantes para la respuesta de anticuerpos a esa infección. Con el fin de conseguir esto, las células B que están implicadas en una respuesta inmune a un patógeno se identificarían realizando un enriquecimiento usando antígenos presentes en el patógeno en cuestión. Estos antígenos podrían ser una única especie de antígeno, un conjunto de especies distintas de antígenos, o una mezcla compleja de antígenos del patógeno incluyendo las células completas del patógeno. Dichos antígenos se inmovilizan entonces en una superficie sólida o se marcan fluorescentemente y el enriquecimiento se lleva a cabo usando bien un protocolo de unión basado en lechos, un protocolo de unión basado en columna, o un método de citometría de flujo en el caso en el que los antígenos estén marcados fluorescentemente. Las células del paciente se perfilan antes y después del enriquecimiento separando moléculas individuales de ADN o ARN del receptor de las células B en dos dimensiones y secuenciando moléculas individuales para formar un perfil de clonotipos de BCR. Las secuencias de los clonotipos que muestran un desplazamiento significativo de la frecuencia entre dos perfiles de clonotipos son entonces candidatas para ser clonotipos que son responsables de una respuesta inmune al o a los antígenos. Pueden desarrollarse opcionalmente

algoritmos adicionales para refinar la predicción de qué clonotipos son probablemente relevantes para esta respuesta inmune específica. Estos algoritmos pueden usar parámetros de secuencia tales como frecuencia, longitud de la secuencia, similitud en la secuencia de aminoácidos, similitud con otras secuencias de clonotipos similares incluyendo aquellos creados por hipermutaciones somáticas, etc.

5 En un caso preferido, la captura del antígeno se hace en un punto de calibración en el tiempo para identificar las células B relevantes y no hecho y posteriormente perfilado de puntos de tiempo en los que todas las células B se perfilan sin enriquecimiento.

En otro caso, se mide la respuesta de las células T. Con el fin de conseguir esto, las células T que están implicadas en una respuesta inmune a un patógeno se identificarían realizando un enriquecimiento usando antígenos presentes en el patógeno en cuestión. Estos antígenos podrían ser una única especie de antígeno, un conjunto de especies distintas de antígenos, o una mezcla compleja de antígenos del patógeno incluyendo las células completas del patógeno. En un caso, los tetrámeros de complejo MHC-antígeno se usan para marcar fluorescentemente las células T. En otro caso, dichos antígenos se añaden a la sangre de un paciente en al menos un primer punto de tiempo y se incuban de manera que se permite que los péptidos antigénicos sean presentados por células presentadoras de antígenos de este individuo. En estos dos casos, se perfilan moléculas de ARN o ADN individuales aisladas espacialmente de muestras de sangre antes y después del enriquecimiento para estas células T usando bien tetrámeros del complejo MHC-antígeno, el método de tinción de citoquinas internas o la captura de citoquinas y separación por FACS. Las secuencias de los clonotipos de células T que muestran un desplazamiento significativo de la frecuencia entre los dos perfiles de clonotipos son entonces candidatas para ser clonotipos que son responsables de una respuesta inmune para el o los antígenos. Pueden desarrollarse opcionalmente algoritmos adicionales para refinar la predicción de qué clonotipos son probablemente relevantes para esta respuesta inmune específica. Estos algoritmos pueden usar parámetros de secuencia tales como frecuencia, longitud de la secuencia, similitud en la secuencia de aminoácidos, similitud con otras secuencias de clonotipos similares, etc.

En un caso preferido, la captura del antígeno se hace en un punto de calibración en el tiempo para identificar las células B relevantes y no hecho y posteriormente perfilado de puntos de tiempo en los que todas las células B se perfilan sin enriquecimiento.

#### IV. Determinación de estados de enfermedad

30 Como el sistema inmune es tan fundamental para la salud humana, la capacidad de medir las respuestas inmunes tiene amplias aplicaciones en medicina. El sistema inmune puede usarse para entender el estado de la enfermedad subyacente cuando está mediado por el sistema inmune. Esto permite un conjunto muy potente de aplicaciones de diagnóstico y pronóstico que usan los perfiles inmunes para informar de los riesgos de una amplia variedad de resultados clínicos y permite a los médicos intervenir de forma más efectiva.

##### A. Utilidad del perfilado inmune en el tratamiento de enfermedades autoinmunes

Los métodos descritos en la presente memoria pueden usarse para diagnosticar y tratar una enfermedad autoinmune en un sujeto. La enfermedad autoinmune implica células inmunes adaptativas que se escapan del proceso habitual confiriendo autoinmunidad y atacando alguna o algunas dianas en tejidos corporales. Las enfermedades autoinmunes incluye, por ejemplo, encefalomiелitis aguda diseminada, enfermedad de Addison, espondilitis anquilosante, síndrome de anticuerpos anti-fosfolípido, anemia hemolítica autoinmune, hepatitis autoinmune, enfermedad del oído interno autoinmune, enfermedad de Behcet, penfigoide ampolloso, enfermedad celíaca, enfermedad de Chagas, enfermedad pulmonar obstructiva crónica, dermatomiositis, diabetes mellitus tipo 1, síndrome de Goodpasture, enfermedad de Graves, síndrome de Guillain-Barré, tiroiditis de Hashimoto, hidradenitis supurativa, púrpura trombocitopénica idiopática, cistitis intersticial, esclerosis múltiple, miastenia grave, neuromiotonía, pénfigo vulgar, anemia perniciosa, polimiositis, cirrosis biliar primaria, artritis reumatoide, escleroderma, lupus eritematoso sistémico, síndrome de Sjögren, y síndromes de vasculitis. Los estados de estas enfermedades autoinmunes pueden diagnosticarse usando los métodos descritos en la presente memoria. Pueden sugerirse tratamientos a un sujeto sobre la base del estado de la enfermedad autoinmune.

La información clínica respecto a un sujeto con una enfermedad autoinmune, o que se sospecha que tiene una enfermedad autoinmune, puede usarse para determinar el estado de la enfermedad (o carga autoinmune). La información clínica puede usarse para identificar patrones de un perfil de clonotipos que se correlacionan con un estado de la enfermedad. La información clínica puede incluir, por ejemplo, altura, peso, color de los ojos, edad, sexo, grupo étnico, presión sanguínea, niveles de colesterol LDL, niveles de colesterol HDL, historial médico familiar, e información sobre marcadores moleculares.

La información clínica puede incluir síntomas de una o más enfermedades autoinmunes. Para la hepatitis autoinmune, los síntomas pueden incluir fatiga, hepatomegalia, ictericia, prurito, erupción cutánea, artralgia, malestar abdominal, angiomas aracniformes, náusea, vómito, anorexia, orina oscura, heces claras o grises. Para la dermatomiositis (DM), los síntomas pueden incluir erupción (decoloraciones irregulares, azul-morado en la cara, cuello, hombros, parte superior del pecho, codos, rodillas, nudillos y espalda) acompañando o precediendo a debilidad muscular, disfagia, mialgia, fatiga, pérdida de peso y fiebre de grado bajo. Para la enfermedad de Graves, los síntomas pueden incluir pérdida de peso debida a gasto energético incrementado, apetito, ritmo cardiaco y presión sanguínea incrementados, y temblores, nerviosismo y sudoración. Para la tiroiditis de Hashimoto, los síntomas pueden incluir ralentización mental y física, mayor sensibilidad al frío, ganancia de peso, engrosamiento de la piel, bocio. Para la enfermedad mixta del tejido conectivo (MCTD), los síntomas pueden incluir características de lupus eritematoso

sistémico (SLE), escleroderma y polimiositis. Para el penfigoide ampoloso (BP), los síntomas pueden incluir verdugones medianamente pruríticos a ampollas graves e infección, ampollas grandes orales o esofágicas. Para el pénfigo, los síntomas pueden incluir presencia de ampollas en la piel y membranas mucosas. Para la anemia perniciosa, los síntomas pueden incluir dificultad para respirar, fatiga, palidez, taquicardia, inapetencia, diarrea, hormigueo y entumecimientos de las manos y pies, boca ulcerada y andar inestable. Para la polimiositis (PM), los síntomas pueden incluir debilidad muscular, disfagia y mialgia. Para la cirrosis biliar primaria (PBC), los síntomas pueden incluir fatiga y prurito. Para el escleroderma (esclerosis sistémica), los síntomas pueden incluir hinchazón y tumefacción de los dedos o manos, espesamiento de la piel, úlceras cutáneas en los dedos, tumefacción articular en las manos, dolor, garganta ulcerada y diarrea. Para el síndrome de Sjögren, los síntomas pueden incluir sequedad de los ojos y boca, glándulas del cuello hinchadas, dificultad para tragar o hablar, sabores u olores inusuales, sed y úlceras en la lengua. Para el lupus eritematoso sistémico (SLE), los síntomas pueden incluir fiebre, pérdida de peso, pérdida de cabello, úlceras en la boca y la nariz, malestar general, fatiga, convulsiones y síntomas de enfermedad mental, inflamación de las articulaciones similar a RA, erupción en forma de mariposa en la nariz y las mejillas, sensibilidad extrema al frío en las manos y pies. Para los síndromes de vasculitis, por ejemplo granulomatosis de Wegener, glomerulonefritis semilunar idiopática (ICGN), poliarteritis microscópica (MPA), síndrome renal pulmonar (PRS), los síntomas pueden incluir fatiga, debilidad, fiebre, artralgia, dolor abdominal, problemas renales y problemas neurológicos. La información clínica puede ser de uno o más sujetos en uno o más puntos de tiempo.

La información clínica puede incluir información respecto a respuestas de un sujeto con una enfermedad autoinmune a uno o más tratamientos que el sujeto ha recibido.

La utilidad clínica de la carga autoinmune se discute para enfermedades autoinmunes específicas más adelante. Otra realización de esta invención contempla la combinación de los ensayos de perfilado inmune con otros marcadores que ya se usan para la detección de la actividad de la enfermedad en estas enfermedades para permitir ensayos con mayor sensibilidad y especificidad. Pueden usarse otros identificadores moleculares o marcadores para computar la carga autoinmune o para determinar el estado de la enfermedad. Los identificadores moleculares pueden incluir ácidos nucleicos, proteínas, carbohidratos, y lípidos, y perfiles de expresión de ácidos nucleicos o proteínas. Los identificadores moleculares pueden ser de origen humano o no humano (por ejemplo, bacteriano). Los identificadores o marcadores pueden determinarse por técnicas que incluyen, por ejemplo, hibridación genómica comparativa (CGH), análisis de micromatrices cromosómicas (CMA), perfilado de expresión, micromatriz de ADN, micromatriz de oligonucleótidos de alta densidad, matriz de expresión de ARN de genoma completo, micromatriz de péptidos, ensayo inmunoabsorbente ligado a enzima (ELISA), secuenciación genómica, análisis del número de copias (CNV), análisis de polimorfismo de nucleótidos pequeños (SNP), inmunohistoquímica, hibridación in-situ, hibridación in-situ fluorescente (FISH), PCR, transferencia Western, transferencia Southern, SDS-PAGE, electroforesis en gel, y transferencia Northern.

Para el lupus eritematoso sistémico, los marcadores pueden incluir los niveles de velocidad de sedimentación de eritrocitos (ESR), niveles de proteína C reactiva (CRP), anti-ADNs, otras titulaciones de autoanticuerpos, niveles de complemento, niveles de proteína en orina, proporción proteína en orina/creatinina, niveles de creatinina, niveles de nitrógeno de urea en la sangre (BUN), niveles de plaquetas, recuentos WBC, hematocrito (Hct), Hb, y resultados de análisis de orina. Otros ensayos que están relacionados por ejemplo con SLE que pueden integrarse incluyen, por ejemplo, nivel de CD27, nivel celular de CD27<sup>++</sup>, genes que responden a INF y puntuación de quimioquinas.

1. Lupus eritematoso sistémico (SLE)

Los métodos descritos en la presente memoria pueden usarse para determinados estados o estadios del lupus eritematoso sistémico (SLE o lupus). SLE es una afección autoinmune seria que aflige frecuentemente a adultos jóvenes (la mayor parte mujeres). Se caracteriza por procesos inflamatorios que pueden afectar a muchos órganos incluyendo la piel, articulaciones, riñones, pulmones, corazón, y sistema nervioso central dando lugar a discapacidades frecuentes y algunas veces a la muerte. La enfermedad sigue un curso muy impredecible marcado por periodos de crisis seguidos de periodos quiescentes de remisión. No obstante, los pacientes diagnosticados con SLE son vistos regularmente por un reumatólogo y tratados con una variedad de medicaciones serias. Estas medicaciones incluyen esteroides tales como prednisona y otros inmunosupresores tales como cellecept (micofenolato mofetil). Aunque estos fármacos pueden reducir el daño en los órganos, contienen efectos secundarios significativos incluyendo riesgo de infección e infertilidad. La falta de fiabilidad para algunos de los síntomas (por ejemplo, dolor y fatiga) y el curso impredecible de la enfermedad hace que la personalización de las dosis de la medicación sea difícil, lo que resulta en un sobretratamiento de algunos pacientes y subtratamiento de otros. Como resultado, el tratamiento de SLE presenta retos terapéuticos significativos para el médico.

Existen varios métodos estándar que puede usar un médico para evaluar la actividad de SLE. El estado de la enfermedad puede medirse observando los síntomas clínicos de la enfermedad. Estos métodos incluyen la evaluación de signos (por ejemplo, erupción cutánea) y síntomas (por ejemplo, dolor articular y fatiga) así como resultados de laboratorio (por ejemplo, proporción de proteína/creatinina en orina, anticuerpo anti-ADNs, y recuentos sanguíneos). Estos marcadores clínicos, sin embargo, pueden ser indicadores atrasados del estado de la enfermedad y como tales los pacientes pueden responder sólo después de semanas o meses de terapia. Además, en algunos casos, los síntomas pueden ser difíciles de evaluar con precisión (por ejemplo, dolor y fatiga). Otros marcadores de la inflamación, por ejemplo,

anticuerpo anti-ADNds, nivel de complemento (por ejemplo, C3), proteína C reactiva (CRP), y velocidad de sedimentación de eritrocitos (ESR) carecen habitualmente de especificidad y/o sensibilidad. Los métodos invasivos tales como biopsia renal son poco prácticos para uso rutinario. Como resultado, los médicos realizan un ensayo bastante frecuente de sus pacientes sin una medida perfecta del estado de la enfermedad. Los síntomas clínicos y evaluación de laboratorio se integran en medidas tales como el índice de actividad de la enfermedad del lupus eritematoso sistémico (SLEDAI) y la evaluación global por el médico (PGA). Estas medidas no se hacen rutinariamente en la práctica clínica y frecuentemente se quedan cortas en varias situaciones clínicas.

Pueden identificarse clonotipos que representan diferentes perfiles inmunes para diferentes estados de la enfermedad. El estado de la enfermedad se sigue entonces comparando el perfil inmune identificado con el perfil inmune actual del paciente. La enfermedad puede ser lupus. Los estados de la enfermedad pueden ser periodos de crisis o periodos no de crisis. El perfil inmune puede usarse como un indicador temprano de un estado de crisis. Esto puede dirigir las decisiones sobre el tratamiento.

Los ejemplos específicos de la utilidad de la carga autoinmune para hacer intervenciones terapéuticas en SLE se discuten con mayor detalle en la sección de los ejemplos junto con estudios habilitadores específicos que determinan la carga autoinmune.

Los clonotipos que se correlacionan con SLE pueden relacionarse con anticuerpos específicos para auto antígenos. De acuerdo con esto, un método para determinar una probabilidad de que un individuo tiene lupus eritematoso sistémico comprende las etapas siguientes: (a) determinar un perfil de clonotipos a partir de una muestra de células B del individuo, comprendiendo la muestra un repertorio de clonotipos de éstas; y (b) comparar los clonotipos del perfil con los clonotipos de una base de datos de clonotipos específicos de antígeno para determinar un nivel de concordancias de clonotipos, determinando de esta manera una probabilidad de lupus eritematoso sistémico, incluyendo la base de datos de clonotipos específicos de antígeno sustancialmente todos los clonotipos de cadenas de inmunoglobulinas humanas específicas para el uno o más antígenos seleccionados del grupo que consiste en ADN bicatenario, malondialdehído, 4-hidroxi-nonenal, superóxido dismutasa, nitrotirosina, cardiolipina, proteína ribosomal P, fosfolípido, proteína central de la proteína ribonuclear nuclear pequeña (antígeno Smith), histona, proteína ribonuclear nuclear pequeña U1, topoisomerasa tipo I, proteínas centroméricas, proteína ribonuclear SS-A, proteína ribonuclear SS-B, e histidina-ARNt ligasa.

## 2. Esclerosis múltiple (MS)

Los métodos descritos en la presente memoria también pueden usarse para determinados estados o estadios de la esclerosis múltiple (MS). MS es una enfermedad autoinmune que afecta el cerebro y la médula espinal (sistema nervioso central). Los síntomas varían, porque la localización y gravedad de cada ataque pueden ser diferentes. Los episodios pueden durar varios días, semanas, o meses. Estos episodios se alternan con periodos de síntomas reducidos o sin síntomas (remisiones). Es común para la enfermedad volver (recidiva). Sin embargo, la enfermedad puede continuar hasta empeorar sin periodos de remisión.

Como los nervios en cualquier parte del cerebro o médula espinal pueden estar dañados, los pacientes con esclerosis múltiple pueden tener síntomas en muchas partes del cuerpo. Los síntomas musculares incluyen, por ejemplo, pérdida de equilibrio, entumecimiento o sensación anormal en cualquier área, dolor debido a espasmos musculares, dolor en los brazos o piernas, problemas al mover los brazos o piernas, problemas al andar, problemas con la coordinación y al hacer pequeños movimientos, habla mal articulada o difícil de entender, temblor en uno o más brazos o piernas, espasmo incontrolable de grupos musculares (espasticidad muscular), y debilidad en uno o más brazos o piernas.

Los síntomas oculares incluyen, por ejemplo, visión doble, malestar ocular, movimientos rápidos de los ojos incontrolables, y pérdida de visión (habitualmente afecta a un ojo a la vez).

Otros síntomas cerebrales y nerviosos incluyen, por ejemplo, periodo disminuido de atención, juicio disminuido, memoria disminuida, depresión o sentimientos de tristeza, mareo y problemas de equilibrio, dolor facial, pérdida de audición, y fatiga.

Los síntomas en el intestino y vejiga incluyen, por ejemplo, estreñimiento, dificultad para empezar a orinar, necesidad frecuente de orinar, escape de heces, urgencia fuerte de orinar, y escape de orina (incontinencia).

No existe una cura conocida para la esclerosis múltiple en este momento. Sin embargo, existen terapias que pueden ralentizar la enfermedad. El objetivo del tratamiento es controlar los síntomas y ayudar al paciente a mantener una calidad de vida normal.

Las medicaciones usadas para ralentizar la progresión de la esclerosis múltiple pueden incluir, por ejemplo, moduladores inmunes para ayudar a controlar el sistema inmune, incluyendo interferones (Avonex, Betaseron, o Rebif), anticuerpos monoclonales (Tysabri), acetato de glatiramer (Copaxona), mitoxantrona (Novantrona), metotrexato, azatioprina (Imuran), ciclofosfamida (Cytoxan), y natalizumab (Tysabri). Los esteroides pueden usarse para disminuir la gravedad de los ataques.

Las medicaciones para controlar los síntomas pueden incluir, por ejemplo, medicinas para reducir los espasmos musculares tales como Lioresal (Baclofen), tizanidina (Zanaflex), o una benzodiazepina, medicaciones colinérgicas para reducir los problemas urinarios, antidepresivos para los síntomas del humor o el comportamiento, y amantadina para la fatiga.

MS afecta a las mujeres más que a los hombres. El trastorno comienza lo más comúnmente entre las edades de 20 y 40, pero puede observarse a cualquier edad. MS es una enfermedad progresiva, lo que significa que el daño nervioso (neurodegeneración) empeora con el tiempo. Lo rápido que empeora MS

varía de persona a persona. La inflamación ocurre cuando las propias células inmunes del cuerpo atacan el sistema nervioso. Los episodios repetidos de inflamación pueden ocurrir a lo largo de cualquier área del cerebro y médula espinal. La gente con un historial familiar de MS y aquellos que viven en un área geográfica con una mayor proporción de incidencia para MS tienen un riesgo mayor de la enfermedad.

5 Los síntomas de MS pueden mimetizar aquellos de muchos otros trastornos del sistema nervioso. La enfermedad se diagnostica descartando otras afecciones. La gente que tiene una forma de MS denominada con recidiva-remisión puede tener un historial de al menos dos ataques, separados por un periodo de síntomas reducidos o sin síntomas. El responsable sanitario puede sospechar MS si hay disminuciones en la función de dos partes diferentes del sistema nervioso central (tales como reflejos anormales) a dos tiempos diferentes. Un examen neurológico puede mostrar una función nerviosa reducida en un área del cuerpo, o diseminación sobre muchas partes del cuerpo.

10 Los ensayos para diagnosticar esclerosis múltiple incluyen, por ejemplo, ensayos de fluido cerebroespinal, incluyendo bandas oligoclonales del CSF, escaneo de cabeza MRI, punción lumbar (punción espinal), estudio de la función nerviosa (ensayo de potencial evocado), y MRI espinal.

15 Como otras enfermedades autoinmunes, MS sigue un curso impredecible con crisis agudas y periodos de remisión. Existe un número creciente de terapias, cada una con efectos secundarios que varían de graves (ganancia de peso y depresión) a potencialmente mortales (pancitopenia e infecciones PML), efectividades variables en diferentes pacientes, y alto coste. Al mismo tiempo, la ausencia de ensayos rutinarios altamente exactos y precisos de la actividad de la enfermedad MS hace que el reto de administrar de manera efectiva una terapia sea complicado. Los episodios clínicos pueden estar separados por periodos de tiempo largos (hasta años en el estadio temprano de la enfermedad) incluso sin tratamiento. Además, las medicaciones disponibles reducen la probabilidad de recidiva pero no la previene completamente. Por lo tanto, la actividad de la enfermedad es difícil de evaluar y así, hay una medida a corto plazo inadecuada de la actividad de la enfermedad que podría usarse para medir si una terapia específica está mostrando eficacia en un paciente dado midiendo la reducción del número o gravedad de las recidivas. El único ensayo adicional disponible para monitorizar la actividad de la enfermedad es MRI cerebral para seguir el estado de las lesiones según se revela con la ayuda de agentes que potencian el contraste tales como gadolinio. Sin embargo, dicha formación de imágenes ofrece sólo una visión integrada del daño cerebral y carece de especificidad y resolución en el tiempo. El intento de usar la formación de imágenes por MRI para seguir el curso de la enfermedad en escalas de tiempo menores de un año no es práctico dados los costes, la ausencia de especificidad y los peligros de exposición excesiva al contraste. Como resultado, los pacientes se tratan frecuentemente con gran coste durante periodos prolongados de tiempo sin ningún resultado efectivo que permitiría al médico modificar la dosificación y/o cambiar a terapias añadidas.

35 Pueden identificarse clonotipos que representan diferentes perfiles inmunes para diferentes estados de la enfermedad. El estado de la enfermedad se sigue entonces comparando el perfil inmune identificado con el perfil inmune actual de los pacientes. La enfermedad puede ser MS. Los estados de la enfermedad pueden ser periodos de remisión y periodos activos. Este perfil inmune puede usarse como un indicador temprano de un periodo de remisión o no remisión. Esto puede dirigir las decisiones sobre el tratamiento.

### 40 3. Artritis reumatoide (RA)

Los métodos pueden usarse para medir el estado de la enfermedad para los pacientes con artritis reumatoide. La artritis reumatoide (RA) es un trastorno inflamatorio crónico, sistémico, que puede afectar muchos tejidos y órganos pero ataca principalmente las articulaciones, produciendo una sinovitis inflamatoria que progresa frecuentemente a la destrucción del cartílago articular y anquilosis de las articulaciones. La artritis reumatoide también puede producir inflamación difusa en los pulmones, pericardio, pleura, y esclera, y también lesiones nodulares, lo más comúnmente en el tejido subcutáneo bajo la piel. Aunque la causa de la artritis reumatoide es desconocida, la autoinmunidad juega un papel crucial en su cronicidad y progresión.

50 Aproximadamente un 1% de la población mundial está afectada por artritis reumatoide, las mujeres tres veces más frecuentemente que los hombres. El inicio es lo más frecuentemente entre las edades de 40 y 50, pero puede verse afectada gente de cualquier edad. Puede ser una afección discapacitante y dolorosa, que puede dar lugar a una pérdida sustancial del funcionamiento y movilidad. RA se diagnostica sobre todo sobre la base de síntomas y signos, pero también puede diagnosticarse con ensayos sanguíneos (especialmente un ensayo denominado factor reumatoide) y rayos X. El diagnóstico y gestión a largo plazo se realizan típicamente por un reumatólogo, un experto en las enfermedades de las articulaciones y tejidos conectivos.

60 Están disponibles varios tratamientos. El tratamiento no farmacológico incluye terapia física, ortosis, y terapia ocupacional. Pueden usarse fármacos analgésicos (contra el dolor) y anti-inflamatorios, incluyendo esteroides, para suprimir los síntomas, mientras que los fármacos antireumáticos modificadores de la enfermedad (DMARD) pueden usarse para inhibir o parar el proceso inmune subyacente y prevenir el daño a largo plazo. Recientemente, el nuevo grupo de biológicos ha incrementado las opciones de tratamiento.

65 Cuando RA se sospecha clínicamente, pueden realizarse estudios inmunológicos, tales como ensayo para la presencia de factor reumatoide (RF, un anticuerpo específico). Un RF negativo no descarta RA; en lugar de esto, la artritis se denomina seronegativa. Éste es el caso en aproximadamente el 15% de los pacientes. Durante el primer año de la enfermedad, es más probable que el factor reumatoide sea negativo con algunos individuos convirtiéndose en un estado seropositivo con el tiempo. RF también se

observa en otras enfermedades, por ejemplo, síndrome de Sjögren, y en aproximadamente el 10% de la población sana, por lo tanto el ensayo no es muy específico.

Debido a esta baja especificidad, se han desarrollado nuevos ensayos serológicos, que ensayan la presencia de los denominados anticuerpos anti-proteína citrulinada (ACPA). Como RF, estos ensayos son positivos sólo en una proporción (67%) de todos los casos de RA, pero son raramente positivos si RA no está presente, dando una especificidad de aproximadamente el 95%. Como son RF, hay evidencia de que ACPA estén presentes en muchos casos incluso antes del inicio de la enfermedad clínica.

Los ensayos más comunes para ACPA son el ensayo anti-CCP (péptido citrulinado cíclico) y el ensayo anti-MCV (anticuerpos frente a vimentina citrulinada mutada). Recientemente, se ha desarrollado un ensayo serológico en el punto de atención (POCT) para la detección temprana de RA. Este ensayo combina la detección del factor reumatoide y anti-MCV para el diagnóstico de la artritis reumatoide y muestra una sensibilidad del 72% y una especificidad del 99,7%.

También, pueden hacerse varios otros ensayos sanguíneos para permitir otras causas de la artritis, tales como lupus eritematoso. La velocidad de sedimentación de los eritrocitos (ESR), proteína C reactiva, recuento sanguíneo completo, función renal, enzimas hepáticas y otros ensayos inmunológicos (por ejemplo, anticuerpo antinuclear/ANA) se realizan todos en este estadio. Los niveles elevados de ferritina pueden revelar hemocromatosis, un mimico de RA, o pueden ser un signo de la enfermedad de Still, una variante de reumatoide seronegativa, habitualmente juvenil.

El término fármaco anti-reumático modificador de la enfermedad (DMARD) significó originalmente un fármaco que afecta las medidas biológicas tales como ESR y niveles de hemoglobina y autoanticuerpo, pero ahora se usa habitualmente para significar un fármaco que reduce la proporción de daño en el hueso y cartílago. Se ha encontrado que los DMARD producen tanto remisiones sintomáticas duraderas como retraso o parada de la progresión. Esto es significativo, ya que dicho daño es habitualmente irreversible. Los anti-inflamatorios y analgésicos mejoran el dolor y la rigidez pero no previenen el daño articular ni ralentizan la progresión de la enfermedad.

Existe un reconocimiento creciente entre los reumatólogos de que el daño permanente en las articulaciones ocurre en un estadio muy temprano en la enfermedad. En el pasado, era común empezar la terapia sólo con un fármaco anti-inflamatorio, y evaluar la progresión clínicamente y usando rayos X. Si había evidencia de que estaba empezando a ocurrir el daño articular, entonces se prescribiría un DMARD más potente. Los ultrasonidos y MRI son métodos más sensibles para formar imágenes de las articulaciones y han demostrado que el daño articular ocurre mucho antes y en más pacientes de lo que se pensaba previamente. La gente con rayos X normales tendrán frecuentemente erosiones detectables por ultrasonidos que los rayos X no pueden demostrar. El objetivo ahora es tratar antes de que se produzca el daño.

Puede haber otras razones por las que el empezar con DMARD pronto es beneficioso para prevenir el daño articular estructural. A partir de los estadios más tempranos de la enfermedad, las articulaciones se infiltran por células del sistema inmune que señalizan entre sí de maneras que pueden implicar una variedad de bucles de retroalimentación positiva (se ha observado durante mucho tiempo que una única inyección de corticosteroide puede abortar la sinovitis en una articulación particular durante periodos largos). La interrupción de este proceso tan pronto como sea posible con un DMARD efectivo (tal como metotrexato) parece mejorar el resultado de la RA durante los años posteriores. El retraso en la terapia durante tan poco como unos pocos meses después del inicio de los síntomas puede resultar en resultados peores a largo plazo. Existe por lo tanto un interés considerable en establecer la terapia más efectiva con la artritis temprana, cuando el paciente responde mejor a terapia y tiene más que ganar.

Los fármacos con masa molecular pequeña tradicionales usados para tratar la artritis incluyen, por ejemplo, DMARD sintetizados químicamente: azatioprina, ciclosporina (ciclosporina A), D-penicilamina, sales de oro, hidroxicloroquina, leflunomida, metotrexato (MTX), minociclina, y sulfasalazina (SSZ). Los fármacos citotóxicos incluyen ciclofosfamida.

Los eventos adversos más comunes están relacionados con toxicidad hepática y en la médula ósea (MTX, SSZ, leflunomida, azatioprina, compuestos de oro, D-penicilamina), toxicidad renal (ciclosporina A, sales de oro parenterales, D-penicilamina), neumonitis (MTX), reacciones cutáneas alérgicas (compuestos de oro, SSZ), autoinmunidad (D-penicilamina, SSZ, minociclina) e infecciones (azatioprina, ciclosporina A). La hidroxicloroquina puede causar toxicidad ocular, aunque es raro, y como la hidroxicloroquina no afecta la médula ósea o el hígado se considera frecuentemente como el DMARD con la toxicidad más baja. Desafortunadamente, la hidroxicloroquina no es muy potente, y habitualmente es insuficiente para controlar los síntomas por sí mismo.

Los agentes biológicos (biológicos) pueden producirse mediante ingeniería genética, e incluyen, por ejemplo, bloqueadores del factor de necrosis tumoral alfa (TNF $\alpha$ )-etanercept (Enbrel), infliximab (Remicade), adalimumab (Humira), bloqueadores de interleuquina 1 (IL-1)-anakinra (Kineret), anticuerpos monoclonales frente a células B-rituximab (Rituxan), bloqueador de la coestimulación de células T-abatacept (Orencia), bloqueadores de interleuquina 6 (IL-6)-tocilizumab (un anticuerpo anti-receptor de IL-6) (RoActemra, Actemra).

Los agentes anti-inflamatorios incluyen, por ejemplo, glucocorticoides, fármacos anti-inflamatorios no esteroideos (NSAID, la mayor parte también actúa como analgésicos). Los analgésicos incluyen, por ejemplo, paracetamol (acetaminofeno en EEUU y Canadá), opiáceos, diprocualona, y lidocaína tópica.

El reto de tratar RA se basa en el hecho de que la enfermedad es una enfermedad crónica a largo plazo que puede resultar en una discapacidad desafiante para la que existe un gran rango de tratamientos cada



uno de los cuales tiene inconvenientes significativos. Muchos de los DMARD someten a los pacientes a efectos secundarios significativos incluyendo un riesgo incrementado de infecciones graves, cáncer, o incluso enfermedad autoinmune. Además, los fármacos derivados biológicamente son muy costosos, y el paciente puede estar sometido a inyecciones frecuentes.

5 Un médico que inicia una terapia para un paciente se enfrenta a muchas opciones posibles. Sería deseable obtener una respuesta rápida una vez el paciente empieza una terapia para entender si el paciente está respondiendo a la terapia que se elige antes de que la manifestación clínica se presente. La formación de imágenes no es sensible y es costosa y muchos marcadores sanguíneos tales como CRP carecen de una sensibilidad suficiente. Un ensayo que permitiera al médico determinar rápidamente el estado de la enfermedad le permitiría a él o ella ajustar la terapia rápidamente a una terapia más efectiva, ahorrando al paciente daño articular adicional y usando de forma más efectiva las terapias costosas disponibles.

10 Un paciente que no ha experimentado ninguna crisis aguda desde que empezó el tratamiento puede de hecho seguir experimentando daño inflamatorio en curso en las articulaciones que no se ha manifestado clínicamente en sí mismo. Un ensayo que permitiera al médico diferenciar este estado del fondo le permitiría ajustar la terapia para intentar llevar al paciente más cerca de un estado en el que no se experimente daño articular en curso.

Los ejemplos específicos de cómo puede usarse la carga autoinmune en la gestión de pacientes con RA se describen con más detalle en la sección de ejemplos de este documento.

15 Pueden identificarse clonotipos que representan diferentes perfiles inmunes para diferentes estados de la enfermedad. El estado de la enfermedad se sigue entonces comparando el perfil inmune identificado con el perfil inmune actual de los pacientes. La enfermedad es RA. Los estados de la enfermedad pueden ser, pero no están limitados a, periodos de alta inflamación y una línea base. Estos perfiles inmunes se usan para dirigir las decisiones sobre el tratamiento.

#### 25 4. Espondilitis anquilosante

Los métodos pueden usarse para detectar la actividad de la enfermedad para espondilitis anquilosante. La espondilitis anquilosante (AS, del griego ankylos, torcido; spondylos, vértebras), conocida previamente como enfermedad de Bechterew, síndrome de Bechterew, y enfermedad de Marie Strümpell, una forma de espondiloartritis, es una enfermedad de artritis inflamatoria crónica y autoinmune. Afecta principalmente las articulaciones de la columna vertebral y el sacroíleo en la pelvis, causando la fusión eventual de la columna vertebral. Es un miembro del grupo de las espondiloartropatías con una predisposición genética fuerte. La fusión completa resulta en una rigidez completa de la columna vertebral, una afección conocida como columna vertebral en forma de bambú.

30 El paciente típico es un hombre joven, de 18-30 años, cuando los síntomas de la enfermedad aparecen por primera vez, con dolor crónico y rigidez en la parte inferior de la columna vertebral o algunas veces la columna vertebral completa, frecuentemente con dolor referido en una u otra nalga o la parte posterior del muslo desde la articulación sacroilíaca. Los hombres están afectados más que las mujeres por una proporción de aproximadamente 3:1, tomando habitualmente la enfermedad un curso más doloroso en hombres que en mujeres. En el 40% de los casos, la espondilitis anquilosante está asociada con una inflamación del ojo (iridociclitis y uveítis), causando enrojecimiento, dolor ocular, pérdida de visión, motas y fotofobia. Otro síntoma común es fatiga generalizada y algunas veces náusea. De forma menos común puede ocurrir aortitis, fibrosis pulmonar apical y ectasia de la vaina de la raíz del nervio sacro. Como con todas las espondiloartropatías seronegativas, puede ocurrir levantamiento de las uñas (onicolisis).

35 No hay ningún ensayo directo para diagnosticar AS. Un examen clínico y estudios de rayos X de la columna vertebral, que muestran cambios característicos en la columna vertebral y sacroiliitis, son las herramientas principales de diagnóstico. Un inconveniente del diagnóstico por rayos X es que los signos y síntomas de AS se han establecido habitualmente tanto como 8-10 años antes de que se produzcan cambios evidentes por rayos X en una película de rayos X simple, lo que significa un retraso de tanto como 10 años antes de que puedan introducirse terapias adecuadas. Las opciones para el diagnóstico temprano son tomografía y formación de imágenes por resonancia magnética de las articulaciones sacroilíacas, pero la fiabilidad de estos ensayos todavía no está clara. El ensayo de Schober es una medida clínica útil de la flexión de la columna vertebral lumbar realizada durante el examen.

40 Durante los periodos de inflamación aguda, los pacientes con AS mostrarán algunas veces un incremento en la concentración sanguínea de proteína C reactiva (CRP) y un incremento en la velocidad de sedimentación de eritrocitos (ESR), pero hay muchos con AS cuyas proporciones de CRP y ESR no se incrementan, de manera que los resultados normales de CRP y ESR no se corresponden siempre con la cantidad de inflamación que realmente tiene una persona. Algunas veces, las personas con AS tienen resultados de nivel normales, y aún así experimentan una cantidad significativa de inflamación en sus cuerpos.

45 La espondilitis anquilosante (AS, del griego ankylos, torcido; spondylos, vértebras), conocida previamente como enfermedad de Bechterew, síndrome de Bechterew, y enfermedad de Marie Strümpell, una forma de espondiloartritis, es una enfermedad de artritis inflamatoria crónica y autoinmune. Afecta principalmente las articulaciones de la columna vertebral y el sacroíleo en la pelvis, causando la fusión eventual de la columna vertebral.

50 Es un miembro del grupo de las espondiloartropatías con una predisposición genética fuerte. La fusión completa resulta en una rigidez completa de la columna vertebral, una afección conocida como columna vertebral en forma de bambú.

Hay tres tipos principales de medicaciones usadas para tratar la espondilitis anquilosante: 1) fármacos anti-inflamatorios, que incluyen NSAID tales como ibuprofeno, fenilbutazona, indometacina, naproxeno e inhibidores de COX-2, que reducen la inflamación y el dolor. También se ha demostrado por evidencia clínica que los analgésicos opioides son muy efectivos para aliviar el tipo de dolor crónico experimentalmente comúnmente por aquellos que padecen AS, especialmente en formulaciones de liberación en el tiempo. 2) DMARD tales como ciclosporina, metotrexato, sulfasalazina, y corticosteroides, usados para reducir la respuesta del sistema inmune a través de inmunosupresión; 3) bloqueadores de TNF $\alpha$  (antagonistas) tales como etanercept, infliximab y adalimumab (también conocidos como biológicos), están indicados para el tratamiento de y son inmunosupresores efectivos como en otras enfermedades autoinmunes.

Se ha mostrado que los bloqueadores de TNF $\alpha$  son el tratamiento más prometedor, ralentizando el progreso de AS en la mayoría de los casos clínicos, ayudando a muchos pacientes a recibir una reducción significativa, aunque no eliminación, de su inflamación y dolor. También se ha mostrado que son altamente efectivos para tratar no sólo la artritis de las articulaciones sino también la artritis espinal asociada con AS. Un inconveniente, además del alto coste frecuente, es el hecho de que estos fármacos incrementan el riesgo de infecciones. Por esta razón, el protocolo para cualquiera de los bloqueadores de TNF- $\alpha$  incluye un ensayo para tuberculosis (como Mantoux o Heaf) antes de empezar el tratamiento. En el caso de infecciones recurrentes, incluso garganta irritada recurrente, la terapia puede suspenderse debido a la inmunosupresión implicada. Se recomienda a los pacientes que toman medicaciones de TNF que limiten su exposición a otros que porten o pueden portar un virus (tal como constipado o gripe) o que pueden tener una infección bacteriana o fúngica.

AS produce síntomas que son muy comunes en las poblaciones sanas. Por ejemplo, un paciente que presenta quejas acerca de dolor severo de espalda puede no estar experimentando una crisis de AS sino que podría tener sólo dolor de espalda rutinario. El médico se ve forzado a tomar una decisión acerca de si tratar estos síntomas con fármacos costosos con efectos secundarios potencialmente graves sin una visión muy precisa del estado de la enfermedad. CRP y ESR no proporcionan una visión muy precisa del estado de la enfermedad. Al mismo tiempo, el curso de la enfermedad no tratada puede resultar en daño espinal debilitante a largo plazo. Este estado de problemas da lugar a un reto clínico difícil y el uso de sobretratamiento significativo. La disponibilidad de una medida objetiva que refleja la actividad de la enfermedad puede resultar de gran ayuda en la gestión de pacientes con AS.

Pueden identificarse clonotipos que representan diferentes perfiles inmunes para diferentes estados de la enfermedad. El estado de la enfermedad se sigue entonces comparando el perfil inmune identificado con el perfil inmune actual de los pacientes. La enfermedad es AS. Los estados de la enfermedad pueden ser, pero no están limitados a, periodos de alta inflamación y una línea base. Estos perfiles inmunes se usan para dirigir las decisiones sobre el tratamiento.

B. Utilidad del perfilado inmune en la detección de cáncer

Estos métodos pueden usarse para medir el riesgo de cáncer. El cáncer se ha convertido en la primera causa de muerte en el mundo industrializado. Por lo tanto, los métodos para el tratamiento del cáncer son muy necesarios. Se están intentando muchas estrategias para el tratamiento del cáncer incluyendo el desarrollo de nuevos fármacos que son moléculas pequeñas así como anticuerpos dirigidos al tumor.

Un conjunto de métodos que se ha propuesto es la inmunoterapia. La vigilancia de los tumores es una de las funciones de las células del sistema inmune. Hay varias categorías de antígenos tumorales que son reconocidos por el sistema inmune. La primera categoría está comprendida por antígenos que son nuevos generados por mutación somática (mutación puntual o una translocación) en el tumor. Otra categoría consiste en antígenos de proteínas que sólo se expresan en células germinales masculinas que no expresan moléculas MHC. La desregulación de la expresión génica en muchos tumores puede permitir que algunos de estos antígenos se expresen. Una tercera categoría incluye antígenos de proteínas que sólo se expresan en tejidos particulares. La cuarta categoría comprende antígenos que están sobreexpresados significativamente en el tejido tumoral. Finalmente, la quinta categoría incluye antígenos que resultan de modificación posterior a la traducción anormal.

Una de las propiedades de los tumores es su capacidad para escapar de la eliminación efectiva por el sistema inmune. Se piensa que nuevas mutaciones adquiridas en el tumor le permiten ir desde la fase de equilibrio (en la que el tumor no es eliminado completamente pero su crecimiento se mantiene controlado) a la fase de escape en la que el tumor crece sin control efectivo por el sistema inmune. Hay muchos mecanismos que el tumor emplea para escapar del sistema inmune. Estos mecanismos incluyen la ausencia de péptidos antigénicos específicos, o las moléculas co-estimuladoras que pueden activar las células T. Otros mecanismos incluyen la secreción por el tumor de factores que inhiben las células T y la creación de un sitio privilegiado inducido por el tumor mediante la creación de una barrera física que separa el tumor de los linfocitos. La inducción del sistema inmune para luchar mejor contra el tumor es una estrategia para tratar el cáncer que se está estudiando y ensayando de múltiples maneras. Una estrategia es la terapia de células T adoptivas. Esta estrategia se centra en la identificación de células T dirigidas a antígenos tumorales a través del aislamiento de células que se infiltran en el tumor y/o reaccionan con antígeno tumoral específico. Estas células T pueden crecerse in vitro en condiciones que aumentan su eficacia, como el uso de IL-2 y/o células presentadoras de antígeno. Las células expandidas se infunden entonces de nuevo en la sangre del paciente. Otra estrategia es usar retrovirus que contiene TCR específico de tumor. Estos retrovirus pueden infundirse en el paciente en células especiales que posteriormente secretan el retrovirus permitiéndole infectar células T que entonces empiezan a expresar el TCR específico de tumor. Finalmente, una estrategia común es el uso de vacunación. La premisa de

esta estrategia de terapia es que la inmunización del paciente con uno o más de los antígenos tumorales estimulará la capacidad del sistema inmune de luchar contra el tumor. La inmunización se hace frecuentemente con el uso de un adyuvante como el bacilo de Calmette-Guerin (BCG). Esta estrategia ha tenido éxito en la prevención de cáncer inducido por virus como es evidente por la capacidad de prevenir

5

cánceres cervicales inducidos por HPV-16 y HPV-18. Sin embargo, ha tenido menos éxito en el tratamiento de otros tumores. Gran parte de la mejora en la mortalidad debida al cáncer se ha producido debido a la disponibilidad de métodos de detección temprana mejores que dan lugar por ejemplo a proporciones reducidas de mortalidad en cáncer de mama y cánceres cervicales. La mutabilidad de los tumores hace que el

10

tratamiento temprano sea mucho más efectivo que cuando se detectan tarde. Tradicionalmente, la búsqueda de biomarcadores para la detección del cáncer implicaba habitualmente buscar marcadores que

15

están altamente expresados en el cáncer y a bajo nivel o ausentes en el tejido normal. Esto ha dado lugar a la identificación de varios marcadores tumorales, como PSA. Un problema con la detección temprana en

20

el cáncer es que el mayor valor de la detección el cáncer se produce cuando la detección del biomarcador es más difícil, es decir, el tumor es muy pequeño. Por lo tanto, con el fin de tener un biomarcador para la

25

detección del cáncer efectivo que pueda distinguir a los pacientes con pequeños tumores de aquellos que no, es necesario que haya una tremenda diferencia en la expresión entre el tumor y el tejido normal debido a la gran diferencia en tamaño entre el tumor y el tejido normal. Además, es necesario que el

30

marcador se "vierta" eficientemente en la sangre u otro fluido corporal para permitir la detección usando una técnica no invasiva. En la presente memoria se describe un nuevo mecanismo para la detección del cáncer usando la

35

respuesta de células inmunes. Desde este punto de vista, la detección del cáncer no se consigue por la detección de un marcador producido por el tumor en sí mismo sino por la respuesta del sistema inmune al

40

tumor. Específicamente, el perfil de TCR y/o BCR puede proporcionar una percepción de si el cuerpo está organizando una respuesta a un tumor o no. Esto puede mejorar algunos de los problemas con los

45

biomarcadores actuales. En primer lugar, la respuesta inmune es una señal de amplificación que puede detectarse más fácilmente. En segundo lugar, los linfocitos pasan a través de la sangre regularmente y por lo tanto el biomarcador relevante puede estar presente y ser detectable fácilmente en la sangre

50

periférica que el biomarcador tumoral tradicional. Finalmente, el problema de material biomarcador "de fondo" generado por el tejido normal se reduce en gran medida. La gran diversidad de células T y/o B proporciona una manera de detectar el biomarcador relevante con alta sensibilidad y especificidad,

55

particularmente con la disponibilidad reciente de métodos de alto rendimiento para la secuenciación del ADN. La estrategia de usar la respuesta del sistema inmune frente al cáncer para su detección se aprovecha de las bases expuestas en este campo por la promesa de la inmunoterapia. Sin embargo, el

60

riesgo de las dos aplicaciones es probablemente bastante diferente. El uso de la respuesta inmune frente al cáncer para su detección no requiere que el clonotipo específico sea efectivo para tratar el tumor sino que esté asociado con la respuesta inmune al tumor. Los ensayos de perfilado inmune pueden combinarse con otros marcadores que ya se usan para la

65

detección del cáncer para permitir ensayos con mayor sensibilidad y especificidad. Pueden usarse otros

identificadores o marcadores moleculares para computar el algoritmo de carga o para la determinación del estado de la enfermedad. Los identificadores moleculares pueden incluir ácidos nucleicos, proteínas, carbohidratos, y lípidos, y perfiles de expresión de ácidos nucleicos o proteínas. Los identificadores moleculares pueden ser de origen humano o no humano (por ejemplo, bacteriano). Los identificadores o marcadores pueden determinarse por técnicas que incluyen, por ejemplo, hibridación genómica comparativa (CGH), análisis de micromatrices cromosómicas (CMA), perfilado de expresión, micromatriz de ADN, micromatriz de oligonucleótidos de alta densidad, matriz de expresión de ARN de genoma completo, micromatriz de péptidos, ensayo inmunoabsorbente ligado a enzima (ELISA), secuenciación genómica, análisis del número de copias (CNV), análisis de polimorfismo de nucleótidos pequeños (SNP), inmunohistoquímica, hibridación in-situ, hibridación in-situ fluorescente (FISH), PCR, transferencia Western, transferencia Southern, SDS-PAGE, electroforesis en gel, y transferencia Northern.

#### C. Utilidad del perfilado inmune en medicina de trasplantes

Estos métodos pueden usarse para detectar rechazo inmune de órganos trasplantados. El trasplante de órganos se ha convertido una parte integral de la medicina produciéndose más de 25.000 trasplantes de órganos sólidos (riñón, hígado, corazón, páncreas, y pulmón) y más de 15.000 trasplantes de médula ósea en los EEUU por año. Éstos son generalmente procedimientos complicados realizados en centros de atención terciaria. Para minimizar el riesgo del rechazo de trasplantes, los pacientes se ponen frecuentemente en inmunosupresión durante periodos prolongados de tiempo sometiéndolos al riesgo de

cáncer e infecciones. Además, muchos trasplantes son rechazados bien de forma aguda o años después del trasplante. A pesar de estos problemas, el trasplante de órganos permanece como una modalidad de

tratamiento esencial ya que los pacientes con fallos orgánicos no tienen muchas más alternativas. El rechazo de trasplante de órganos sólidos se produce principalmente debido a la respuesta del sistema

inmune adaptativo al órgano trasplantado. Esto se debe a la presencia de aloantígenos en el injerto que son reconocidos por el sistema inmune del huésped. El rechazo puede ocurrir en tres fases diferentes. La

primera es la fase hiperaguda en minutos después del trasplante en la que anticuerpos preformados

organizan una respuesta frente al injerto. La segunda es el rechazo agudo que se produce en las primeras

semanas o meses después del trasplante. La última es el rechazo crónico que puede ocurrir años

después del trasplante. Dados estos riesgos, se ha tenido cuidado de minimizar las diferencias

inmunogénicas entre el donante y el receptor. Por ejemplo, el riesgo de la reacción hiperaguda se reduce en gran medida cuando el donante y el receptor son concordantes para sus subtipos ABO así como se ensayan para concordancia cruzada (determinando si el receptor tiene anticuerpos que reaccionan con los leucocitos del donante). De forma similar, se hace una concordancia cuidadosa del complejo mayor de histocompatibilidad (MHC) para reducir el rechazo agudo. Sin embargo, dado que las moléculas de MHC son muy polimórficas, es muy difícil encontrar la identidad de una concordancia perfecta. Los gemelos monocigotos tienen una concordancia de MHC perfecta. De forma similar, se espera que los 1/4 de los hermanos tengan una concordancia de MHC perfecta. Los individuos no relacionados que tienen los mismos alelos detectados según el ensayo clínico tienen frecuentemente diferencias debidas a otros sitios polimórficos que no se ensayan en la práctica clínica rutinaria. Si embargo, incluso con concordancia de MHC perfecta de hermanos, todavía existe un riesgo significativo de rechazo debido a la existencia de antígenos de histocompatibilidad menor, y de hecho el rechazo agudo es muy común, produciéndose en más de la mitad de los injertos.

Se podría imaginar que el ensayo más agresivo del locus de MHC así como la identificación y concordancia de los antígeno de histocompatibilidad menores mejoraría significativamente el rechazo de injertos y posiblemente las proporciones de supervivencia. Aunque esto podría ser cierto, el número limitado de órganos donantes disponible hace que esta tarea sea poco práctica ya que un ensayo más agresivo puede retrasar significativamente la identificación de un injerto apropiado para usarse para cada paciente. Por lo tanto, gran parte del progreso que se ha producido en el campo de trasplantes fue en el uso de agentes inmunosupresores para prevenir y tratar el rechazo. Actualmente, se utilizan muchos fármacos para este propósito, incluyendo: azatioprina, corticosteroides, ciclosporina, tacrolimus, ácido micofenolato, sirolimus, muromonab-CD3, anticuerpo monoclonal anti-CD25, anticuerpo monoclonal anti-CD20, e inhibidores de calcineurina.

El trasplante de médula ósea se usa lo más frecuentemente para el tratamiento de la leucemia y linfoma. Típicamente, el receptor experimenta un régimen agresivo de radiación y/o quimioterapia para disminuir la carga del tumor antes del trasplante. Las células T maduras del donante pueden atacar parte de los tejidos del huésped en el rechazo inverso que se denomina enfermedad de injerto frente a huésped (GVHD). Esto se manifiesta frecuentemente por erupción, diarrea, y enfermedad hepática. La concordancia cuidadosa de MHC puede mejorar pero no eliminar este problema. Una solución es la depleción de la médula ósea del donante in vitro de células T maduras que son finalmente responsables de GVHD. Un problema con esto es que el mismo fenómeno que causa GVHD puede ser responsable de parte del efecto terapéutico del trasplante de médula ósea a través del efecto del injerto frente a la leucemia en el que las células T del donante atacan las células de cáncer remanentes. Además, la depleción de las células T del donante puede exponer al paciente al riesgo de ser inmunodeficiente. Por lo tanto, tienen que equilibrarse los riesgos y beneficios cuando se consideran estas estrategias. Los pacientes se tratan por lo tanto frecuentemente con inmunosupresores para prevenir así como tratar GVHD.

La gestión actual del trasplante de médula ósea pero incluso más para el trasplante de órganos sólidos se basa fuertemente en el tratamiento con agentes inmunosupresores fuertes. Sin embargo, dado que estos fármacos tienen riesgos significativos, se usan de una manera en la que se equilibre el riesgo y el beneficio. Sin embargo, dado que el riesgo para un paciente específico en un momento particular no se entiende bien, los pacientes se tratan con la dosis en la que el riesgo y los beneficios están equilibrados para el paciente promedio. Los ensayos que pueden predecir los eventos de rechazo futuros pueden ser potencialmente muy útiles para personalizar el tratamiento para los pacientes en los momentos apropiados en los que los necesitan. Esto puede resultar en la reducción de la dosis de inmunosupresores en algunos de los pacientes a la vez que se mejora la proporción de rechazo y con suerte la supervivencia del injerto.

Los ensayos de perfilado inmune pueden combinarse con otros marcadores que ya se usan para la detección del rechazo de trasplantes para permitir ensayos con mayor sensibilidad y especificidad. Pueden usarse otros identificadores moleculares o marcadores para computar el algoritmo de carga o para determinar el estado de la enfermedad. Los identificadores moleculares pueden incluir ácidos nucleicos, proteínas, carbohidratos, y lípidos, y perfiles de expresión de ácidos nucleicos o proteínas. Los identificadores moleculares pueden ser de origen humano o no humano (por ejemplo, bacteriano). Los identificadores o marcadores pueden determinarse por técnicas que incluyen, por ejemplo, hibridación genómica comparativa (CGH), análisis de micromatrices cromosómicas (CMA), perfilado de expresión, micromatriz de ADN, micromatriz de oligonucleótidos de alta densidad, matriz de expresión de ARN de genoma completo, micromatriz de péptidos, ensayo inmunoabsorbente ligado a enzima (ELISA), secuenciación genómica, análisis del número de copias (CNV), análisis de polimorfismo de nucleótidos pequeños (SNP), inmunohistoquímica, hibridación in-situ, hibridación in-situ fluorescente (FISH), PCR, transferencia Western, transferencia Southern, SDS-PAGE, electroforesis en gel, y transferencia Northern.

D. Utilidad del perfilado inmune en el tratamiento de infección  
Estos métodos tienen utilidad en la guía del tratamiento de infecciones particularmente cuando estas infecciones pueden existir en estados activos y latentes. La llegada de antibióticos para el tratamiento de enfermedades infecciosas durante el siglo pasado ha producido un gran impacto en la esperanza de vida. Durante la década pasada, las técnicas de diagnóstico molecular han tomado un papel rápidamente creciente en el diagnóstico y gestión de las enfermedades infecciosas. La excelente sensibilidad y especificidad proporcionada por la amplificación de ácidos nucleicos ha permitido la aplicación de estas

técnicas a un número creciente de aplicaciones. Muchas de las aplicaciones se usan para la evaluación de diagnóstico de la presencia o ausencia de agentes infecciosos. Por ejemplo, el ensayo de enfermedades de transmisión sexual se hace frecuentemente por un ensayo molecular que emplea una técnica de amplificación de ácido nucleico. Otro conjunto de aplicaciones implica la evaluación de la "carga" de la infección en un paciente con un agente infeccioso ya diagnosticado. Un ejemplo de esto es la evaluación de la carga viral de VIH en pacientes ya diagnosticados con SIDA. Este ensayo ayuda al médico a determinar el estado de la enfermedad del paciente y por lo tanto puede proporcionar una guía sobre la eficacia del régimen de tratamiento que se está usando.

Algunas veces es útil no sólo considerar el nivel del agente infeccioso sino también la respuesta inmune frente al agente infeccioso. Un ejemplo en el que la respuesta inmune frente a la infección se usa rutinariamente en la práctica clínica es en la hepatitis B. Un aspecto del ensayo de la hepatitis B se basa en detectar el agente infeccioso a través de la detección de antígenos de la hepatitis B por un ensayo de amplificación de ácido nucleico. Además, es común en la práctica clínica rutinaria ensayar para la presencia de diferentes anticuerpos dirigidos al virus de la hepatitis B. La presencia de IgM anti-HBc ocurre habitualmente en un entorno de infección aguda, la aparición de IgG anti-HBc indica que la infección es crónica. De forma similar, el surgimiento de anticuerpo anti-HBs señala el aclaramiento de la infección.

El valor de evaluar la respuesta inmune frente a una infección puede emplearse junto con la sensibilidad y especificidad del ensayo molecular. Esto puede ser particularmente útil para enfermedades infecciosas que son crónicas en las que el agente infeccioso permanece latente en el cuerpo. El perfil del TCR y/o BCR puede usarse para evaluar la respuesta inmune frente a una infección. La secuenciación puede usarse para obtener un perfil del TCR y/o BCR lo que permite la detección de clonotipos particulares con alta sensibilidad y especificidad. Para determinar los clonotipos específicos que se correlacionan con la enfermedad se conciben varias estrategias.

Los ensayos de perfilado inmune pueden combinarse con otros marcadores que ya se usan para la detección de los agentes infecciosos para permitir ensayos con mayor sensibilidad y especificidad. Pueden usarse otros identificadores moleculares o marcadores para computar el algoritmo de carga o para determinar el estado de la enfermedad. Los identificadores moleculares pueden incluir ácidos nucleicos, proteínas, carbohidratos, y lípidos, y perfiles de expresión de ácidos nucleicos o proteínas. Los identificadores moleculares pueden ser de origen humano o no humano (por ejemplo, bacteriano). Los identificadores o marcadores pueden determinarse por técnicas que incluyen, por ejemplo, hibridación genómica comparativa (CGH), análisis de micromatrices cromosómicas (CMA), perfilado de expresión, micromatriz de ADN, micromatriz de oligonucleótidos de alta densidad, matriz de expresión de ARN de genoma completo, micromatriz de péptidos, ensayo inmunoabsorbente ligado a enzima (ELISA), secuenciación genómica, análisis del número de copias (CNV), análisis de polimorfismo de nucleótidos pequeños (SNP), inmunohistoquímica, hibridación in-situ, hibridación in-situ fluorescente (FISH), PCR, transferencia Western, transferencia Southern, SDS-PAGE, electroforesis en gel, y transferencia Northern. E. Utilidad del perfilado inmune en el tratamiento de pacientes ancianos

Estos métodos tienen utilidad en la monitorización del estado del sistema inmune en los ancianos. Los ancianos padecen una disminución del sistema inmune denominado inmunosenescencia que afecta su capacidad de responder a infecciones y para incitar respuestas efectivas frente a vacunas (Weinberg et al., 2008). Esto es aparente a partir de las altas proporciones de mortalidad debidas a neumonía en los ancianos (Office for National Statistics, 2005), y su susceptibilidad a infecciones adquiridas en hospitales, tales como *Clostridium difficile* y *Staphylococcus aureus* resistente a meticilina (Health Protection Agency, 2008). Además, se piensa que la disminución en la capacidad del sistema inmune explica la proporción incrementada de cánceres en los ancianos. Además, la inmunosenescencia puede contribuir a otras enfermedades importantes de los ancianos con un componente significativo de los procesos inflamatorios, como la enfermedad de Alzheimer y cardíaca. Una capacidad para predecir qué individuos presentan un riesgo mayor para estos resultados mortales sería útil para los geriatras ya que toman decisiones clínicas acerca de vacunación, tratamiento agresivo de infecciones y hospitalización.

Muchos aspectos del sistema inmune innato y adaptativo están alterados en la inmunosenescencia. Las células T pierden capacidad de respuesta, los macrófagos tienen una capacidad disminuida de presentación de antígeno y secreción alterada de citoquinas, las células asesinas naturales tienen toxicidad reducida, las células dendríticas foliculares no pueden presentar antígenos tan eficientemente, y los neutrófilos pierden capacidad fagocítica. Existe un conjunto menor de células T y B sin estimular y un incremento en el conjunto de memoria y efector dando lugar a una diversidad reducida de repertorios de células T y B dando lugar a la reducción de la capacidad del sistema inmune adaptativo de responder a nuevos antígenos. En particular, los repertorios de célula T que están asociados con citomegalovirus (CMV) están incrementados en gran medida y tanto como el 45% del repertorio de células T total puede estar dedicado a ello. Se ha observado que estas expansiones son menos pronunciadas en los centenarios.

Los estudios han sugerido que los marcadores inmunes pueden predecir la supervivencia en los ancianos. Se ha mostrado que el grado de diversidad del repertorio de células B predice la supervivencia en los ancianos al menos en una población. Aunque se mostró que estas diferencias globales en la diversidad de TCR y BCR predicen resultados clínicos, estos marcadores carecen de especificidad. Un análisis más profundo de los datos de repertorios puede proporcionar una exactitud en la predicción significativamente mayor. Por ejemplo, las expansiones que responden a CMV pueden tener una significancia diferente que

otras expansiones.

El ARN de las células T y B encontrado en sangre periférica puede recogerse de una cohorte longitudinal de pacientes ancianos cuyos historiales clínicos se siguen durante varios años. Los genes de TCR $\alpha$  y TCR $\beta$  y los genes de IgH, IgK e IgL pueden amplificarse en cada una de estas cohortes a varios puntos de tiempo en sus historiales clínicos. Los perfiles de pacientes con una larga supervivencia se comparan con pacientes con una supervivencia corta. En primer lugar, puede obtenerse la medida global de la diversidad. Esto incluirá no sólo el número de diferentes clonotipos identificados sino también su diversidad. Por ejemplo, ¿es el uso del segmento V, D, J el mismo en los dos grupos o en un grupo está más restringido su uso? Por ejemplo, dos muestras pueden tener el mismo número de clonotipos independientes pero los clonotipos para una de las dos muestras no abarca muchos de los segmentos V. Es lógico esperar que esta muestra sería menos versátil en su respuesta a un nuevo antígeno comparada con la otra muestra cuyos clonotipos están distribuidos entre todos los segmentos V.

Además de la diversidad global, se determina si los clonotipos expandidos en los pacientes que tuvieron una supervivencia larga pueden distinguirse sobre la base de algún parámetro de secuencia comparado con los clonotipos en pacientes que tuvieron una supervivencia corta. Esta estrategia puede suplementarse observando los clonotipos que responden a antígenos específicos. Por ejemplo, dada la evidencia disponible, la identificación de clonotipos que responden a CMV puede tener un poder predictivo. La captura de clonotipos de células T que son reactivos frente a CMV en un estudio de descubrimiento puede hacerse a partir de un conjunto de ancianos así como pacientes sanos. Las secuencias de estos clonotipos pueden estudiarse para identificar parámetros que los distinguen de otros clonotipos. Usando este algoritmo predictivo de clonotipos de CMV con la cohorte longitudinal descrita anteriormente, puede evaluarse si la adición de esta información puede añadir a la capacidad de predecir el paciente que sobrevive durante un tiempo largo del que no lo hace.

Los ensayos de perfilado inmune pueden combinarse con otros marcadores que ya se usan para la detección de salud en la población anciana para permitir ensayos con mayor sensibilidad y especificidad. Pueden usarse otros identificadores moleculares o marcadores para computar el algoritmo de carga o para determinar el estado de la enfermedad. Los identificadores moleculares pueden incluir ácidos nucleicos, proteínas, carbohidratos, y lípidos, y perfiles de expresión de ácidos nucleicos o proteínas. Los identificadores moleculares pueden ser de origen humano o no humano (por ejemplo, bacteriano). Los identificadores o marcadores pueden determinarse por técnicas que incluyen, por ejemplo, hibridación genómica comparativa (CGH), análisis de micromatrices cromosómicas (CMA), perfilado de expresión, micromatriz de ADN, micromatriz de oligonucleótidos de alta densidad, matriz de expresión de ARN de genoma completo, micromatriz de péptidos, ensayo inmunoabsorbente ligado a enzima (ELISA), secuenciación genómica, análisis del número de copias (CNV), análisis de polimorfismo de nucleótidos pequeños (SNP), inmunohistoquímica, hibridación in-situ, hibridación in-situ fluorescente (FISH), PCR, transferencia Western, transferencia Southern, SDS-PAGE, electroforesis en gel, y transferencia Northern.

F. Utilidad del perfilado inmune en la administración de vacunas

Estos métodos tienen utilidad en la administración de vacunas. El uso de vacunación ha dado lugar a una gran reducción en la proporción de infecciones de múltiples organismos. Una enfermedad infecciosa que continúa teniendo un impacto en la salud significativo con más de 30.000 muertes al año en EEUU es la gripe. La vacunación de la gripe debe hacerse anualmente ya que la cepa muta rápidamente. La mayor parte de las secuelas graves de la enfermedad ocurren en los ancianos. Desafortunadamente, los ancianos experimentan frecuentemente inmunosenescencia lo que les hace responder inadecuadamente a la vacunación.

Con el fin de distinguir a los pacientes que responden a la vacunación de aquellos que no, es necesario realizar un estudio de descubrimiento. En esta población están disponibles muestras de sangre anteriores y (en uno o más momentos dados) posteriores a la vacunación para una cohorte de pacientes vacunados de la gripe con resultado a la gripe conocido (es decir, si están protegidos posteriormente frente a la infección o no). La secuencia de TCR y/o BCR puede obtenerse de estas muestras. Se determinan los clonotipos que están enriquecidos después de la vacunación en cada paciente. Los clonotipos enriquecidos en los pacientes que respondieron a la vacunación se comparan con un conjunto control de clonotipos (por ejemplo, el resto de los clonotipos en el mismo conjunto de pacientes) para distinguir los clonotipos que se correlacionan de los demás clonotipos. El algoritmo para predecir estos clonotipos se usa entonces para predecir clonotipos que se correlacionan entre pacientes que no respondieron a la vacunación. Los pacientes que no respondieron pueden generar el mismo tipo de clonotipos que aquellos que respondieron pero a niveles menores. Alternativamente, podría ser que los no respondedores generaran una clase distinta de clonotipos. El número de clonotipos que se correlacionan identificado en el no respondedor puede distinguir estas dos posibilidades.

La capacidad de respuesta de un individuo a una vacunación puede monitorizarse obteniendo en primer lugar una muestra de linfocitos del individuo después de la vacunación de la que se aíslan los linfocitos reactivos a la vacuna. Dicho aislamiento se lleva a cabo fácilmente para las células B usando purificación por afinidad convencional con material antigénico de o relacionado con la vacuna como restos de captura unidos a un soporte sólido tal como lechos magnéticos. El aislamiento de las células T también puede llevarse a cabo con medios convencionales, por ejemplo, las patentes U.S. 7.776.562; 7.125.964; 5.635.363; o semejantes.

Se genera un perfil de clonotipos a partir de la muestra aislada de linfocitos para obtener un conjunto de clonotipos que se correlacionan. A puntos de tiempo posteriores, se obtienen muestras de sangre

periférica del individuo y se generan los perfiles de clonotipos. La proporción de cambio de la frecuencia de los clonotipos que se correlacionan en las muestras posteriores se monitoriza para determinar la capacidad de respuesta del individuo a la vacunación. Dicho método de monitorizar la capacidad de respuesta a una vacunación puede implementarse con las etapas siguientes: (a) enriquecer después de una vacunación una muestra de linfocitos de sangre periférica de un individuo para obtener una muestra de linfocitos que responden a la vacuna; (b) determinar un perfil de clonotipos de la muestra de linfocitos que responden a la vacuna para identificar uno o más clonotipos que se correlacionan específicos de paciente con la respuesta a la vacuna; y (c) determinar un nivel de cada uno de los uno o más clonotipos específicos de paciente en un perfil de clonotipos de una muestra de células de sangre periférica obtenida a uno o más tiempos posteriores para monitorizar la capacidad de respuesta del individuo a la vacunación. La capacidad de respuesta puede determinarse por el incremento en la cantidad o frecuencia del uno o más clonotipos específicos de paciente en perfiles de clonotipos medidos posteriormente.

Con los clonotipos que se correlacionan identificados, se construye un algoritmo para generar una puntuación para predecir la probabilidad de inmunización. Los datos de los perfiles de los respondedores a la vacuna y aquellos que no responden se utilizan para generar este algoritmo. Este algoritmo puede usarse entonces para predecir la probabilidad de inmunización en el paciente siguiente usando los clonotipos que se correlacionan predichos de una muestra obtenida después de la inmunización. La predicción se hace a través de la aplicación de otro algoritmo que también se ha generado en el estudio de descubrimiento. Opcionalmente, puede verse ayudado (o sustituido) por los datos de la pre-calibración para limitar la búsqueda para clonotipos que se correlacionan a aquellos que se enriquecieron después de la inmunización.

Los ensayos de perfilado inmune pueden combinarse con otros marcadores que ya se usan para la detección de respuesta a la vacunación para permitir ensayos con mayor sensibilidad y especificidad. Pueden usarse otros identificadores moleculares o marcadores para computar el algoritmo de carga o para determinar el estado de la enfermedad. Los identificadores moleculares pueden incluir ácidos nucleicos, proteínas, carbohidratos, y lípidos, y perfiles de expresión de ácidos nucleicos o proteínas. Los identificadores moleculares pueden ser de origen humano o no humano (por ejemplo, bacteriano). Los identificadores o marcadores pueden determinarse por técnicas que incluyen, por ejemplo, hibridación genómica comparativa (CGH), análisis de micromatrices cromosómicas (CMA), perfilado de expresión, micromatriz de ADN, micromatriz de oligonucleótidos de alta densidad, matriz de expresión de ARN de genoma completo, micromatriz de péptidos, ensayo inmunoabsorbente ligado a enzima (ELISA), secuenciación genómica, análisis del número de copias (CNV), análisis de polimorfismo de nucleótidos pequeños (SNP), inmunohistoquímica, hibridación in-situ, hibridación in-situ fluorescente (FISH), PCR, transferencia Western, transferencia Southern, SDS-PAGE, electroforesis en gel, y transferencia Northern.

G. Utilidad del perfilado inmune en la monitorización de hipersensibilidad inmune (alergia)  
El sistema inmune adaptativo ha evolucionado para responder a antígenos que están asociados con patógenos. Como en el caso de las enfermedades autoinmunes, el sistema inmune puede tener a veces la diana errónea. Mientras en las enfermedades autoinmunes el sistema inmune tiene como diana un auto antígeno, en las reacciones de hipersensibilidad organiza una respuesta frente a estímulos no dañinos como medicaciones, polvo, y alimentos. La hipersensibilidad es muy común con hasta tanto como el 50% de la población de EEUU teniendo alergia a un estímulo medioambiental, y está causada por mecanismos. La hipersensibilidad se divide en 4 tipos. La hipersensibilidad de tipo I es la hipersensibilidad de tipo inmediato y está mediada por IgE. El tipo II se debe frecuentemente a la unión del anticuerpo IgG a un antígeno asociado a la superficie celular. Por ejemplo, una fármaco no dañino que se une a la superficie de la célula puede hacer que la célula sea una diana para IgG anti-fármaco en pacientes que tienen estos anticuerpos. El tipo III está causado por la deposición de complejos antígeno-anticuerpo en los tejidos. Esto ocurre, por ejemplo, cuando la cantidad de antígeno es grande lo que resulta en complejos inmunes pequeños que no pueden aclararse de forma eficiente y en lugar de esto se depositan en las paredes de los vasos sanguíneos. La sensibilidad de tipo IV es una hipersensibilidad de tipo retrasado mediada por células T. El tipo I y tipo IV tienen el mayor impacto en la salud humana.

En la reacción de hipersensibilidad de tipo I, el paciente se sensibiliza a un antígeno no dañino (alérgeno) mediante la producción de anticuerpo IgE frente a él. La exposición posterior al alérgeno induce la activación de células que se unen a IgE, tales como mastocitos y basófilos. Una vez activadas estas células causan la reacción alérgica a través de la inducción de un proceso inflamatorio mediante la secreción de químicos almacenados y síntesis de citoquinas, leucotrienos, y prostaglandinas. La dosis y la ruta de entrada del alérgeno determinan la magnitud de la reacción alérgica que puede variar de síntomas de rinitis alérgica al colapso circulatorio potencialmente mortal en la anafilaxis. Frecuentemente, la reacción de tipo I aguda está seguida posteriormente por otra fase tardía que juega un papel en muchos de los procesos patológicos resultantes. La fase tardía de reclutamiento de células T auxiliares y otras células inflamatorias es esencialmente una reacción de hipersensibilidad de tipo IV. Algunas reacciones alérgicas de tipo I incluyen rinoconjuntivitis estacional (fiebre del heno), alergia alimentaria, anafilaxis inducida por fármaco, dermatitis atópica (eccema), y asma. Éstas son afecciones muy comunes con una importancia creciente causando costes significativos así como morbilidad y mortalidad. Por ejemplo, el asma es una enfermedad crónica que inflige al ~7% de la población de EEUU causando ~4.000 muertes al año. Algunas de estas enfermedades tienen algunos aspectos relacionados. Por ejemplo, los pacientes con dermatitis atópica presentan un riesgo incrementado de tener asma. La alergia alimentaria puede causar vómito y diarrea pero también puede resultar en anafilaxis en un número significativo de pacientes-

30.000 casos resultando en ~200 muertes al año en EEUU. Algunos de los mismos alérgenos que activan mastocitos submucosales en los síntomas nasales causados por la rinitis alérgica también pueden activar los mastocitos en las vías aéreas inferiores causando construcción bronquial, un síntoma típico de asma. Algunas de las reacciones de hipersensibilidad de tipo IV son dermatitis de contacto (por ejemplo, hiedra venenosa), rinitis crónica, asma crónico, y enfermedad celíaca. La enfermedad celíaca es una enfermedad crónica causada por una alergia alimentaria no mediada por IgE. Es una enfermedad del intestino delgado causada por la respuesta alérgica frente al gluten, un componente presente en el trigo y otros alimentos. Más del 95% de los pacientes celíacos tienen un alelo específico de MHC clase II, el HLA-DQ2.

El tratamiento de las reacciones de hipersensibilidad varía, pero frecuentemente tiene dos aspectos: el tratamiento agudo y la gestión crónica o prevención. Algunas de estas afecciones pueden ser potencialmente mortales (anafilaxis, y asma agudo) e implican una atención médica inmediata. La gestión crónica implica en general intentar evitar el alérgeno específico. Esto puede ser efectivo cuando el alérgeno puede identificarse claramente (por ejemplo, alergia a frutos secos), pero esto puede ser difícil cuando el alérgeno está presente ampliamente en el medioambiente, como el polen o el polvo. Por lo tanto, el tratamiento crónico con medicaciones se utiliza frecuentemente para algunas de estas enfermedades (por ejemplo, asma y rinitis alérgica). El nivel de eficacia de la gestión del tratamiento se ensaya finalmente cuando el paciente se re-expone al o a los alérgenos. Por lo tanto, algunos pacientes pueden estar sometidos a sobre o sub-tratamiento. Idealmente, estaría disponible un ensayo que evalúa la actividad de la enfermedad y el grado al que el paciente es propenso de organizar una respuesta de hipersensibilidad. Dicho ensayo permitiría la personalización del tratamiento a las necesidades individuales del paciente.

#### H. Detección de neoplasmas linfoides

Un aspecto de esta invención utilizará tecnologías de secuenciación de nueva generación para evaluar el nivel de reorganizaciones específicas de TCR o BCR en cánceres de los linfocitos. Estas tecnologías de secuenciación pueden obtener lecturas de secuencia a partir de 1 millón o más de moléculas de TCR o BCR aisladas espacialmente individualmente a un coste razonable. Una secuencia presente a 1/1.000.000 o menos todavía puede detectarse de una manera específica usando estas tecnologías permitiendo así detectar una célula de cáncer asociada con una reorganización particular de TCR o BCR a ese nivel. La amplificación múltiple para amplificar todos los tipos diferentes de secuencias para un gen específico puede hacerse a partir de ADN de sangre o médula ósea. Por ejemplo, para amplificar secuencias de IgH, pueden usarse varios cebadores complementarios a todos los segmentos V conocidos y alelos junto con varios cebadores complementarios a todos los segmentos J y alelos. Es importante que se produzca un sesgo pequeño en la amplificación entre las diferentes secuencias. Hemos mostrado que podemos amplificar a partir de ARN los genes de TCR $\beta$  e IgH sólo con pequeñas diferencias en las eficiencias de los diferentes cebadores de V, y de esta manera validar la posibilidad de hacer lo mismo a partir de ADN lo que permitirá la evaluación de células de cáncer incluso cuando el TCR o BCR no se está expresando.

La sensibilidad de esta invención se determina mediante estadística de recuento de las moléculas individuales aisladas espacialmente teniendo en cuenta cualquier sesgo en la amplificación de los diferentes clonotipos. Por lo tanto, se anticipa que esta estrategia ofrecerá más sensibilidad y estará menos sometida a diferencias de sensibilidad para diferentes secuencias de TCR o BCR de las células tumorales cuando se compara con los ensayos de PCR en tiempo real desarrollados para el mismo propósito. Además, para obtener más sensibilidad simplemente se pueden obtener más lecturas de secuenciación. Como los costes de secuenciación continúan cayendo, anticipamos que la sensibilidad a un coste dado continuará mejorando. Con lecturas de secuenciación suficientes, la sensibilidad está limitada por el número de linfocitos en la muestra. Por el contrario, la sensibilidad para el ensayo de PCR en tiempo real está limitada por los fondos causados por las amplificaciones e hibridaciones no específicas de cualesquiera sondas que pueden ser sustanciales.

Con el fin de usar esta invención para monitorizar los cánceres linfoides los clones específicos de un paciente pueden determinarse secuenciando una muestra de diagnóstico de leucemia o linfoma; esto es, los clonotipos específicos de paciente se determinan generando un perfil de clonotipos de una muestra de un tejido relacionado con la enfermedad, en el que los linfocitos que responden a la enfermedad se encuentran en un estado enriquecido. Una vez se determina el perfil de clonotipos, pueden determinarse los niveles de clonotipos relacionados con la enfermedad por comparación con un perfil de clonotipos de una muestra de un tejido que no está asociado con la enfermedad. Después de esto, se determinan los niveles de los clonotipos relacionados con la enfermedad en perfiles de clonotipos de muestras de ese paciente a puntos de tiempo posteriores en el curso de la enfermedad. Preferiblemente, dichas muestras posteriores se toman de tejidos a los que se puede acceder convenientemente, tal como sangre periférica. Pueden usarse las células en la muestra de sangre y alternativamente, puede usarse el ADN o ARN del plasma sin células. No existe ningún requerimiento para la utilización de sondas específicas de un paciente o la utilización de moldes específicos de paciente como estándares, como se requiere para la tecnología actual. En esta realización de la invención, los clonotipos específicos de paciente que se identifican pueden seguirse mediante la obtención de repertorios de secuencia completa y midiendo informáticamente los clonotipos que se correlacionan relevantes sobre la base de la concordancia de secuencias obtenidas con datos almacenados relacionados con las secuencias relevantes para cada paciente. Los tejidos relacionados con la enfermedad para linfomas pueden incluir tejidos linfoides, médula ósea, sangre periférica, y semejantes.

Identificación de clonotipos de cáncer. Con el fin de usar los métodos de secuenciación para monitorizar el



cáncer, es crítico definir los clonotipos del cáncer en cada individuo. Para el ensayo secundario (recurrencia y aplicación en pronóstico para pacientes diagnosticados con neoplasmas linfoides), la identificación del clonotipo de cáncer puede ser frecuentemente bastante directa. Por ejemplo, las muestras de sangre o médula ósea de pacientes con leucemia en el diagnóstico presentan típicamente el clonotipo de cáncer como el clonotipo más frecuente en la muestra. En otros casos de ensayo secundario (por ejemplo, biopsia de algunas muestras de linfoma), el clon de cáncer puede no estar presente a niveles muy altos. Otros clonotipos que son reactivos frente a una variedad de antígenos incluyendo aquellos que están atacando el cáncer se encuentran a una frecuencia mayor. Si el nivel del clonotipo no es en sí mismo suficiente para determinar el clonotipo del cáncer pueden usarse otros criterios. Varios métodos descritos más adelante pueden usarse para identificar los clonotipos del cáncer.

Reorganización cruzada de linaje. Algunos tipos de reorganizaciones de otra manera infrecuentes son comunes en algunos cánceres y así pueden usarse para asociarlas con tumor. Por ejemplo, son comunes las reorganizaciones cruzadas de linaje, como receptor de células T ( $\alpha$ ,  $\beta$ ,  $\gamma$ , y/o  $\delta$ ) en células B o receptor de células B (IgH, IgK, y/o IgL) en células T, especialmente en ALL. La presencia de reorganizaciones cruzadas de linaje es probable que apoye un origen maligno del clonotipo.

Secuenciación de reorganización cruzada de linaje usando secuenciación de tipos celulares específicos. Alternativamente, las células de un tipo pueden aislarse a través de métodos estándar (por ejemplo, lechos magnéticos y FACS) que utilizan la presencia de determinados antígenos en la superficie de las células. La secuenciación puede hacerse para evaluar la presencia de reorganización cruzada de linaje. Por ejemplo, las células B pueden aislarse y puede realizarse la secuenciación de TCR $\beta$ . La presencia de una secuencia de TCR $\beta$  específica enriquecida sería consistente con cáncer. La secuenciación puede realizarse antes y después del enriquecimiento permitiendo la determinación del nivel sin enriquecimiento y el grado de enriquecimiento.

Receptor inmune inactivo. Otra característica que puede ser útil para distinguir células malignas de otras es que las células no malignas necesitan tener un receptor de células inmunes activo. Los linfocitos proliferan en reacción a antígenos y pueden alcanzar altos niveles. Por lo tanto, el alcanzar altos niveles para un clonotipo normal (no canceroso) depende de un receptor inmune activo. La identificación en una biopsia de un alto nivel de secuencia no funcional no es suficiente para identificar cáncer ya que las reorganizaciones no funcionales pueden encontrarse en la misma célula como funcionales debido a reorganizaciones no funcionales del segundo alelo en una célula que contiene una reorganización funcional. El uso de ARN puede permitir la desambiguación de este punto ya que las células tumorales pueden continuar expresando reorganizaciones no funcionales mientras es improbable que esto ocurra en las células normales pero en general son útiles métodos más definitivos. Hay características adicionales que pueden distinguir secuencias no funcionales en el cáncer. Por ejemplo, algunos de los cánceres inmaduros, por ejemplo, ALL, tienen frecuentemente sólo la IgH reorganizada pro no IgK o IgL. Es improbable que este patrón alcance una alta frecuencia en ausencia de cáncer. Estas reorganizaciones no funcionales pueden evaluarse usando las técnicas siguientes.

Asociación estadística en una serie de muestras diluidas. Una alternativa a la asociación física es buscar asociación de mapeo como se ha descrito anteriormente. En este caso, sólo un gen (por ejemplo, IgH) se está ensayando y la cuestión es cuáles son las secuencias no funcionales asociadas, es decir, cuál es la secuencia del otro alelo de IgH en la misma célula. Los alelos no funcionales con alta frecuencia asociados a un segundo alelo no funcional son consistentes con alcanzar una alta frecuencia debido al cáncer.

Secuenciación de un tipo celular específico. Este patrón también puede identificarse capturando células que portan un marcador y evaluando la secuencia del otro. Por ejemplo, se puede usar FCM para capturar células que son negativas para IgK e IgL. La secuenciación de IgH antes y después del enriquecimiento FCM puede identificar clones que están enriquecidos en esta población. No se espera que las células que son negativas para IgK e IgL alcancen una alta frecuencia y su presencia es consistente con un cáncer inmaduro, como ALL.

Hipermutaciones somáticas inactivadoras. Un patrón alternativo puede encontrarse en linfomas de células B cuyos clonotipos han experimentado hipermutaciones somáticas. Algunos de estos clones pueden tener variabilidad inter-clonal, en la que las células de cáncer comprenden varios clones con diferentes mutaciones en ellos. Algunos de los clones resultantes pueden tener posiblemente mutaciones inactivadoras. Para una hipermutación somática normal dirigida por antígeno es improbable que un clon con una hipermutación somática inactivadora se seleccione y expanda. La presencia de dichos clones es consistente con cáncer.

Secuenciación de fracciones celulares que portan marcadores de cáncer. En los casos en los que se conocen los marcadores del cáncer, puede realizarse FCM para enriquecer para células de cáncer usando este marcador. La secuenciación del repertorio de receptor inmune puede realizarse antes y después del enriquecimiento. Los clonotipos que se enriquezcan probablemente están asociados con el clonotipo del cáncer. Por ejemplo, las células de linfoma pueden enriquecerse usando FACS para aislar las células que portan los marcadores particulares (lo más convenientemente, marcadores de superficie) relevantes para el tumor. La secuenciación de BCR antes y después del enriquecimiento identificaría fácilmente el clonotipo enriquecido y por extensión el clonotipo del cáncer.

Alternativamente, la asociación de un marcador puede evaluarse a nivel de ADN o ARN. Esto puede conseguirse por varios medios incluyendo PCR asociada o asociación estadística con células diluidas de forma seriada como se ha descrito anteriormente. La cuantificación de la asociación de estos marcadores

permitirá un mejor rendimiento del ensayo ya que muchos marcadores están sobreexpresados en las células de cáncer pero todavía están presentes a algún nivel en las células normales. Para tener en cuenta esto, la PCR asociada puede hacerse usando tres genes: el gen del receptor inmune, del marcador de cáncer, y el control. El gen del receptor inmune puede asociarse con cualquiera de los otros dos genes y la fracción de moléculas asociadas que son el resultado de una asociación entre la reorganización del receptor y el marcador del cáncer puede indicar el nivel de expresión de este marcador del cáncer.

Detección de translocaciones. Además de servir como un marcador de células que se han vuelto cancerosas, IgH es frecuentemente una de las dos parejas de translocación patológica en los neoplasmas linfoides. Un ejemplo es la t(11:14) que pone el segmento J de IgH muy cerca del gen de la ciclina D1 (CCND1) lo que resulta en su sobreexpresión. Esta reorganización que es referida como BCL1-IgH ocurre en tanto como el 60-70% de los linfomas de células del manto así como otros neoplasmas linfoides (por ejemplo, 20% del mieloma múltiple). Otro ejemplo es t(14:18) que pone el segmento J de IgH muy cerca de BCL2 lo que resulta en su sobreexpresión. Esta reorganización ocurre en hasta el 90% del linfoma folicular y 20% del linfoma de células B grandes. Estas reorganizaciones se identifican típicamente por citogenética, transferencia Southern, o FISH. La PCR tiene el potencial de identificar reorganizaciones con una sensibilidad y especificidad muy altas como se muestra por BCR-ABL para la detección del cromosoma Filadelfia. Se han usado diferentes técnicas de PCR para la evaluación de translocaciones relevantes para el linfoma, siendo probablemente la recientemente introducida PCR en tiempo real (por ejemplo, BCL2-IgH) la más avanzada. Hay unas pocas características de BCL1-IgH y BCL2-IgH que hacen que su detección sea menos sensible y específica que la de BCR-ABL. En primer lugar, a diferencia de BCR-ABL, BCL1-IgH y BCL2-IgH no generan una proteína de fusión, y no hay un evento de corte y empalme que genere una estructura molecular predecible. En lugar de esto, los puntos de rotura pueden abarcar una región grande. Existen puntos de rotura comunes que permiten la detección de hasta el 88% de BCL2-IgH usando una combinación de cebadores y ~40% del BCL1-IgH. Esto resulta en la pérdida de algunos pacientes que tienen la translocación. En segundo lugar, estas reorganizaciones pueden estar presentes en los individuos normales que nunca tendrán cáncer. Por ejemplo, la translocación BCL2-IgH se ha encontrado al nivel de  $\sim 10^{-5}$  en una gran fracción de los individuos normales con más del ~4% portando BCL2-IgH a una frecuencia de  $>1/25K$ . La frecuencia de BCL2-IgH es mayor al incrementarse la edad. También se establece la hipótesis de que diferentes personas pueden tener diferentes niveles de translocación "de fondo". Presumiblemente, la presencia de esta translocación en una muestra normal se debe al hecho de que la tumorigénesis es un proceso con múltiples etapas y la BCL2-IgH no es suficiente para que surjan los tumores. La presencia de este fondo de bajo nivel establece un límite en la sensibilidad de la detección.

La amplificación de un conjunto de los cebadores de J complementarios a todos los segmentos J y cebadores complementarios a las regiones aguas arriba de los puntos de rotura de las translocaciones BCL1 o BCL2. pueden secuenciarse. Esto puede generar un método para la detección sensible de estas translocaciones y de las células de cáncer que aparecen. En primer lugar, la secuenciación profunda de moléculas individuales aisladas (por ejemplo, 100K o 1 millón de lecturas) puede permitir la detección de las secuencias apropiadas a partir de un pequeño número de células en un fondo de amplificaciones de otros loci. Además, el problema de las translocaciones de fondo en individuos normales puede mejorar el problema que experimenta la PCR en tiempo real. Existe evidencia de que, al menos en algunos casos, las translocaciones de fondo no son clonales sino que aparecen repetidamente en el mismo paciente. Usando la secuenciación, se pueden distinguir los diferentes eventos de translocación para obtener la frecuencia de los eventos de translocación independientes. Como el punto de rotura de las diferentes translocaciones probablemente sea distinto, pueden distinguirse eventos de translocación uno de otro. Alternativamente o adicionalmente, una PCR de asociación usando la translocación con un gen de receptor de células B o T puede hacerse para proporcionar un código de barras único para la translocación. La asociación también puede hacerse estadísticamente usando un conjunto de muestras de dilución como se ha descrito anteriormente.

La monitorización seriada del nivel de la translocación para detectar puntos en los que se incrementa en frecuencia puede ser útil en la detección temprana del cáncer así como para la detección de recurrencia. En este último caso, el punto de rotura específico relevante para el paciente puede identificarse a partir de la biopsia de diagnóstico. Este concepto de distinguir translocaciones por sus puntos de rotura a través de secuenciación y por lo tanto distinguir el fondo del cáncer puede extenderse a otras translocaciones que implican IgH (por ejemplo, t(8:14)) o todas las demás translocaciones en neoplasma linfóide u otros cánceres.

Cambio en los niveles de células de cáncer y probabilidad de cáncer. La mera presencia de secuencias que apuntan a la existencia de células tumorales remanentes puede no predecir en sí misma una recidiva clínica. Por ejemplo, un estado estacionario del nivel tumoral puede conseguirse por las fuerzas de equilibrio de la proliferación de células tumorales y la respuesta inmune al tumor. Se anticipa que además del nivel absoluto de un clonotipo, su proporción de cambio puede ser informativa para predecir la probabilidad de recidiva. Por ejemplo, considérense dos pacientes cada uno con un nivel X de sus clonotipos de cáncer específicos de paciente respectivos. Si el nivel en ensayos previos para uno de los pacientes ha sido de manera consistente X y el nivel para el otro paciente en ensayos previos ha sido considerablemente menor entonces la probabilidad de que el segundo paciente desarrolle una recidiva podría ser mayor que para el primer paciente.

De forma similar, los datos adicionales respecto al estado de la célula que contiene el clonotipo relacionado con cáncer pueden usarse para predecir la probabilidad de recurrencia. Por ejemplo, la presencia de determinados marcadores (de superficie o no de superficie) puede ser una indicación del estado funcional de la célula y por lo tanto la probabilidad de recurrencia. La secuenciación antes y después de la captura de las células con los marcadores relevantes puede determinar la fracción de células con el clonotipo del cáncer que porta los marcadores relevantes. De forma similar, algunos marcadores relevantes para la probabilidad de recurrencia (por ejemplo, expresión de algún gen relacionado con el crecimiento celular) pueden evaluarse a nivel del ARN. Esto puede hacerse por varios métodos incluyendo PCR de asociación o estadísticamente por dilución celular como se ha descrito anteriormente. Finalmente, es posible que el nivel del ARN específico del receptor inmune en la célula tumoral pueda tener una consecuencia funcional y asociación con la probabilidad de recurrencia. Este nivel puede evaluarse haciendo PCR de asociación entre un gen control 1 que puede asociarse bien a la reorganización del receptor inmune o gen control 2. La fracción relativa de los dos productos puede ser indicativa de la cantidad relativa del ARN en la célula. Otro método implica comparar el nivel de ARN con el nivel de ADN de la reorganización del receptor inmune. La frecuencia del clonotipo específico de cáncer en el ADN identifica el nivel relativo del clonotipo específico de cáncer. La frecuencia del mismo clonotipo puede evaluarse a partir del ARN, y la frecuencia relativa en el ARN y en el ADN puede seguirse. Un cambio en esta frecuencia relativa puede ser indicativo de un cambio en la probabilidad de recurrencia. Reacción inmune frente a cáncer linfóide. Además de monitorizar el clonotipo del cáncer y su progenie potencial, también se puede evaluar la respuesta inmune frente al tumor. Se pueden identificar los clonotipos que probablemente organicen una respuesta frente al tumor. Por ejemplo, los clonotipos de células B o T que están enriquecidos en la biopsia de la biopsia de nodo linfático de diagnóstico pueden ser el resultado de la respuesta inmune frente al tumor. Además, puede hacerse el ensayo funcional para identificar células T que interactúan con algún antígeno tumoral. Éstos pueden ser antígenos específicos o pueden ser las células tumorales en sí mismas. Por ejemplo, la secuenciación del TCR antes y después de la estimulación con el antígeno o células tumorales puede identificar las células T relevantes gracias a su enriquecimiento después de la estimulación. El nivel de estas células T en muestras de sangre posteriores del paciente puede ser útil para predecir la recidiva. Por ejemplo, considérese el caso mencionado anteriormente en el que un nivel X específico del clonotipo del cáncer se ha detectado de manera estable. Esto es probablemente el resultado de equilibrar el crecimiento tumoral con la respuesta inmune frente al tumor. Si en algún punto la respuesta inmune frente al tumor se reduce es posible anticipar que el tumor recidivará. La respuesta inmune frente al tumor puede cuantificarse a través de la determinación del nivel de los clonotipos de células T que se han determinado que son capaces de atacar el tumor.

Integración de secuenciación con otros marcadores celulares. La detección del clonotipo específico del cáncer puede hacerse a través de la secuenciación de reorganizaciones de receptor inmune como se ha descrito anteriormente. La presencia de marcadores (de superficie o no de superficie) que son relevantes para las células de cáncer puede usarse para capturar subpoblaciones celulares que posteriormente se secuencian. La combinación del uso de captura específica de marcador y secuenciación antes y después de la captura puede proporcionar información adicional. En primer lugar, esto puede usarse para identificar los clonotipos que están enriquecidos y por lo tanto probablemente son clonotipos específicos de cáncer como se ha descrito anteriormente. Además, la secuenciación de células que tienen marcadores específicos de cáncer puede dar lugar a sensibilidades mayores. Con un marcador perfecto, sólo es necesario hacer pocas lecturas para detectar clonotipos específicos de cáncer en millones de otras células. La mayor parte de los marcadores no tienen un rendimiento perfecto y se genera un fondo (es decir, células no cancerosas) significativo por su captura. Sin embargo, el enriquecimiento de células de cáncer por estos marcadores puede dar lugar bien a sensibilidad equivalente con menos lecturas de secuenciación o a mayor sensibilidad que hacer el mismo número de lecturas de secuenciación sin enriquecimiento. Por ejemplo, con 1 millón de lecturas de secuenciación se puede evaluar ~1 millón de células que se capturan con un marcador específico de cáncer. Esto corresponde a más células de las estaban presentes antes de la captura y por lo tanto a una mejor sensibilidad. Finalmente, el uso de marcadores puede proporcionar aspectos funcionales que se relacionan con la biología tumoral y el pronóstico. Puede estar presente algún nivel de células tumorales en la sangre de diferentes muestras, pero el marcador funcional en las células puede distinguir las muestras que indican una alta probabilidad de recurrencia de aquellas que predicen baja probabilidad de recurrencia. Por ejemplo, pueden secuenciarse muestras antes y después de la captura con un marcador relevante y puede evaluarse el porcentaje de las secuencias de clonotipos con el marcador específico. Puede predecirse que dos muestras con el mismo nivel total de clonotipos específicos de cáncer pero diferentes fracciones de aquellas células que portan el marcador relevante tienen una probabilidad diferente de recurrencia.

**Evaluación de clones y detección de clones filogenéticos (clanes) y otras mutaciones relacionadas con cáncer**

Como se ha mencionado anteriormente, los métodos de la invención monitorizan un nivel de un clan de clonotipos en lugar de un clonotipo individual. Esto se debe al fenómeno de evolución clonal, por ejemplo, Campbell et al, Proc. Natl. Acad. Sci., 105: 13081-13086 (2008); Gerlinger et al, Br. J. Cancer, 103: 1139-1143 (2010). La secuencia de un clon que está presente en la muestra de diagnóstico puede no permanecer exactamente la misma que la de una en una muestra posterior, tal como una tomada después de una recidiva de la enfermedad. Por lo tanto, si se está siguiendo la secuencia de clonotipo exacta que

concuera con la secuencia de la muestra de diagnóstico, la detección de una recidiva puede fracasar. Dicho clon evolucionado se detecta e identifica fácilmente por secuenciación. Por ejemplo, muchos de los clones evolucionados surgen por el reemplazo de la región V (denominado reemplazo VH). Estos tipos de clones evolucionados se pierden por las técnicas de PCR en tiempo real ya que los cebadores están dirigidos al segmento V erróneo. Sin embargo, dado que la unión D-J permanece intacta en el clon evolucionado, puede detectarse e identificarse en esta invención usando la secuenciación de moléculas individuales aisladas espacialmente. Además, la presencia de estos clonotipos relacionados a una frecuencia apreciable en la muestra de diagnóstico incrementa la probabilidad de la relevancia del clonotipo. De forma similar, el desarrollo de hipermutaciones somáticas en la secuencia del receptor inmune puede interferir con la detección de la sonda de PCR en tiempo real, pero los algoritmos apropiados aplicados al resultado de la secuenciación (como se ha descrito anteriormente) todavía pueden reconocer un clonotipo como un clonotipo evolucionado. Por ejemplo, pueden reconocerse las hipermutaciones somáticas en los segmentos V o J. Esto se hace mapeando los clonotipos hasta las secuencias de V y J de la línea germinal más cercana. Las diferencias respecto a las secuencias de la línea germinal pueden atribuirse a hipermutaciones somáticas. Por lo tanto, los clonotipos que evolucionan a través de hipermutaciones somáticas en los segmentos V o J pueden detectarse e identificarse fácilmente. Pueden predecirse las hipermutaciones somáticas en la región NDN. Cuando el segmento D remanente es lo suficientemente largo como para ser reconocido y mapeado, cualquier mutación somática en él puede reconocerse fácilmente. Las hipermutaciones somáticas en las bases N+P (o en el segmento D que no es mapeable) no pueden reconocerse para algunas ya que estas secuencias pueden modificarse en células recién recombinadas que pueden no ser la progeñe del clonotipo canceroso. Sin embargo, se construyen fácilmente algoritmos para identificar cambios de bases que tienen una alta probabilidad de deberse a mutación somática. Por ejemplo, un clonotipo con los mismos segmentos V y J y 1 diferencia de base en la región NDN respecto al o a los clones originales tiene una alta probabilidad de ser el resultado de recombinación somática. Esta probabilidad puede incrementarse si hay otras hipermutaciones somáticas en los segmentos V y J porque esto identifica a este clonotipo específico como uno que ha sido objeto de hipermutación somática. Por lo tanto, la probabilidad de que un clonotipo sea el resultado de hipermutación somática respecto a un clonotipo original puede computarse usando varios parámetros: el número de diferencias en la región NDN, la longitud de la región NDN, así como la presencia de otras hipermutaciones somáticas en los segmentos V y/o J.

Los datos de evolución clonal pueden ser informativos. Por ejemplo, si el clon principal es un clon evolucionado (uno que está ausente previamente, y por lo tanto, no registrado previamente) entonces esto es una indicación de que el tumor ha adquirido nuevos cambios genéticos con ventajas selectivas potenciales. Esto no quiere decir que los cambios específicos en el receptor de las células inmunes sean la causa de la ventaja selectiva sino que pueden representar un marcador para ésta. Los tumores cuyos clonotipos han evolucionado pueden estar asociados potencialmente con un pronóstico diferencial. En un aspecto de la invención, un clonotipo o clonotipos que se están siendo usados como un biomarcador específico de paciente de una enfermedad, tal como un neoplasma linfoide, por ejemplo, una leucemia, incluyen clonotipos no registrados previamente que son mutantes somáticos del clonotipo o clonotipos que se están monitorizando. En otro aspecto, siempre que cualquier clonotipo no registrado previamente es al menos noventa por ciento homólogo respecto a un clonotipo existente o grupo de clonotipos que sirven como biomarcadores específicos de paciente, entonces dicho clonotipo homólogo se incluye con o en el grupo de clonotipos que se están monitorizando hacia delante. Esto es, si uno o más clonotipos específicos de paciente se identifican en un neoplasma linfoide y se usan para monitorizar periódicamente la enfermedad (por ejemplo, haciendo medidas en muestras de sangre adquiridas de forma menos invasiva) y si en el curso de una de dichas medidas se detecta un nuevo (no registrado previamente) clonotipo que es una mutación somática de un clonotipo del conjunto actual, entonces se añade al conjunto de clonotipos específico de paciente que se monitorizan para medidas posteriores. En una realización, si dicho clonotipo no registrado previamente es al menos noventa por ciento homólogo respecto a un miembro del conjunto actual, entonces se añade al conjunto específico de paciente de biomarcadores clonotipo para el siguiente ensayo que se lleva a cabo en el paciente; esto es, dicho clonotipo no registrado previamente se incluye en el clan del miembro del conjunto actual de clonotipos del que derivó (sobre la base del análisis anterior de los datos del clonotipo). En otra realización, dicha inclusión se lleva a cabo si el clonotipo no registrado previamente es al menos noventa y cinco por ciento homólogo respecto a un miembro del conjunto actual. En otra realización, dicha inclusión se lleva a cabo si el clonotipo no registrado previamente es al menos noventa y ocho por ciento homólogo respecto a un miembro del conjunto actual.

También es posible que una célula evolucione a través de un proceso que reemplaza la región NDN pero conserva el segmento V y V junto con sus mutaciones acumuladas. Dichas células pueden identificarse como clonotipos del cáncer no registrados previamente por la identificación del segmento V y J común siempre que contengan un número suficiente de mutaciones para volver la posibilidad de estas mutaciones independientemente derivada pequeña. Una limitación adicional puede ser que la región NDN tenga un tamaño similar al clon secuenciado previamente.

**Evaluación del número suficiente de células**

La sensibilidad del ensayo está limitada por el número de células que generan el molde de ácido nucleico que se usa en la reacción de amplificación. Típicamente, están presentes ~6 pg de ADN en cada célula. Por lo tanto, para tener una sensibilidad de 1/1.000.000, es necesario usar ~6 µg de ADN. Sin embargo,

en la sangre periférica sólo una fracción de las células son células B y por lo tanto ~6 µg de ADN de sangre periférica pueden tener sólo ~100.000 células B. Para obtener una mayor sensibilidad pueden usarse cantidades mayores de ADN. Un problema puede ser cuanto más ADN se usa el efecto de los inhibidores purificados con el ADN puede ser más profundo y puede observarse una variación de una muestra a otra. La obtención de una población más pura de células puede mejorar este problema. La generación de células mononucleares de sangre periférica (PBMC) se hace frecuentemente en los entornos clínicos. ~6 µg de ADN de PBMC pueden tener ~250.000-300.000 células B. La captura de células B específicamente también puede hacerse para obtener más células B por µg de ADN usado.

Puede seguirse más de una reorganización de receptor inmune para maximizar la sensibilidad y mejorar el problema de evolución clonal. Por lo tanto, si se están siguiendo 3 reorganizaciones la división de las células disponibles entre ellas disminuiría la sensibilidad en el análisis de cada reorganización. Por lo tanto, la amplificación del ADN (o ARN) de manera que los 3 loci de las reorganizaciones se amplifiquen antes de la división mejora este problema. Las metodologías de amplificación del genoma completo se han empleado previamente y pueden usarse aquí para conseguir la amplificación de los tres loci antes de la división. Alternativamente, puede emplearse la amplificación para los loci específicos en una reacción para conseguir lo mismo. En este caso, es opcional una división posterior para amplificar cada reorganización separadamente. La amplificación del genoma completo antes de la amplificación de la reorganización de receptor inmune específica también puede ser útil cuando se emplea para evaluar sólo una reorganización de receptor inmune. Por ejemplo, la evaluación de IgH frecuentemente es complicada con las hipermutaciones somáticas haciendo que el uso de múltiples conjuntos de cebadores sea frecuentemente deseable. En este caso, la amplificación del genoma completo antes de la división del ácido nucleico de entrada entre diferentes reacciones con diferentes conjuntos de cebadores no da lugar necesariamente a una sensibilidad mejorada para detectar el clonotipo específico del cáncer. En este caso, las diferentes (por ejemplo, 3) reacciones evalúan el repertorio completo de ácido nucleico de entrada disponible y por lo tanto no habría ninguna ventaja de la amplificación del genoma completo (aparte de conservar el ADN para otra pregunta). Sin embargo, esto no es cierto cuando la hipermutación somática ocurre en la secuencia complementaria a un cebador. Por ejemplo, si sólo hay una molécula de ADN que representa el clonotipo específico del cáncer entonces iría a una de las tres reacciones. Una hipermutación somática evita que este clonotipo específico se amplifique. Por otra parte, la amplificación del genoma completo evita esto ya que la única molécula inicial de secuencia específica de cáncer se amplifica y por lo tanto estaría presente en los 3 tubos. De manera que aunque la frecuencia del clonotipo específico del cáncer no se incrementa en el molde de entrada para la amplificación del locus, el hecho de que esté presente en los 3 tubos es una ventaja. En lugar de la amplificación del genoma completo, puede hacerse la amplificación específica de locus usando estrategias como PCR de rango largo o usando cebadores de los tres conjuntos de cebadores o amplificación preliminar.

Cribado para neoplasmas linfoides. Los métodos anteriores se aplican a la monitorización de pacientes después del diagnóstico inicial; sin embargo, esta invención también se aplica al cribado de cáncer. El cribado para cáncer primario ha sido una fuerza mayor para reducir la mortalidad. La detección temprana de neoplasmas linfoides puede dar lugar a una gran mejora en las proporciones de supervivencia de estos cánceres. Se ha mostrado al menos en leucemia linfocítica aguda y crónica que los clonotipos específicos del cáncer pueden detectarse años antes del diagnóstico final. Es posible que los clonotipos de linfoma también puedan detectarse antes de lo que es posible usando las metodologías de diagnóstico actuales. La detección de los clonotipos del cáncer para propósitos de cribado (es decir, antes de que se produzca un tumor primario) puede hacerse usando esta invención mediante la secuenciación de receptores de células inmunes como se ha descrito anteriormente. El clonotipo del cáncer en cada paciente es probablemente único, y está claro que no se sabe *a priori* la secuencia que se va a cribar en cada paciente. Muchos de los métodos listados anteriormente pueden usarse, sin embargo, en la sangre de un paciente que todavía no ha sido diagnosticado con cáncer con el fin de identificar clonotipos que están asociados probablemente con cáncer y sus niveles y cambios en estos niveles pueden usarse para evaluar el riesgo de que un paciente desarrolle cáncer clínico.

Tipos de neoplasmas linfoides. Los métodos de la invención proporcionada pueden usarse para monitorizar neoplasmas linfoides, por ejemplo, linfoma o leucemia. Los neoplasmas de células B maduras pueden incluir, por ejemplo, leucemia linfocítica crónica/linfoma linfocítico pequeño, leucemia prolinfocítica de células B, linfoma linfoplasmacítico (tal como macroglobulinemia de Waldenström), linfoma de la zona marginal esplénica, neoplasmas de células plasmáticas (mieloma de células plasmáticas, plasmacitoma, enfermedades de deposición de inmunoglobulina monoclonal, y enfermedades de cadena pesada), linfoma de células B de la zona marginal extranodal (linfoma MALT), linfoma de células B de la zona marginal nodal (NMZL), linfoma folicular, linfoma de células del manto, linfoma de células B grandes difuso, linfoma de células B grandes mediastinal (tímico), linfoma de células B grandes intravascular, linfoma de efusión primaria, y linfoma/leucemia de Burkitt.

Los neoplasmas de células T maduras pueden incluir, por ejemplo, leucemia prolinfocítica de células T, leucemia linfocítica granular grande de células T, leucemia/linfoma de células T adultas, linfoma de células T extranodal (tipo nasal), linfoma de células T tipo enteropatía, linfoma de células T hepatoesplénico, micosis fungoides/síndrome de Sezary, trastornos linfoproliferativos de células T cutáneos primarios positivos para CD30, linfoma de células grandes anaplásico cutáneo primario, papulosis linfomatoide, linfoma de células T angioinmunoblástico, linfoma de células T periférico (no especificado), o linfoma de células grandes anaplásico.

Los métodos de la invención proporcionada pueden usarse para monitorizar leucemia aguda o leucemia crónica. La leucemia puede ser leucemia linfoblástica aguda (ALL) (por ejemplo, leucemia linfoblástica aguda de precursor de B, leucemia linfoblástica aguda de precursor de T, leucemia de Burkitt, y leucemia bifenotípica aguda); leucemia linfocítica crónica (CLL) (por ejemplo, leucemia prolinfocítica de células B); leucemia mielógena aguda (AML) (por ejemplo, leucemia promielocítica aguda, leucemia mieloblástica aguda, y leucemia megacarioblástica aguda); leucemia mielógena crónica (CML) (por ejemplo, leucemia de monocitos crónica); leucemia de células pilosas; leucemia prolinfocítica de células T (T-PLL); o leucemia linfocítica granular grande.

#### I. Cribado de cáncer

Otro ejemplo de una situación en la que el perfilado de células T y/o B específicas de antígeno puede usarse para informar sobre situaciones clínicas es el uso de autoantígenos de cáncer específicos en el cribado del cáncer. Las células de cáncer producen frecuentemente moléculas aberrantes que son antigénicas y provocan una respuesta inmune. Dichas moléculas pueden cribarse directamente en la sangre o fluidos corporales con el fin de evaluar la probabilidad de que un paciente vaya a desarrollar cáncer. Se ha mostrado que la sensibilidad de estos métodos es un factor limitante, sin embargo, ya que estos antígenos se encuentran en concentraciones muy pequeñas en la sangre u otros fluidos corporales cuando los cánceres están en un estadio temprano cuando lo más probablemente es que sean tratables y asintomáticos. Estas moléculas antigénicas pueden provocar, sin embargo, una respuesta inmune que, aunque puede ser insuficiente para controlar el crecimiento tumoral, puede tener una fuerza suficiente para ser mensurable como una manera de detectar el cáncer temprano. Los anticuerpos frente a antígenos específicos, por ejemplo, de células de cáncer de pulmón y mama se han detectado y pueden usarse como un medio para cribar para estos cánceres usando los antígenos en sí mismos para capturar y detectar los anticuerpos (M. Nesterova et al, *Biochimica et Biophysica Acta* 2006; 1762: 398-403). Como se ha descrito anteriormente, esos ensayos carecen de la capacidad de diferenciar entre los diferentes clones de anticuerpos y pierden cualquier respuesta inmune de células T potencial frente a los antígenos tumorales. Los métodos descritos anteriormente pueden usarse para enriquecer para células T y/o células B que expresan TCR o BCR que se unen a antígenos que se sabe que existen en células de cáncer. Usando los métodos descritos en la presente memoria se pueden generar frecuencias de clonotipos antes y después del enriquecimiento en T/BCR específicos de antígeno con el fin de identificar células T o B que probablemente formen parte de una respuesta inmune frente a estos antígenos en un individuo particular. Hay varias maneras en las que este método se podría implementar clínicamente. En el primer caso, la sangre de un individuo cuyo riesgo de desarrollar un tumor peligroso se va a evaluar se muestrea y perfila para perfiles de clonotipos de células T y B completos antes y después del enriquecimiento para células que se unen a un conjunto específico de antígenos de cáncer. Estos antígenos pueden ser una única especie de antígeno, un conjunto de antígenos, o una mezcla compleja de antígenos o podría ser la mezcla completa de material de un único tumor de una mezcla de tumores. Estos antígenos podrían incluir p53, c-myc, NY-ESO-1, BRCA1, BRCA2, HER2, MUC1, CAGE, Sox2, GBU4-5, Anexina1, cox-2. Los clones que probablemente estén asociados con una respuesta inmune frente a estos antígenos podrían identificarse por un desplazamiento significativo en la frecuencia antes y después del enriquecimiento, además de otros parámetros incluyendo parámetros de secuencia de ácido nucleico y aminoácidos, longitud, uso de segmentos, etc. El riesgo de desarrollar cáncer podría determinarse sobre la base de una medida en un único punto de tiempo de este tipo o de al menos un segundo punto de tiempo en el que los cambios en la frecuencia de estos clonotipos se usan para calcular un riesgo de cáncer. El enriquecimiento podría hacerse en un primer punto de tiempo para identificar los clonotipos específicos de antígeno que se correlacionan y estos clonotipos específicos de antígeno que se correlacionan cuantificarse en puntos de tiempo posteriores en los que el enriquecimiento no se lleva a cabo con el fin de calcular una puntuación de riesgo.

También podría usarse un estudio poblacional para generar un algoritmo para predecir clonotipos que se correlacionan a partir del enriquecimiento específico de antígeno. En este estudio poblacional se obtiene más de un individuo con riesgos de cáncer conocidos y los clonotipos específicos de antígeno se identifican usando las técnicas de esta invención. Esta población se usa para generar un algoritmo que puede usarse para predecir clonotipos específicos de antígeno que son relevantes para la enfermedad en un nuevo individuo cuyo riesgo de cáncer es desconocido.

#### J. Reacciones adversas a fármacos

Los beneficios del tratamiento con fármacos se equilibran frecuentemente con sus efectos adversos. La mayoría de reacciones adversas a los fármacos (ADR) son bastante predecibles y dependientes de la dosis. Otras ADR son idiosincráticas, y muchas de éstas están causadas por mecanismos inmunológicos. La predisposición a varias de estas ADR está asociada con genotipos HLA específicos. Idealmente, los pacientes sabrían antes de tomar la medicación si están predispuestos a tener una ADR. Recientemente, la FDA añadió en la etiqueta de la medicación de VIH, Abacavir, una recomendación para ensayar para el alelo HLA-B\*5701 antes de iniciar el tratamiento ya que los pacientes con este genotipo están predispuestos a la reacción de hipersensibilidad al fármaco. En los casos en los que no es posible decir al paciente que está predispuesto a tener una ADR, es deseable detectar la evidencia de la ADR antes de que aparezca cualquier síntoma usando ensayos en sangre. Hay varios métodos para diagnosticar ADR relacionadas con inmunidad. Hay varios métodos *in vivo* (ensayo cutáneo, intradérmico, ensayo de parche, y ensayos de provocación de fármaco) que reproducen la alergia a fármacos exponiendo al paciente al fármaco. Los métodos *in vitro* incluyen evaluar la activación de basófilos, IgE específica de

fármaco, y ensayo de estimulación de linfocitos específicos de fármaco. Las diferentes versiones de los ensayos de activación de linfocitos específicos de fármaco se usan para evaluar diferentes propiedades de la activación de los linfocitos. Éstas incluyen ensayos de estimulación de linfocitos, migración de linfocitos, toxicidad de linfocitos, y transformación de linfocitos. Algunas variantes en los ensayos incluyen la evaluación de marcadores de activación como CD69 o el nivel de citoquinas liberadas. Generalmente, todos estos métodos se usan para diagnosticar a pacientes que ya tienen una alergia y no para predecir la reacción de hipersensibilidad. Existen problemas adicionales que afectan a las diferentes técnicas. Por ejemplo, algunos de los ensayos *in vivo* tienen algún riesgo de respuesta alérgica grave en el paciente. La activación de basófilos carece de la especificidad para el antígeno relevante e IgE específica de fármaco son sólo relevantes para aquellos tipos de alergia que implican IgE (por ejemplo, anemia hemolítica y anafilaxis). Por lo tanto, existe una necesidad de un método que sea capaz de predecir ADR bien antes de la administración del fármaco o antes de que aparezcan los síntomas. La evaluación del repertorio de células T y/o B puede generar dicho ensayo. Algunos de los métodos *in vitro* pueden usarse para identificar los clonotipos que interactúan con el fármaco. Por ejemplo, el ensayo de estimulación de linfocitos puede hacerse con el fármaco específico para identificar los clonotipos que interactúan con el fármaco.

#### K. Detección de daño tisular

Se ha mostrado que el uso de marcadores moleculares en la sangre y fluidos y/o tejidos corporales es capaz de proporcionar información crítica sobre órganos potencialmente dañados lo que puede orientar a un diagnóstico de la enfermedad e intervenciones terapéuticas. Un ejemplo de dicho marcador es la detección de la proteína troponina en la sangre como parte del diagnóstico de enfermedad cardíaca. La troponina es una molécula que es altamente específica para los tejidos del corazón que está contenida en gran medida en las células del tejido del corazón y se encuentra a niveles muy bajos en la circulación para individuos con corazones sanos. Cuando se produce la enfermedad cardíaca, sin embargo, la muerte y apoptosis celulares resultan en derrame de ésta y otras moléculas en la corriente sanguínea en la que la detección sensible usando ensayos ELISA puede revelar niveles elevados de troponina que están claramente asociados con enfermedad cardíaca.

Este paradigma puede extenderse fácilmente a otros tejidos que podrían padecer un daño similar pero dichas técnicas están limitadas por la capacidad de los investigadores de identificar marcadores que sean lo suficientemente específicos para un tejido dado como para proporcionar información relevante desde un punto de vista de diagnóstico y suficientemente abundantes en las fases tempranas del daño tisular como para proporcionar información en un tiempo clínicamente útil en el curso de una enfermedad que podría resultar en daño de órganos.

Se ha mostrado que aunque los marcadores de superficie presentes en las células humanas no resultan en reacciones inmunes, las moléculas contenidas internamente en las células pueden ser inmunogénicas cuando se liberan en la corriente sanguínea. Como en el caso de la detección de autoantígenos del cáncer, dichos autoantígenos relacionados con daño orgánico podrían ser detectados de forma más sensible indirectamente a través de la detección de células inmunes que son reactivas para estos antígenos de lo que sería posible por detección directa del antígeno del tipo que se hace en el caso de la troponina.

Esto puede usarse para proporcionar una percepción diagnóstica en los niveles de daño orgánico en un individuo. Las células T y/o B que son reactivas para antígenos que son específicos para moléculas encontradas en un tipo particular de tejido humano pueden enriquecerse usando los métodos descritos anteriormente. Los desplazamientos en la frecuencia en los clonotipos de células T y/o B antes y después del enriquecimiento pueden usarse para averiguar qué clonotipos están reaccionando probablemente con estos antígenos. Este método puede combinarse con algoritmos de secuencia que usan parámetros de secuencia para averiguar cuáles de estos clonotipos enriquecidos están reaccionando lo más probablemente con estos antígenos. También pueden usarse estudios poblacionales que implican más de un individuo con daño orgánico conocido cuyos clonotipos que se correlacionan específicos de antígeno se han identificado empíricamente para informar un algoritmo que puede usarse para refinar estas predicciones identificando características de secuencia que están asociadas frecuentemente con estos clonotipos que se correlacionan en individuos posteriores.

La correlación de clonotipos específicos de antígeno puede hacerse al menos en un primer punto de tiempo para identificar y predecir clonotipos que se correlacionan. La sangre o fluidos corporales muestreados a puntos de tiempo posteriores puede perfilarse con o sin enriquecimiento específico de antígeno para medir los niveles de estos clonotipos que se correlacionan que puede usarse para generar una puntuación de daño orgánico que se correlaciona con el grado de daño a un órgano específico en ese individuo en ese momento. Los niveles de estos clonotipos pueden usarse para establecer esta puntuación como asimismo los desplazamientos en estos niveles con el tiempo.

Los antígenos para tejidos específicos pueden usarse en esta realización. Los tejidos podrían ser: corazón, pulmón, hígado, intestino, páncreas, esófago, estómago, riñón, nervios, testículos, ovario, próstata, timo, placenta, útero, etc. Los antígenos para cada uno de estos tejidos podrían ser un conjunto seleccionado de productos génicos que se sabe que se expresa específicamente en estos tejidos. Estos productos génicos específicamente expresados podrían ser el resultado de la observación de expresión génica diferencial entre estos órganos y otros órganos. Los antígenos podrían ser un único antígeno, un conjunto de antígenos, o mezclas complejas de materiales hasta e incluyendo material de células completas de los tejidos en cuestión.

## L. Identificación de la exposición a antígenos locales

Los métodos descritos pueden usarse para generar una base de datos de perfiles inmunes relacionados con una localización geográfica particular sobre la base de antígenos locales. Estos antígenos podrían ser pero no están limitados a pólenes locales. Estos antígenos podrían tener un componente estacional. Una vez se generan los perfiles inmunes en localizaciones geográficas, un sujeto tendrá su perfil inmune actual comparado con la base de datos. Dicha comparación se usa para determinar si un sujeto ha estado expuesto recientemente a los antígenos locales. En un caso, este método se usa para ensayar si un sujeto estaba en una localización sospechada. En otro caso, este método se usa para identificar localizaciones en las que es probable que el sujeto haya estado sin una sospecha preconcebida de dónde ha estado el sospechoso. De acuerdo con esto, un método para determinar la exposición de un individuo a uno o más antígenos puede comprender las etapas de (a) determinar un perfil de clonotipos de una muestra de células B y/o células T del individuo, comprendiendo dicha muestra un repertorio de clonotipos de éstos; y (b) comparar los clonotipos del perfil con clonotipos de una base de datos de clonotipos específicos de antígeno para determinar un nivel de concordancias de clonotipos, determinando de esta manera un nivel de exposición al antígeno, incluyendo la base de datos de clonotipos específicos de antígeno sustancialmente todos los clonotipos de cadenas de TCR y/o inmunoglobulinas humanas específicas para el uno o más antígenos. En un caso, el uno o más antígenos consisten en antígenos de un patógeno. En otro caso, el patógeno es un virus. En otro caso, dicho virus es un virus influenza, un virus de viruela, un virus de hepatitis C, un coronavirus, un virus de dengue, o un lentivirus. En un caso, dicha base de datos de clonotipos específicos de antígeno consiste en clonotipos que consisten en sustancialmente todas las regiones CDR3 de cadenas de TCR $\beta$  e IgH humanas.

## M. Identificación de la exposición a antígenos relacionados con bio-terrorismo

Los métodos descritos pueden usarse para generar una base de datos de perfiles inmunes relacionados con antígenos que probablemente estén relacionados con la producción de compuestos relacionados con el bioterrorismo. Estos antígenos podrían ser, pero no están limitados a, vectores virales que pueden convertirse en armas. Una vez se generan los perfiles inmunes, un sujeto tendrá su perfil inmune actual comparado con la base de datos. Dicha comparación se usa para determinar si un sujeto se ha expuesto recientemente a los compuestos relacionados con bioterrorismo. En un caso, este método se usa para ensayar si un sujeto se ha expuesto a un compuesto sospechoso particular. En otro caso, este método se usa para identificar si es probable que el sujeto se haya expuesto a una lista de compuestos potenciales sin una sospecha preconcebida de qué compuestos se esperan. En un caso, el perfil inmune que se va a ensayar se genera después de que se ha encontrado un agente biológico. Por ejemplo, un individuo sospechoso de un ataque de bio-terrorismo se identifica por las autoridades. Se obtiene una muestra del sospechoso y se obtiene un perfil inmune usando los métodos descritos anteriormente. Este perfil se compara estadísticamente con una base de datos que contiene muchos perfiles de muestras. Los perfiles de muestras incluyen perfiles inmunes que representan determinados antígenos relacionados con bioterrorismo. Los perfiles de muestras incluyen perfiles inmunes que representan determinados antígenos o combinaciones de antígenos que sólo existen en localizaciones geográficas particulares en momentos particulares del año. La comparación del perfil inmune del sospechoso con esta base de datos proporciona evidencia de que el sospechoso estaba en una localización geográfica particular durante un marco de tiempo particular y que el sospechoso estuvo expuesto a determinados antígenos relacionados con bioterrorismo. Esta evidencia se usa para dirigir adicionalmente la investigación y durante la acusación.

**Kits**

En la comercialización de los métodos descritos en la presente memoria, los kits para la amplificación de regiones reorganizadas somáticamente específicas o partes de éstas son particularmente útiles. Dichos kits pueden ser para llevar a cabo una o dos PCR por etapas (como se ha descrito anteriormente) para amplificar una región reorganizada somáticamente predeterminada o parte de ésta para el propósito de preparar una muestra de clonotipos para análisis de secuencia. Un kit comprende típicamente uno o más reactivos, tales como, sin limitación, cebadores de ácido nucleico, envasados en un contenedor, tal como, sin limitación, una vial, tubo o botella, en un envase adecuado para la distribución comercial, tal como, sin limitación, una caja, una bolsa cerrada, un envase blíster o un bote.

El envase contiene típicamente una etiqueta o prospecto que indica que los reactivos envasados pueden usarse en un método para generar un perfil de clonotipos a partir de una muestra de tejido de un paciente. Tal y como se usa en la presente memoria, "materiales de envasado" incluye cualquier artículo usado en el envasado para la distribución de los reactivos en un kit, incluyendo sin limitación, contenedores, viales, tubos, botellas, bolsas, envases blíster, etiquetas, marcapagos, hojas de instrucciones y prospectos. Un ejemplo de dicho kit incluye los reactivos necesarios para la amplificación en un tubo de secuencias de TCR $\beta$  a partir de ADN o ARN extraído de células T de un paciente o linfocitos de sangre periférica de un paciente o médula ósea de un paciente, como se ha descrito anteriormente. Otro ejemplo de dicho kit incluye los reactivos necesarios para la amplificación en una pluralidad de tubos de secuencias de IgH a partir de ADN o ARN extraído de células B de un paciente o linfocitos de sangre periférica de un paciente o médula ósea de un paciente, como se ha descrito anteriormente. En el último ejemplo, los reactivos necesarios incluyen una pluralidad de conjuntos de cebadores para generar conjuntos anidados de moldes, como se ha descrito anteriormente. Típicamente, dicha pluralidad es 2 ó 3 ó 4. Para el último ejemplo, en un caso, se proporcionan tres conjuntos de cebadores; y más específicamente, se proporcionan los tres conjuntos de cebadores siguientes: el conjunto 1 que comprende cebadores directos



de la Tabla 5 y cebadores inversos de la Tabla 8; el conjunto 2 que comprende cebadores directos de la Tabla 6 y cebadores inversos de la Tabla 8; el conjunto 3 que comprende cebadores directos de la Tabla 7 y cebadores inversos de la Tabla 8. En otro ejemplo, el kit incluirá los reactivos descritos anteriormente, incluyendo uno o más conjuntos de cebadores de PCR y una ADN polimerasa termoestable, tal como polimerasa Taq, y si las secuencias se amplifican a partir de ARN, una transcriptasa inversa. Los cebadores pueden estar presentes en cantidades que rendirían una amplificación equilibrada de secuencias de clonotipos individuales en la muestra de un paciente, como se ha descrito anteriormente. Las cantidades de cebadores pueden proporcionarse para asegurar una amplificación equilibrada de los clonotipos. Dicho equilibrio de PCR múltiples es muy conocido para los expertos en la técnica e incluye, pero no está limitado a, ajustar las concentraciones de cebadores en la reacción y/o seleccionar las posiciones y longitudes de los cebadores en una región de interés para incrementar o disminuir la proporción de hibridación de los cebadores individuales. En un caso, las cantidades de cebadores se seleccionan de manera que en la PCR sus concentraciones son tales que la proporción a la que cada cebador hibrida a su sitio de unión de cebador es sustancialmente idéntica. En otro caso, las cantidades de cebadores se seleccionan de manera que cada secuencia en una muestra se amplifica hasta una cantidad que es 2 veces la cantidad de amplificación media de una muestra aleatoria de clonotipos. En otro caso más, dicha muestra aleatoria contiene al menos 100 clonotipos.

Las ADN polimerasas y transcriptasas termoestables están disponibles comercialmente en una variedad de fabricantes. Los materiales adicionales en el kit pueden incluir: tubos o viales de reacción adecuados, una composición de barrera, típicamente un lecho de cera, opcionalmente incluyendo magnesio; mezclas de reacción (frecuentemente concentradas, por ejemplo 2X, 5X, 10X ó 20X) para las etapas de PCR, incluyendo tampones y reactivos necesarios tales como dNTP; agua sin nucleasa o ARNasa; inhibidor de ARNasa; ácido o ácidos nucleicos control (es decir, tales como estándares internos), y/o cualesquiera tampones, compuestos, co-factores, constituyentes iónicos, proteínas y enzimas, polímeros adicionales, y semejantes que pueden usarse en reacciones de PCR múltiple.

Los componentes de un kit se envasan de cualquier manera que sea comercialmente practicable. Por ejemplo, los cebadores de PCR y/o transcriptasa inversa pueden envasarse individualmente para facilitar la flexibilidad en la configuración del ensayo, o conjuntamente para incrementar la facilidad de uso y para reducir la contaminación. De manera similar, los tampones, sales y co-factores pueden envasarse separadamente o conjuntamente. Los kits también pueden incluir reactivos y componentes mecánicos adecuados para la extracción manual o automatizada de ácido nucleico de una muestra de tejido. Estos reactivos son conocidos para los expertos en la técnica y típicamente dependen de la elección de diseño. Por ejemplo, en un caso de un proceso automatizado, el tejido se rompe por ultrasonidos en una disolución de lisis adecuada proporcionada en el kit.

### 35 **Ejemplo 1**

#### **Determinación del estado de una enfermedad autoinmune**

Se toma una muestra de fluido cerebroespinal (CSF) y sangre de un paciente con un pico de episodio de esclerosis múltiple. Las células CD4+ se aíslan del CSF y sangre, y la CDR3 del gen del receptor  $\beta$  de las células T se amplifica por PCR. Los fragmentos amplificados se amplifican adicionalmente para añadir sitios de unión de cebador para amplificación en puente y sitios de unión de cebador de secuenciación para secuenciación Solexa. Las regiones variables del gen del receptor  $\beta$  de las células T se secuencian para identificar los clonotipos. La información de la secuencia se usa para generar un perfil de clonotipos para el paciente.

Se toma otra muestra de sangre cuando el paciente está en un estado relativamente inactivo de la esclerosis múltiple. Se repite el mismo procedimiento que anteriormente para generar un perfil de clonotipos. Los clonotipos patológicos se identifican como aquellos que están altos en el episodio de pico y disminuyen significativamente en el estado inactivo. Se toma otra muestra de sangre del paciente en un estado posterior. A este tiempo, sólo se amplifica una fracción de las regiones CDR3 del gen del receptor  $\beta$  de las células T y se secuencia. Este subconjunto contiene los clonotipos patológicos. El nivel de los distintos clonotipos se determina para evaluar el estado de la enfermedad del paciente.

### 50 **Ejemplo 2**

#### **Análisis del repertorio de TCR $\beta$ : estrategia de amplificación y secuenciación**

En este ejemplo, se analizan las cadenas de TCR $\beta$ . El análisis incluye amplificación, secuenciación, y análisis de las secuencias de TCR $\beta$ . Un cebador AGCGACCTCGGGTGGGAACA (SEQ ID NO: 1) es complementario a una secuencia común en C $\beta$ 1 y C $\beta$ 2, y hay 34 cebadores de V (Tabla 1) capaces de amplificar todos los 48 segmentos V. C $\beta$ 1 o C $\beta$ 2 se diferencian entre sí en la posición 10 y 14 de la unión J/C. El cebador para C $\beta$ 1 y C $\beta$ 2 termina en la posición 16 pb y no tiene preferencia para C $\beta$ 1 o C $\beta$ 2.

Los 34 cebadores de V se modifican de un conjunto original de cebadores descrito en Van Dongen et al, publicación de patente U.S. 2006/0234234.

60 Tabla 1.

Secuencias de cebadores complementarios a las diferentes familias de V.

ES 2 593 614 T3

| Familia de segmento V        | Secuencia del cebador   | SEQ |
|------------------------------|-------------------------|-----|
| V20-1                        | AACTATGTTTTGGTATCGTCAGT | 2   |
| V29-1                        | TTCTGGTACCGTCAGCAAC     | 3   |
| V9, 5-1, 5-6, 5-5, 5-8, 5-4A | AGTGTATCCTGGTACCAACAG   | 4   |
| V9, 5-1, 5-6, 5-5, 5-8, 5-4B | AGTGTGTACTGGTACCAACAG   | 5   |
| V9, 5-1, 5-6, 5-5, 5-8, 5-4C | ACTGTGTCCTGGTACCAACAG   | 6   |
| V9, 5-1, 5-6, 5-5, 5-8, 5-4D | AGTGTGTCCTGGTACCAACAG   | 7   |
| V9, 5-1, 5-6, 5-5, 5-8, 5-4E | TCTGTGTACTGGTACCAACAG   | 8   |

ES 2 593 614 T3

|                                      |                             |    |
|--------------------------------------|-----------------------------|----|
| V7-3, 7-6, 7-9, 7-2, 7-4A            | CCCTTTACTGGTACCGACAG        | 9  |
| V7-3, 7-6, 7-9, 7-2, 7-4B            | GCCTTTACTGGTACCGACAG        | 10 |
| V7-3, 7-6, 7-9, 7-2, 7-4C            | CCCTTTACTGGTACCGACAAA       | 11 |
| V7-8, 16A                            | TTTTGGTACCAACAGGTCC         | 12 |
| V7-8, 16B                            | TTTTGGTACCAACAGGCC          | 13 |
| V 7-7                                | AACCCTTTATTGGTATCAACAG      | 14 |
| V4-1, 4-3, 4-2A                      | CGCTATGTATTGGTACAAGCA       | 15 |
| V4-1, 4-3, 4-2B                      | CGCTATGTATTGGTACAAGCA       | 16 |
| V12-3, 12-4, 12-5                    | TTTCTGGTACAGACAGACCATGA     | 17 |
| V3-1                                 | TACTATGTATTGGTATAAACAGGACTC | 18 |
| V25-1                                | CAAAATGTACTGGTATCAACAA      | 19 |
| V28, 10-3, 6-2, 6-3, 6-1, 6-6, 24-1A | CAAAATGTACTGGTATCAACAA      | 20 |
| V28, 10-3, 6-2, 6-3, 6-1, 6-6, 24-1B | ATGTACTGGTATCGACAAGACC      | 21 |
| V6-4, 6-9A                           | TGCCATGTACTGGTATAGACAAG     | 22 |
| V6-4, 6-9B                           | ATACTTGTCTGGTATCGACAAG      | 23 |
| V10-1, 10-2, 6-5, 6-9, 6-8, 27A      | ATATGTTCTGGTATCGACAAGA      | 24 |
| V10-1, 10-2, 6-5, 6-9, 6-8, 27B      | ATATGTCCTGGTATCGACAAGA      | 25 |
| V10-1, 10-2, 6-5, 6-9, 6-8, 27C      | ACATGTCCTGGTATCGACAAGA      | 26 |
| V14                                  | TAATCTTTATTGGTATCGACGTGT    | 27 |
| V19                                  | GCCATGTACTGGTACCGACA        | 28 |
| V18                                  | TCATGTTTACTGGTATCGGCAG      | 29 |
| V30                                  | CAACCTATACTGGTACCGACA       | 30 |
| V11-1, 11-3, 11-2A                   | CATGCTACCCTTTACTGGTACC      | 31 |
| V11-1, 11-3, 11-2B                   | CACAATACCCTTTACTGGTACC      | 32 |
| V2                                   | ATACTTCTATTGGTACAGACAAATCT  | 33 |
| V13                                  | CACTGTCTACTGGTACCAGCA       | 34 |
| V15                                  | CGTCATGTACTGGTACCAGCA       | 35 |

El analizador de genoma Illumina se usa para secuenciar el amplicón producido por los cebadores anteriores. Se realiza una amplificación en dos etapas en transcritos de ARN mensajero (200), como se ilustra en las Figs. 2A-2B, empleando la primera etapa los cebadores anteriores y una segunda etapa para añadir cebadores comunes para amplificación en puente y secuenciación. Como se muestra en la FIG. 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 con Cβ1 (203) y los dos alelos de Cβ2. En la región V (206) de los transcritos de ARN (200), se proporciona un conjunto de cebadores (212) que contiene secuencias de cebador complementarias a las diferentes secuencias de la región V (34 en un caso). Los cebadores del conjunto (212) también contienen una cola no complementaria (214) que produce el amplicón (216) que tiene un sitio de unión de cebador (218) específico para cebadores P7 (220). Después de una PCR múltiple convencional, se forma el amplicón (216) que contiene la parte

5  
10

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 añadir una etiqueta de muestra (221) y cebadores (220 y 222) para la formación de agrupaciones por PCR en puente. En la PCR secundaria, en el mismo lado del molde, se usa un cebador (222 en la Fig. 2B y se refiere en la presente memoria como "C10-17-P5") que tiene en su extremo 3' la secuencia de las 10 bases más cercanas a la unión J/C, seguido de 17 pb con la secuencia de las posiciones 15-31 de la unión J/C, seguido de la secuencia P5 (224), que juega un papel en la formación de agrupaciones por PCR en puente en la secuenciación Solexa. (Cuando el cebador C10-17-P5 (222) hibrida con el molde generado a partir de la primera PCR, se crea un bucle de 4 pb (posición 11-14) en el molde, al hibridar el cebador con la secuencia de las 10 bases más cercanas a la unión J/C y bases en las posiciones 15-31 de la unión J/C. La formación de bucle de las posiciones 11-14 elimina la amplificación diferencial de los moldes que portan C $\beta$ 1 o C $\beta$ 2. La secuenciación se hace entonces con un cebador complementario a la secuencia de las 10 bases más cercanas a la unión J/C y bases en las posiciones 15-31 de la unión J/C (este cebador se denomina C'). El cebador C10-17-P5 puede purificarse por HPLC con el fin de asegurar que todo el material amplificado tiene extremos intactos que pueden utilizarse de forma eficiente en la formación de agrupaciones.

En la FIG. 2A, la longitud de la protuberancia en los cebadores de V (212) es preferiblemente 14 pb. La PCR primaria se ayuda por una protuberancia más corta (214). Alternativamente, para facilitar la PCR secundaria, la protuberancia en el cebador de V se usa en la PCR primaria tanto sea posible porque la PCR secundaria se ceba a partir de esta secuencia. Se investigó un tamaño mínimo de la protuberancia (214) que apoya una PCR secundaria eficiente. Se prepararon dos series de cebadores de V (para dos segmentos V diferentes) con tamaños de protuberancia de 10 a 30 con pasos de 2 pb. Usando las secuencias sintéticas apropiadas, la primera PCR se realizó con cada uno de los cebadores en la serie y se realizó una electroforesis en gel para mostrar que todos se amplificaron. Con el fin de medir la eficiencia de la segunda amplificación por PCR se realizó una PCR en tiempo real de SYBR verde usando como molde los productos de PCR de las diferentes primeras reacciones de PCR y como cebadores Read2-tag1-P7 y Read2-tag2-P7. Surgió una imagen consistente usando las 4 series de datos de tiempo real (2 PCR primarias con dos segmentos V diferentes y dos PCR secundarias con diferentes cebadores que contenían dos etiquetas diferentes). Hubo una mejora en la eficiencia entre los tamaños de protuberancia 10 y 14 pb. Sin embargo, hubo poca o ninguna mejora en la eficiencia con una protuberancia mayor de 14 pb. La eficiencia permaneció alta al disminuir la protuberancia hasta 14 pb debido a la alta concentración de cebadores que permitió que la 14pb fuera un molde de cebado suficiente a una temperatura mucho más alta que su temperatura de fusión. Al mismo tiempo, la especificidad se mantuvo porque el molde no era todo el ADNc sino un producto de PCR de baja complejidad en el que todas las moléculas tenían la protuberancia de 14 pb.

Como se ilustra en la Fig. 2A, la PCR primaria usa 34 cebadores diferentes de V (212) que hibridan con la región V (206) de moldes de ARN (200) y contienen una protuberancia de 14 pb común en la cola 5'. Las 14 pb es la secuencia parcial de uno de los cebadores de secuenciación de Illumina (denominado el cebador Read 2). El cebador de la amplificación secundaria (220) en el mismo lado incluye la secuencia P7, una etiqueta (221), y la secuencia del cebador Read 2 (223 (este cebador se denomina Read2\_tagX\_P7). la secuencia P7 se usa para la formación de agrupaciones. El cebador Read 2 y su complemento se usan para la secuenciación del segmento V y la etiqueta, respectivamente. Se crea un conjunto de 96 de estos cebadores con etiquetas numerados de 1 a 96 (véase más adelante). Estos cebadores se purifican por HPLC con el fin de asegurar que todo el material amplificado tiene extremos intactos que pueden utilizarse de forma eficiente en la formación de agrupaciones.

Como se ha mencionado anteriormente, el cebador de la segunda etapa, C-10-17-P5 (222, FIG. 2B) tiene homología interrumpida con el molde generado en la primera etapa de PCR. La eficiencia de la amplificación usando este cebador se ha validado. Un cebador alternativo a C-10-17-P5, denominado CsegP5, tiene una homología perfecta con el cebador C de primera etapa y una cola 5' que porta P5. La eficiencia de usar C-10-17-P5 y CsegP5 en la amplificación de moldes de la primera etapa de PCR se comparó realizando PCR en tiempo real. En varios replicados, se encontró que la PCR usando el cebador C-10-17-P5 presentaba poca o ninguna diferencia en eficiencia comparada con la PCR usando el cebador CsegP5.

El amplicón (300) que resulta de la amplificación en 2 etapas ilustrada en las Figs. 2A-2B tiene la estructura usada típicamente con el secuenciador Illumina como se muestra en la FIG. 3A. Se usan dos cebadores que hibridan con la parte más exterior de la molécula, los cebadores Illumina P5 (AATGATACGGCGACCACCGAG) (SEQ ID NO: 36) y P7 (CAAGCAGAAGACGGCATAACGAGAT) (SEQ ID NO: 37) para la amplificación en fase sólida de la molécula (formación de agrupaciones). Se hacen tres lecturas de secuencia por molécula. La primera lectura de 100 pb se hace con el cebador C', que tiene una temperatura de fusión que es apropiada para el proceso de secuenciación Illumina. La segunda lectura tiene una longitud de sólo 6 pb y es únicamente para el propósito de identificar la etiqueta de la muestra. Se genera usando el cebador Tag de Illumina (AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC) (SEQ ID NO: 38). La lectura final es el cebador Read 2, como cebador de Illumina con la secuencia GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT (SEQ ID NO: 39). Usando este cebador, se genera una lectura de 100 pb en el segmento V empezando con la secuencia del cebador de V de la 1ª PCR.

Se diseñó un conjunto de etiquetas de secuencia de 6 pb para distinguir diferentes muestras operadas en

el mismo carril de secuenciación, en las que cada etiqueta es diferente de todas las demás etiquetas en el conjunto por al menos 2 diferencias. Las 2 diferencias evitan la asignación errónea de una lectura a la muestra errónea si hay un error en la secuenciación. El alineamiento hecho para comparar las etiquetas permitió huecos y por lo tanto un error de delección o inserción por secuenciación tampoco asignará la lectura a la muestra errónea. Las características adicionales para seleccionar las etiquetas fueron limitar operaciones de bases únicas (4 A o T y 3 G o C) así como sin similitud con los cebadores de Illumina. En total, se generaron 143 etiquetas, 96 de ellas se usan.

Secuenciación de TCR $\beta$ . Se usaron seis amplificaciones múltiples con los oligos combinados y una muestra de ADNc como un molde. Tres de cada una de las amplificaciones se hicieron con Accuprime y otras tres con Taq de alta fidelidad. Dos amplificaciones con cada enzima usaron ADNc que corresponde a 500 ng de ARN inicial, y una amplificación con cada enzima usó 10 veces menos ADNc. Para cada una de las seis reacciones se realizó una PCR primaria y secundaria y el material amplificado se secuenció usando la plataforma Illumina y el esquema descrito anteriormente. Se obtuvo una secuencia de 100 pb de cada lado. El análisis primario de los datos se hizo usando los mismos conceptos descritos más adelante.

Para evaluar la reproducibilidad del ensayo, se determinó si los niveles de clonotipos son consistentes en los experimentos en duplicado. Como se muestra en las FIGs. 5A-5C, se obtiene una alta correlación cuando se usó la misma enzima y cantidad de ADNc introducido de partida (cada una de las 2 comparaciones tuvo  $r^2 = 0,944$ ). Cuando se usaron diferentes enzimas, la correlación es peor (correlación mediana para las 4 combinaciones posibles  $r^2 = 0,931$ ), y sólo se reduce modestamente ( $r^2 = 0,924$ ) cuando las 2 enzimas se usaron para amplificar ADNc introducido más pequeño (correspondiente sólo a 50 ng de ARN).

En las Figs. 5A-5C, se identificaron secuencias idénticas en cada muestra. Después, para abordar los errores de secuenciación, algunos clonotipos se coalescieron para formar clonotipos mayores usando las estrategias generales descritas en la sección de análisis primario de secuencia. Los recuentos de clonotipos se computaron entonces en cada muestra. Una fracción de los clonotipos (no mostrado en la figura) estuvo presente en una muestra pero no en la otra, debido probablemente a que el algoritmo les coalesció con otro clonotipo en una muestra pero no en la otra. La frecuencia de los clonotipos en una muestra se computa entonces como su número de cuentas dividido por el número total de lecturas obtenido para esa muestra. Por ejemplo, si se observan 1.000 cuentas para un clonotipo en una muestra con 1.000.000 lecturas, su frecuencia se computa como 0,1%. La FIG. 7A muestra el  $\log_{10}$  de la frecuencia de cada clonotipo en dos muestras en duplicado usando Accuprime y ADNc correspondiente a 500 ng de ARN como molde de entrada. La correlación ( $r^2$ ) entre estos duplicados es 0,944. La FIG. 7B representa el  $\log_{10}$  de la frecuencia de cada clonotipo usando ADNc correspondiente a 500 ng de ARN como molde de entrada y Accuprime (eje de las X) o Taq de alta fidelidad (eje de las Y). Hay 4 comparaciones con esta combinación con una correlación mediana  $r^2 = 0,931$ . La mostrada en la figura tiene  $r^2 = 0,929$ . La FIG. 7C muestra el  $\log_{10}$  de la frecuencia de cada clonotipo usando ADNc correspondiente a 50 ng de ARN como molde de entrada y Accuprime (eje de las X) o Taq de alta fidelidad (eje de las Y). La correlación observada  $r^2 = 0,924$ .

#### 40 **Ejemplo 3**

##### **Análisis del repertorio de IgH: estrategia de amplificación y secuenciación**

En este ejemplo, se usan tres cebadores para amplificar regiones V de moléculas de IgH. Preferiblemente, los cebadores están en regiones que evitan las CDR, que tienen la mayor frecuencia de mutaciones somáticas. Se realizan tres reacciones de amplificación diferentes. En cada reacción, cada uno de los segmentos V se amplifica por uno de los tres cebadores y todos usarán los mismos cebadores de segmento C. Los cebadores en cada una de las reacciones separadas están aproximadamente a la misma distancia de la unión V-D y distancias diferentes respecto a los cebadores en diferentes reacciones, de manera que los cebadores de las tres reacciones están espaciados entre sí a lo largo del segmento V. Asumiendo la última posición del segmento V como 0, entonces el primer conjunto de cebadores (marco A) tiene el extremo 3' a aproximadamente -255, el segundo conjunto (marco B) tiene el extremo 3' a aproximadamente -160, y el tercer conjunto (marco C) tiene el extremo 3' a aproximadamente -30. Dada la homología entre varios segmentos V, para amplificar todos los 48 segmentos V y los muchos alelos conocidos (como se define por el sistema de información internacional ImMunoGeneTics <<<http://imgt.cines.fr/>>>) se necesitan 23, 33, y 32 cebadores en los marcos A, B, y C, respectivamente. La lista de los cebadores se muestra en las Tablas 2, 3, y 4.

Tabla 2  
Cebadores Marco A

| Segmento | Secuencia                    | SEQ ID NO |
|----------|------------------------------|-----------|
| IGHV1_1  | CCTCAGTGAAGGTCTCCTGCAAGG     | 40        |
| IGHV1_2  | CCTCGGTGAAGGTCTCCTGCAAGG     | 41        |
| IGHV1_3  | CCTCAGTGAAGGTTTCTCCTGCAAGG   | 42        |
| IGHV1_4  | GGGCTACAGTGAAAATCTCCTGCAAGG  | 43        |
| IGHV2_1  | AAACCCACACAGACCCTCACGCTGAC   | 44        |
| IGHV2_2  | AAACCCACAGAGACCCTCACGCTGAC   | 45        |
| IGHV2_3  | AAACCCACACAGACCCTCACACTGAC   | 46        |
| IGHV3_1  | CTGGGGGGTCCCTGAGACTCTCCTG    | 47        |
| IGHV3_2  | CTGGGGGGTCCCTTAGACTCTCCTG    | 48        |
| IGHV3_3  | CAGGGCGGTCCCTGAGACTCTCCTG    | 49        |
| IGHV3_4  | CAGGGCCGTCCCTGAGACTCTCCTG    | 50        |
| IGHV3_5  | CTGGGGGGTCCCTGAAACTCTCCTG    | 51        |
| IGHV3_6  | CTGGCAGGTCCCTGAGACTCTCCTG    | 52        |
| IGHV3_7  | CTGGAGGGTCCCTGAGACTCTCCTG    | 53        |
| IGHV3_8  | CTGGGAGGTCCCTGAGACTCTCCTG    | 54        |
| IGHV3_9  | TGGGGGGGCCCTGAGACTCTCCT      | 55        |
| IGHV4_1  | CTTCGGAGACCCTGTCCCTCACCTG    | 56        |
| IGHV4_2  | CTTCGGACACCCTGTCCCTCACCTG    | 57        |
| IGHV4_3  | CTTCACAGACCCTGTCCCTCACCTG    | 58        |
| IGHV4_4  | CTTCGGAGACCCCGTCCCTCACCTG    | 59        |
| IGHV4_5  | CGGGGACCCTGTCCCTCACCTG       | 60        |
| IGHV5_1  | GATCTCCTGTAAGGGTCTGGATACAGCT | 61        |
| IGHV6    | TCGCAGACCCTCTCACTCACCTGTG    | 62        |

Tabla 3  
Cebadores Marco B

5

| Segmento | Secuencia                 | SEQ ID NO |
|----------|---------------------------|-----------|
| IGHV6    | TGGATCAGGCAGTCCCCATCGAGAG | 63        |
| IGHV5_1  | GCTGGGTGCGCCAGATGCCC      | 64        |
| IGHV2_1  | TGGATCCGTCAGCCCCCAGG      | 65        |
| IGHV2_2  | TGGATCCGTCAGCCCCCGGG      | 66        |
| IGHV1_1  | GTGCGACAGGCCCTGGACAA      | 67        |
| IGHV1_2  | GGGTGCGACAGGCCACTGGACAA   | 68        |
| IGHV1_3  | GTGCGCCAGGCCCCCGGACAA     | 69        |

|          |                          |    |
|----------|--------------------------|----|
| IGHV1_4  | GGGTGCGACAGGCTCGTGGACAA  | 70 |
| IGHV1_5  | GGGTGCAACAGGCCCTGGAAAA   | 71 |
| IGHV1_6  | GGGTGCGACAGGCTCCTGGAAAA  | 72 |
| IGHV1_7  | GTGCGACAGGCCCCCGGACAA    | 73 |
| IGHV1_8  | GTGCGACAGGCCCCCAGACAA    | 74 |
| IGHV4_1  | TCCGCCAGCCCCCAGGGAAGG    | 75 |
| IGHV4_2  | TCCGGCAGCCCCCAGGGAAGG    | 76 |
| IGHV4_3  | TCCGGCAGCCACCAGGGAAGG    | 77 |
| IGHV4_4  | TCCGCCAGCACCCAGGGAAGG    | 78 |
| IGHV4_5  | TCCGGCAGCCCGCCGGGAA      | 79 |
| IGHV4_6  | TCCGGCAGCCGCGGGGAA       | 80 |
| IGHV4_7  | TCCGGCAGCCCGCTGGGAAGG    | 81 |
| IGHV4_8  | TCCGCCAGCCCCCTAGGGAAGG   | 82 |
| IGHV3_1  | GGTCCGCCAGGCTCCAGGGAA    | 83 |
| IGHV3_2  | GTTCCGCCAGGCTCCAGGGAA    | 84 |
| IGHV3_3  | GGTCCGCCAGGCTTCCGGGAA    | 85 |
| IGHV3_4  | GGTCCGTCAAGCTCCGGGGAA    | 86 |
| IGHV3_5  | GATCCGCCAGGCTCCAGGGAA    | 87 |
| IGHV3_6  | GGTCCGCAAGCTCCAGGGAA     | 88 |
| IGHV3_7  | GGTCCGCCAGGCTCCAGGCAA    | 89 |
| IGHV3_8  | GGTCCGCCAGGCTCCAGGCAA    | 90 |
| IGHV3_9  | GGTCCGCCAGGCTCCGGGCAA    | 91 |
| IGHV3_10 | GGGTCCGTCAAGCTCCAGGGAAGG | 92 |
| IGHV3_11 | CTGGGTCCGCCAAGCTACAGGAAA | 93 |
| IGHV3_12 | GGTCCGCCAGCCTCCAGGGAA    | 94 |
| IGHV3_13 | GGTCCGGCAAGCTCCAGGGAA    | 95 |

Tabla 4  
Cebadores Marco C

| Segmento | Secuencia                    | SEQ ID NO |
|----------|------------------------------|-----------|
| IGHV7    | CTAAAGGCTGAGGACACTGCCGTGT    | 96        |
| IGHV6    | CTCTGTGACTCCCAGGACACGGCT     | 97        |
| IGHV5_1  | AGTGGAGCAGCCTGAAGGCCTC       | 98        |
| IGHV2_1  | TGACCAACATGGACCCTGTGGACAC    | 99        |
| IGHV1_1  | ACATGGAGCTGAGCAGCCTGAGATC    | 100       |
| IGHV1_2  | ACATGGAGCTGAGCAGGCTGAGATC    | 101       |
| IGHV1_3  | ACATGGAGCTGAGGAGCCTGAGATC    | 102       |
| IGHV1_4  | ACATGGAGCTGAGGAGCCTAAGATCTGA | 103       |
| IGHV4_1  | GAGCTCTGTGACCGCCGCGGAC       | 104       |
| IGHV4_2  | GAGCTCTGTGACCGCCGTGGACA      | 105       |
| IGHV4_3  | GAGCTCTGTGACCGCTGCAGACACG    | 106       |

|          |                                 |     |
|----------|---------------------------------|-----|
| IGHV4_4  | GAGCTCTGTGACCGCTGCCGACA         | 107 |
| IGHV4_5  | GAGCTCTGTGACTGCCGCAGACACG       | 108 |
| IGHV4_6  | GAGCTCTGTGACTGCAGCAGACACG       | 109 |
| IGHV4_7  | GAGCTCTGTGACTGCCGCGGACA         | 110 |
| IGHV4_8  | GAGCTCTGTGACCGCGGACGCG          | 111 |
| IGHV4_9  | GGCTCTGTGACCGCCGCGGAC           | 112 |
| IGHV4_10 | GAGCTCTGTGACCGCCGCAGACA         | 113 |
| IGHV4_11 | GAGCTCTGTGACCGCTGACACGG         | 114 |
| IGHV3_1  | CAAATGAACAGCCTGAGAGCCGAGGACA    | 115 |
| IGHV3_2  | CAAATGAACAGCCTGAAAACCGAGGACA    | 116 |
| IGHV3_3  | CAAATGAACAGTCTGAAAACCGAGGACA    | 117 |
| IGHV3_4  | CAAATGATCAGCCTGAAAACCGAGGACA    | 118 |
| IGHV3_5  | CAAATGAACAGTCTGAGAAGTGGAGGACACC | 119 |
| IGHV3_6  | CAAATGAACAGTCTGAGAGCCGAGGACA    | 120 |
| IGHV3_7  | CAAATGAACAGCCTGAGAGCTGAGGACA    | 121 |
| IGHV3_8  | CAAATGAGCAGCCTGAGAGCTGAGGACA    | 122 |
| IGHV3_9  | CAAATGAACAGCCTGAGAGACGAGGACA    | 123 |
| IGHV3_10 | CAAATGGGCAGCCTGAGAGCTGAGGACA    | 124 |
| IGHV3_11 | CAAATGAACAGCCTGAGAGCCGGGGA      | 125 |
| IGHV3_12 | CAAATGAACAGTCTGAGAGCTGAGGACA    | 126 |
| IGHV3_13 | CAAATGAGCAGTCTGAGAGCTGAGGACA    | 127 |

En el lado del segmento C, dos secuencias con una diferencia de base entre ellas (GCCAGGGGGAAGACCGATGG (SEQ ID NO: 128), y GCCAGGGGGAAGACGGATGG) (SEQ ID NO: 129) abarcan los cuatro segmentos y los múltiples alelos conocidos de IgG. Se usa un esquema similar para las dos etapas de PCR para los genes de TCR $\beta$ .

En el lado de V, se usa la misma protuberancia de 14 pb 5' en cada uno de los cebadores de V. En la PCR secundaria, se emplea el mismo cebador Read2-tagX-P7 en el lado de V. En el lado de C, se usa una estrategia similar a la usada con la amplificación de TCR $\beta$  para evitar variantes entre los diferentes segmentos de IgG y sus alelos conocidos. La secuencia de cebador (AATGATACGGCGACCACCGAGATCTGGGAAGACGATGGGCCCTTGGTGGGA) (SEQ ID NO: 130) comprende la secuencia del segmento C desde las posiciones 3-19 y 21-28 y se salta la posición 20 que tiene una base diferente en al menos uno de los diferentes alelos de IgG y la secuencia para P5 como se puede usar para la formación de las agrupaciones como se muestra en la FIG. 4A.

Se llevó a cabo una PCR múltiple usando tres combinaciones de cebadores correspondientes a los tres marcos usando ADNc como un molde. Después de las PCR primaria y secundaria, los productos se corrieron en un gel de agarosa. Se obtuvieron bandas simples con los tamaños relativos apropiados a partir de las tres combinaciones.

En un caso, se mezclan tres reacciones diferentes de una única muestra a proporción equimolar y se someten a secuenciación. La secuenciación se hace desde ambas direcciones usando los dos cebadores de Illumina, tal como se ha descrito anteriormente. Se secuencian 100 pb desde cada lado. Las secuencias de línea germinal máximas que engloban los segmentos D+J son ~30 pb más largas para BCR que para TCR. Por lo tanto, si el resultado neto de la eliminación y adición de nucleótidos en las uniones (nucleótidos N y P) generan una distribución similar de IgH y TCR $\beta$ , de media 90 pb y como máximo 120 pb de secuencia después del segmento C es suficiente para alcanzar el 3' del segmento V. Por lo tanto, en la mayor parte de los casos, la secuencia del cebador de C es suficiente para alcanzar el segmento V. La secuenciación de uno de los adaptadores de Illumina identifica el segmento V usado así como hipermutaciones somáticas en los segmentos V. Se secuencian partes diferentes de los segmentos V dependiendo de a partir de cuál de las tres reacciones de amplificación se origine la secuencia. La secuencia completa del BCR puede alinearse a partir de diferentes lecturas que se originaron a partir de diferentes reacciones de amplificación. La reacción de secuenciación desde uno de los extremos que muestra la secuencia completa de CDR3 facilita en gran medida el alineamiento exacto de lecturas diferentes.

#### Ejemplo 4

##### Análisis del repertorio de TCR e IgH en muestras de pacientes con SLE

En primer lugar, se ensayará si hay clonotipos que se correlacionan con la actividad de la enfermedad en los pacientes. En segundo lugar, se define un conjunto de características de secuencia y/o marcadores de la superficie celular que distinguen a los clonotipos que se correlacionan con la enfermedad de aquellos que no lo hacen. En tercer lugar, se mide el grado en el que el análisis de los clonotipos proporciona información clínicamente útil, tal como la correlación con un resultado a corto plazo (por ejemplo, 3



meses).

#### 1. Presencia de clonotipos que se correlacionan con la enfermedad

Hay dos tareas principales: identificar los clonotipos que se correlacionan y medir la actividad de la enfermedad a partir de su nivel. Estas tareas pueden hacerse en un entorno clínico en dos etapas para cada paciente:

1) Puede hacerse un ensayo de calibración para determinar la identidad de los clonotipos que se correlacionan para el paciente específico. Esto puede hacerse secuenciando el ARN de IgH y TCR $\beta$  (o secuencia asociada de TCR $\alpha$ -TCR $\beta$  de una única célula) para cada paciente en un momento de un pico de un episodio, momento en el que el nivel de clonotipos que se correlacionan puede alcanzar sus niveles más altos.

2) Puede hacerse un ensayo de monitorización para determinar el nivel de los clonotipos que se correlacionan en un punto de tiempo posterior al ensayo de calibración. Esto puede hacerse secuenciando el ARN de IgH y TCR $\beta$  y determinando el nivel de los clonotipos que se correlacionan específicos que se habían identificado en la muestra de calibración del mismo paciente. El nivel de los clonotipos que se correlacionan se usa para computar la actividad de la enfermedad en estos puntos.

El desarrollo de la amplificación, secuenciación, y análisis primario como se ha descrito anteriormente se usa para evaluar las muestras del paciente. Específicamente, se evalúa un conjunto de pacientes con lupus eritematoso sistémico (SLE) que tiene un periodo de seguimiento de un año y muestras de sangre seriadas durante este periodo. Estos pacientes fueron atendidos por el Dr. Michele Petri en la Johns Hopkins Medical School cada tres meses durante un año, y las medidas clínicas de la actividad de la enfermedad incluyendo índice de la actividad de la enfermedad de lupus eritematoso sistémico (SLEDAI), evaluación global por el médico (PGA), así como múltiples ensayos de laboratorio incluyendo niveles de C3 (complemento 3) y anti-ADNs están disponibles para todas las visitas de todos los pacientes. Los fármacos que se estaban administrando a los pacientes, incluyen prednisona, plaquenil, NSAID, tipo NSAID, dosis de ácido acetilsalicílico (ASA), plavix, diurético, inhibidores de ACE o bloqueadores del receptor de angiotensina (ARB), bloqueador del canal de Ca, Triam y, solumedrol. Se estudian los pacientes que tuvieron al menos en un momento durante el seguimiento un cambio significativo en la actividad de la enfermedad como se define por un cambio de 3 puntos en el SLEDAI o un cambio de 1 punto en PGA. Globalmente, hay 181 pacientes (con un total de 815 muestras de sangre) que se ajustan a estos criterios. El ARN de todas estas muestras de sangre se somete a PCR múltiple usando los cebadores descritos anteriormente para amplificar las secuencias que engloban CDR3 en IgH y TCR $\beta$ . Todos los materiales amplificados se secuencian (hasta un millón de lecturas) y se determina la abundancia de los diferentes clonotipos.

Usando los datos clínicos, secuenciación, se identifican las características que distinguen los clonotipos cuyo nivel se correlaciona con la actividad de la enfermedad de aquellos que no lo hacen. En segundo lugar, se desarrolla un algoritmo para determinar la actividad de la enfermedad usando el perfil sanguíneo de IgH y TCR $\beta$ .

#### 2. Identificación de características de los clonotipos que se correlacionan

Se anticipa que los clonotipos que son relevantes para la enfermedad se incrementan en el momento de alta actividad de la enfermedad. Sin embargo, no todos los clonotipos enriquecidos en un punto de alta actividad de la enfermedad se correlacionan necesariamente con la enfermedad. Por ejemplo, en un paciente particular podría haber 10 clonotipos enriquecidos en el punto de alta actividad de la enfermedad, pero sólo 5 se correlacionan con la enfermedad. Con el fin de identificar estos clonotipos relevantes, se estudia un subconjunto de clonotipos que se correlacionan claramente con la enfermedad y otro conjunto que no se correlacionan claramente con la enfermedad. Se investigan las características que distinguen estas dos clases de clonotipos.

Todos los pacientes tendrán al menos un cambio significativo en la actividad de la enfermedad durante el seguimiento de un año en este diseño experimental. Se analizan los clonotipos de IgH y TCR obtenidos en el pico de la actividad de la enfermedad en cada paciente. Se seleccionan los conjuntos de clonotipos que se correlacionan y no se correlacionan entre los clonotipos con un nivel más alto. Por lo tanto, la primera etapa es definir los clonotipos que están a un nivel alto. Los criterios específicos para elegir los clonotipos que entrarán en el análisis incluirán una combinación de rango de frecuencia del clonotipo y el nivel del clonotipo (número de lecturas de clonotipo por millón), así como evidencia de que el clonotipo no pertenece a la distribución de clonotipos de baja frecuencia.

Este conjunto de clonotipos para cada muestra de paciente, denominados clonotipos con alta prevalencia (HPC) se analiza adicionalmente. Se evalúa la correlación del nivel de cada uno de estos clonotipos con las medidas clínicas. Se computa la correlación de la puntuación SLEDAI con el nivel de clonotipo. Para cada paciente, hay 4-5 puntos de estudio que pueden usarse para evaluar la correlación de SLEDAI con el nivel de cada HPC. Se investiga la distribución de estas correlaciones obtenidas. Se anticipa que la mayor parte de los HPC tendrán una baja correlación con SLEDAI. Se investiga si en el extremo de alta correlación hay un exceso de lo que se espera que se genere aleatoriamente. Por ejemplo, con 4 y 5 puntos de datos se espera que ~2,5% y ~0,6% de los niveles de correlación ( $r^2$ ) sea  $>0,9$  por azar. Mayores proporciones de HPC con  $r^2 > 0,9$  indican la presencia de clonotipos que se correlacionan con la enfermedad. Además de comparar el número de clonotipos que se correlacionan con la expectación aleatoria, se realiza un análisis de permutación en el que se calcula la correlación de puntuaciones SLEDAI de un paciente y el nivel de HPC individuales de otro. La distribución de correlaciones generada a partir de esta permutación puede usarse como la correlación "de fondo". (Para asegurar su validez, se

confirma que hay poca correlación entre SLEDAI entre diferentes pacientes). La correlación en exceso en el extremo de alta correlación, por ejemplo,  $r^2 > 0,9$ , indicará la presencia de clonotipos que se correlacionan con la enfermedad. Se seleccionan los clonotipos que más se correlacionan como el conjunto de clonotipos que se correlacionan. Como se sabe el número de HPC que tiene una correlación por azar mayor que un umbral ajustado (a partir del cálculo usando asunción aleatoria o mediante el análisis de permutación descrito anteriormente), el umbral para definir el clonotipo que se correlaciona puede ajustarse de manera que se tiene un 10% de proporción de descubrimientos falsos, es decir, el 10% del conjunto de los clonotipos que se correlacionan se correlaciona por azar. Se selecciona un conjunto de HPC que tiene muy poca correlación con la puntuación SLEDAI. Estos servirán como el conjunto de clonotipos que no se correlacionan. Estos 2 conjuntos de clonotipos pueden analizarse adicionalmente para identificar características que puedan distinguirlos. Estas características pueden buscarse entonces en muestras nuevas para identificar los clonotipos que probablemente se correlacionen con la actividad de la enfermedad en estas muestras. Los niveles sanguíneos de estos clonotipos pueden seguirse entonces para determinar la actividad de la enfermedad.

Una complicación surge de la premisa de que el nivel de clonotipos puede cambiar antes de que lo haga la actividad de la enfermedad. Por lo tanto, es posible que mediante el intento de estudiar sólo los HPC que tienen una alta correlación con SLEDAI, puedan eliminarse los clonotipos clínicamente útiles que cambian antes que el SLEDAI. Se selecciona otro conjunto de clonotipos que se correlaciona con una puntuación SLEDAI modificada (MSLEDAI). MSLEDAI es lo mismo que SLEDAI en todos los puntos del estudio excepto en aquellos justo antes de un cambio significativo. Para estos puntos de datos, la puntuación MSLEDAI es la media entre la puntuación SLEDAI en ese punto y el siguiente punto de estudio. Los clonotipos que cambian antes de SLEDAI probablemente muestren una mejor correlación con MSLEDAI que con SLEDAI. Es informativo computar el número en exceso de HPC que tienen una alta correlación con MSLEDAI de lo esperado por expectativas generadas aleatoriamente o por permutación. Se identificarán entonces las características que distinguen los clonotipos que se correlacionan de los que no se correlacionan. El análisis se hace de una manera exacta para aquellos clonotipos que se correlacionan con SLEDAI o MSLEDAI. En cualquier caso, el objetivo sería para este conjunto de características recapitular correctamente esta clasificación permitiendo la identificación de clonotipos que se correlacionan en el siguiente conjunto de muestras. Se espera que cada paciente tendrá un único conjunto de clonotipos que se correlacionan, pero el estudio de entrenamiento se diseña para generar las reglas que predicen los clonotipos que se correlacionan a partir de una muestra de calibración (a alta actividad de la enfermedad). Pueden ensayarse dos tipos generales de parámetros: aquellos que se obtienen a partir de los datos de secuenciación en sí mismos, y aquellos que pueden usar experimentación adicional. La experimentación adicional puede incluir la evaluación de diferentes células con diferentes marcadores de superficie celular u otros. Aquí hay unos pocos tipos de parámetros que se investigan: 1) resto de secuencia; el resto puede ser una región V o J específica, una combinación VJ, o secuencias cortas en una región DJ que está asociada con un clonotipo que se está correlacionando. 2) Tamaño del clonotipo. 3) Nivel: nivel absoluto (número de lecturas por millón) o nivel de rango. 4) Similitud con otros clonotipos: la presencia de otros clonotipos altamente relacionados, como aquellos con cambios silenciosos (diferencias de nucleótidos que codifican los mismos aminoácidos) o aquellos con cambios de aminoácidos conservativos. 5) Para los BCR el nivel de las mutaciones somáticas en el clonotipo y/o el número de clonotipos distintos que se diferencian por mutaciones somáticas de algunos clonotipos de la línea germinal.

Cada uno de estos parámetros se estudia individualmente para asociación con clonotipos que se correlacionan. Se ajusta un umbral de 0,05 (no corregido para ensayo múltiple) para eliminar factores que probablemente no contribuyan a la predicción de clonotipos que se correlacionan. Dados los múltiples parámetros, se realizan muchos ensayos para generar múltiples resultados positivos por azar. Sin embargo, el objetivo principal de esta etapa es filtrar los parámetros hasta un conjunto más pequeño. El conjunto de parámetros positivos se usará entonces para crear un algoritmo para clasificar los dos conjuntos de clonotipos. Se emplea un algoritmo de aprendizaje de máquina que usa los diferentes parámetros para clasificar los dos conjuntos de clonotipos. Con el fin de minimizar el riesgo de sobreajuste, se usa la técnica de validación cruzada. Usando este algoritmo cada clonotipo obtendrá una puntuación que corresponde a la probabilidad de que sea un clonotipo que se correlaciona. Se situará entonces un umbral para clasificar a los clonotipos por encima de éste como que se correlacionan y aquellos por debajo de éste como que no se correlacionan. La exactitud de la clasificación puede estimarse por la técnica de la validación cruzada; por ejemplo, los clonotipos se ponen en grupos iguales y el algoritmo usando todos los clonotipos excepto un grupo. Los clonotipos en el último grupo (grupo de ensayo) se clasifican entonces usando el algoritmo que se obtuvo usando el resto de los clonotipos. Esto se itera tantas veces como el número de grupos, y en cada iteración todos los grupos excepto uno se usan para entrenamiento y un grupo se clasifica. La exactitud del algoritmo puede estimarse a partir de la exactitud media de las diferentes clasificaciones en las diferentes iteraciones. Debe indicarse que en todas estas iteraciones el algoritmo exacto será ligeramente diferente. La exactitud de la clasificación es entonces una estimación ya que no es en el algoritmo final sino en un conjunto de algoritmos generados con los datos de entrenamiento de todos los clonotipos excepto uno.

Finalmente, se generan dos algoritmos entrenados en dos conjuntos de clonotipos que se correlacionan diferentes: uno que se correlaciona con SLEDAI y el otro que se correlaciona con MSLEDAI. Incluso si los clonotipos en el conjunto de entrenamiento son diferentes, el algoritmo resultante puede ser o no muy

diferente, dependiendo de si estos clonotipos de hecho vienen de dos poblaciones distintas. Los algoritmos se comparan. Además, estos algoritmos se usan para identificar clonotipos que se correlacionan que no estaban inicialmente en el conjunto de entrenamiento. Los clonotipos identificados en los dos algoritmos se comparan, y si los clonotipos iniciales en los dos conjuntos de entrenamiento eran de la misma población, los clonotipos identificados probablemente sean muy similares. A no ser que los resultados del algoritmo sean bastante similares, ambos algoritmos se llevan a cabo para identificar clonotipos que se correlacionan con el fin de medir la actividad de la enfermedad de lupus.

Otras estrategias experimentales pueden añadir al poder de secuenciación en la identificación de clonotipos que se correlacionan con enfermedades. Los clonotipos que se correlacionan pueden enriquecerse en células con algún marcador de superficie u otros. Por ejemplo, las células B con altos niveles de CD27 son conocidas en pacientes con lupus activo, y por lo tanto podría ser que los clonotipos que se correlacionan podrían enriquecerse en la población CD27 de células. Si esto fuera cierto, la predicción de los clonotipos que se correlacionan puede mejorarse haciendo un enriquecimiento para células con altos niveles de CD27. Específicamente, puede realizarse una reacción de secuenciación en las secuencias de IgH de todas las células B en la muestra de sangre así como aquellas células B con CD27 alto. Los clonotipos que se correlacionan se espera que estén presentes a una frecuencia más alta en la población con CD27 alto que en la muestra de sangre completa.

3. Uso de los perfiles de IgH y TCR $\beta$  para determinar la actividad de la enfermedad de lupus

La sección anterior ha descrito análisis basados en clonotipos para identificar características de los clonotipos que se correlacionan. Además, para este análisis se usó sólo una fracción de todos los HPC para designar claramente los clonotipos como que se correlacionan o no se correlacionan. Esta sección describe análisis a nivel de paciente con el objetivo de computar una medida de la actividad de la enfermedad, que se denomina puntuación autoinmune (AI). El algoritmo desarrollado según la sección anterior se aplica para identificar clonotipos que se correlacionan entre todos los HPC. Se determina el nivel de estos HPC que se correlacionan. El nivel de los clonotipos que se correlacionan puede normalizarse respecto al número total de clonotipos de TCR así como a los HPC que se predice que no se correlacionan con la enfermedad. El nivel de estos clonotipos que se correlacionan a diferentes puntos de tiempo se usa para computar la puntuación AI a estos diferentes puntos.

En pacientes con más de un clonotipo que se correlaciona, la información respecto al nivel de estos diferentes clonotipos se combina. Además, se integran los datos de los clonotipos de IgH y TCR $\beta$ . Se intentan diferentes algoritmos para preparar la combinación. Por ejemplo, se estudia la media, mediana, suma, y nivel de clonotipo que se correlaciona más alto. El nivel de clonotipo puede ser sus cuentas de lectura lineales simples, el algoritmo de ésta o alguna otra conversión. Potencialmente, puede ser la diferencia entre los clonotipos que se correlacionan y no se correlacionan. Además, pueden utilizarse métodos para media ponderada. La ponderación puede basarse en la probabilidad de que un clonotipo se esté correlacionado.

Con el fin de evaluar cuál de los modelos es el óptimo, todos los modelos se evalúan para identificar el que genera la correlación más alta entre la puntuación AI y la puntuación SLEDAI. Para este análisis, se hace la correlación de las puntuaciones SLEDAI y AI a lo largo de todos los datos obtenidos de todos los puntos de estudio de todos los pacientes. Con el fin de estimar y mejorar el grado de sobreajuste, se usa la técnica de la validación cruzada. El nivel de correlación medido refleja la relación "transversal" entre las puntuaciones AI y SLEDAI. Además de SLEDAI, se estudia la correlación con otras medidas clínicas como niveles de C3 y anticuerpos anti-ADNs así como proteína en orina/creatinina en suero para pacientes con manifestación renal y recuentos sanguíneos para pacientes con implicación hematológica. La correlación puede deberse a la clasificación de pacientes en actividad de la enfermedad alta y baja, y no es necesariamente un reflejo de que AI se correlacione con la puntuación SLEDAI en un paciente. Para demostrar esto, se hace una evaluación "longitudinal".

4. Análisis longitudinal

En el análisis longitudinal, se evalúan dos cuestiones generales: si la puntuación AI en un punto del estudio predice la actividad de la enfermedad en el mismo punto, y si la puntuación AI en un punto del estudio predice la actividad de la enfermedad en un punto posterior, por ejemplo, el siguiente punto del estudio 3 meses después.

La relación entre las puntuaciones AI y SLEDAI en el mismo punto del estudio se evalúa de dos maneras. En primer lugar, se calcula la correlación de la AI y SLEDAI en cada paciente, y después se computa la media y mediana del nivel de correlación del paciente. Si la correlación observada en el análisis transversal anterior se debe a la clasificación de pacientes con actividad de la enfermedad alta y baja y no cambia la actividad de la enfermedad en pacientes individuales, entonces la correlación longitudinal en pacientes individuales probablemente sea baja. Una mediana alta del nivel de correlación del paciente sugiere que AI no refleja la puntuación SLEDAI a nivel de un paciente individual. Además de la correlación de puntuaciones AI y SLEDAI, se evalúa la correlación de AI con otras medidas relevantes como C3 y anticuerpo anti-ADNs así como proteína en orina/creatinina en suero para pacientes con manifestación renal y recuentos sanguíneos para pacientes con implicación hematológica.

Otra manera de demostrar la capacidad de la puntuación AI para medir los cambios en la actividad de la enfermedad en pacientes individuales es mediante la determinación de su exactitud para distinguir estados de actividad de la enfermedad altos de bajos en los mismos pacientes. Para cada uno de los 181 pacientes, se seleccionan los dos puntos del estudio cuando la SLEDAI estaba a los niveles más altos (denominado HDAP para punto de la actividad de la enfermedad alta) y más bajos (denominado LDAP

para punto de la actividad de la enfermedad baja). Se compara la distribución del AI de todos los HDAP con la del AI de todos los LDAP, y se computa el valor p en el que son diferentes. Además, se evalúa la frecuencia que la AI a HDAP es mayor que LDAP en cada paciente. Si AI no cambia con la actividad de la enfermedad en un paciente individual, entonces se espera que AI a HDPA sea mayor que a LDAP sólo el 50% de las veces. Se hace otro análisis en el que se determina la fracción de veces en la que AI a HDAP es mayor que la de a LDAP por una diferencia significativa (es decir, por encima de la variación probable de AI). Para medir la fluctuación de AI, se usan todos los puntos del estudio de todos los pacientes, y puede computarse la desviación estándar (y desviación estándar relativa) de AI en los diferentes grupos de valores SLEDAI. Esto generará desviación estándar relativa a lo largo de todos los pacientes (AI-RSDall) y este valor puede depender o no de SLEDAI (es decir, el AI-RSDall puede ser diferente a diferentes valores de SLEDAI). La proporción de pacientes en la que AI a HDAP es mayor que AI a LDAP por un número específico (por ejemplo, 2) de AI-RSDall puede computarse. Puede haber algún sesgo sistemático en el que el AI computado en algunos pacientes sea consistentemente mayor (o menor) de lo que se espera de la puntuación SLEDA. Por lo tanto, AI-RSDall es una combinación de la fluctuación intrínseca de AI en un paciente así como la diferencia sistemática de AI para pacientes con SLEDAI similar. La fluctuación intrínseca de AI puede computarse en un paciente calculando la desviación estándar (y desviación estándar relativa) de las puntuaciones AI entre los puntos del estudio con valores SLEDAI similares (diferencia <2 puntos) en un paciente. Puede computarse la mediana entre todos los pacientes de la desviación estándar relativa (AI-RSDpt-med). La proporción de pacientes en la que AI a HDAP es mayor que AI a LDAP por un número específico (por ejemplo, 2) de AI-RSDpt-med puede entonces evaluarse.

Después de la demostración de que AI no fluctúa de hecho con SLEDAI en pacientes individuales se evalúa si AI puede predecir SLEDAI en el siguiente punto del estudio, 3 meses después. Para evaluar esto, puede cuantificarse el nivel de correlación entre la puntuación AI en el tiempo 0 y la puntuación SLEDAI en el tiempo + 3 meses. La correlación puede computarse a nivel de paciente y después puede obtenerse la mediana de la correlación del paciente. Otra manera de demostrar la capacidad de AI para predecir la actividad de la enfermedad en un futuro cercano es evaluar la sensibilidad y especificidad de AI para predecir la actividad de la enfermedad 3 meses en el futuro. Clínicamente, aquellos pacientes con un buen comportamiento en su gestión actual pueden distinguirse de aquellos que no lo tienen. El estado de un paciente en un tiempo particular se clasifica en una de dos clases: mal control (PC) e incluye pacientes que en 3 meses tendrán una actividad de la enfermedad alta (SLEDAI > 6 puntos) y/o una crisis (incremento de SLEDAI de 3 puntos), y buen control (GC) e incluye pacientes que en 3 meses tendrán una actividad de la enfermedad baja o moderada (SLEDAI < 6) y/o una reducción significativa de la actividad de la enfermedad (disminución de SLEDAI de 3 puntos). La clasificación de sensibilidad puede evaluarse y obtenerse la especificidad usando diferentes umbrales de AI. Puede generarse una curva ROC que describe el comportamiento de AI en la predicción del estado del paciente (PC o GC) con 3 meses de antelación. El comportamiento obtenido por este ensayo se compara con el de las medidas clínicas estándar incluyendo SLEDAI, niveles de anti-ADNs y de C3.

También se llevará a cabo un análisis para evaluar la capacidad de AI para predecir cambios en las puntuaciones SLEDAI 3 meses después. Usando los datos de todos los puntos del estudio de todos los pacientes, la relación entre las puntuaciones AI y SLEDAI puede representarse para identificar el nivel de correlación "transversal" como se ha discutido anteriormente. Esto determina la relación entre SLEDAI y AI en el mismo punto del estudio. Esta relación se ajusta con una ecuación que permite la predicción de la puntuación SLEDAI dada una puntuación AI (o viceversa). Si AI predice crisis, entonces los cambios en SLEDAI en algún punto del estudio 1 están precedidos de cambios en AI en el punto 0. Por lo tanto, si se produce una crisis entre el punto 0 y 1, la puntuación AI en el punto 0 (denominado Almeas) es mayor de lo que se espera (denominada Alexp) dado el SLEDAI en el punto de estudio 0. Por otra parte, sin cambio en la actividad de la enfermedad entre el punto del estudio 0 y el punto del estudio 1, la puntuación AI en el punto 0 es muy similar de que se espera dado el SLEDAI en el punto del estudio 0. El cambio relativo de AI (Rel-AI-diff) puede computarse dividiendo la diferencia de Almeas y Alexp por Almeas. La sensibilidad y especificidad de AI para predecir un cambio significativo en SLEDAI 3 meses después puede evaluarse usando diferentes umbrales de Rel-AI-diff. Los umbrales pueden ser bidireccionales de manera que si Rel-AI-diff en un punto del estudio específico es mayor que un umbral específico, se predice una crisis, y de forma similar si es menor que el negativo del umbral específico se espera una reducción significativa en SLEDAI. Por otra parte, cuando el Rel-AI-diff en un punto del estudio está entre el umbral y su negativo, no se esperan cambios significativos en la actividad de la enfermedad. Puede generarse una curva ROC que muestra el cambio de sensibilidad y falsos positivos usando muchos umbrales diferentes de Rel-AI-diff. Pueden generarse curvas ROC similares usando medidas clínicas estándar incluyendo SLEDAI, niveles de anti-ADNs y de C3.

Si la fluctuación de AI varía a diferentes valores de SLEDAI, el análisis anterior se refina. Una sección anterior describió la computación de AI-RSDall y AI-RSDpt-med y mencionó la evaluación si cambiaban a diferentes valores de SLEDAI. Si lo hacen, entonces el análisis ROC puede hacerse como se ha descrito anteriormente pero en lugar de usar diferentes umbrales de Rel-AI-diff, se usan diferentes umbrales de AI-RSDall y AI-RSDpt-med. Se compara el comportamiento obtenido por el ensayo con el de medidas clínicas estándar incluyendo SLEDAI, niveles de anti-ADNs y de C3.

En el análisis anterior, se hacen intentos para predecir el SLEDAI en el punto 1 a partir de la puntuación AI en el punto 0. Es probable que además del nivel absoluto en el punto 0, el cambio de AI desde el punto -1

a 0 sea informativo para predecir SLEDAI en el punto 1. Por ejemplo, considérese un paciente que tiene en el punto del estudio -1 una puntuación AI de X-1, y en el punto 0 la puntuación AI se incrementa a un nuevo valor X0 que es apreciablemente mayor que X-1. Este paciente puede tener una mayor probabilidad de una crisis en el punto 1 que un paciente cuyo AI ha sido estable en X0 en los puntos del estudio -1 y 0. Este concepto de cambio de AI o velocidad se incorpora para generar una puntuación AI modificada (MAI). Para generar una MAI en el punto 0, se necesita la puntuación AI en el punto -1 y en el punto 0, y por lo tanto un punto de datos por paciente no tendrá una MAI asociada con él. La fórmula específica para incorporar la velocidad en el cálculo de AI para obtener MAI se optimiza. Esta optimización puede hacerse a través de la maximización de la correlación de MAI y SLEDAI tres meses después. El diseño de validación cruzada se usa para evaluar y controlar el grado de sobreajuste. La correlación puede hacerse para puntos de datos de todas las muestras, pero también puede hacerse a nivel de paciente y puede evaluarse la mediana de la correlación entre todos los pacientes. La última estrategia mejora el problema de algunos pacientes que tienen un sesgo sistemático de puntuación AI demasiado baja o demasiado alta. Usando MAI, puede realizarse el mismo tipo de análisis ROC que se mencionó para AI para evaluar su capacidad de predecir SLEDAI 3 meses después. En primer lugar, de forma análoga a la descrita para AI, puede hacerse un análisis para mostrar la capacidad de MAI en el punto 0 de distinguir los estados PC y GC en el punto 1. Además, puede realizarse un análisis similar al que se describió para AI para evaluar la capacidad de MAI en el punto 0 para predecir cambio significativo en la actividad de la enfermedad (cambio de 3 puntos en SLEDAI) entre los puntos 0 y 1. Para este último análisis, pueden usarse diferentes umbrales de Rel-AI-diff, AI-RSDall o AI-RSDpt-med. El comportamiento de MAI se compara con el AI para determinar si es útil la adición del factor de velocidad.

Una complicación del estudio descrito es que los cambios en el tratamiento se hacen para diferentes pacientes durante el periodo de seguimiento del estudio. Esto complica probablemente la predicción de la actividad de la enfermedad. Por ejemplo, considérense dos pacientes con la misma puntuación AI en el punto 0 y uno de esos pacientes tuvo una reducción en la medicación en el mismo tiempo. La probabilidad de que este paciente tenga una elevación en la actividad de la enfermedad en el punto 1 probablemente es entonces más alta que para el paciente que no cambió las medicaciones en el punto 0. Esto da lugar probablemente a una subestimación del comportamiento de AI. Una manera de aliviar esto es eliminar todos los puntos con cambios significativos en la medicación del estudio. Otra es modificar la puntuación AI para incluir si un paciente tiene un cambio en la medicación y crear un AI modificado por la medicación. Por lo tanto, en el ejemplo anterior con los dos pacientes, el que tiene el cambio en la medicación tendrá un AI modificado por la medicación más alto.

#### 5. Integración con otros marcadores predictivos

La capacidad predictiva del marcador de la actividad de la enfermedad puede maximizarse. Por lo tanto, se ensaya la capacidad predictiva de la información del repertorio TCR/BCR integrada con otros marcadores. Estos marcadores incluyen marcadores estándar usados en clínica como niveles de anti-ADNds y C3. También incluirá otros marcadores que están publicados. Por ejemplo, ya se ha mostrado que un panel de quimioquinas tiene alguna capacidad predictiva usando el mismo conjunto de pacientes que se usa. Se evalúa si este panel incrementará la capacidad predictiva del repertorio de TCR y BCR. La primera etapa es integrar la puntuación AI con la medida adicional para generar una puntuación AI expandida (EAI). Pueden evaluarse las diferentes formas para hacer la integración, y esto puede optimizarse mediante la maximización de la correlación de EAI y SLEDAI tres meses después. El diseño de validación cruzada se usa para evaluar y controlar el grado de sobreajuste. Usando EAI la capacidad para predecir la actividad de la enfermedad 3 meses después se evalúa por su capacidad para distinguir GC de PC y para predecir cambios en la actividad de la enfermedad. El comportamiento en la medida de la actividad de la enfermedad y cambio en la actividad de la enfermedad puede describirse a través de análisis ROC como se ha descrito anteriormente.

#### 6. Validación

El número de variables que se está ensayando es alto comparado con el número de muestras. Esto puede dar lugar por sí mismo a sobreajuste, con resultados inicialmente prometedores que no pueden validarse en estudios posteriores. Se usa una estrategia de validación cruzada en el entrenamiento para obtener una medida del grado de sobreajuste. Sin embargo, una validación en un conjunto independiente de muestras está implicada en el trabajo posterior. Esto no es parte de esta propuesta, pero este marcador puede ser clínicamente aplicable. Usando los datos obtenidos anteriormente, puede determinarse si AI, MAI, o EAI, deberían validarse y la manera específica de computar la medida de interés. Se toma un algoritmo específico para la validación. Además se especifican uno o más puntos finales específicos. La sensibilidad y especificidad de AI pueden evaluarse en la capacidad de distinguir GC de PC 3 meses después para evaluar la capacidad de AI para predecir la actividad de la enfermedad. En otro ejemplo, puede evaluarse la sensibilidad y especificidad de AI para predecir cambios significativos en la actividad de la enfermedad en 3 meses usando un umbral de Rel-AI-diff específico.

#### Ejemplo 5

##### Medición de la respuesta de un paciente con SLE a la terapia con fármacos

Los métodos descritos en la presente memoria se usan para medir la respuesta de un paciente con SLE a la terapia con fármacos. La determinación de si un paciente con SLE al que se ha administrado un fármaco costoso con efectos secundarios graves está respondiendo al fármaco juega un papel tanto en el cuidado del paciente como también para hacer que la administración de dicho cuidado sea rentable. Muchos indicadores clínicos de la actividad de la enfermedad responden al tratamiento de forma

imprecisa y después de un periodo de tiempo de hasta varios meses. Durante este tiempo, la enfermedad puede progresar y los efectos secundarios pueden añadir complicaciones a la terapia. Un conocimiento temprano de la respuesta al fármaco permitiría cambiar a los pacientes a terapias más efectivas más rápidamente.

5 En este Ejemplo, una mujer afroamericana de 35 años de edad con un diagnóstico previo de lupus se presenta a su reumatólogo habitual. El estado de la enfermedad de la paciente se evalúa trimestralmente a través de una evaluación clínica completa además del ensayo de laboratorio que incluye medida de C3, niveles de anticuerpo anti-ADNs, recuentos sanguíneos, y análisis de orina. Durante una visita, la paciente se queja de lesiones cutáneas y fatiga, y el análisis de orina muestra evidencia de proteinuria y/o cilindros celulares. El reumatólogo deriva la paciente a un nefrólogo para una biopsia de riñón para evaluar el estado inflamatorio del riñón y ordena creatinina en suero y proporción proteína a creatinina en orina de 24 horas para evaluar el grado de la alteración de la función renal. Una biopsia de riñón muestra evidencia de nefritis lúpica difusa, mientras el ensayo de proteína a creatinina en orina revela la evidencia de síndrome nefrótico (proporción de proteína a creatinina en orina de 3,6). Sobre la base de esta información, se proporciona un diagnóstico de nefritis lúpica aguda y la paciente se incluye en un curso de terapia con fármaco. Hay varios fármacos posibles que pueden elegirse en este punto. Los inmunomoduladores tales como micofenolato mofetil (Cellcept) se usan frecuentemente aunque algunas veces en casos graves se prescriben fármacos tales como metotrexato, azatiopurina (Imuran), ciclofosfamida (citoxan). Rituximab (Rituxan) también se usa algunas veces como una segunda o tercera elección. Uno de estos fármacos se usa frecuentemente en combinación con un esteroide sistémico tal como prednisona o metilprednisolona con el fin de suprimir los síntomas agudos. Aquí, el micofenolato mofetil se prescribe a 150 mg al día junto con 60 mg de prednisona. Dados los muchos efectos secundarios de los esteroides, incluyendo el riesgo de osteoporosis, hiperglucemia, ganancia de peso, y otros síntomas relacionados con Cushing a largo plazo, la dosis de prednisona de la paciente se disminuye durante ~6 semanas si la evaluación clínica lo permite.

La primera cuestión que se determina es si la paciente está respondiendo a terapia, y como resultado, si la dosis de esteroide puede disminuirse apropiadamente. Por lo tanto, durante este periodo la creatinina en suero de la paciente así como la proteína y creatinina en orina se siguen para asegurar que la paciente está respondiendo a las medicaciones. Puede hacerse una biopsia de riñón frecuente para detectar si el daño por inflamación se está revirtiendo, sin embargo, el uso rutinario de biopsia de riñón conlleva un riesgo demasiado alto y es demasiado invasivo para ser práctico. Los marcadores actuales basados en sangre que se están usando para evaluar el estado inflamatorio tienen un uso limitado para tomar esta decisión ya que no se correlacionan lo suficientemente bien con la enfermedad subyacente como para tomarlos como base para calcular el riesgo de los efectos secundarios incrementados que acompañan a las altas dosis de esteroides. Los marcadores de la función en suero y orina pueden presentar algún retraso para detectar la mejoría en el estado inflamatorio y por lo tanto los esteroides pueden disminuirse antes de que estos marcadores muestren un cambio definitivo y por lo tanto extendiendo el periodo de la crisis renal. Una disminución más lenta, informada por marcadores más sensibles, en estos casos podría haber acertado el periodo de crisis previniendo daño adicional al tejido renal. Después de la reducción del esteroide hasta una dosis de mantenimiento de aproximadamente 10 mg, la paciente puede mostrar niveles persistentemente elevados de proteína en la orina y la alta proporción de proteína a creatinina en orina de 2, y el médico debe decidir ahora si cambiar de Cellcept a otra medicación. Los argumentos a favor de esto son la evidencia continuada de pérdida de función renal pero sin una medida exacta del estado inflamatorio del riñón, puede ser difícil saber si la enfermedad en sí misma está en remisión habiendo causado sin embargo algún nivel de daño renal irreversible que está resultando en estos niveles persistentes de proteinuria. Aquí, de nuevo, los marcadores basados en sangre existentes informan de manera imperfecta y una biopsia de riñón adicional no es práctica. Esta decisión estaría ayudada en gran medida por una medida basada en sangre exacta del estado de la enfermedad.

La carga autoinmune sería muy útil en esta situación para evaluar la respuesta a la terapia midiendo la actividad de la enfermedad bien sola o en combinación con otros marcadores de la actividad de la enfermedad. Un algoritmo para carga autoinmune se desarrolla usando el estudio descrito anteriormente. Los clonotipos que se correlacionan que se usan para calcular la carga autoinmune se miden usando un ensayo de calibración. Este ensayo de calibración se hace usando sangre de un paciente en un tiempo de pico de la actividad de la enfermedad, por ejemplo, al inicio de la terapia. El ensayo de calibración se realiza usando sangre o alternativamente usando el tejido que está afectado (por ejemplo, biopsia de riñón o biopsia de piel). En un tiempo posterior en el que se va a evaluar la respuesta a la terapia, se toma una muestra de sangre y se usa junto con el ensayo de calibración para medir la carga autoinmune. Esto se usa para tomar una decisión sobre el tratamiento. Si los clonotipos que se correlacionan derivan de un estudio poblacional, no hay necesidad del ensayo de calibración y un ensayo de sangre en el momento en el que se va a evaluar la respuesta a la terapia es suficiente para medir la carga autoinmune con el fin de informar la decisión sobre el tratamiento.

#### **Ejemplo 6**

##### **Determinación del momento apropiado para disminuir o parar la terapia para un paciente con SLE**

Los métodos descritos en la presente memoria pueden usarse para determinar el momento apropiado para disminuir o parar la terapia para un paciente con SLE. Además del periodo de tiempo que pueden presentar las medidas clínicas de la actividad de la enfermedad, una dificultad añadida se basa en la ausencia de sensibilidad de estas medidas. La enfermedad subclínica puede resultar no obstante en una

re-activación de la enfermedad si la terapia se disminuye demasiado pronto. Como resultado de esto, los cursos de terapia inmunosupresora se administran típicamente durante un periodo de tiempo que es mucho más largo de lo necesario para el paciente medio para asegurar que el riesgo de re-activación es bajo para el paciente medio aunque aún así todavía puede ser lo suficientemente largo para a cola final de la distribución. Por lo tanto, en la mayor parte de los pacientes se está produciendo un sobre-tratamiento significativo, que causa efectos secundarios y costes, mientras se produce el sub-tratamiento de algunos pacientes lo que causa re-activaciones potencialmente prevenibles. Un método que pudiera medir la actividad subclínica que fuera predictivo del riesgo de re-activación permitiría disminuir la terapia sobre la base de dichas medidas en lugar de basarse en sobretratamiento por diseño.

En este ejemplo, la paciente del Ejemplo 7 está en prednisona y microfenolato de mofetil durante un periodo de 6 meses y la proporción proteína a creatinina en orina vuelve a un nivel de 0,5. Este nivel permanece por encima del nivel de línea base esperado en individuos sanos pero no está claro que este nivel no se deba a algún daño renal que no es reversible. Otras medidas clínicas de inflamación son normales y la paciente no reporta ningún otro síntoma. Al mismo tiempo, la paciente está experimentando niveles moderados de náusea y ganancia de peso como efectos secundarios posibles de las medicaciones que además tienen efectos secundarios graves a largo plazo. El médico se enfrenta con una decisión difícil: equilibrar el miedo de disminuir el Cellcept y/o esteroide demasiado rápido, lo que podría resultar en inflamación renal renovada y probablemente daño renal irreversible adicional a largo plazo y las reacciones adversas que pueden ocurrir debido a las medicaciones. Aquí, de nuevo, una evaluación inequívoca del estado de la enfermedad sin tener que realizar una biopsia de riñón jugaría un papel para tomar esta decisión. El intento de reducir los esteroides se recomienda a través de ensayos repetidos con esteroides que dan lugar a la recurrencia del mismo dilema clínico. De hecho, esta cuestión surge cada momento en el que el paciente está en remisión y el paciente está en esteroides o inmunomoduladores.

La carga autoinmune Load sería muy útil en esta situación para evaluar si disminuir o no la terapia midiendo la actividad de la enfermedad bien sola o en combinación con otros marcadores de la actividad de la enfermedad. Un algoritmo para la carga autoinmune se desarrolla usando el estudio descrito anteriormente. Los clonotipos que se correlacionan que se usan para calcular la carga autoinmune se miden usando un ensayo de calibración. Este ensayo de calibración se hace usando sangre de un paciente en un tiempo de pico de la actividad de la enfermedad, por ejemplo, al inicio de la terapia. El ensayo de calibración podría realizarse usando sangre o alternativamente usando el tejido que está afectado (por ejemplo, biopsia de riñón o biopsia de piel). En un tiempo posterior en el que se va a evaluar el nivel de la actividad de la enfermedad, se puede tomar una muestra de sangre y se usa junto con el ensayo de calibración para medir la carga autoinmune. Esto se usa para tomar una decisión sobre el tratamiento y para evaluar si el paciente tiene alguna actividad de la enfermedad detectable. Si los clonotipos que se correlacionan derivan de un estudio poblacional, no hay necesidad del ensayo de calibración y un ensayo de sangre en el momento en el que se va a evaluar la respuesta a la terapia es suficiente para medir la carga autoinmune con el fin de informar la decisión sobre el tratamiento.

#### **Ejemplo 7**

#### **Predicción de crisis en un paciente con SLE**

Un reto en el tratamiento de pacientes con SLE es el representado por el hecho de que las crisis surgen sin aviso, frustrando así los esfuerzos de los médicos para tratar la enfermedad de forma preventiva. La espera para se produzcan crisis antes de empezar el tratamiento somete a los pacientes a síntomas clínicos potencialmente destructivos, puede implicar hospitalización costosa e inconveniente, y puede causar que se produzca daño orgánico a largo plazo mientras también se necesitan intervenciones terapéuticas agresivas que en sí mismas están cargadas de efectos secundarios. Un paradigma mucho más deseable sería un paradigma terapéutico en el que las crisis se detectan en una fase subclínica momento en el que la terapia podría administrarse proactivamente ahorrando un sufrimiento significativo al paciente, resultando en hospitalizaciones menos costosas y finalmente permitiendo un mejor pronóstico a largo plazo para los pacientes.

La paciente del Ejemplo 7 se está recuperando de la crisis aguda descrita anteriormente, y la paciente se quita de todas las terapias excepto Plaquinil y una dosis baja de 5 mg de prednisona. No obstante, este paciente permanece con un alto riesgo de tener otro episodio inflamatorio. Como resultado, esta paciente permanecerá bajo el cuidado de un reumatólogo que continuará siguiendo los síntomas clínicos del paciente y ensayos de laboratorio. Desafortunadamente, estos síntomas y ensayos no proporcionan un aviso temprano de una crisis inminente hasta que los pacientes han presentado realmente los síntomas clínicos de una crisis y la secuencia se repite. Un marcador altamente específico del incremento de la actividad subclínica podría incluirse en la evaluación clínica rutinaria del paciente con el fin de detectar signos inequívocos de una crisis que puede alcanzar un estadio clínicamente detectable en los 1-3 meses posteriores. El comienzo más temprano de las terapias podría hacer que la crisis fuera menos grave y puede permitir que se consiga el tratamiento con menos daño orgánico a largo plazo o usar menos esteroides de lo que es actualmente el caso.

La carga autoinmune sería muy útil en esta situación para evaluar la probabilidad de una crisis incipiente midiendo la actividad de la enfermedad bien sola o en combinación con otros marcadores de la actividad de la enfermedad. Esta puntuación bien en sí misma o la proporción de incremento (velocidad) o aceleración de esta puntuación puede usarse para evaluar la probabilidad de progresión a una crisis. Podría desarrollarse un algoritmo para la carga autoinmune usando el estudio descrito anteriormente. Los

clonotipos que se correlacionan que se usan para calcular la carga autoinmune podrían medirse usando un ensayo de calibración. Este ensayo de calibración podría hacerse usando sangre de un paciente en un tiempo de pico de la actividad de la enfermedad, por ejemplo, al inicio de la terapia. El ensayo de calibración podría realizarse usando sangre o alternativamente usando el tejido que está afectado (por ejemplo, biopsia de riñón o biopsia de piel). En un tiempo posterior en el que se va a evaluar la respuesta a la terapia, se puede tomar una muestra de sangre y se usa junto con el ensayo de calibración para medir la carga autoinmune. Esto puede usarse para tomar una decisión sobre el tratamiento. Si los clonotipos que se correlacionan derivan de un estudio poblacional, no hay necesidad del ensayo de calibración y un ensayo de sangre en el momento en el que se va a evaluar el riesgo de crisis es suficiente para medir la carga autoinmune con el fin de informar la decisión sobre el tratamiento.

#### **Ejemplo 8**

##### **Medida objetiva para evaluar síntomas subjetivos de pacientes con SLE**

SLE afecta muchos órganos y produce muchos síntomas potenciales que incluyen unos que son muy comunes en las poblaciones sanas. Por ejemplo, si un paciente con SLE se queja de un dolor de cabeza, el dolor de cabeza puede ser un signo de lupus en SNC o puede deberse al dolor de cabeza común. De forma similar, si los pacientes con SLE se quejan de empeoramiento en la fatiga durante un periodo de tiempo, el empeoramiento en la fatiga puede deberse al deterioro de su enfermedad o puede deberse a depresión u otras causas. La disponibilidad de una medida objetiva que refleje la actividad de la enfermedad puede ser de gran ayuda en la gestión de los pacientes con SLE.

La paciente del Ejemplo 7 se presenta al reumatólogo con quejas principales de dolor de cabeza, fatiga, y dificultad para concentrarse. El dolor de cabeza de la paciente es recurrente y sólo mejora transitoriamente por tratamiento con motrina. El SLE de la paciente está por lo demás con buen control. Los factores estresantes psicosociales relevantes en la vida de la paciente incluyen que se está divorciando. Los médicos tienen un dilema cuando se enfrentan a pacientes con SLE con síntomas que no son específicos de SLE y son comunes en la población general. ¿Está sufriendo la paciente de lupus en el SNC? O ¿podría estar sufriendo otras causas comunes de sus síntomas, como depresión? Los ensayos de laboratorio actuales carecen actualmente de la sensibilidad y especificidad para servir de base para distinguir entre estas posibilidades. Un ensayo fiable para medir la actividad de la enfermedad SLE puede utilizarse rutinariamente para ayudar a distinguir entre las dos posibilidades.

La carga autoinmune sería muy útil en esta situación para evaluar objetivamente la actividad de la enfermedad bien sola o en combinación con otros marcadores de la actividad de la enfermedad. Un algoritmo para la carga autoinmune se desarrolla usando el estudio descrito anteriormente. Los clonotipos que se correlacionan que se usan para calcular la carga autoinmune se medirán usando un ensayo de calibración. Este ensayo de calibración se hace usando sangre de un paciente en un tiempo de pico de la actividad de la enfermedad, por ejemplo, al inicio de la terapia. El ensayo de calibración se hace usando sangre o alternativamente usando el tejido que está afectado (por ejemplo, biopsia de riñón o biopsia de piel). En un tiempo posterior en el que se va a evaluar la actividad de la enfermedad objetiva, se puede tomar una muestra de sangre y se usa junto con el ensayo de calibración para medir la carga autoinmune. Esto se usa para tomar una decisión sobre el tratamiento. Si los clonotipos que se correlacionan derivan de un estudio poblacional, no hay necesidad del ensayo de calibración y un ensayo de sangre en el momento en el que se va a evaluar la actividad de la enfermedad objetiva es suficiente para medir la carga autoinmune con el fin de informar la decisión sobre el tratamiento.

#### **Ejemplo 9**

##### **Medición de la respuesta a la terapia con fármacos de un paciente con MS**

Como se ha indicado anteriormente, uno de los retos principales en la terapia de MS es medir cómo de bien y si un paciente está respondiendo a una terapia con fármacos. Durante la enfermedad progresiva y en estadio tardío hay evaluaciones clínicas tales como la puntuación del estado de discapacidad expandida (EDSS) que miden el grado de alteración física que ha resultado de la enfermedad. Sin embargo, estas evaluaciones no son útiles en la enfermedad en estadio temprano o recidiva/remisión. Los parámetros clínicos acerca de las recidivas pueden usarse para evaluar la progresión de la enfermedad, pero éstos con indicadores vulgares y atrasados, ya que los pacientes pueden estar varios años entre recidivas, durante los cuales pueden deducirse pocas evidencias a partir de las evaluaciones clínicas. Por último, la formación de imágenes cerebrales tales como MRI potenciada con gadolinio puede usarse para examinar las lesiones cerebrales. Los pacientes con MS se someten típicamente a dicha MRI anualmente. Sin embargo, dichas imágenes carecen de especificidad. Además, como una medida de daño cerebral integrado, no son buenas medidas de la actividad de la enfermedad actual sino que reflejan el historial de la enfermedad y su impacto en el cerebro.

Aunque es cierto que el paradigma del tratamiento clínico actual para MS es que los pacientes diagnosticados con enfermedad recidiva remitente deben estar bajo terapia continua con el fin de retrasar el inicio de la enfermedad progresiva, el repertorio creciente de fármacos aprobados para tratar MS hace que la ausencia de repercusión biológica sea cada vez más problemática. La lista mostrada anteriormente de fármacos aprobados para tratar MS continúa creciendo al empezar a dar fruto la inversión sustancial en terapias para MS. Cada uno de estos fármacos tiene efectos secundarios graves y es muy costoso de administrar, con costes de \$30.000-\$100.000 por año de tratamiento. Los pacientes que no están bien gestionados pronto pasarán a enfermedad progresiva que es debilitante y causa intervenciones sanitarias costosas incluyendo hospitalizaciones y cuidado a largo plazo. Por lo tanto, puede permitirse que el paciente reciba una terapia óptima antes en el tratamiento.



## Ejemplo de utilidad clínica

Perfil del paciente: una mujer de 30 años de edad viene al hospital con alteración visual monocular con dolor. Se le proporciona una evaluación neurológica y una punción lumbar para obtener fluido cerebroespinal que se usa para evaluar si están presentes células T clonales. También se refiere para un MRI cerebral. Sobre la base de estos ensayos, se hace un diagnóstico de MS. Se le prescribe Betaseron 250 mcg por inyección para ser auto-administrado subcutáneamente en días alternos. En una visita de seguimiento seis meses después, la paciente se queja de depresión y ganancia de peso. No se han reportado al médico eventos neurológicos adicionales. El médico se enfrenta ahora con un dilema clínico. ¿Debe el médico mantener la terapia como se está administrando? ¿Debería usarse otra terapia? ¿Debería el médico mandar un MRI incurriendo en coste y sometiendo al paciente a exposición adicional a contraste? ¿Debería el médico esperar hasta que el siguiente MRI programado muestre nuevas lesiones? ¿Debería el médico esperar para ver si vuelve a ocurrir una crisis? Todas estas decisiones se beneficiarían de una medida inequívoca de si la enfermedad está activa o no.

La carga autoinmune sería muy útil en esta situación para evaluar la respuesta a la terapia midiendo la actividad de la enfermedad bien sola o en combinación con otros marcadores de la actividad de la enfermedad. Un algoritmo para la carga autoinmune se desarrolla usando los estudios descritos en la presente memoria. Los clonotipos que se correlacionan que se usan para calcular la carga autoinmune se miden usando un ensayo de calibración. Este ensayo de calibración se hace usando sangre de un paciente en un tiempo de pico de la actividad de la enfermedad, por ejemplo, al inicio de la terapia. El ensayo de calibración podría realizarse usando sangre o alternativamente usando el tejido que está afectado (por ejemplo, CSF). En un tiempo posterior en el que se va a evaluar la respuesta a la terapia, se puede tomar una muestra de sangre y se usa junto con el ensayo de calibración para medir la carga autoinmune. Esto se puede usar para tomar una decisión sobre el tratamiento. Si los clonotipos que se correlacionan derivan de un estudio poblacional, no hay necesidad del ensayo de calibración y un ensayo de sangre en el momento en el que se va a evaluar la respuesta a la terapia es suficiente para medir la carga autoinmune con el fin de informar la decisión sobre el tratamiento.

**Ejemplo 10****Predicción de crisis de MS**

Como en todas las enfermedades autoinmunes, la mejoría de las crisis es un objetivo principal de la terapia. No sólo son las crisis debilitantes para el paciente y costosas de tratar, sino que cada vez más se cree que cada crisis contribuye a una progresión de la enfermedad no reversible a largo plazo. Pueden usarse varias terapias para controlar las crisis incipientes tales como metilprednisolona IV o prednisona oral. Dichas medicaciones tienen efectos secundarios significativos y como tales no se prescriben sin evidencia de una crisis activa. Una medida de la actividad subclínica creciente que se correlacionara con crisis clínicas posteriores podría usarse para informar de esta clase de tratamiento de crisis proactivo que podría resultar en crisis más cortas y menos dañinas. Además, hay terapias que demuestran una alta eficacia clínica para la reducción de las crisis que porta riesgos de efectos secundarios muy significativos y letales. Uno de dichos fármacos es Tysabri, un fármaco que se ha mostrado que resulta tanto en resultados clínicos mejorados como en incremento del riesgo de infecciones cerebrales mortales tales como PML. Estos riesgos han reducido el valor de dichos fármacos a terapia de última línea cuando se demuestra que otros fármacos no controlan más la progresión y limitan el valor de estos fármacos como tratamientos crónicos. Un ensayo que pudiera predecir cuando un estado de crisis es incipiente podría incrementar la utilidad de dichos fármacos ya que podrían usarse de una manera similar a los esteroides para controlar periodos de crisis aguda mientras se minimizan los riesgos de efectos secundarios letales.

## Ejemplo de utilidad clínica

La paciente del Ejemplo 11 está en Betaseron durante 3 años y reporta una crisis clínica que dura una semana. El MRI de la paciente al final del año muestra nuevas lesiones significativas (múltiples lesiones ovoides de tamaño variable discreto hiperintensas T2W y FLAIR dirigidas perpendicularmente (placas), que aparecen iso-hipointensas en las imágenes T1W e hiperintensas en las imágenes T2W implicando regiones de materia blanca periventriculares y subcorticales bilaterales, incluyendo la interfase calloso-septal). El médico está preocupado de que la paciente presente alto riesgo de crisis durante el curso de los 12 meses siguientes. Un dilema clínico se presenta en sí mismo. ¿Espera el médico para síntomas clínicos adicionales para intervenir con terapia adicional? ¿Debería el médico cambiar las terapias? Si lo hiciera, ¿debería usarse otra clase de inyectable tal como copaxona o debería usarse una nueva clase de terapia tal como Tysabri? ¿Deberían prescribirse esteroides? Un ensayo que pudiera monitorizar la actividad de la enfermedad subclínica y mostrar cuando la enfermedad se incrementa y cuando es probable que resulte una crisis podría usarse para ayudar a tomar estas decisiones clínicas.

La carga autoinmune sería muy útil en esta situación para evaluar el riesgo de crisis midiendo la actividad de la enfermedad bien sola o en combinación con otros marcadores de la actividad de la enfermedad. Un algoritmo para la carga autoinmune podría desarrollarse usando los estudios descritos en la presente memoria. Los clonotipos que se correlacionan que se usan para calcular la carga autoinmune podrían medirse usando un ensayo de calibración. Este ensayo de calibración podría hacerse usando sangre de un paciente en un tiempo de pico de la actividad de la enfermedad, por ejemplo, al inicio de la terapia. El ensayo de calibración podría realizarse usando sangre o alternativamente usando el tejido que está afectado (por ejemplo, CSF). En un tiempo posterior en el que se va a evaluar el riesgo de crisis, se puede tomar una muestra de sangre y se usa junto con el ensayo de calibración para medir la carga autoinmune. Esto se puede usar para tomar una decisión sobre el tratamiento. Si los clonotipos que se correlacionan

derivan de un estudio poblacional, no hay necesidad del ensayo de calibración y un ensayo de sangre en el momento en el que se va a evaluar el riesgo de crisis es suficiente para medir la carga autoinmune con el fin de informar la decisión sobre el tratamiento.

#### **Ejemplo 11**

##### **5 Monitorización del cumplimiento con la terapia para MS**

Debido a la infrecuencia relativa de síntomas clínicos en los estadios tempranos de la enfermedad, las interacciones entre un paciente y su médico no son muy frecuentes. Al mismo tiempo, las terapias que están siendo prescritas son tanto costosas como inconvenientes para el paciente, implicando autoinyecciones que pueden causar reacción dolorosa y efectos secundarios. Como resultado hay un grado significativo de falta de cumplimiento con los regímenes terapéuticos que son difíciles de monitorizar por el médico ya que las interacciones entre el paciente y el médico no son rutinarias. Un ensayo que pudiera medir el estado de la enfermedad subclínica permitiría tanto al médico como al paciente ver rutinariamente cómo de bien controlada está la enfermedad subyacente. Dichos métodos se han mostrado muy efectivos en pacientes con VIH motivándolos a continuar la terapia de forma efectiva. Un ensayo de sangre que se realizara trimestralmente permitiría al médico ver al paciente y medir el estado de la enfermedad.

La carga autoinmune sería muy útil en esta situación para evaluar el cumplimiento con la terapia midiendo la actividad de la enfermedad bien sola o en combinación con otros marcadores de la actividad de la enfermedad. Un algoritmo para la carga autoinmune se desarrolla usando los estudios descritos en la presente memoria. Los clonotipos que se correlacionan que se usan para calcular la carga autoinmune se miden usando un ensayo de calibración. Este ensayo de calibración se hace usando sangre de un paciente en un tiempo de pico de la actividad de la enfermedad, por ejemplo, al inicio de la terapia. El ensayo de calibración podría realizarse usando sangre o alternativamente usando el tejido que está afectado (por ejemplo, CSF). En un tiempo posterior en el que se va a evaluar el cumplimiento con la terapia, se puede tomar una muestra de sangre y se usa junto con el ensayo de calibración para medir la carga autoinmune. Esto se usa para tomar una decisión sobre el tratamiento y para guiar mejor al paciente hacia un mejor cumplimiento. Si los clonotipos que se correlacionan derivan de un estudio poblacional, no hay necesidad del ensayo de calibración y un ensayo de sangre en el momento en el que se va a evaluar el cumplimiento con la terapia es suficiente para medir la carga autoinmune con el fin de informar la decisión sobre el tratamiento.

#### **30 Ejemplo 12**

##### **Amplificación de las secuencias de TCR $\beta$ e IgH de ratón**

Se desarrolla un esquema de amplificación y secuenciación para TCR $\beta$  e IgH de ratón que es similar al desarrollado para seres humanos. Se aplican métodos similares para minimizar las diferencias en la eficiencia de la amplificación de diferentes secuencias y técnicas de validación similares usando adiciones y la técnica 5' RACE descrita anteriormente. La cantidad de entrada mínima de ADNc se determina en una metodología similar como se ha descrito para muestras humanas. Una diferencia en el esquema de amplificación entre ratón y seres humanos es que los dos segmentos C para TCR $\beta$  en el ratón no tienen ningún polimorfismo en las 50 pb más cercanas a la unión J/C. Por lo tanto, en el esquema el cebador para la primera etapa de amplificación se pone en las posiciones 25-50 y para la segunda etapa de amplificación el cebador se pone en las posiciones 1-25, y el cebador tendrá una cola en 5' para el último cebador que contiene la secuencia P5. Las diferentes secuencias mejorarán la especificidad y es similar a la estrategia usada en los seres humanos excepto en que no hay necesidad de "excluir del bucle" a ninguna base para polimorfismos.

#### **45 Ejemplo 13**

##### **Análisis primario de los datos de secuencia de ratón**

El marco de análisis que se usa para el análisis de los datos de ratón es similar al descrito anteriormente para los datos humanos. Una diferencia es que las muestras de ratón se secuencian a una profundidad menor que las muestras humanas. Se anticipa que las muestras de sangre del ratón es 100  $\mu$ l. En 100  $\mu$ l de sangre hay ~100K linfocitos y por lo tanto la secuenciación a una profundidad mucho mayor que 100K no mejora significativamente la precisión. Por lo tanto, sólo se obtienen 100K lecturas para cada muestra de ratón. Aunque el número de lecturas es menor para ratón que para seres humanos, se muestrea una fracción mayor de linfocitos totales y de sangre de ratón. El número de linfocitos totales de ratón se espera que sea más de 3 órdenes de magnitud menor que el de los seres humanos. De forma similar, 100  $\mu$ l de sangre proporcionarán un muestreo mejor (~10%) de los linfocitos en la sangre de ratón en el momento cuando se compara con el muestreo obtenido usando 10 ml de sangre humana (0,2%).

#### **55 Ejemplo 14**

##### **Análisis del repertorio de IgH y TCR en el modelo de SLE en ratón**

Se usó un modelo de ratón de SLE para estudiar la relación entre el repertorio de TCR/BCR y la actividad de la enfermedad. El modelo de ratón es el B6 con los loci sle1 y sle3 de NZM2410. Estos ratones B6.sle1.sle3 (BSS) desarrollan nefritis semejante a SLE de una manera espontánea. Se estudian tres tipos de cohortes. Para todos los puntos del estudio, se obtiene BUN sanguíneo, nivel de creatinina, y anticuerpos anti-nucleares, proteína en orina, y creatinina. Se determina si una puntuación generada del repertorio de TCR/BCR en sangre se correlaciona bien con estos índices medidos de enfermedad renal. La primera cohorte es similar a la cohorte humana descrita en la que se recogen muestras de sangre longitudinales junto con evaluación de la función renal. Específicamente, 7 ratones BSS se siguen mensualmente hasta el mes 8. Al final, estos ratones se sacrifican y además de la sangre, se analizan el tejido de bazo y de riñón. Como un control, se evalúan 5 ratones B6 de una manera similar. La segunda

cohorte es transversal en el que diferentes cohortes de animales se sacrifican a tiempos específicos y se analizan muestras de bazo, riñón, y sangre en ese momento. Específicamente, 5 ratones BSS se sacrifican cada mes y se analiza la sangre, bazo, y riñón. Como un control, se evalúan dos ratones B6 control de una manera similar. Finalmente, una tercera cohorte se trata con esteroides después del inicio de la enfermedad y se obtiene una evaluación de la nefritis y muestras de sangre de manera regular después de esto. Específicamente, a los 4 meses de edad, 20 ratones que tienen la enfermedad se tratan con esteroides y después bisemanalmente durante los 4 meses siguientes se toma sangre para análisis del repertorio TCR/BCR y evaluación de la función renal. Como un control, 5 ratones BSS se tratan con placebo y se siguen de manera similar. El análisis del repertorio de TCR/BCR se realiza a partir de todos los puntos del estudio (es decir, diferentes puntos de tiempo y diferentes tejidos para el mismo punto de tiempo). El análisis implicará 2 etapas de PCR, procesamiento de secuenciación, y análisis de los datos primarios como se ha descrito anteriormente.

#### **Ejemplo 15**

##### **Identificación y dinámicas de los clonotipos que se correlacionan con SLE en ratón**

En primer lugar, se identifica un conjunto de clonotipos que se correlacionan con la función renal. Como una medida de la función renal, puede usarse proporción de proteína en orina/creatinina, creatinina en suero, o niveles de BUN. En la primera y tercera cohortes, puede evaluarse la correlación del nivel sanguíneo de cada clonotipo HPC con cada una de las tres medidas. De una manera similar a la descrita en los seres humanos, puede evaluarse si hay un gran incremento en el número de clonotipos con alta correlación para 1, 2, o las 3 de las medidas de la función renal sobre la expectación aleatoria (o ensayo de permutación). Dada esta expectación aleatoria, se selecciona el umbral de correlación en el que sólo el 10% de los clonotipos con un nivel de correlación por encima de ese umbral se espera que tenga el nivel de correlación observado por azar (10% de descubrimiento falso). Se centra la atención en estos clonotipos, y este conjunto se define como "clonotipos que se correlacionan".

Además de este método estadístico para identificar clonotipos que se correlacionan, los clonotipos podrían identificarse como relevantes para la enfermedad por un método "funcional" de enriquecimiento de clonotipos específicos en el tejido renal. Mediante el método funcional un conjunto de clonotipos puede identificarse en la cohorte 2 que puede ser relevante para la enfermedad, y éstos se denominan clonotipos que se correlacionan identificados funcionalmente. Puede evaluarse el grado de superposición entre la definición "estadística" y la definición "funcional" de los clonotipos que se correlacionan. Las cohortes 1 y 3 tienen muestras de riñón recogidas en el último punto de tiempo. Puede evaluarse si los clonotipos enriquecidos en estas muestras de riñón están presentes en la sangre y si están entre los clonotipos con mayor correlación con la función renal.

Las dinámicas de los clonotipos que se correlacionan (identificados estadísticamente y funcionalmente) pueden evaluarse. Por ejemplo, usando los datos de la cohorte 2, se evalúa el curso de tiempo de la elevación y caída (si existe) de sus niveles en los tres compartimentos: riñón, sangre, y bazo.

En los clonotipos que se correlacionan identificados estadísticamente, un subconjunto de clonotipos que se correlacionan se identificaría gracias a su correlación con la función renal. Los clonotipos que se correlacionan pueden identificarse sin conocer los datos de la función renal. En otras palabras, pueden entenderse las características que distinguen los clonotipos que se correlacionan de aquellos que son irrelevantes para la enfermedad. Con el fin de hacer esto, se identifica un conjunto de clonotipos con baja correlación con la función renal como clonotipos que no se correlacionan control.

Características de los clonotipos que se correlacionan con la enfermedad. Después de la identificación de los dos conjuntos de clonotipos, que se correlacionan y no se correlacionan, se buscan las características que distinguen estos dos conjuntos. Se realiza el análisis separado y combinado usando los clonotipos que se correlacionan identificados estadísticamente y funcionalmente. Se evalúa el mismo tipo de características estudiadas en los seres humanos, por ejemplo, el nivel del clonotipo, la presencia de restos de secuencia particulares, y la secuencia de otros clonotipos relacionados. Como se ha descrito para el estudio en seres humanos, hay un riesgo significativo de sobreajuste y por lo tanto es necesario emplear una técnica de validación cruzada o conjuntos separados de entrenamiento y ensayo.

Una utilidad para el experimento en ratón es la disponibilidad de células lo que permite la evaluación de si los clonotipos que se correlacionan están enriquecidos en un subtipo específico de células. Se estudia si los clonotipos que se correlacionan están enriquecidos en algunos subtipos de células; puede hacerse la secuenciación del conjunto completo de linfocitos y del subtipo específico en el que los clonotipos que se correlacionan están enriquecidos, y estos criterios de enriquecimiento pueden usarse como una característica adicional para distinguir los clonotipos que se correlacionan de otros clonotipos irrelevantes para la enfermedad. Con el fin de saber qué clonotipos de subtipos de células están enriquecidos se adoptan dos estrategias: dirigida por hipótesis y sin hipótesis. La primera es intentar doce marcadores de superficie candidatos en células T o B en un conjunto de muestras. Por ejemplo, un candidato es CD69 en las células T para seleccionar células T activadas. Para las células B, los estudios han mostrado el incremento de células altas en CD27 en SLE activo, y por lo tanto, es un buen candidato para un marcador de células que pueden tener enriquecimiento de los clonotipos que se correlacionan. En cada uno de estos experimentos, los subtipos de células específicos se purifican mediante FACS. Después, se hace una reacción de secuenciación para ADNc del complemento completo de los linfocitos así como para ADNc de los linfocitos que se purificaron por FACS a partir de una colección de diferentes muestras. Se evalúa si los dos conjuntos de clonotipos que se correlacionan y no se correlacionan están presentes en diferentes proporciones en el complemento completo de linfocitos comparado con el subconjunto

purificado por FACS. Los marcadores que tienen una gran diferencia pueden ser útiles para identificar los clonotipos que se correlacionan. El enriquecimiento de los clonotipos en subtipos de células con estos marcadores se usa además de los parámetros de secuencia para detectar clonotipos que se correlacionan.

5 En la estrategia sin hipótesis, se buscan marcadores que se expresan de manera diferencial en células con un clonotipo que se correlaciona de otras células. Se eligen unos pocos casos en los que un clonotipo de TCR específico se correlaciona claramente con la enfermedad, y se seleccionan casos en los que ese clonotipo está altamente enriquecido ya que representa la mayoría de los clonotipos con el mismo segmento V. Se hace FACS usando anticuerpo frente al segmento V específico (los anticuerpos frente a todos los segmentos V están disponibles comercialmente) para seleccionar una población que está altamente enriquecida para células que portan el clonotipo que se correlaciona. El ARN puede prepararse a partir de estas células y puede estudiarse la expresión de todos los genes realizando un experimento en matriz. Como un control, puede usarse ARN total de linfocitos y/o ARN de células purificadas por FACS que portan otro segmento V irrelevante. Pueden buscarse los marcadores que distinguen de forma máxima la muestra obtenida del segmento V purificado por FACS con el clonotipo que se correlaciona de los controles. Pueden encontrarse los marcadores, incluyendo los marcadores de superficie (ya que es mucho más fácil hacer FACS con proteínas de superficie) que distinguen las dos poblaciones. Si se observa un marcador de ARN consistente de muestras de varios ratones se valida a nivel de proteína. Usando las mismas muestras, se usan anticuerpos frente a la proteína marcadora en un ensayo FACS para purificar las células que portan la proteína marcadora. Puede ensayarse más de un marcador para incrementar la oportunidad de validar uno de ellos. Se secuencian el TCR y/o BCR de las células purificadas. Si los resultados de ARN se mantienen a nivel de proteína, entonces los clonotipos que se correlacionan deben estar enriquecidos en el subconjunto purificado de células. Después de validar esto, esos resultados de ARN todavía mantenidos a nivel de proteína, los resultados se validan en otras muestras. Las muestras que no eran objeto del análisis en matriz se someten a análisis FACS usando el anticuerpo frente a la o las proteínas marcadoras. Se secuencian el TCR y/o BCR de las células purificadas. Se evalúa si los clonotipos que se correlacionan están enriquecidos en las células purificadas usando anticuerpo frente al o a los marcadores específicos. Esto validará la utilidad del o de los marcadores en la identificación de clonotipos que se correlacionan.

#### 30 **Ejemplo 16**

##### **Uso del repertorio de IgH y TCR $\beta$ para medir la actividad de la enfermedad**

El algoritmo para los clonotipos que se correlacionan de anteriormente puede aplicarse para identificar en todas las muestras de cohortes 1 y 3 clonotipos que se correlacionan gracias a su secuencia y/o marcadores. Usando el nivel de los clonotipos que se correlacionan en cada paciente, puede generarse una puntuación AI que se correlaciona con una medida de la función renal. Como se ha descrito anteriormente, hay un riesgo de sobreajuste y es necesario emplear la técnica de validación cruzada y/o conjunto separado de entrenamiento y ensayo. La correlación de AI y las medidas de la función renal puede evaluarse de una manera transversal (todos los puntos del estudio en todos los ratones). La cuestión de si los cambios en la puntuación AI en un ratón individual también pueden evaluarse cuando la función renal cambia. Esto puede evaluarse comparando el AI de la función renal alta y baja en el mismo animal de una manera similar a lo que se ha descrito en losseres humanos.

#### 40 **Ejemplo 17**

##### **Monitorización de recurrencia metastásica en pacientes con cáncer de colon**

Muchos cánceres que se detectan en un estadio tratable todavía poseen un alto riesgo para el paciente de recurrencia tumoral metastásica. Dichas recurrencias se detectan frecuentemente tarde y en estadios intratables y pueden ser mortales para los pacientes. Un ejemplo de dicha situación es la de cáncer de colon recurrente. A pesar de los programas de cribado cada vez más agresivos de cáncer de colon, el cáncer de colon representa una de las malignidades más comunes en los EEUU. Aproximadamente 150.000 pacientes al año se diagnostican con cáncer de colon en estadios graves pero tratables (Estadio II y Estadio III). Estos pacientes se tratan por resección tumoral seguido de un curso de quimioterapia. Aunque estos tratamientos son generalmente efectivos, hay sin embargo una posibilidad significativa de que estos pacientes tendrán recurrencias metastásicas del tumor primario en los años siguientes al tratamiento. El 50% de los pacientes en Estadio III por ejemplo tendrán recurrencia en los 5 años posteriores a la cirugía. Estas recurrencias pueden ser bien aisladas (por ejemplo, en el colon o el hígado) o multifocales. En cualquier caso, pero particularmente si son aisladas, su detección en un estadio temprano puede jugar un papel para maximizar las posibilidades de terapia exitosa (cirugía y/o quimioterapia).

Actualmente hay dos ensayos usados en el seguimiento posterior al tratamiento. El escaneo CT del abdomen y pecho se usa para identificar tumores visibles en estas imágenes. Típicamente, estos escaneos se hacen a intervalos de 6-12 meses durante los primeros 5 años posteriores a la terapia. Aunque estos escaneos pueden revelar malignidades en estadio temprano, su eficacia clínica está siendo discutida. Los inconvenientes de estos escaneos incluyen el hecho de que someten a los pacientes a cantidades significativas de radiación que pueden en sí mismas causar tumores adicionales y el coste significativo. Se ha mostrado que otro ensayo basado en sangre tiene algún valor: ensayo de CEA. Este ensayo de anticuerpo mide el nivel de una proteína en suero que es específica para algunos tumores de colon. El inconveniente del ensayo de CEA es su falta de sensibilidad (<60% de los pacientes con escaneos CT positivos tiene un ensayo de CEA positivo).

Los linfocitos obtenidos del tumor primario reseccionado se usan para desarrollar un perfil inmune que puede usarse para añadir sensibilidad a un ensayo basado en sangre para recurrencia de cáncer temprana. Los TCR (y/o BCR) de los linfocitos encontrados en el tumor reseccionado pueden amplificarse y secuenciarse. Los clonotipos que están enriquecidos en la muestra de tumor son probablemente

5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55  
60

relevantes para la respuesta inmune al tumor. Las tomas posteriores de sangre del paciente pueden usarse para evaluar el nivel de estos clonotipos. Una elevación en el nivel de estos clonotipos puede señalar una respuesta inmune a una recurrencia tumoral. En este caso, la detección de la respuesta inmune puede ser más sensible que la detección del marcador tumoral en sí mismo. Estudio de descubrimiento para la detección de recurrencia de cáncer usando un ensayo de calibración. Se realiza un estudio de descubrimiento para determinar la probabilidad de la detección de la recurrencia dado el perfil de TCR (y/o BCR) en sangre. Las muestras de muestras de tumor reseccionado así como las muestras de sangre de seguimiento de pacientes con resultado conocido se usan para este estudio. Se secuencia TCR (y/o BCR) de todas estas muestras. Los candidatos para los clonotipos que se correlacionan son aquellos que están presentes en los datos de TCR (y/o BCR) de las muestras de tumor. Dados los resultados conocidos en este estudio de entrenamiento, se considera uno que usa las técnicas de validación cruzada estándar, un modelo que genera una puntuación (riesgo de recurrencia) dado el nivel de los diferentes clonotipos. Esta puntuación de recurrencia se calcula así en un nuevo paciente midiendo los clonotipos en el tumor reseccionado (punto de calibración) y los datos de los clonotipos encontrados en la sangre del mismo paciente en un momento posterior durante el seguimiento para recurrencia. El uso de los datos del tumor permite una gran reducción en el número de clonotipos presentes en la sangre que se consideran en este análisis.

Estudio de descubrimiento para la detección de recurrencia de cáncer usando un ensayo de calibración y un estudio poblacional. Es probable que no todos los clonotipos que están enriquecidos en la muestra de tumor sean relevantes para la respuesta inmune al tumor. Podría haber algunos linfocitos que se expandan localmente debido a una afección inflamatoria favorable. En otro caso, el estudio de descubrimiento se hace usando las mismas muestras pero el estudio se usa para identificar parámetros que distinguen clonotipos "que se correlacionan" de clonotipos "que no se correlacionan". Estos parámetros pueden incluir 1) Resto de secuencia: el resto puede ser una región V o J específica, una combinación VJ, o secuencias cortas en una región DJ que están asociadas con un clonotipo que se está correlacionando; 2) Tamaño del clonotipo; 3) Nivel: nivel absoluto (número de lecturas por millón) o nivel de rango; 4) Similitud con otros clonotipos: la presencia de otros clonotipos altamente relacionados, como aquellos con cambios silenciosos (diferencias de nucleótidos que codifican los mismos aminoácidos) o aquellos con cambios de aminoácidos conservativos; 5) Para los BCR el nivel de las mutaciones somáticas en el clonotipo y/o el número de clonotipos distintos que se diferencian por mutaciones somáticas de algunos clonotipos de la línea germinal. 6) Presencia en una célula que porta un marcador específico. Este estudio resulta entonces en un algoritmo que puede predecir qué clonotipos probablemente se correlacionen con recurrencia de cáncer en sangre dado un conjunto específico de clonotipos presente en una muestra de tumor dada. Estos clonotipos se usan entonces para desarrollar una puntuación de riesgo de recurrencia de la misma manera que se ha descrito anteriormente.

Estudio de descubrimiento para la detección de recurrencia de cáncer usando un estudio poblacional. En otro caso, los clonotipos medidos en el tumor reseccionado se usan para generar un modelo que predice clonotipos que se correlacionan en muestras hasta el momento no observadas. Este modelo también puede usarse para generar una puntuación de riesgo de recurrencia de una manera análoga a la descrita anteriormente. En este modelo, no habría necesidad de medir los clonotipos en el tejido canceroso reseccionado en un nuevo paciente sometido a seguimiento de recurrencia sino que el riesgo de recurrencia podría evaluarse simplemente midiendo los clonotipos en una muestra de sangre dada.

Estudio de descubrimiento para la detección de cáncer de colon primario usando un estudio poblacional. Como una extensión, la detección de cánceres primario se consigue usando la misma metodología. Con los cánceres primarios no hay tumor reseccionado que pueda usarse para enriquecer para clonotipos relevantes. Sin embargo, incluso en presencia de datos de resección tumoral la secuencia adicional y otros parámetros necesarios usados para identificar los clonotipos relevantes y generar una puntuación para la probabilidad de la detección del cáncer. Por lo tanto, por extensión si el algoritmo es lo suficientemente predictivo se puede detectar el cáncer a partir de la sangre (u otro fluido corporal) sin los datos del tumor reseccionado. En este caso, es necesario que esté disponible un estudio de descubrimiento con muestras de sangre de pacientes antes de su diagnóstico de cáncer primario.

De una manera análoga a la descrita anteriormente, pueden identificarse parámetros (secuencia y otros) para predecir los clonotipos que se correlacionan con la respuesta del sistema inmune al tumor. Puede usarse entonces un modelo para generar una puntuación de riesgo de cáncer que predice el riesgo de progresión para cáncer de colon. Este algoritmo puede aplicarse a la muestra de sangre de un paciente nuevo para medir el riesgo de cáncer de colon primario.

#### **Ejemplo 18**

##### **Monitorización del rechazo en pacientes con trasplante de corazón**

Los trasplantes de corazón son un procedimiento relativamente poco común ya que el suministro de órganos está muy limitado. En el mundo se realizan cada año 3.500 trasplantes de corazón. Cada procedimiento es muy costoso y los órganos que se usan no tienen precio. Como resultado, los pacientes que reciben estos órganos se tratan de una manera extremadamente proactiva. Con el fin de medir el estado de la reacción inmune al órgano donado en un momento en el que las intervenciones con

65

inmunosupresores pueden ser efectivas, los pacientes son sometidos a biopsias de corazón periódicas para medir la inflamación del órgano. Sobre la base de estos ensayos, pueden proporcionarse cursos agresivos de inmunosupresores. Estos procedimientos tienen varias limitaciones. Como procedimientos quirúrgicos invasivos, tienen riesgos para el paciente. Además, son costosos y sólo pueden hacerse a intervalos infrecuentes. Se ha mostrado que un ensayo basado en sangre basado en el perfilado de la expresión de un panel de 11 genes de ensayo (Allomap) es bastante sensible en la detección del rechazo de órganos pero carece de la sensibilidad suficiente como para usarse como un reemplazo de la biopsia y se usa en lugar de esto para decidir cuando hacer una biopsia. En un caso, los perfiles de TCR (y/o BCR) se usan para evaluar el estado de "rechazo" y generar una puntuación de riesgo de rechazo que predice la probabilidad de rechazo en un marco de tiempo específico. Se concibe que puede realizarse un estudio de descubrimiento para determinar la probabilidad de rechazo dado el perfil de TCR (y/o BCR) en sangre. Esto puede usarse en la clínica para informar de las terapias inmunosupresoras que se están usando.

Descubrimiento de clonotipos que se correlacionan usando un estudio poblacional. En este caso, se usa un estudio población de pacientes post-trasplantados con muestras de sangre con resultado clínico conocido. Se secuencian TCR (y/o BCR) de todas estas muestras y la correlación de clonotipos individuales con el resultado de rechazo se usa para distinguir los clonotipos que se correlacionan de los que no se correlacionan. Posteriormente, se derivan parámetros que distinguen estas dos clases de clonotipos. Estos parámetros pueden incluir 1) Resto de secuencia: el resto puede ser una región V o J específica, una combinación VJ, o secuencias cortas en una región DJ que están asociadas con un clonotipo que se está correlacionando; 2) Tamaño del clonotipo; 3) Nivel: nivel absoluto (número de lecturas por millón) o nivel de rango; 4) Similitud con otros clonotipos: la presencia de otros clonotipos altamente relacionados, como aquellos con cambios silenciosos (diferencias de nucleótidos que codifican los mismos aminoácidos) o aquellos con cambios de aminoácidos conservativos; 5) Para los BCR el nivel de las mutaciones somáticas en el clonotipo y/o el número de clonotipos distintos que se diferencian por mutaciones somáticas de algunos clonotipos de la línea germinal. 6) Presencia en una célula que porta un marcador específico. Un método alternativo o suplementario para definir el clonotipo que se correlaciona y no se correlaciona vendría si las muestras de estudio tienen muestras de biopsia del injerto, particularmente si estaba en rechazo activo. Se espera que en ese momento haya un gran enriquecimiento de los clonotipos que se correlacionan. Los parámetros para distinguir éstos de los demás clonotipos se identifican como se ha discutido anteriormente.

Los datos del perfil de las muestras de sangre se usan entonces para predecir la probabilidad de rechazo. Dados los resultados conocidos en este estudio de entrenamiento se puede considerar un modelo que usa las técnicas de validación cruzada estándar que genera una puntuación de riesgo de rechazo dado el nivel de los diferentes clonotipos. Dado el perfil en una nueva muestra de sangre de TCR (y/o BCR) en un punto específico puede generarse una puntuación de riesgo de rechazo relacionado con la probabilidad de rechazo.

Descubrimiento de clonotipos que se correlacionan usando un ensayo de calibración. En otro caso, se implementa un método para identificar clonotipos que se correlacionan usando un ensayo de calibración para cada paciente. Este método implica tomar una primera muestra de biopsia post trasplante. La presencia de material de biopsia del injerto post trasplante ofrece la posibilidad de analizar los TCR de la muestra de biopsia para identificar los clonotipos que se correlacionan como se define por aquellos que son predominantes en la muestra. Este conjunto de clonotipos se sigue en la sangre y se genera una puntuación para la probabilidad de rechazo. El algoritmo para generar la puntuación de riesgo de rechazo se deriva a través de un estudio de descubrimiento que es similar al descrito anteriormente que utiliza los datos clínicos disponibles y los niveles de los clonotipos que se correlacionan para generar una puntuación de riesgo de rechazo que aproxima la probabilidad de rechazo.

En este caso, se hace un ensayo de calibración específico usando material de una primera biopsia post trasplante pero biopsias adicionales podrían ser reemplazadas por el uso de muestras de sangre cuyos clonotipos podrían usarse junto con este ensayo de calibración para medir una puntuación de riesgo de rechazo.

Además de la biopsia del injerto, las muestras de sangre antes del trasplante sirven como otro punto de calibración. Los clonotipos que son predominantes en esta muestra es improbable que estén relacionados con el rechazo representando en su lugar el historial de antígenos anteriores a los que ha estado expuesto el paciente. Por lo tanto, cuando se consideran las muestras de sangre después del trasplante se pueden sustraer los clonotipos que estaban presentes antes del trasplante para determinar los clonotipos que se correlacionan. Estos clonotipos se usan entonces para generar un modelo de riesgo de rechazo.

En este caso, se usarían dos ensayos de calibración: uno anterior al trasplante y uno a partir de una biopsia después del trasplante. Estas calibraciones se podrían usar entonces junto con los clonotipos derivados de un ensayo en sangre para medir el riesgo de rechazo.

Descubrimiento de clonotipos que se correlacionan usando un ensayo de calibración y un estudio poblacional. En otro caso, la identificación de clonotipos que se correlacionan puede conseguirse a través de una combinación de las estrategias anteriores. Específicamente, esto puede conseguirse usando el estudio poblacional para generar un algoritmo para predecir los clonotipos que se correlacionan. Además, puede conseguirse a través de los datos de calibración del mismo paciente usando biopsia de injerto y/o muestras de sangre pre-trasplante. Un caso más preferido empleará ambas estrategias: algoritmo construido de población y calibración individual para identificar de manera más exacta los clonotipos que

se correlacionan. Se genera entonces una puntuación de riesgo de rechazo usando el nivel de estos clonotipos para predecir la probabilidad de rechazo a través del uso del estudio poblacional como un conjunto de entrenamiento.

En este caso, pueden usarse dos ensayos de calibración: uno anterior al trasplante y uno a partir de una biopsia después del trasplante. Estas calibraciones podrían usarse entonces junto con los clonotipos derivados de un ensayo en sangre para medir el riesgo de rechazo.

La predicción de GVHD puede hacerse de una manera muy similar con el mismo concepto del estudio poblacional para generar un algoritmo para predecir clonotipos que se correlacionan. También, la calibración "negativa" puede generarse a partir de la muestra del donante pre-trasplante. Una estrategia usando tanto el algoritmo como la calibración probablemente sea más predictiva de los clonotipos que se correlacionan. Un algoritmo para computar una puntuación de la probabilidad de GVHD dado el nivel de clonotipos que se correlacionan puede generarse usando un estudio poblacional de una manera como se ha descrito anteriormente. Este algoritmo puede usarse entonces para la predicción de la probabilidad de GVHD en el siguiente conjunto de pacientes.

#### 15 **Ejemplo 19**

##### **Monitorización para infección PML en pacientes con MS tratados con natalizumab**

El perfil de TCR y/o BCR puede usarse para detectar leucoencefalopatía multifocal progresiva (PML) subclínica en pacientes con MS. PML es una enfermedad grave y frecuentemente mortal que causa frecuentemente enfermedad desmielinante rápidamente progresiva a través de la muerte de oligodendrocitos que sintetizan mielina. Está causada por el virus JC que está presente en una fase latente en la mayoría de la población. En una fracción de la población inmunosuprimida (por ejemplo, SIDA) el virus se reactiva dando lugar al desarrollo de esta grave enfermedad. Además, algunos pacientes que están siendo inmunosuprimidos a través del uso de medicación como los pacientes post trasplante también pueden desarrollar PML. Alguna medicación específica se ha ligado con el riesgo de PML en poblaciones específicas de pacientes. Por ejemplo, natalizumab (Tysabri) se asoció con el desarrollo de más de 10 casos de PML entre pacientes con esclerosis múltiple (MS) dando lugar a su retirada del mercado durante un periodo de tiempo. Se acepta que natalizumab es más efectivo que las demás medicaciones aprobadas por la FDA para la esclerosis múltiple, pero su uso se ha limitado por el miedo de desarrollo de PML. Una vez se sospecha PML, puede realizarse plasmaféresis para reducir la concentración del fármaco en el paciente. La superposición entre los síntomas de MS y PML puede retrasar algunas veces la detección de PML. La detección temprana de PML subclínica se necesita urgentemente.

Estos clonotipos pueden discernirse de las muestras de sangre de una población en la que algunos pacientes desarrollaron PML. Esta población puede usarse para identificar clonotipos que se correlacionan con el desarrollo posterior de PML. Con la disponibilidad de estos clonotipos puede generarse un algoritmo para identificar los parámetros que distinguen éstos de otros clonotipos.

Descubrimiento de clonotipos que se correlacionan usando un estudio poblacional. En este caso, se genera un algoritmo para predecir los clonotipos que son relevantes para el surgimiento de PML. El algoritmo puede entrenarse en un conjunto de clonotipos que se considera que se correlacionan con la enfermedad. Pueden usarse muestras de sangre (u otro fluido corporal) en un estudio de descubrimiento de una población de pacientes con una infección latente con virus JC algunos de los cuales continúan hasta el desarrollo de PML. Puede secuenciarse el TCR (y/o BCR) de todas estas muestras y puede usarse la correlación de clonotipos individuales con el resultado de la reactivación del agente infeccioso para distinguir clonotipos que se correlacionan de clonotipos que no se correlacionan. Pueden identificarse los parámetros que distinguen estas dos clases de clonotipos. Estos parámetros pueden incluir 1) Resto de secuencia: el resto puede ser una región V o J específica, una combinación VJ, o secuencias cortas en una región DJ que están asociadas con un clonotipo que se está correlacionando; 2) Tamaño del clonotipo; 3) Nivel: nivel absoluto (número de lecturas por millón) o nivel de rango; 4) Similitud con otros clonotipos: la presencia de otros clonotipos altamente relacionados, como aquellos con cambios silenciosos (diferencias de nucleótidos que codifican los mismos aminoácidos) o aquellos con cambios de aminoácidos conservativos; 5) Para los BCR el nivel de las mutaciones somáticas en el clonotipo y/o el número de clonotipos distintos que se diferencian por mutaciones somáticas de algunos clonotipos de la línea germinal, 6) Presencia en una célula que porta un marcador específico. Un método alternativo o suplementario para definir el clonotipo que se correlaciona y no se correlaciona vendría de un conjunto de pacientes que están organizando una respuesta inmune frente al mismo agente infeccioso. Los clonotipos enriquecidos (particularmente aquellos que están a un nivel significativamente mayor que antes de la respuesta inmune) en estos pacientes pueden considerarse que se correlacionan y pueden identificarse los parámetros que los distinguen de los demás clonotipos.

De forma similar, puede identificarse los clonotipos que se correlacionan de muestras de pacientes con PML activa, o de estudios in vitro para identificar clonotipos que responden al antígeno del virus JC. Los clonotipos que responden pueden originarse de uno o una pluralidad de sujetos que pueden ser sanos o estar infectados con el agente infeccioso. Puede considerarse que estos clonotipos se correlacionan y pueden identificarse los parámetros que los distinguen de los demás clonotipos.

Los datos del perfil de las muestras en el estudio de descubrimiento se usan entonces para predecir la probabilidad de reactivación. Dados los resultados conocidos en este estudio de entrenamiento se puede considerar el uso de técnicas de validación cruzada estándar, un modelo que genera una puntuación de riesgo de PML dado el nivel de los diferentes clonotipos. De esta manera, dado el perfil en una muestra de

sangre de TCR (y/o BCR) en un punto específico puede generarse una puntuación respecto a la probabilidad de reactivación. Este algoritmo puede usarse ahora con datos de un paciente nuevo para predecir los clonotipos que se correlacionan del paciente así como para generar una puntuación de riesgo de PML para la probabilidad de reactivación.

5 De una manera muy similar pueden estudiarse otros resultados relacionados con infección. Por ejemplo, además de la reactivación de la infección latente, se puede evaluar el aclaramiento de la infección. Además, dado el repertorio de TCR y/o BCR se puede evaluar la probabilidad de tener inmunidad para un agente infeccioso específico.

#### **Ejemplo 20**

##### **Monitorización para la reactivación de infecciones latentes**

10 El perfilado de TCR y BCR puede usarse para monitorizar infecciones que tienen periodos de infección aguda seguidos de latencia y reactivación. Los ejemplos de dichas enfermedades incluyen hepatitis B y C así como virus herpes. Sería deseable la predicción de infecciones en un estadio temprano.

15 Descubrimiento de clonotipos que se correlacionan usando un ensayo de calibración. Puede implementarse un método para identificar clonotipos que se correlacionan usando un ensayo de calibración para cada paciente. La presencia de una muestra biológica del mismo paciente en un punto de tiempo previo cuando el paciente estaba organizando una respuesta inmune frente al agente infeccioso puede servir para identificar los clonotipos que se correlacionan. Este conjunto de clonotipos puede seguirse en la sangre y se genera una puntuación de riesgo de reactivación para la probabilidad de reactivación. El algoritmo para generar la puntuación se deriva a través de un estudio de descubrimiento que es similar al descrito anteriormente que utiliza los datos clínicos disponibles y las cuentas de los clonotipos que se correlacionan para generar una puntuación de riesgo de reactivación que aproxima la probabilidad de reactivación. Para usar esta puntuación se toma una muestra de un nuevo paciente en la práctica clínica durante un periodo de infección aguda. Estos datos se usarán junto con una muestra posterior tomada durante el periodo latente para medir el riesgo de reactivación para propósitos clínicos.

20 Descubrimiento de clonotipos que se correlacionan usando un ensayo de calibración y un estudio poblacional. La identificación de los clonotipos que se correlacionan puede conseguirse a través de una combinación de las estrategias anteriores. Específicamente, esto puede conseguirse usando el estudio poblacional para generar un algoritmo para predecir los clonotipos que se correlacionan. Los clonotipos que se correlacionan pueden obtenerse de un estudio poblacional de pacientes con resultado conocido de la infección y/o un conjunto de pacientes con respuesta inmune activa frente al agente infeccioso, y/o de experimentos in vitro para identificar clonotipos reactivos con el agente infeccioso. Además, puede conseguirse a través de los datos de calibración del mismo paciente usando puntos de datos más antiguos en el momento de una respuesta inmune activa frente al agente infeccioso relevante.

25 Preferiblemente, se emplearán ambas estrategias: algoritmo construido sobre población y calibración individual para identificar de la forma más exacta los clonotipos que se correlacionan. Se genera entonces una puntuación de riesgo de reactivación usando el nivel de estos clonotipos para predecir la probabilidad de reactivación a través del uso del estudio poblacional como un conjunto de entrenamiento. Para usar esta puntuación se perfila una muestra tomada de un paciente nuevo en la clínica durante un periodo de infección aguda. Estos datos se usarán junto con una muestra posterior tomada durante el periodo latente para medir el riesgo de reactivación para propósitos clínicos. Puede emplearse una estructura similar para estudiar el aclaramiento del agente infeccioso y/o la inmunidad frente a éste.

#### **Ejemplo 21**

##### **Monitorización para reacción alérgica durante inmunoterapia**

45 La rinitis alérgica es una afección común que afecta al ~11% de la población de EEUU. Ésta es típicamente una alergia al polen o polvo. La eliminación de la exposición es difícil e implica un esfuerzo de vigilancia. Los tratamientos más comunes usados en la rinitis crónica son descongestionantes, antihistamínicos, y esteroides nasales. En los casos graves, se emplea inmunoterapia. El objetivo de la inmunoterapia es desensibilizar al paciente. En primer lugar, se hace un pulso con muchos alérgenos potenciales para identificar el alérgeno específico frente al que está reaccionado el paciente. Después, se inyecta al paciente una cantidad creciente de alérgeno durante un periodo de meses a años hasta que se consigue una dosis de mantenimiento, y el tratamiento se continúa durante varios años. Típicamente, el paciente puede sentir una mejoría en los síntomas en 3-6 meses, pero también puede ocurrir tan tarde como 12-18 meses, pero una gran fracción de los pacientes no se beneficia del tratamiento o tiene recidivas. Una razón para la lenta escalada de la dosis es el riesgo de anafilaxis si se proporciona al paciente una alta dosis de alérgeno antes de que esté suficientemente desensibilizado.

50 Los perfiles de TCR (y/o BCR) pueden usarse para evaluar el estado de la enfermedad en rinitis alérgica y generar una puntuación de alergia que predice la predisposición del paciente para organizar una respuesta alérgica si se expone al alérgeno relevante. Se concibe que puede realizarse un estudio de descubrimiento para determinar la probabilidad de la respuesta alérgica dado el perfil de TCR (y/o BCR) en sangre. Esto puede usarse para personalizar el tratamiento de inmunoterapia. La posible decisión clínica puede ser interrumpir el tratamiento si se considera ineficaz, continuar el régimen de inyecciones, o acelerar el tratamiento para alcanzar la dosis de mantenimiento más rápidamente.

60 Descubrimiento de clonotipos que se correlacionan usando un estudio poblacional. En este caso, puede usarse una población de pacientes con rinitis alérgica en inmunoterapia con muestras de sangre con resultado clínico conocido. Puede secuenciarse TCR (y/o BCR) de todas estas muestras y puede usarse la correlación de clonotipos individuales con resultado de alergia para distinguir los clonotipos que se



correlacionan de los que no se correlacionan. Posteriormente, se pueden derivar parámetros que distinguen estas dos clases de clonotipos. Estos parámetros pueden incluir 1) Resto de secuencia: el resto puede ser una región V o J específica, una combinación VJ, o secuencias cortas en una región DJ que están asociadas con un clonotipo que se está correlacionando; 2) Tamaño del clonotipo; 3) Nivel: nivel absoluto (número de lecturas por millón) o nivel de rango; 4) Similitud con otros clonotipos: la presencia de otros clonotipos altamente relacionados, como aquellos con cambios silenciosos (diferencias de nucleótidos que codifican los mismos aminoácidos) o aquellos con cambios de aminoácidos conservativos; 5) Para los BCR el nivel de las mutaciones somáticas en el clonotipo y/o el número de clonotipos distintos que se diferencian por mutaciones somáticas de algunos clonotipos de la línea germinal, 6) Presencia en una célula que porta un marcador específico. Un método alternativo o suplementario para definir el clonotipo que se correlaciona y no se correlaciona usaría la biopsia de material de ensayo positivo en alergia de pacientes positivos para un alérgeno específico. En el sitio de la inyección del alérgeno se espera que haya un enriquecimiento de los clonotipos que se correlacionan. Los parámetros para distinguir éstos de los demás clonotipos pueden identificarse como se ha discutido anteriormente.

Los datos del perfil de las muestras de sangre se usan entonces para predecir el estado de la alergia. Dados los resultados conocidos en este estudio de entrenamiento se puede considerar un modelo usando las técnicas de validación cruzada estándar que genera una puntuación de alergia dado el nivel de los diferentes clonotipos. Dado el perfil en una nueva muestra de sangre de TCR (y/o BCR) en un punto específico, puede generarse una puntuación de alergia para estimar el grado en el que este paciente está predispuesto para organizar una respuesta alérgica.

Descubrimiento de clonotipos que se correlacionan usando un ensayo de calibración. Puede implementarse un método para identificar clonotipos que se correlacionan usando un ensayo de calibración para cada paciente. Este método implica tomar del paciente una muestra de biopsia de un sitio con una respuesta positiva al alérgeno. Ésta puede ser del ensayo de alergia inicial que se realizó para determinar el alérgeno específico frente al que responde el paciente o una muestra del sitio de cualquiera de las inyecciones de tratamiento adicionales. Esto puede hacerse más de una vez para asegurar que se estén siguiendo los clonotipos apropiados en el caso de que haya algo de diseminación de epítomos. Puede usarse TCR y/o BCR de estas muestras de biopsia para identificar los clonotipos que se correlacionan como se define por aquellos que son predominantes en esta muestra. Este conjunto de clonotipos puede seguirse en la sangre y se genera una puntuación para la probabilidad de respuesta alérgica. El algoritmo para generar la puntuación de alergia se deriva a través de un estudio de descubrimiento que es similar al descrito anteriormente que utiliza los datos clínicos disponibles y los niveles de los clonotipos que se correlacionan para generar una puntuación de alergia que estima el estado de la alergia.

Descubrimiento de clonotipos que se correlacionan usando un ensayo de calibración y un estudio poblacional. La identificación de los clonotipos que se correlacionan puede conseguirse a través de una combinación de las estrategias anteriores. Específicamente, esto puede conseguirse usando el estudio poblacional para generar un algoritmo para predecir los clonotipos que se correlacionan. Además, puede conseguirse a través de los datos de calibración del mismo paciente usando una biopsia de un sitio con una respuesta a alérgeno positiva. Preferiblemente, se emplearán ambas estrategias: algoritmo construido sobre población y calibración individual para identificar de la forma más exacta los clonotipos que se correlacionan. Se genera entonces una puntuación de alergia usando el nivel de estos clonotipos para predecir el estado de la alergia a través del uso del estudio poblacional como un conjunto de entrenamiento.

#### **Ejemplo 22**

##### **Amplificación de secuencias de IgH a partir de ADN genómico**

En este ejemplo, se describe la amplificación de secuencias de IgH a partir de ADN genómico. Dicha amplificación es ventajosa porque (1) el nivel de un clonotipo en el ADN genómico puede convertirse fácilmente en número de células, y (2) en algunos neoplasmas linfoides, el ARN puede no expresarse para la reorganización de receptor inmune relevante.

La amplificación de reorganización de receptor inmune es importante para la detección de neoplasmas linfoides. Los neoplasmas de células B son más comunes que los tumores de células T e IgH es el receptor inmune reorganizado más común en neoplasmas de células B. Debido a hipermutación somática, la fiabilidad de amplificar IgH a partir de ADN genómico puede incrementarse mediante la amplificación con múltiples cebadores para cada segmento V, aunque hay un riesgo de amplificación diferencial. En la amplificación a partir de ADN genómico, se usaron los mismos cebadores de V que los que se usaron en la amplificación a partir de ADNc. Cada segmento V se amplifica por 3 cebadores (en 3 regiones distintas del segmento V: A, B, y C) en tres reacciones diferentes (Tablas 5-7, respectivamente) (véase la Fig. 4A).

Tabla 5

Cebadores del segmento V de IgH humana para la reacción A

(Todos los cebadores tienen 14 pb comunes (AGATCGGAAGAGCA) (SEQ ID NO: 165) unidas a su extremo 5')

| Secuencia                     | SEQ ID NO |
|-------------------------------|-----------|
| TCGCAGACCCTCTCACTCACCTGTG     | 62        |
| GATCTCCTGTAAGGGTTCTGGATACAGCT | 61        |
| AAACCCACACAGACCCTCACGCTGAC    | 44        |
| AAACCCACAGAGACCCTCACGCTGAC    | 45        |
| AAACCCACACAGACCCTCACACTGAC    | 46        |
| CCTCAGTGAAGGTCTCCTGCAAGG      | 40        |
| CCTCGGTGAAGGTCTCCTGCAAGG      | 41        |
| CCTCAGTGAAGGTTTCTCCTGCAAGG    | 42        |
| GGGCTACAGTGAAAATCTCCTGCAAGG   | 43        |
| CTTCGGAGACCCTGTCCCTCACCTG     | 56        |
| CTTCGGACACCCTGTCCCTCACCTG     | 57        |
| CTTCACAGACCCTGTCCCTCACCTG     | 58        |
| CTTCGGAGACCCCGTCCCTCACCTG     | 59        |
| CGGGGACCCTGTCCCTCACCTG        | 60        |
| CTGGGGGGTCCCTGAGACTCTCCTG     | 47        |
| CTGGGGGGTCCCTTAGACTCTCCTG     | 48        |
| CAGGGCGGTCCCTGAGACTCTCCTG     | 49        |
| CAGGGCCGTCCCTGAGACTCTCCTG     | 50        |
| CTGGGGGGTCCCTGAAACTCTCCTG     | 51        |
| CTGGCAGGTCCCTGAGACTCTCCTG     | 52        |
| CTGGAGGGTCCCTGAGACTCTCCTG     | 53        |
| CTGGGAGGTCCCTGAGACTCTCCTG     | 54        |
| TGGGGGGGCCCTGAGACTCTCCT       | 55        |

Tabla 6  
Cebadores del segmento V de IgH humana para la reacción B

5 (Todos los cebadores tienen 14 pb comunes (AGATCGGAAGAGCA) (SEQ ID NO: 165) unidas a su extremo 5')

| Secuencia                 | SEQ ID NO |
|---------------------------|-----------|
| TGGATCAGGCAGTCCCCATCGAGAG | 63        |
| GCTGGGTGCGCCAGATGCC       | 64        |
| GTGTGAGCTGGATCCGTCAGCC    | 131       |
| GTGTGGGCTGGATCCGTCAGCC    | 132       |
| GTGCGACAGGCCCTGGACAA      | 67        |
| GGGTGCGACAGGCCACTGGACAA   | 68        |
| GTGCGCCAGGCCCGGACAA       | 69        |
| GGGTGCGACAGGCTCGTGGACAA   | 70        |
| GGGTGCAACAGGCCCTGGAAAA    | 71        |
| GGGTGCGACAGGCTCCTGGAAAA   | 72        |
| GTGCGACAGGCCCGGACAA       | 73        |
| GTGCGACAGGCCCGGACAA       | 74        |
| TCCGCCAGCCCCAGGGAAGG      | 75        |
| TCCGGCAGCCCCAGGGAAGG      | 76        |
| TCCGGCAGCCACCAGGGAAGG     | 77        |
| TCCGCCAGCACCAGGGAAGG      | 78        |
| TCCGGCAGCCCGCGGGAA        | 79        |
| TCCGGCAGCCCGCGGGAA        | 80        |
| TCCGGCAGCCCGCTGGGAAGG     | 81        |
| TCCGCCAGCCCCTAGGGAAGG     | 82        |
| GGTCCGCCAGGCTCCAGGGAA     | 83        |

ES 2 593 614 T3

|                          |     |
|--------------------------|-----|
| GTTCCGCCAGGCTCCAGGGAA    | 84  |
| GGTCCGCCAGGCTTCCGGGAA    | 85  |
| GGTCCGTCAAGCTCCGGGGAA    | 86  |
| GATCCGCCAGGCTCCAGGGAA    | 87  |
| GGTCCGCCAAGCTCCAGGGAA    | 88  |
| GGTCCGCCAGGCTCCAGGCAA    | 89  |
| GGTCCGCCAGGCCCCAGGCAA    | 90  |
| GGTCCGCCAGGCTCCGGGGCAA   | 91  |
| GGGTCCGTCAAGCTCCAGGGAAGG | 92  |
| CTGGGTCCGCCAAGCTACAGGAAA | 93  |
| GGTCCGCCAGCCTCCAGGGAA    | 94  |
| GGTCCGGCAAGCTCCAGGGAA    | 95  |
| GTGCGAGCTGGATCCGTCAGCC   | 133 |

Tabla 7  
Cebadores del segmento V de IgH humana para la reacción C  
(Todos los cebadores tienen 14 pb comunes (AGATCGGAAGAGCA) (SEQ ID NO: 165) unidas a su extremo 5')

5

| Secuencia                    | SEQ ID NO |
|------------------------------|-----------|
| GCAGCCTAAAGGCTGAGGACACTG     | 134       |
| CTCTGTGACTCCCGAGGACACGGCT    | 97        |
| AGTGGAGCAGCCTGAAGGCCTC       | 98        |
| TGACCAACATGGACCCTGTGGACAC    | 99        |
| ACATGGAGCTGAGCAGCCTGAGATC    | 100       |
| ACATGGAGCTGAGCAGGCTGAGATC    | 101       |
| ACATGGAGCTGAGGAGCCTGAGATC    | 102       |
| ACATGGAGCTGAGGAGCCTAAGATCTGA | 103       |
| GAGCTCTGTGACCGCCGCGGAC       | 104       |
| GAGCTCTGTGACCGCCGTGGACA      | 105       |
| GAGCTCTGTGACCGCTGCAGACACG    | 106       |
| GAGCTCTGTGACCGCTGCGGACA      | 107       |
| GAGCTCTGTGACTGCCGCAGACACG    | 108       |
| GAGCTCTGTGACTGCAGCAGACACG    | 109       |
| GAGCTCTGTGACTGCCGCGGACA      | 110       |
| GAGCTCTGTGACCGCGGACGCG       | 111       |
| GGCTCTGTGACCGCCGCGGAC        | 112       |
| GAGCTCTGTGACCGCCGCGAGACA     | 113       |
| GAGCTCTGTGACCGCTGACACGG      | 114       |
| CAAATGAACAGCCTGAGAGCCGAGGACA | 115       |
| CAAATGAACAGCCTGAAAACCGAGGACA | 116       |

|                                 |     |
|---------------------------------|-----|
| CAAATGAACAGTCTGAAAACCGAGGACA    | 117 |
| CAAATGATCAGCCTGAAAACCGAGGACA    | 118 |
| CAAATGAACAGTCTGAGAAGCTGAGGACACC | 119 |
| CAAATGAACAGTCTGAGAGCCGAGGACA    | 120 |
| CAAATGAACAGCCTGAGAGCTGAGGACA    | 121 |
| CAAATGAGCAGCCTGAGAGCTGAGGACA    | 122 |
| CAAATGAACAGCCTGAGAGACGAGGACA    | 123 |
| CAAATGGGCAGCCTGAGAGCTGAGGACA    | 124 |
| CAAATGAACAGCCTGAGAGCCGGGGA      | 125 |
| CAAATGAACAGTCTGAGAGCTGAGGACA    | 126 |
| CAAATGAGCAGTCTGAGAGCTGAGGACA    | 127 |
| GCACGCTAAAGGCTGAGGACACTG        | 135 |

La amplificación de IgH a partir de ADN genómico tiene varias diferencias de su amplificación a partir de ADNc. El segmento C se obtiene unido a la región VDJ a través de corte y empalme y por lo tanto las secuencias del segmento C pueden usarse para la amplificación a partir de ADNc pero no de ADN genómico. El uso del segmento C permite el uso de dos cebadores distintos en la 1ª y 2ª amplificaciones incrementando la especificidad. Para la amplificación a partir de ADN genómico hemos optado por el uso de cebadores que son complementarios a las secuencias J (Tabla 8).

Tabla 8

Cebadores del segmento J de IgH humana\*

## Cebador del segmento J

## SEQ ID NO

|  |     |
|--|-----|
| ACGAGCCTCATGCGTAGANct cacCTGAGGAGACGGTGACC | 136 |
| ACGAGCCTCATGCGTAGANct cacCTGAGGAGACAGTGACC | 137 |
| ACGAGCCTCATGCGTAGANct tacCTGAAGAGACGGTGACC | 138 |
| ACGAGCCTCATGCGTAGANct tacCTGAGGAGACGGTGACC | 139 |

\* Los cebadores del segmento J usados. Las 18 pb en el 5' son secuencias comunes que se unen a la secuencia complementaria al segmento J con el fin de permitir la amplificación de segunda etapa. La posición N significa una posición aleatoria con el fin de obtener diversidad en las agrupaciones secuenciadas. Las secuencias en minúsculas están en el intrón, y las secuencias en mayúsculas en el 3' de la secuencia están en el exón. Las letras en *italica* enfatizan las bases que se diferencian entre los cebadores.

Estos cebadores abarcan los límites exón-intrón, y los cuatro cebadores utilizados amplifican los diferentes segmentos J y los alelos descritos en la base de datos IMGT. Los cebadores de la segunda etapa no tienen ninguna secuencia complementaria a secuencias genómicas.

El uso de los cebadores de J sobre cebadores de región constante complementarios a la región constante de IgG permite la evaluación de las otras clases (IgM, IgD, IgA, e IgE).

**00411** En el caso del ADNc, existe la elección de si usar el cebador de J o los cebadores de la región constante. Pueden usarse varios cebadores de la región constante para amplificar todas las clases y la secuencia de alguna de la región constante antes de entrar en la secuencia de J con el fin de asociar la información sobre el clonotipo y su clase específica. Las lecturas de secuenciación para muchas de las tecnologías de secuenciación son cortas y sería difícil conseguir esto. Una de las plataformas actuales en el mercado (454 Roche) tiene una lectura más larga pero tiene un rendimiento menor que otras plataformas. Al irse desarrollando estas tecnologías adicionalmente esta opción se hace posible. Con las lecturas cortas actuales (<100 pb), nuestro trabajo sobre el ensayo de ADN genómico sugiere que para la amplificación a partir de ADNc puede hacerse usando tanto la estrategia de cebado de J como de C. Podemos implementar la amplificación usando cebadores de J a partir de ADNc. Sin embargo, dado que el segmento exónico de estos cebadores puede ser demasiado corto para la amplificación específica a partir de ADNc, la primera etapa PCR puede hacerse potencialmente usando un conjunto de cebadores de la región constante que engloban todas las clases diferentes (y los cebadores del segmento V en el otro lado como hemos demostrado). Después, puede hacerse la segunda etapa de PCR con los cebadores de J que son lo suficientemente largos como para tener especificidad para la segunda etapa de PCR que está utilizando un molde de baja complejidad. Los productos se secuencian. Como se ha indicado anteriormente, la desventaja comparado con el esquema demostrado para IgG es que las mutaciones somáticas en la secuencia de J pueden inhibir la amplificación. La ventaja es que se evalúan todas las clases diferentes, aunque la información sobre la clase de cada clonotipo no se determina completamente. Potencialmente, se puede hacer amplificación específica de clase IgG, IgM, IgD, IgA, o IgE, y comparar con la imagen global obtenida a partir del uso de todos los cebadores seguido de cebador

de J. Por ejemplo, se puede comparar el perfil de clonotipos obtenido a partir de la amplificación de IgG con el que usa todos los cebadores seguido de cebador de J. La diferencia se debería presumiblemente a las mutaciones somáticas en el cebador de J (lo que puede identificarse fácilmente en la reacción usando el cebador de IgG) y los clonotipos de las demás clases, que pueden cuantificarse.

5 El uso de cebadores de J en el ADNc también permite la comparación directa entre los resultados de ADNc y ADN genómico. Esto proporcionaría información sobre el nivel de expresión a nivel de clonotipo y puede tener de hecho relevancia funcional. La comparación del perfil de clonotipos de ADNc y ADN genómico a partir de la misma muestra de sangre u otra muestra biológica identifica los clonotipos que tienen diferentes frecuencias indicando una expresión infrecuentemente alta o baja por célula. Esta información funcional puede utilizarse para predecir si un clonotipo se correlaciona probablemente con la enfermedad o no. Además, el nivel de expresión por célula de un clonotipo que se correlaciona con una enfermedad puede usarse para determinar la actividad de la enfermedad o la probabilidad de un resultado de enfermedad. Por ejemplo, la obtención del mismo nivel en el ensayo de ADNc para un clonotipo que se correlaciona en dos individuos todavía puede indicar que los pacientes tienen diferentes actividades de la enfermedad si el nivel de expresión de los clonotipos por célula (según se determina por comparación con el perfilado de clonotipos en el ADN genómico) es diferente.

La segunda etapa de PCR es para unir las secuencias necesarias para la amplificación. Los cebadores usados en la segunda etapa se listan en la Tabla 9.

Tabla 9

20 Cebadores comunes\*

| Secuencia del cebador   | Utilización  | SEQ ID NO |
|---|--|-----------|
| AATGATACGGCGACCACCGAGATCT   | Tercera etapa de PCR                                   | 140       |
| CAAGCAGAAGACGGCATACGAGAT  | Tercera etapa de PCR                                   | 37        |
| CAAGCAGAAGACGGCATACGAGATNNNNNNNGTGAC<br>TGGAGTTCAGACGTGTGCTCTTCCGATCT | Segunda etapa de PCR común                             | 141       |
| TGATGGCTCAAACAAGGAGACCT   | Primera etapa de PCR para TCR $\beta$ de ratón         | 142       |
| AATGATACGGCGACCACCGAGATCTGACCTTGGGT<br>GGAGTCACATTTCTCAGATCCT         | Segunda etapa de PCR para TCR $\beta$ de ratón         | 143       |
| AATGATACGGCGACCACCGAGATCTACACTCTTTC<br>CCTACACGAGCCTCATGCGTAGA        | Segunda etapa para IgH humana a partir de ADN genómico | 144       |

\* La tercera etapa es una etapa de amplificación opcional para todos los ensayos (por ejemplo, TCR $\beta$  de ratón e IgH humana). Se hace para asegurar la integridad de las secuencias de extremo que hibridan con los oligonucleótidos unidos a la celda de flujo. El cebador de la segunda etapa común se usa en todos los ensayos (por ejemplo, TCR $\beta$  de ratón e IgH humana). Obsérvese que el uso de N en el cebador de la segunda etapa común es para indicar el hecho de que cada uno de estos cebadores contiene una única etiqueta de 6 pares de bases para permitir identificar posteriormente a las muestras.

25 La amplificación es posible usando los cebadores anteriores y otras secuencias que son sustancialmente similares. Las Figs. 8A-8B muestran ejemplos de dichas amplificaciones, que tuvieron éxito al menos en el intervalo de ADN genómico 50-2.000  $\mu$ g en 20  $\mu$ l de ADN genómico de entrada.

30 El ensayo necesita acomodar un gran rango dinámico de ADN. Las muestras de biopsia pueden no tener una gran cantidad de material, pero dado que el tumor esté probablemente enriquecido en gran medida no hay necesidad de una gran cantidad de material de partida. Por otra parte, un millón de células tendrá ~6  $\mu$ g de ADN genómico. Las PBMC que contienen 1 millón de células B tendrán probablemente ~20  $\mu$ g de ADN genómico. Para poder evaluar 1 millón de células B, se usan ~6,6  $\mu$ g de ADN genómico en cada una de las 3 reacciones de PCR. Debe indicarse que si hay una mutación somática en la secuencia complementaria a uno de los cebadores entonces en este ejemplo sólo se están evaluando ~660K células B. Es útil si el ensayo funciona sobre el intervalo de 50 a 10.000 ng. Se ha demostrado que el ensayo funciona en el intervalo de 50-2.000 ng de ADN en 20  $\mu$ l. Mediante el aumento de escala de la reacción hasta 100  $\mu$ l, pueden usarse 10  $\mu$ g de ADN.

### Ejemplo 23

#### Monitorización de leucemia linfoblástica aguda (ALL)

La enfermedad residual mínima (MRD) es un factor de pronóstico importante para la estratificación de ALL en la infancia. MRD se ensaya típicamente en la médula ósea en pocas semanas después de la terapia de inducción. La detección más sensible de células leucémicas puede permitir la monitorización de recurrencia de cáncer en la sangre.

El perfilado de clonotipos para evaluar el nivel del clonotipo tumoral en la sangre se usa para detectar las células leucémicas de una forma sensible.

50 La calibración se identifica a través de la investigación de una muestra con alta carga leucémica. Las células leucémicas están presentes típicamente a alta frecuencia en las muestras de diagnóstico (sangre o médula ósea). La muestra de diagnóstico se secuencía frecuentemente para varias reorganizaciones.

Si el tumor es de células B, pueden evaluarse la IgH completamente reorganizada, la IgH parcialmente reorganizada D-J; IgK incluyendo las reorganizaciones de Kde.

Las reorganizaciones cruzadas de linaje ocurren frecuentemente, y lo más frecuente es el (V-D) o (D-D) parcialmente reorganizado de TCR $\beta$ . Para las células T, ocurre una reorganización frecuente para TCR $\alpha$  y TCR $\beta$ , y con menor frecuencia TCR $\alpha$ . La secuenciación del repertorio de clonotipos para estas diferentes reorganizaciones identifica el tumor particular la reorganización presente en el tumor. El nivel sanguíneo de la secuencia específica puede monitorizarse.

El ensayo de monitorización puede implicar sólo los tipos de reorganizaciones relevantes. Por ejemplo, si las reorganizaciones tumorales identificadas en la muestra de diagnóstico son IgH e IgK, entonces IgH e IgK se amplificarán y secuenciarán en muestras de sangre posteriores. El ADN de hasta aproximadamente 1 millón de células B de estas muestras puede usarse para amplificar IgH e IgK y pueden obtenerse hasta aproximadamente 1 millón o más de lecturas de secuenciación, lo que proporciona una sensibilidad al ensayo de 1 en un millón de células B es ~ a una sensibilidad de 1 en 10 millones de células sanguíneas blancas. Con esta gran sensibilidad, las células leucémicas se detectan probablemente significativamente antes de una recidiva clara.

La evolución clonal se ha descrito en ALL. Esto puede ocurrir a través del reemplazo de V u otros mecanismos. Para detectar la evolución, se identificarán clonotipos que están relacionados con aquellos presentes en la muestra de diagnóstico. Por ejemplo, se identificarán los clonotipos con la misma unión D-J, pero con un V diferente. La presencia de estos clonotipos relacionados a una frecuencia apreciable en la muestra de diagnóstico incrementa la probabilidad de la relevancia del clonotipo. El seguimiento de más de una reorganización (para este ejemplo, IgH e IgK tumoral) también puede mejorar este problema.

La mera presencia de células leucémicas puede no ser necesariamente suficiente para predecir una recidiva. Se realiza un estudio de descubrimiento con muestras de sangre longitudinales para pacientes con resultado conocido. Se evaluará en estas muestras si la mera presencia de clonotipo leucémico es suficiente para predecir una recidiva algún tiempo después. Además, se evaluará el cambio en la frecuencia de estos clonotipos como un pronosticador de recidiva. Además de la frecuencia de los clonotipos, los marcadores en las células leucémicas pueden ser indicativos de recidiva. Puede realizarse secuenciación antes y después del enriquecimiento de células con el marcador relevante. Por lo tanto, se determina la frecuencia total del clonotipo. Además, se determina la fracción de estas células con el marcador relevante lo que permite una estimación más precisa del riesgo de recidiva.

Según lo anterior, en un aspecto, los métodos de la invención proporcionan ensayos para MRD sobre la base de medidas del perfil de clonotipos. Dichos ensayos incluyen el uso de perfiles de clonotipos para (i) monitorizar la presencia y abundancia de clonotipos específicos de paciente asociados con la enfermedad, incluyendo clones evolucionados, (ii) proporcionar cuentas de linfocitos a partir de las que derivan los clonotipos, y (iii) proporcionar una medida de la clonalidad (esto es, una medida de cómo de "sesgado" es un perfil para uno o unos pocos clonotipos). En un aspecto, se proporciona un ensayo que proporciona valores para (i) y (ii) simultáneamente; y, en otro aspecto, se proporciona un ensayo que proporciona las tres cantidades simultáneamente a partir de una única medida del perfil de clonotipos de un paciente.

#### 40 **Ejemplo 24**

##### **Monitorización del rechazo de trasplante de un órgano sólido**

El rechazo del trasplante de un órgano sólido puede ocurrir a través de dos rutas distintas: presentación directa e indirecta. La ruta directa usa las células presentadoras de antígeno del donante que se transfieren con el trasplante. Los receptores de las células T están reconociendo en este caso el HLA del donante. La ruta indirecta, por otra parte, ocurre algún tiempo después. En este caso, los péptidos del donante se presentan por HLA del receptor a las células T.

Pueden usarse muestras de biopsia del órgano trasplantado para identificar secuencias relevantes para la calibración de los receptores de las células T (o B) relevantes en el rechazo del trasplante. Los clonotipos enriquecidos en una biopsia del órgano trasplantado pueden compararse con la sangre en el momento del rechazo para identificar los clonotipos que son relevantes para el rechazo. El nivel de estos clonotipos en la sangre se monitoriza entonces para predecir el estado del rechazo.

También se realiza la calibración específica de antígeno. Para identificar los clonotipos relevantes para el rechazo por la ruta directa, los linfocitos del donante se irradian y se mezclan con PBMC del receptor. Las PBMC del receptor capaces de reconocer los linfocitos del donante se activan. El aislamiento de estas células activadas se hace por cualquiera de varias técnicas. Por ejemplo, estas células se aíslan gracias a la liberación de citoquinas por tinción intracelular de citoquinas o técnicas de captura de citoquinas. En lugar de (o además de) el aislamiento, se deja que las células se repliquen in vitro. La comparación de las secuencias de las células aisladas (y/o las células replicadas) con las secuencias pre-activación identifica los clonotipos que interactúan con las PBMC del donante. También se realizan los mismos procedimientos de activación de células T sin la adición de antígeno seguido de secuenciación con el fin de sustraer el fondo potencial de activación independiente de antígeno. El nivel sanguíneo de los clonotipos relevantes para el rechazo del trasplante se monitoriza para evaluar la actividad de rechazo de la ruta directa. Se usa entonces una medida de la diversidad global para monitorizar el nivel de rechazo por la ruta directa.

Para identificar los clonotipos relevantes para el rechazo por la ruta indirecta, los antígenos del donante necesitan ser presentados en el contexto del HLA del receptor. Como el HLA del donante es frecuentemente un antígeno importante, el HLA del donante se incuba con células presentadoras de

antígeno del receptor que son capaces de presentar péptidos del HLA del donante en el contexto del HLA del receptor a células T del receptor. De una manera similar a la descrita anteriormente, estas células se aíslan y replican para identificar los clonotipos que están interaccionando con el HLA del donante en la ruta indirecta. Alternativamente, se realizan los mismos procedimientos de la activación de las células T sin la adición de antígeno seguido de secuenciación con el fin de sustraer el fondo potencial de activación independiente de antígeno. Alternativamente, las células del donante y no sólo el HLA se usan como una fuente de antígeno. Las células del donante se preparan de una manera que facilita a las células presentadoras de antígeno del receptor presentar el antígeno del donante en el contexto del HLA del receptor. Esto se hace por varios métodos alternativos incluyendo lisis usando varios ciclos de congelación y descongelación o por sonicación antes de la adición a las células presentadoras de antígeno. Los clonotipos que se activan por estos antígenos en el contexto del HLA del receptor, se identifican entonces por la secuenciación de las células aisladas y/o replicadas como se ha descrito anteriormente. Se realizan alternativamente los mismos procedimientos de la activación de las células T sin la adición de antígeno seguido de secuenciación con el fin de sustraer el fondo potencial de activación independiente de antígeno. Una vez se identifican estos clonotipos, su nivel sanguíneo se monitoriza para evaluar la actividad de rechazo de la ruta indirecta.

**Ejemplo 25**

**Recurrencia del cáncer**

La recurrencia del cáncer se detecta por la detección de la respuesta inmune al tumor. El nivel de clonotipos de células T y B que son relevantes para el tumor se usa para detectar la recurrencia del cáncer. Se detecta el incremento en el nivel sanguíneo de las células T y B relevantes (o la frecuencia del ADNc de los clonotipos relevantes obtenido en la sangre) e indica el reconocimiento por el sistema inmune de la recurrencia del tumor.

La reducción de estos niveles también se detecta e indica el éxito del tumor para escapar del sistema inmune y por lo tanto el desarrollo de recurrencia del cáncer.

Se detecta el cambio de los marcadores celulares en las células que contienen clonotipos relevantes e indican recurrencia del tumor en ausencia de un cambio en la frecuencia de los clonotipos relevantes. Estos últimos casos pueden ser un reflejo del efecto del tumor en las células inmunes para hacerlas ineficaces o anérgicas.

Con el fin de determinar los clonotipos relevantes en un individuo, se usan muestras del tumor original así como algoritmos desarrollados en estudios poblacionales. Alternativamente, se usan antígenos específicos del tumor para definir clonotipos que interaccionan con las células tumorales. Por ejemplo, las células B o T que interaccionan con algunos antígenos específicos de tumor se capturan y secuencian antes y después de este enriquecimiento para determinar los clonotipos específicos que interaccionan con los antígenos específicos.

La experimentación in vitro se realiza usando técnicas como las discutidas anteriormente (por ejemplo, unión de tetrámero, tinción intracelular de citoquinas, o captura de citoquinas) para determinar en una muestra del paciente los clonotipos específicos que interaccionan con los antígenos específicos de tumor particulares. Una vez se definen estos clonotipos, se monitoriza su nivel en otras muestras de sangre. El cambio en el nivel de estos clonotipos indica una recurrencia del tumor.

Los clonotipos relevantes para el cáncer en un individuo se identifican usando un ensayo in vitro para determinar los clonotipos que interaccionan con el tumor. Las células tumorales se lisan usando ciclos repetidos de congelación y descongelación o sonicación. Esta preparación se añade a células presentadoras de antígeno autólogas (o a PBMC autólogas que contienen células presentadoras de antígeno y células T). La mezcla se añade a células T autólogas y los clonotipos que se activan por el antígeno se identifican secuenciando células T aisladas y/o replicadas y comparando con la secuencia del material no enriquecido como se ha descrito anteriormente. Se realiza el mismo procedimiento de la activación de las células T sin la adición de antígeno seguido de secuenciación con el fin de sustraer opcionalmente el fondo potencial de activación independiente de antígeno. Una vez se determinan los clonotipos relevantes, se monitoriza su nivel en la sangre con el fin de evaluar la probabilidad de recurrencia.

Cuando el cáncer es de células que son capaces de presentación de antígeno, las células tumorales pueden no lisarse opcionalmente ya que pueden servir como las células presentadoras de antígeno. El linfoma, un tumor de células B que puede ser capaz de presentar antígenos en algunos casos, sirve como una célula presentadora de antígeno. Las células tumorales se activan opcionalmente in vitro para mejorar su capacidad de presentar antígenos. Estos tumores se mezclan entonces con células T (o PBMC) autólogas. La secuenciación de los clonotipos de células T antes y después del enriquecimiento identifica los clonotipos relevantes para el cáncer. El nivel de estos clonotipos se monitoriza entonces en la sangre para determinar el riesgo de recurrencia.

El nivel del clonotipo en un tipo específico de células, por ejemplo, aquellas con marcadores de superficie específicos, se monitoriza para detectar células de cáncer que pueden escapar del sistema inmune. De esta manera, para dos pacientes con el mismo nivel del clonotipo relevante, dependiendo de los marcadores contenidos en la célula que contiene los clonotipos, un paciente puede tener una mayor probabilidad de recurrencia comparado con el otro. Con el fin de obtener la información, puede hacerse la secuenciación antes y después del enriquecimiento de células por el marcador particular. Por lo tanto, puede medirse el total así como la fracción y número de células de clonotipos con el marcador particular.

**Ejemplo 26**

**Monitorización de infección por hepatitis C**

La infección aguda de hepatitis C está acompañada frecuentemente de una respuesta inmune que es capaz de aclarar la infección en el ~15% de los casos. Se ha mostrado que la capacidad de aclarar la infección está asociada con determinados genotipos de HLA. En la mayoría de los casos, el virus no se aclara y ocurre una infección crónica. Durante esta infección crónica, el virus es capaz de evadir la respuesta inmune lo que es probablemente responsable de mucho del daño hepático resultante. El tratamiento más efectivo para la enfermedad es interferón. Este tratamiento mata al virus al menos parcialmente a través de la activación de la respuesta inmune. La monitorización de la respuesta inmune puede ser por lo tanto útil en diferentes estadios durante el curso de la enfermedad. Durante la fase aguda, la evaluación del grado de reacción inmune puede ser útil para predecir quién es probable de aclarar el virus. Durante la fase crónica, la medida del nivel de la respuesta inmune puede proporcionar una indicación del grado de la inflamación hepática. Finalmente, la evaluación de la reacción inmune durante el tratamiento con interferón puede proporcionar una indicación temprana de si el tratamiento está siendo efectivo. La evaluación de la respuesta inmune puede hacerse por la medida del repertorio de células T y B por secuenciación como se ha descrito anteriormente.

La identificación de los clonotipos relevantes para la hepatitis C en cada individuo se hace por varios métodos. Los antígenos de la hepatitis C se usan como péptidos individuales, una mezcla de péptidos, proteínas, o el virus completo. Las células T y/o células B que interactúan con el antígeno se identifican por evidencia de enriquecimiento de los clonotipos en las células activadas por antígenos comparado con su nivel en el resto de las células. Además, se hace opcionalmente biopsia de hígado para estos pacientes durante el curso del tratamiento. Esto proporciona medios adicionales o alternativos para identificar clonotipos relevantes para la hepatitis C. Los clonotipos significativamente más enriquecidos en el hígado comparado con la sangre probablemente sean relevantes al menos respecto al proceso inflamatorio en el hígado. Por lo tanto, sus niveles en la sangre en puntos tardíos se monitorizan para evaluar la actividad inflamatoria en el hígado. Finalmente, los estudios de descubrimiento en una población de pacientes pueden indicar un conjunto de secuencias o restos que son relevantes para la hepatitis C. En este estudio poblacional se identifican clonotipos relevantes gracias a su correlación con la enfermedad o enriquecimiento en biopsias de hígado y se descubren algoritmos para distinguir estos clonotipos de otros. Algunos de los criterios incluyen frecuencia de clonotipos, rango, similitud de secuencia de múltiples clonotipos, o resto de secuencia así como la presencia de algún marcador celular. La tipificación de HLA se usa como un método de estratificación. Los restos específicos sólo son predictivos en el contexto de tipos de HLA específicos.

Las células infectadas por virus pueden evadir algunas veces el sistema inmune. Por lo tanto, se realiza la monitorización del nivel del clonotipo en un tipo específico de células, por ejemplo, aquellas con marcadores de superficie específicos. Por lo tanto, para dos pacientes con el mismo nivel del clonotipo relevante, dependiendo de los marcadores contenidos en la célula que contiene los clonotipos, un paciente organiza una respuesta más rigurosa que el otro. Por ejemplo, el efecto del tratamiento con interferón puede estar en el cambio cualitativo así como cuantitativo en las células del clonotipo. Por lo tanto, es importante obtener el nivel del clonotipo y definir si tienen marcadores celulares específicos. Con el fin de obtener la información, se hace secuenciación antes y después del enriquecimiento de células por el marcador particular. Por lo tanto, puede medirse el total así como la fracción y número de células de clonotipo con el marcador particular.

**Ejemplo 27****Hipersensibilidad a fármacos**

La identificación de clonotipos que son relevantes para una hipersensibilidad a fármaco específica se hace usando estudios poblacionales. En estos estudios, los clonotipos que se correlacionan con la ADR se identifican y se identifican las características que los distinguen de los demás clonotipos por diferentes criterios como frecuencia, rango, cambio relativo antes y después del tratamiento, similitud de secuencia de múltiples clonotipos, resto de secuencia, así como la presencia de marcador celular. El resto de secuencia puede ser dependiente de HLA en el que diferentes restos se determinan como relevantes para diferentes secuencias de HLA correspondientes.

Otro método para la identificación de clonotipos relevantes para la hipersensibilidad a fármacos es por interacción con antígeno. Los fármacos y/o su o sus metabolitos se usan para capturar células B que interactúan con ellos. De forma similar, el fármaco o sus metabolitos se incuban con células presentadoras de antígeno autólogas óptimamente antes o al mismo tiempo de la adición de células T. Las células T activadas se aíslan o replican usando algunos de los métodos discutidos anteriormente para obtener células enriquecidas en antígeno. Estas células enriquecidas en antígeno se secuencian entonces y los clonotipos enriquecidos en estas células comparado con las células no enriquecidas se identifican como relevantes para la interacción con el fármaco.

También se realizan los mismos procedimientos de activación de células T sin la adición de antígeno seguido de secuenciación con el fin de sustraer el fondo potencial de activación independiente de antígeno del fármaco.

Este ensayo de calibración se hace antes o después de tomar el fármaco. Se usan muestras de sangre antes y después de tomar el fármaco y el incremento in vivo en el nivel del clonotipo como un criterio adicional para definir los clonotipos relevantes. Una vez se identifican estos clonotipos, se monitorizan para predecir la probabilidad de una hipersensibilidad a fármaco. La activación de linfocitos con fármacos se utiliza meramente para definir los clonotipos relevantes y el nivel sanguíneo de estos clonotipos puede



monitorizarse en muestras posteriores por secuenciación generando un método sensible y específico para predecir ADR. Los fármacos que se usan son moléculas pequeñas o biológicos, como anticuerpos. De forma similar, se usa un metabolito del fármaco o una combinación de metabolitos para identificar células T que interactúan con éste como se ha descrito anteriormente. El metabolito se genera por síntesis química o se purifica de una muestra biológica. Por ejemplo, el fármaco se introduce en un organismo y los metabolitos del fármaco se purifican para uso en el ensayo. Los metabolitos también se obtienen por procesamiento del fármaco por las células *in vitro*.

Se monitoriza el nivel del clonotipo en un tipo específico de células, por ejemplo, aquellas con marcadores de superficie específicos. El incremento en el nivel de un clonotipo en células que portan marcadores de activación después de la introducción del fármaco es más indicativo de hipersensibilidad al fármaco que si no está presente el marcador de activación. Con el fin de obtener la información, se hace la secuenciación antes y después del enriquecimiento de células con el marcador particular. Por lo tanto, se mide el total así como la fracción y número de células de clonotipo con el marcador particular.

Una ADR se predice sin tomar el fármaco. Se detecta un alto nivel de clonotipos que interactúan con el fármaco o la presencia de resto de secuencia que probablemente organice una fuerte respuesta al fármaco para predecir la ADR sin administrar el medicamento al paciente.

Se identifican características de respuesta inmune similares de fármacos con ADR relacionadas con inmunidad idiosincrática. Estos fármacos pueden tener clonotipos de alta frecuencia que interactúan con los fármacos o sus metabolitos. Los fármacos que probablemente tengan ADR relacionadas con inmunidad se identifican después de la administración en sólo un pequeño número de pacientes.

La respuesta subclínica de un fármaco que causa ADR se detecta por la determinación de un incremento en los clonotipos específicos del fármaco (y/o su metabolito) después de la administración de la medicación.

Los fármacos que probablemente tengan una ADR relacionada con inmunidad se identifican siguiendo el incremento en los clonotipos correspondientes a ADR después de la administración en un pequeño número de pacientes.

#### **Ejemplo 28**

##### **Métodos para la estratificación del riesgo en enfermedad vascular carotídea**

Se detecta la inflamación implicada en la formación y estabilidad de las placas en un paciente. La respuesta inmune específica para la inflamación vascular se usa para indicar el riesgo de la desestabilización de las placas. Los antígenos específicos relevantes en la reacción inmune en la ICA (incluyendo LDL modificada u oxidada y proteína de choque térmico) se utilizan para identificar los clonotipos específicos relevantes para la reacción inmune en ICA. Usando procedimientos similares a los descritos anteriormente, se identifican los clonotipos de células T o B que interactúan con los antígenos específicos. Los niveles de los clonotipos identificados se monitorizan para evaluar el riesgo de desestabilización de placas de ICA.

Los clonotipos que son relevantes para la desestabilización de placas de ICA también se identifican usando un algoritmo que se genera a partir de un estudio poblacional que identifica características que distinguen los clonotipos relevantes de otros. Estos clonotipos relevantes se identifican en el estudio poblacional gracias a su correlación con la desestabilización de las placas o por su enriquecimiento significativo en la placa de ICA (por ejemplo, como se obtiene de endarterectomía carotídea) comparado con la sangre. Los restos de secuencia pueden ser específicos para genotipos de HLA distintos correspondientes. El algoritmo desarrollado se usa entonces para predecir clonotipos relevantes para la desestabilización de placas de ICA en otros pacientes.

#### **Ejemplo 29**

##### **Análisis del repertorio de TCR en ratones con EAE**

Se trataron 10 ratones de la cepa SJL con el péptido 139-151 junto con adyuvante completo de Freund (CFA) usando un protocolo disponible comercialmente, por ejemplo, Hooke Laboratories (Lawrence, MA). Ocho de estos ratones desarrollaron encefalitis autoinmune experimental (EAE), un modelo de ratón de esclerosis múltiple, y los otros dos no lo hicieron. Además, 2 ratones de la misma cepa se trataron sólo con CFA. Para cada ratón, se obtuvieron puntuaciones de la enfermedad diariamente durante 61 días después de la inyección. El intervalo de puntuación fue de 0 a 5. Se obtuvieron muestras de sangre antes de la inyección y en días específicos posteriormente. Globalmente, se obtuvieron 11 muestras de sangre de cada ratón y los ratones se sacrificaron en el día 62 ó 63 y se obtuvo el bazo, nodos linfáticos, y médula espinal. La sangre y el tejido se almacenaron inmediatamente en reactivo para proteger el ARN sanguíneo animal y ARNlater, respectivamente. Se extrajo el ARN de las muestras de sangre usando el kit de extracción de sangre animal protegido de Qiagen y las muestras de tejido se homogeneizaron mecánicamente, y el ARN se preparó usando el minikit de ARN Qiagen Plus. Se produjo ADNc de cada muestra usando el kit de síntesis de ADNc Vilo (Life technologies). El repertorio de TCR $\beta$  de cada una de las muestras se amplificó usando los cebadores de la Tabla 10.

Tabla 10

Cebadores del segmento V de ratón\*

| Cebadores del segmento V         | SEQ ID NO |
|----------------------------------|-----------|
| CAAAGAGGTCAAATCTCTCCCG           | 145       |
| CTTATGGACAATCAGACTGCCTCA         | 146       |
| GTCATGGAGAAGTCTAAACTGTTAAGG      | 147       |
| GTAACGAAACAGTTCCAAGGCG           | 148       |
| GGTGCCAGTCGTTTATACCTGAAT         | 149       |
| CCCAGCAGATTCTCAGTCCAACAGT        | 150       |
| AGATATCCCTGATGGATACAAGGC         | 151       |
| AGATATCCCTGATGGGTACAAGGC         | 152       |
| AGATGTCCCTGATGGGTACAAGGC         | 153       |
| GATAATTCACAGTTGCCCTCGGAT         | 154       |
| GATGGTGGGGCTTCAAGGATC            | 155       |
| CAAGCTCTATAGATGATTCAGGG          | 156       |
| CTATGATAAGATTTGAACAGGGAAGC       | 157       |
| GATCTACTATTCAATAACTGAAAACGATCTTC | 158       |
| TAGCACTTTCTACTGTGAACTCAGCA       | 159       |
| CTTGATCAAATAGACATGGTCAAGG        | 160       |
| AGAGATTCTCAGCTAAGTGTTCCTCG       | 161       |
| GTTCTTCAGCAAATAGACATGACTG        | 162       |
| AGCGAAGGAGACATCCCTAAAGGAT        | 163       |
| CGAGAGTGGATCACCAAGGACAAG         | 164       |

\* Cada uno de los cebadores tiene 14 pb comunes (AGATCGGAAGAGCA) (SEQ ID NO: 165) unidas a su extremo 5'.

5 Se realizó una segunda PCR en cada una de las muestras usando los cebadores de la Tabla 7. Cada muestra se amplificó con una pareja de cebadores que amplificaron el conjunto completo de amplicones de la primera etapa de PCR pero que también incluyeron una etiqueta de secuencia individual que pudo permitir después identificar a las muestras individuales en una mezcla de datos de secuencia. Se hace una tercera PCR usando una pareja de cebadores para todas las muestras (tabla 7). El propósito de la tercera PCR fue asegurar la integridad de las secuencias terminales. La estructura del producto de PCR se muestra en las Figs 2A-2B y la Fig. 3A. Estas reacciones de PCR se combinaron en conjuntos de 64 de manera que cada muestra en el combinado se indexó por una de las etiquetas únicas incorporadas en la segunda etapa de PCR. Las moléculas de estos combinados se desnaturalizaron y separaron en dos dimensiones en una superficie sólida mediante hibridación con una celda de flujo que contenía oligonucleótidos que hibridan con las secuencias terminales de los productos amplificados. Las moléculas hibridadas se amplificaron por una amplificación en puente para formar agrupaciones en la superficie bidimensional de manera que cada agrupación contenía aproximadamente 1.000 moléculas cada una el resultado de una amplificación de una única molécula del combinado de PCR. Se usó entonces un método para escindir y liberar una de las dos cadenas de cada una de estas moléculas dejando un molde monocatenario. Se hibridó un cebador de secuenciación a las agrupaciones. Se llevaron a cabo rondas iterativas de secuenciación que implicaron: la introducción de 4 nucleótidos marcados fluorescentemente terminados químicamente de cada base de ácido nucleico y una polimerasa y tampones de manera que ocurriría una incorporación de un único nucleótido para la posición activa del producto de extensión que era complementario a cada nucleótido; una etapa de lavado, un escaneo fluorescente de la superficie para medir qué agrupaciones incorporaban qué tinte fluorescente, la introducción de un químico de escisión que liberó las moléculas de terminación de las bases incorporadas junto con el marcador fluorescente para permitir ciclos posteriores; una etapa de lavado. Estas etapas se iteraron ~100 veces para revelar la secuencia de cada agrupación (lectura 1). La cadena sintetizada se retiró por desnaturalización. Se introdujo un segundo cebador y el proceso de secuenciación se reiteró para leer una etiqueta de 6 bases. Esta cadena sintetizada se retiró por desnaturalización. La cadena molde original se dejó hibridar de nuevo con los cebadores inmovilizados en superficie que se extendieron para formar de nuevo las agrupaciones bicatenarias. En este momento, la cadena original se escindió resultando en agrupaciones formadas de los complementos de las agrupaciones monocatenarias originales. Se introdujo un tercer cebador de secuenciación y se hibridó con esta cadena y la secuenciación se iteró ~60 ciclos para obtener secuencias de la cadena inversa de los amplicones (lectura 2). Las secuencias resultantes contribuyeron a ~100.000-200.000 lecturas por muestra después de haberse separado sobre la base de las secuencias de etiqueta.

Las secuencias obtenidas se mapearon en primer lugar a regiones específicas de los segmentos V y J. Específicamente, las primeras 27 pb de la lectura 1 se mapearon a las últimas 27 pb de los diferentes segmentos J de ratón. De forma similar, las secuencias iniciales de la lectura 2 se mapearon a las secuencias de los cebadores usados. Cuando una secuencia se mapea a uno de los cebadores entonces

la secuencia de la lectura 2 después del cebador se mapea a diferentes segmentos V de ratón. Como se espera que la lectura 1 alcance el segmento V, también intentamos mapearla con el fin de determinar el marco y la secuencia de aminoácidos. Las posiciones 81-95 de la lectura 1 se mapearon a las 60 pb más 3' del segmento V específico en el que se mapea la lectura 2. Si una secuencia no tiene identidad sustancial con ninguna de las secuencias J y V sería eliminada. Además, los mapeos de J y V se extienden en la lectura 1. Como los J y V específicos se mapean para cada agrupación se puede evaluar si las bases fuera de la secuencia originalmente mapeada son consistentes con la secuencia de V y J mapeada. Estas secuencias dejarán de ser consistentes cuando se alcancen las bases que resultan de la replicación no de molde (bases N) o segmento D. Las diferencias entre las secuencias leídas y la secuencia V o J en las regiones originalmente mapeadas o extendidas se consideran debidas a error y se "corrigen".

Las lecturas que tienen secuencias idénticas para las posiciones 28-80 se consideran un clonotipo. Los clonotipos que son muy similares en secuencia se evalúan para ver si probablemente son clonotipos independientes o clonotipos que se separaron debido a error de PCR y/o de secuenciación. Hemos ideado un algoritmo que incorpora la frecuencia de los dos clonotipos en consideración, el número de diferencias entre ellos y la probabilidad de error en esa posición para determinar si los dos clonotipos deben coalescer en un clonotipo o no. Cuando uno de los dos clonotipos está a una frecuencia muy alta y el otro es raro y sólo hay una diferencia de una base hacia el final de la lectura (que generalmente tiene más errores) entonces los dos clonotipos probablemente son un resultado de error y se coalescen. Por otra parte, la presencia de dos clonotipos a una frecuencia similar y con tres diferencias entre ellos indica que estos dos clonotipos son genuinamente independientes y no se coalescen. Entonces se computa la frecuencia de cada clonotipo de TCR $\beta$ .

En todos los 12 ratones estaba presente un clonotipo público. Este clonotipo tenía la firma de un clonotipo reactivo para CFA. Fue indetectable en ninguno de los ratones antes de la inyección o en el día 5, pero se incrementó dramáticamente a alta frecuencia a puntos de tiempo posteriores. Múltiples secuencias de nucleótidos en los ratones y en un único ratón codificaban la misma secuencia de aminoácidos de este clonotipo. Globalmente y después de eliminar clonotipos observados en menos de 3/10 puntos de tiempo para un ratón (el punto de tiempo 8 se eliminó de todos los ratones ya había características inusuales en algunos de los ratones), hay 23 observaciones de clonotipos con esta secuencia de aminoácidos (19 de los cuales presentan una frecuencia media  $>10^{-4}$ ) con 10 secuencias de nucleótidos distintas en los 12 ratones (8 secuencias distintas para los 19 clonotipos de alta frecuencia) se obtuvieron (véase la tabla 10). Además, hay clonotipos adicionales relacionados (sólo con una diferencia de un aminoácido) que también tienen el mismo patrón de estar bajos antes de la inyección y altos después de 10 días de la inyección. Esto proporciona una evidencia abrumadora de un clonotipo que responde a CFA.

Tabla 11

Clonotipos públicos de TCR $\beta$  de ratón en respuesta a CFA\*

ES 2 593 614 T3

| ID del ratón | Log10 frecuencia media del clon | Secuencia de nucleótidos                                     | SEQ ID NO | Secuencia de proteína | SEQ ID NO |
|--------------|---------------------------------|--|-----------|-----------------------|-----------|
| 62           | -2,73861                        | CAAATACAGCGTTTCTGCACTA<br>CCCCAAGCTATACTGCTGGCACA<br>GAGAAAA | 171       | FLCASSIAWGSAETLYF     | 193       |
| 62           | -3,78089                        | CAAATACAGCGTTTCTGCACTC<br>CCCAGGCTATACTGCTGGCACA<br>GAGAAAA  | 172       | FLCASSIAWGSAETLYF     | 193       |
| 34           | -3,75307                        | CAAATACAGCGTTTCTGCACTC<br>CCCAGGCTATACTGCTGGCACA<br>GAGAAAA  | 172       | FLCASSIAWGSAETLYF     | 193       |
| 34           | -3,2046                         | CAAATACAGCGTTTCTGCACTA<br>CCCAGGCTATACTGCTGGCACA<br>GAGAAAA  | 173       | FLCASSIAWGSAETLYF     | 193       |
| 64           | -2,95504                        | CAAATACAGCGTTTCTGCACTC<br>CCCCAAGCTATACTGCTGGCACA<br>GAGAAAA | 174       | FLCASSIAWGSAETLYF     | 193       |
| 14           | -3,10634                        | CAAATACAGCGTTTCTGCACTC<br>CCCAGGCTATACTGCTGGCACA<br>GAGAAAA  | 172       | FLCASSIAWGSAETLYF     | 193       |
| 61           | -3,07614                        | CAAATACAGCGTTTCTGCACTA<br>CCCCAAGCTATACTGCTGGCACA<br>GAGAAAA | 171       | FLCASSIAWGSAETLYF     | 193       |
| 61           | -4,62344                        | CAAATACAGCGTTTCTGCACTA<br>CCCAGGCTATACTGCTGGCACA<br>GAGAAAA  | 173       | FLCASSIAWGSAETLYF     | 193       |
| 61           | -4,6644                         | CAAATACAGCGTTTCTGCACTT<br>CCCAGGCTATACTGCTGGCACA<br>GAGAAAA  | 175       | FLCASSIAWGSAETLYF     | 193       |
| 55           | -2,70719                        | CAAATACAGCGTTTCTGCACTC<br>CCCAGGCTATACTGCTGGCACA<br>GAGAAAA  | 172       | FLCASSIAWGSAETLYF     | 193       |
| 55           | -3,05281                        | CAAATACAGCGTTTCTGCACTA<br>CCCAGGCAATACTGCTGGCAC              | 176       | FLCASSIAWGSAETLYF     | 193       |

ES 2 593 614 T3

|    |          |   |     |                    |     |
|----|----------|---|-----|--------------------|-----|
|    |          | AGAGAAAA  |     |                    |     |
| 55 | -3,44137 | CAAAATACAGCGTTTCTGCACTC<br>CCCCAGGCAATACTGCTGGCAC<br>AGAGAAAA | 177 | FLCASSIAWGS AETLYF | 193 |
| 65 | -3,73155 | CAAAATACAGCGTTTCTGCACTG<br>CCCCAAGCTATACTGCTGGCACA<br>GAGAAAA | 178 | FLCASSIAWGS AETLYF | 193 |
| 65 | -3,42266 | CAAAATACAGCGTTTCTGCACTA<br>CCCCAGGCTATACTGCTGGCACA<br>GAGAAAA | 173 | FLCASSIAWGS AETLYF | 193 |
| 35 | -4,35749 | CAAAATACAGCGTTTCTGCACTG<br>CCCCAGGCTATACTGCTGGCACA<br>GAGAAAA | 179 | FLCASSIAWGS AETLYF | 193 |
| 35 | -2,97796 | CAAAATACAGCGTTTCTGCACTC<br>CCCCAGGCTATACTGCTGGCACA<br>GAGAAAA | 172 | FLCASSIAWGS AETLYF | 193 |
| 23 | -3,20311 | CAAAATACAGCGTTTCTGCACTC<br>CCCCAGGCTATACTGCTGGCACA<br>GAGAAAA | 172 | FLCASSIAWGS AETLYF | 193 |
| 23 | -4,70989 | CAAAATACAGCGTTTCTGCACTA<br>CCCCAGGCTATACTGCTGGCACA<br>GAGAAAA | 173 | FLCASSIAWGS AETLYF | 193 |
| 11 | -2,8685  | CAAAATACAGCGTTTCTGCACTC<br>CCCCAGGCTATACTGCTGGCACA<br>GAGAAAA | 172 | FLCASSIAWGS AETLYF | 193 |
| 63 | -3,38278 | CAAAATACAGCGTTTCTGCACTA<br>CCCCAGGCTATACTGCTGGCACA<br>GAGAAAA | 173 | FLCASSIAWGS AETLYF | 193 |
| 63 | -3,21617 | CAAAATACAGCGTTTCTGCACTC<br>CCCCAAGCTATACTGCTGGCACA<br>GAGAAAA | 174 | FLCASSIAWGS AETLYF | 193 |
| 45 | -3,79581 | CAAAATACAGCGTTTCTGCACTC<br>CCCCAGGCTATACTGCTGGCACA<br>GAGAAAA | 172 | FLCASSIAWGS AETLYF | 193 |
| 45 | -3,86157 | CAAAATACAGCGTTTCTGCGCTC<br>CCCCATGCTATACTGCTGGCACA<br>GAGAAAA | 180 | FLCASSIAWGS AETLYF | 193 |

\* El log10 de la frecuencia media describe la frecuencia media entre todos los 10 puntos de tiempo para el ratón específico.

5 El clonotipo público descrito en el ejemplo anterior se validó mirando en repertorios de TCR de ratones adicionales. Específicamente, se evaluó la presencia del clonotipo en 11 ratones tratados con PLP y 3 tratados de forma simulada. El clonotipo estaba presente en 12/14 ratones. Además, el clonotipo relacionado mencionado en el ejemplo anterior (con diferencia de un aminoácido) también se observó repetidamente. Hubo 15 observaciones (diferentes secuencias de nucleótidos en el mismo animal o en diferentes animales) para el clon con diferencia de un aminoácido. El clonotipo no se observó en ninguno de los animales antes de la vacunación. El patrón de frecuencia también fue bastante similar al observado en el primer conjunto de animales. Esto valida que el clonotipo público identificado es de hecho común para la mayoría de los ratones que se trataron con el adyuvante de Freund.

10 El uso de tejido afectado para ayudar a determinar clonotipos relevantes para la enfermedad. Se encontraron los clonotipos que se correlacionan con la actividad de la enfermedad en el modelo EAE anterior. Sin embargo, no es fácilmente obvio si los clonotipos eran específicos para el aspecto del adyuvante de la vacunación o el péptido PLP. La presencia o ausencia de los clonotipos en las muestras tratadas de forma simulada fue un método para distinguir las dos posibilidades. Como se ha descrito

previamente, el tejido afectado puede usarse para determinar los clonotipos relevantes para la enfermedad. En este caso, estaba disponible tejido de la médula espinal. Los clonotipos se filtraron a partir de los datos que tenían una frecuencia que es al menos 3 mayor en la médula espinal que cada una de las demás tres muestras obtenidas al mismo tiempo (bazo, nodo linfático, y sangre). Se examinaron los clonotipos que se correlacionaban con la puntuación de la enfermedad. Se identificó un clonotipo que aparecía más de una vez. La secuencia del clonotipo era LYCTCSALGGSSYEQYF (secuencia A) (SEQ ID NO: 194). Esta secuencia se buscó e identificó en todos los 8 ratones con la enfermedad, dos ratones que se trataron con PLP pero no desarrollaron la enfermedad y no se detectó en los dos ratones tratados de forma simulada. El patrón de frecuencia del clonotipo en las muestras de sangre seriadas fue diferente entre ratones. Sin embargo, en todos los ratones el clonotipo estuvo ausente antes de la vacunación. Además, el patrón de mayor frecuencia en la médula espinal que el otro tejido en el mismo punto de tiempo se demostró en todos excepto un ratón.

Dado el riesgo de sobreajuste, se examinó un conjunto independiente de ratones para validar este descubrimiento. Se examinó la presencia de la secuencia A en ratones adicionales. Específicamente, la presencia de la secuencia A se ensayó en las circunstancias siguientes: en ocho ratones adicionales tratados con PLP que desarrollaron EAE, tres ratones tratados con PLP que no desarrollaron EAE, y tres ratones tratados de forma simulada. Entre los 11 ratones tratados con PLP, el clonotipo estuvo presente en ocho ratones. Dos de tres que no tenían el clonotipo no desarrollaron la enfermedad. El clonotipo no se encontró en los tres ratones tratados de forma simulada ni se encontró en ninguno de los ratones tratados con PLP en el punto anterior a la vacunación. Además, el patrón de mayor frecuencia en la médula espinal que en los otros tejidos del mismo punto de tiempo se demostró en todos los animales. En todas las muestras (el descubrimiento inicial y la validación posterior) entre 21 animales tratados con PLP, 18 tienen los clones, pero ninguno antes de la vacunación. Además, ninguno de los cinco ratones tratados de forma simulada tiene la secuencia A. Esto proporciona una fuerte evidencia para un clonotipo de EAE público que es específico para la enfermedad.

Los clonotipos adicionales que se correlacionan con la enfermedad y que están enriquecidos en la médula espinal parecen tener secuencias muy similares entre sí. Estos clonotipos semi-privados son clonotipos específicos de la enfermedad adicionales. Éstos pueden ser identificados más exhaustivamente mediante la utilización de un algoritmo de búsqueda de resto para determinar secuencias compartidas entre los clonotipos enriquecidos en la médula espinal y que se correlacionan con la enfermedad.

### **Ejemplo 30**

#### **Hipermutaciones somáticas identificadas en el perfilado de clonotipos de IgH**

Se realizaron tres conjuntos de amplificaciones usando los cebadores del segmento V de IgH listados en la tabla 5 y los cebadores complementarios a la secuencia constante de IgG que también se han descrito anteriormente. Se usaron los ADNc de 7 muestras normales, 7 muestras de pacientes con esclerosis múltiple, y 4 muestras con SLE como moldes para la amplificación. Después de una segunda etapa de PCR que introdujo una única etiqueta para cada muestra, los productos se aislaron espacialmente y se cometieron a secuenciación.

Las secuencias se mapearon a segmentos V y J individuales y se ensamblaron en clonotipos usando las metodologías descritas anteriormente. Buscamos evidencia de clonotipos frecuentes altamente relacionados que pueden ser el resultado de hipermutación somática en estas muestras. Los datos en la tabla 11 muestran un ejemplo llamativo en uno de los pacientes con esclerosis múltiple. En este ejemplo, se identificaron 12 clonotipos distintos. Estos 12 clonotipos nucleótidos codifican 3 secuencias de aminoácidos altamente relacionadas. Dos de las secuencias de aminoácidos están a muy alta frecuencia (>1%) y se diferencian en un aminoácido conservado (lisina frente a arginina).

Tabla 12  
Lista de clonotipos relacionados en un paciente con esclerosis múltiple\*

| Frecuencia (%) | Secuencia consenso  | SEQ ID NO | Secuencia de proteína | SEQ ID NO |
|----------------|---|-----------|-----------------------|-----------|
| 0,095834689    | CTGGCCCCAATTCCATCTGCCTG<br>TAAAGCATGTACAGTAATACACA<br>GCCGTGT | 181       | TAVYYCTCFTGRWNWGQ     | 195       |
| 2,098569825    | CTGGCCCCAGTTCCATTTGGTAG<br>TAAACATGTACAGTAATACACA<br>GCCGTGT  | 182       | TAVYYCTCFTTKWNWGQ     | 196       |
| 0,856216488    | CTGGCCCCAATTCCAATTGGTAG<br>TAAACATGTACAGTAATACACA<br>GCCGTGT  | 183       | TAVYYCTCFTTKWNWGQ     | 196       |
| 0,785914399    | CTGGCCCCAATTCCAATTGGTGG<br>TAAACATGTACAATAATAGACA<br>GCCGTGT  | 184       | TAVYYCTCFTTKWNWGQ     | 196       |
| 0,395930174    | CTGGCCCCAGTTCCATTTGGTAG<br>TAAACATGTACAGTAATATACA<br>GCAGTGT  | 185       | TAVYYCTCFTTKWNWGQ     | 196       |
| 0,128712283    | CTGGCCCCAGTTCCATTTGGTAG<br>TAAACATGTACAGTAATATACA<br>GCTGTGT  | 186       | TAVYYCTCFTTKWNWGQ     | 196       |
| 0,094435642    | CTGGCCCCAGTTCCATTTGGTAG<br>TAAACATGTACAGTAATATACA<br>GCGGTGT  | 187       | TAVYYCTCFTTKWNWGQ     | 196       |
| 0,053513531    | CTGGCCCCAATTCCAATTGGTAG<br>TAAACATGTACAGTAATAGAC<br>AGCAGTGT  | 188       | TAVYYCTCFTTKWNWGQ     | 196       |
| 0,019936413    | CTGGCCCCAGTTCCATTTGGTAG<br>TAAACATGTACAGTAATATACG<br>GCCGTGT  | 189       | TAVYYCTCFTTKWNWGQ     | 196       |
| 0,014340227    | CTGGCCCCAGTTCCATTTGGTAG<br>TAAACATGTACAGTAATATACA<br>GCCGTTG  | 190       | TAVYYCTCFTTKWNWGQ     | 196       |
| 1,305310431    | CTGGCCCCAGTTCCATCTCGTAG<br>TAAACATGTACAATAATACACA<br>GCCGTGT  | 191       | TAVYYCTCFTRWNWGQ      | 197       |
| 0,199364133    | CTGGCCCCAGTTCCATCTGGTAG<br>TAAACACGTACAATAATACACA<br>GCCGTGT  | 192       | TAVYYCTCFTRWNWGQ      | 197       |

5 \* La secuencia de las posiciones 28-80 se muestra con la secuencia de aminoácidos. La columna de la frecuencia indica la frecuencia del clonotipo en la muestra. Por ejemplo, los segundos clonotipos representan más del 2% de todas las secuencias en esta muestra.

**Ejemplo 31**

10 **Uso forense del perfilado inmune**

Los perfiles de clonotipos para receptores de células T y/o B pueden emplearse para la identificación humana y animal. La tremenda diversidad de estos perfiles de clonotipos proporciona una firma extremadamente única del individuo. Esto se ejemplifica por los perfiles de la Fig. 10. En este ejemplo, las secuencias de TCRβ se amplificaron por PCR con transcriptasa inversa a partir de ARNm extraído de la

5 sangre de dos individuos diferentes. Los cebadores complementarios a las regiones V se listan en la  
 10 Tabla 13. Los cebadores de la segunda etapa de amplificación son los mismos que los del Ejemplo 2. Los  
 productos se secuenciaron y se determinó la frecuencia de la frecuencia de cada clonotipo. Como puede  
 observarse en la Fig. 10, la gran mayoría de clonotipos de los individuos era diferente, los clonotipos de  
 un individuo (muestra 122) posicionados casi exclusivamente a lo largo del eje de las X y los clonotipos  
 del otro individuo (muestra 140) posicionados casi exclusivamente a lo largo del eje de las Y. Sólo  
 aproximadamente 25 clonotipos parecían compartirse, como se indica por los puntos de datos fuera del  
 eje. Esta disparidad de uso de clonotipos entre individuos ha sido corroborada por Warren et al, Genome  
 Research, Epub (24 febrero 2011)). En un aspecto de la invención, esta disparidad en el uso de clonotipos  
 proporciona un método para distinguir individuos que comprende las etapas: (a) obtener un perfil de  
 clonotipos de una primera muestra, (b) obtener un perfil de clonotipos de una segunda muestra, y (c)  
 determinar si la primera muestra y la segunda muestra son del mismo individuo midiendo el grado en el  
 que se superpone el uso de clonotipos.

Tabla 13

15 Cebadores del segmento V de TCR $\beta$  humano

| Cebador del segmento V  | SEQ ID NO |
|-------------------------|-----------|
| AACTATGTTTTGGTATCGTCAGT | 2         |
| TTCTGGTACCGTCAGCAAC     | 3         |
| AGTGTATCCTGGTACCAACAG   | 4         |
| AGTGTGTA CTGGTACCAACAG  | 5         |
| ACTGTGTCCTGGTACCAACAG   | 6         |
| AGTGTGTCCTGGTACCAACAG   | 7         |
| TCTGTGTA CTGGTACCAACAG  | 8         |
| CCCTTTACTGGTACCGACAG    | 9         |
| GCCTTTATTGGTACCGACAG    | 166       |
| CCCTTTACTGGTACCGACAAA   | 11        |
| CCCTTTATTGGTACCGACAG    | 167       |
| TTTTGGTACCAACAGGTCC     | 12        |
| TTTTGGTACCAACAGGCC      | 13        |
| AACCCTTTATTGGTATCAACAG  | 14        |
| CGCTATGTATTGGTACAAGCA   | 15        |
| GGCAATGTATTGGTACAAGCA   | 16        |



|                             |     |
|-----------------------------|-----|
| GGCTATGTATTGGTACAAGCA       | 168 |
| TTTCTGGTACAGACAGACCATGA     | 17  |
| TACTATGTATTGGTATAAACAGGACTC | 18  |
| CAAAATGTACTGGTATCAACAA      | 19  |
| ATGTTCTGGTATCGACAAGACC      | 20  |
| ATGTAAGTGGTATCGACAAGACC     | 21  |
| TGCCATGTACTGGTATAGACAAG     | 22  |
| GTATCGACAAGACCCAGGCA        | 169 |
| ATGTCCTGGTATCGACAAGACC      | 170 |
| TAATCTTTATTGGTATCGACGTGT    | 27  |
| GCCATGTACTGGTACCGACA        | 28  |
| TCATGTTTACTGGTATCGGCAG      | 29  |
| CAACCTATACTGGTACCGACA       | 30  |
| CATGCTACCCTTTACTGGTACC      | 31  |
| CACAATACCCTTTACTGGTACC      | 32  |
| ATACTTCTATTGGTACAGACAAATCT  | 33  |
| CACTGTCTACTGGTACCAGCA       | 34  |
| CGTCATGTACTGGTACCAGCA       | 35  |

\* Todos los cebadores tienen 14 pb comunes (AGATCGGAAGAGCA) (SEQ ID NO: 165) unidas a su extremo 5'.

De los 342 clonotipos en la sangre de la primera muestra (muestra 122) a una frecuencia  $>10^{-4}$  sólo uno se detecta en el segundo individuo (muestra 140). Recíprocamente, de 505 clonotipos presentes a frecuencia  $>10^{-4}$  en la muestra 144, 3 se detectan en la muestra 122. Como un control de fluctuación aleatoria de la medida, de los mismos 505 clonotipos. 504 clonotipos estaban presentes en una muestra de amplificación replicada. Esto demuestra el potencial de que un perfil de clonotipos completo sea potencialmente un identificador extremadamente específico. Por supuesto, la naturaleza de estos perfiles es que no son estables con el tiempo ya que nuevas reacciones inmunes añadirán nuevos clonotipos al espectro. Estos procesos, sin embargo, no alteran estos perfiles muy rápidamente. Puede observarse que aunque la frecuencia precisa de clonotipos individuales se desplaza con el tiempo, el conjunto de clonotipos presente a una frecuencia mensurable es probablemente mucho mayor del que se encontraría en un segundo individuo. Pueden desarrollarse algoritmos para definir la fracción y número de clonotipos idénticos que serían necesarios para determinar que dos muestras son del mismo individuo. Además, como esta diversidad se encuentra en un gen con un gran grado de función activa a diferencia de diversidad de microsatélite, es posible extraer información de identificación potencialmente relevante sin la necesidad de tener una muestra que concuerde de una muestra de donante identificada. La información sobre la salud de este individuo, su historial de vacunación, etc. es mensurable a partir de la información de los clonotipos.

#### Definiciones

A no ser que se defina específicamente otra cosa en la presente memoria, los términos y símbolos de química de ácidos nucleicos, bioquímica, genética, y biología molecular usados en la presente memoria siguen aquellos de los tratados y textos estándar en el campo, por ejemplo, Kornberg y Baker, DNA Replication, Segunda Edición (W.H. Freeman, Nueva York, 1992); Lehninger, Biochemistry, Segunda Edición (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).

"Amplicón" significa el producto de una reacción de amplificación de polinucleótidos; esto es, una población clonal de polinucleótidos, que pueden ser monocatenarios o bicatenarios, que se replican a partir de una o más secuencias de partida. La una o más secuencias de partida pueden ser una o más copias de la misma secuencia, o pueden ser una mezcla de diferentes secuencias. Preferiblemente, los amplicones se forman por la amplificación de una única secuencia de partida. Los amplicones pueden producirse por una variedad de reacciones de amplificación cuyos productos comprenden replicados del uno o más ácidos nucleicos de partida, o diana. En un aspecto de los métodos de la invención, las reacciones de amplificación que producen amplicones están "dirigidas por molde" ya que el emparejamiento de bases de los reactantes, bien nucleótidos u oligonucleótidos, tienen complementos en un polinucleótido molde que se requiere para la creación de los productos de la reacción. En un aspecto de los métodos de la invención, las reacciones dirigidas por molde son extensiones de cebador con una

polimerasa de ácido nucleico o ligaciones de oligonucleótidos con una ligasa de ácidos nucleicos. Dichas reacciones incluyen, pero no están limitadas a, reacciones en cadena de la polimerasa (PCR), reacciones de polimerasa lineal, amplificación basada en la secuencia de ácidos nucleicos (NASBA), amplificaciones por círculo rodante, y semejantes, descritas en las referencias siguientes: Mullis et al, patentes U.S. 4.683.195; 4.965.188; 4.683.202; 4.800.159 (PCR); Gelfand et al, patente U.S. 5.210.015 (PCR en tiempo real con sondas "taqman"); Wittwer et al, patente U.S. 6.174.670; Kacian et al, patente U.S. 5.399.491 ("NASBA"); Lizardi, patente U.S. 5.854.033; Aono et al, publicación de patente japonesa JP 4-262799 (amplificación por círculo rodante); y semejantes. En un aspecto de los métodos de la invención, los amplicones se producen por PCR. Una reacción de amplificación puede ser una amplificación en "tiempo real" si está disponible una química de detección que permite medir el producto de la reacción al progresar la reacción de amplificación, por ejemplo, "PCR en tiempo real" descrita más adelante, o "NASBA en tiempo real" como se describe en Leone et al, *Nucleic Acids Research*, 26: 2150-2155 (1998), y referencias semejantes. Tal y como se usa en la presente memoria, el término "amplificar" significa realizar una reacción de amplificación. Una "mezcla de reacción" significa una disolución que contiene todos los reactantes necesarios para realizar una reacción, que puede incluir, pero no está limitado a, agentes tamponadores para mantener el pH a un nivel seleccionado durante una reacción, sales, co-factores, secuestradores, y semejantes.

"Clonalidad" tal y como se usa en la presente memoria significa una medida del grado con el que la distribución de abundancias de clonotipos entre clonotipos de un repertorio está sesgada para un único o unos pocos clonotipos. Aproximadamente, la clonalidad es una medida inversa de la diversidad de los clonotipos. Muchas medidas o estadísticas están disponibles de ecología que describen relaciones especies-abundancia que pueden usarse para las medidas de clonalidad según la invención, por ejemplo, Capítulos 17 y 18, en Pielou, *An Introduction to Mathematical Ecology*, (Wiley-Interscience, 1969). En un aspecto, una medida de clonalidad usada con los métodos de la invención es una función de un perfil de clonotipos (esto es, el número de clonotipos distintos detectados y sus abundancias), de manera que después de medir un perfil de clonotipos, la clonalidad puede computarse a partir de éste para proporcionar un único número. Una medida de clonalidad es la medida de Simpson, que es simplemente la probabilidad de que dos clonotipos cogidos aleatoriamente sean el mismo. Otras medidas de clonalidad incluyen medidas basadas en información e índice de diversidad de McIntosh, descrito en Pielou (citado anteriormente).

"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 parte de éstos. En un aspecto de los métodos de la invención, una colección de todos los distintos clonotipos de una población de linfocitos de un individuo es un repertorio de dicha población, por ejemplo, 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 semejantes. Tal y como se usa en la presente memoria, "perfil de clonotipos", o "perfil de repertorio", es una tabulación de los clonotipos de una muestra de células T y/o células B (tal como una muestra de sangre periférica que contiene dichas células) que incluye sustancialmente todos los clonotipos del repertorio y sus abundancias relativas. "Perfil de clonotipos", "perfil de repertorio", y "repertorio" se usan en la presente memoria indistintamente. (Esto es, el término "repertorio", tal y como se discute más completamente más adelante, significa un repertorio medido de una muestra de linfocitos). En un aspecto de los métodos de la invención, los clonotipos comprenden partes de una cadena pesada de inmunoglobulina (IgH) y una cadena  $\beta$  de TCR. En otros aspectos de los métodos de la invención, los clonotipos pueden estar basados en otras moléculas recombinadas, tales como cadenas ligeras de inmunoglobulina o cadenas  $\alpha$  de TCR, o partes de éstas.

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

"Estándar interno" significa una secuencia de ácido nucleico que se amplifica en la misma reacción de amplificación que uno o más polinucleótidos diana con el fin de permitir la cuantificación absoluta o relativa de los polinucleótidos diana en una muestra. Un estándar interno puede ser endógeno o exógeno. Esto es, un estándar interno puede aparecer de forma natural en la muestra, o puede añadirse a la muestra antes de la amplificación. Pueden añadirse múltiples secuencias de estándar interno exógenas a una mezcla de reacción en una serie de concentraciones predeterminadas para proporcionar una calibración con la que puede compararse un amplicón diana para determinar la cantidad de su polinucleótidos diana correspondiente en una muestra. La selección del número, secuencias, longitudes, y otras características de los estándares internos exógenos es una elección de diseño rutinaria para un experto en la técnica. Preferiblemente, los estándares internos endógenos, también referidos en la presente memoria como "secuencias de referencia" son secuencias naturales en la muestra que corresponden a genes mínimamente regulados que presentan un nivel de transcripción constante e independiente del ciclo celular, por ejemplo, Selvey et al, *Mol. Cell Probes*, 15: 307-311 (2011). Las secuencias de referencia ejemplares incluyen, pero no están limitadas a, secuencias de los genes siguientes: GAPDH,  $\beta_2$ -microglobulina, ARN ribosomal 18S, y  $\beta$ -actina (aunque véase Selvey et al, citado anteriormente).

"Kit" se refiere a cualquier sistema de administración para administrar materiales o reactivos para llevar a cabo un método de la invención. En el contexto de los ensayos de reacción, dichos sistemas de administración incluyen sistemas que permiten el almacenamiento, transporte, o administración de los reactivos de reacción (por ejemplo, cebadores, enzimas, etc., en los contenedores apropiados) y/o materiales de soporte (por ejemplo, tampones, instrucciones escritas para realizar el ensayo etc.) de una localización a otra. Por ejemplo, los kits incluyen uno o más recipientes (por ejemplo, cajas) que contienen los reactivos de reacción relevantes y/o materiales de soporte. Dichos contenidos pueden administrarse al receptor pretendido conjuntamente o separadamente. Por ejemplo, un primer contenedor puede contener una enzima para uso en un ensayo, mientras un segundo contenedor contiene cebadores.

"Neoplasma linfoide" significa una proliferación anormal de linfocitos que puede ser maligna o no maligna. Un cáncer linfoide es un neoplasma linfoide maligno. Los neoplasmas linfoides son el resultado de, o están asociados con, trastornos linfoproliferativos, incluyendo pero no limitado a, linfoma folicular, leucemia linfocítica crónica (CLL), leucemia linfocítica aguda (ALL), leucemia de células pilosas, linfomas, mieloma múltiple, trastorno linfoproliferativo post-trasplante, linfoma de células del manto (MCL), linfoma de células grandes difuso (DLBCL), linfoma de células T, o semejantes, por ejemplo, Jaffe et al, Blood, 112: 4384-4399 (2008); Swerdlow et al, Clasificación de tumores de tejidos hematológicos y linfoides de la OMS (e. 4<sup>a</sup>) (IARC Press, 2008).

"Enfermedad residual mínima" significa células cancerosas remanente después del tratamiento. El término se usa lo más frecuentemente en conexión con el tratamiento de linfomas y leucemias.

"Porcentaje de homología", porcentaje de identidad", o términos semejantes usados en referencia a la comparación de una secuencia de referencia y otra secuencia ("secuencia de comparación") significan que en un alineamiento óptimo entre las dos secuencias, la secuencia de comparación es idéntica a la secuencia de referencia en un número de posiciones de subunidades equivalente al porcentaje indicado, siendo las subunidades nucleótidos para comparaciones de polinucleótidos o aminoácidos para las comparaciones de polipéptidos. Tal y como se usa en la presente memoria, un "alineamiento óptimo" de secuencias que se están comparando es uno que maximiza las concordancias entre las subunidades y minimiza el número de huecos empleado para construir un alineamiento. El porcentaje de identidad puede determinarse con implementaciones de algoritmos disponibles comercialmente, tales como el descrito por Needleman y Wunsch, J. Mol. Biol., 48: 443-453 (1970) (programa "GAP" de Wisconsin Sequence Analysis Package, Genetics Computer Group, Madison, WI), o semejantes. Otros paquetes de software en la técnica para construir alineamientos 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) (Wisconsin Sequence Analysis Package, 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 puede delecionarse o sustituirse con otro nucleótido, o un número de nucleótidos de hasta el cinco por ciento del número total de nucleótidos en la secuencia de referencia puede insertarse en la secuencia de referencia.

"Reacción en cadena de la polimerasa" o "PCR", significa una reacción para la amplificación in vitro de secuencias de ADN específicas por la extensión simultánea por cebador de cadenas complementarias de ADN. En otras palabras, la PCR es una reacción para preparar múltiples copias o réplicas de un ácido nucleico diana flanqueado por sitios de unión de cebador, comprendiendo dicha reacción una o más repeticiones de las etapas siguientes: (i) desnaturalizar el ácido nucleico diana, (ii) hibridar cebadores a los sitios de unión de cebador, y (iii) extender los cebadores por una polimerasa de ácido nucleico en presencia de nucleósidos trifosfato. Habitualmente, la reacción se cicla a través de diferentes temperaturas optimizadas para cada etapa en un instrumento termociclador. Las temperaturas particulares, duraciones de cada etapa, y velocidades de cambio entre etapas dependen de muchos factores muy conocidos para los expertos en la técnica, por ejemplo, ejemplificado por las referencias: McPherson et al, editores, PCR: A Practical Approach y PCR2: A Practical Approach (IRL, Press, Oxford, 1991 y 1995, respectivamente). Por ejemplo, en una PCR convencional usando ADN polimerasa Taq, un ácido nucleico bicatenario diana puede desnaturalizarse a una temperatura de >90°C, los cebadores se hibridan a una temperatura en el intervalo 50-75°C, y los cebadores se extienden a una temperatura en el intervalo 72-78°C. El término "PCR" engloba formas derivadas de la reacción, incluyendo pero no limitado a, RT-PCR, PCR en tiempo real, PCR anidada, PCR cuantitativa, PCR múltiple, y semejantes. Los volúmenes de la reacción varían de unos pocos cientos de nanolitros, por ejemplo, 200 nL, a unos pocos cientos de µL, por ejemplo, 200 µL. "PCR por transcripción inversa", o "RT-PCR", significa una PCR que está precedida por una reacción de transcripción inversa que convierte un ARN diana en un ADN monocatenario complementario, que se amplifica, por ejemplo, Tecott et al, patente U.S. 5.168.038. "PCR en tiempo real" significa una PCR para la que la cantidad de producto de la reacción, es decir, amplicón, se monitoriza al proceder la reacción. Existen muchas formas de PCR en tiempo real que se diferencian principalmente en las químicas de detección usadas para monitorizar el producto de la reacción, por ejemplo, Gelfand et al, patente U.S. 5.210.015 ("taqman"); Wittwer et al, patentes U.S. 6.174.670 y 6.569.627 (tintes intercalantes); Tyagi et al, patente U.S. 5.925.517 (balizas moleculares). Las químicas de detección para PCR en tiempo real se revisan por MacKay et al, Nucleic Acids Research, 30: 1292-1305 (2002). "PCR anidada" significa una PCR en dos etapas en la que el amplicón de una primera PCR se vuelve la muestra para una segunda PCR usando un nuevo conjunto de cebadores, al menos uno de los cuales se une a una localización interior del primer amplicón. Tal y como se usa en la presente memoria,

"cebadores iniciales" en referencia a una reacción de amplificación anidada significa los cebadores usados para generar un primer amplicón, y "cebadores secundarios" significa el uno o más cebadores usados para generar un segundo, o anidado, amplicón. "PCR múltiple" significa una PCR en la que múltiples secuencias dianas (o una única secuencia diana 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). Habitualmente, se emplean distintos conjuntos de cebadores para cada secuencia que se está amplificando. Típicamente, el número de secuencias diana en una PCR múltiple 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 diana específicas en una muestra o espécimen. La PCR cuantitativa incluye tanto la cuantificación absoluta como la cuantificación relativa de dichas secuencias diana. Las medidas cuantitativas se hacen usando una o más secuencias de referencia o estándares internos que pueden ensayarse separadamente o junto con una secuencia diana. La secuencia de referencia puede ser endógena o exógena para la muestra o espécimen, y en el último caso, puede comprender uno o más moldes competidores. Las secuencias de referencia endógenas típicas incluyen segmentos de los transcritos de los genes siguientes:  $\beta$ -actina, GAPDH,  $\beta_2$ -microglobulina, ARN ribosomal, y semejantes. Las técnicas para la PCR cuantitativa son muy conocidas para los expertos en la técnica, como se ejemplifica en las referencias siguientes: 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 semejantes.

"Cebador" significa un oligonucleótido, bien natural o sintético, que es capaz, después de formar un dúplex con un molde de polinucleótido, de actuar como un punto de inicio de la síntesis de ácidos nucleicos y extenderse desde su extremo 3' a lo largo del molde de manera que se forma un dúplex extendido. La extensión de un cebador se lleva a cabo habitualmente con una polimerasa de ácido nucleico, tal como una polimerasa de ADN o ARN. La secuencia de nucleótidos añadida en el proceso de extensión está determinada por la secuencia del polinucleótido molde. Habitualmente, los cebadores se extienden por una polimerasa de ADN. Los cebadores tienen habitualmente una longitud en el intervalo de de 14 a 40 nucleótidos, o en el intervalo de de 18 a 36 nucleótidos. Los cebadores se emplean en una variedad de reacciones de amplificación de ácidos nucleicos, por ejemplo, reacciones de amplificación lineales usando un único cebador, o reacciones en cadena de la polimerasa, empleando dos o más cebadores. La guía para seleccionar las longitudes y secuencias de los cebadores para aplicaciones particulares es muy conocida para los expertos en la técnica, como se evidencia por las referencias siguientes: Dieffenbach, editor, *PCR Primer: A Laboratory Manual*, 2ª Edición (Cold Spring Harbor Press, Nueva York, 2003).

"Puntuación de calidad" significa una medida de la probabilidad de que una asignación de base en una localización de la secuencia particular sea correcta. Una variedad de métodos son muy conocidos para los expertos en la técnica para calcular puntuaciones de calidad para circunstancias particulares, tales como, para bases llamadas como un resultado de diferentes químicas de secuenciación, sistemas de detección, algoritmos de llamada de bases, etc. Generalmente, los valores de puntuación de calidad están relacionados monotónicamente con probabilidades de llamada de bases correctas. Por ejemplo, una puntuación de calidad, o Q, de 10 puede significar que hay un 90 por ciento de posibilidad de que una base se llame correctamente, un Q de 20 puede significar que hay un 99 por ciento de posibilidad de que una base se llame correctamente, etc. Para algunas plataformas de secuenciación, particularmente aquellas que usan químicas de secuenciación por síntesis, las puntuaciones de calidad medias disminuyen como una función de la longitud de la lectura de la secuencia, de manera que las puntuaciones de calidad al principio de una lectura de secuencia son mayores que aquellas al final de una lectura de secuencia, debiéndose estas disminuciones fenómenos tales como extensiones incompletas, extensiones que se llevan adelante, pérdida de molde, pérdida de polimerasa, fallos en la terminación de la cadena, fallos en desprotección, y semejantes.

"Repertorio", o "repertorio inmune", significa un conjunto de secuencias de nucleótidos recombinadas distintas que codifican receptores de células T (TCR) o receptores de células B (BCR), o fragmentos de éstos, respectivamente, en una población de linfocitos de un individuo, en el que las secuencias de nucleótidos del conjunto tienen una correspondencia de una a una con distintos linfocitos o sus subpoblaciones clonales para sustancialmente todos los linfocitos de la población. Una población de linfocitos de la que se determina un repertorio puede tomarse de una o más muestras de tejido, tal como una o más muestras de sangre. Una secuencia de nucleótidos miembro de un repertorio se refiere en la presente memoria como un "clonotipo". En un aspecto de los métodos de la invención, los clonotipos de un repertorio comprenden cualquier segmento de ácido nucleico común a una población de células T o células B que ha experimentado recombinación somática durante el desarrollo de los TCR o BCR, incluyendo moléculas precursoras de éstos normales o aberrantes (por ejemplo, asociadas con cánceres), incluyendo, pero no limitado a, cualquiera de los siguientes: una cadena pesada de inmunoglobulina (IgH) o subconjuntos de ésta (por ejemplo, una región variable de IgH, región CDR3, o semejantes), moléculas de IgH incompletas, una cadena ligera de inmunoglobulina o subconjuntos de ésta (por ejemplo, una región variable, región CDR, o semejantes), cadena  $\alpha$  del receptor de células T o subconjuntos de ésta, cadena  $\beta$  del receptor de células T o subconjuntos de ésta (por ejemplo, región variable, CDR3, región V(D)J, o semejantes), una CDR (incluyendo CDR1, CDR2 o CDR3, de bien TCR o BCR, o combinaciones de dichas CDR), regiones V(D)J de bien TCR o BCR, regiones hipermutadas de las regiones variables de

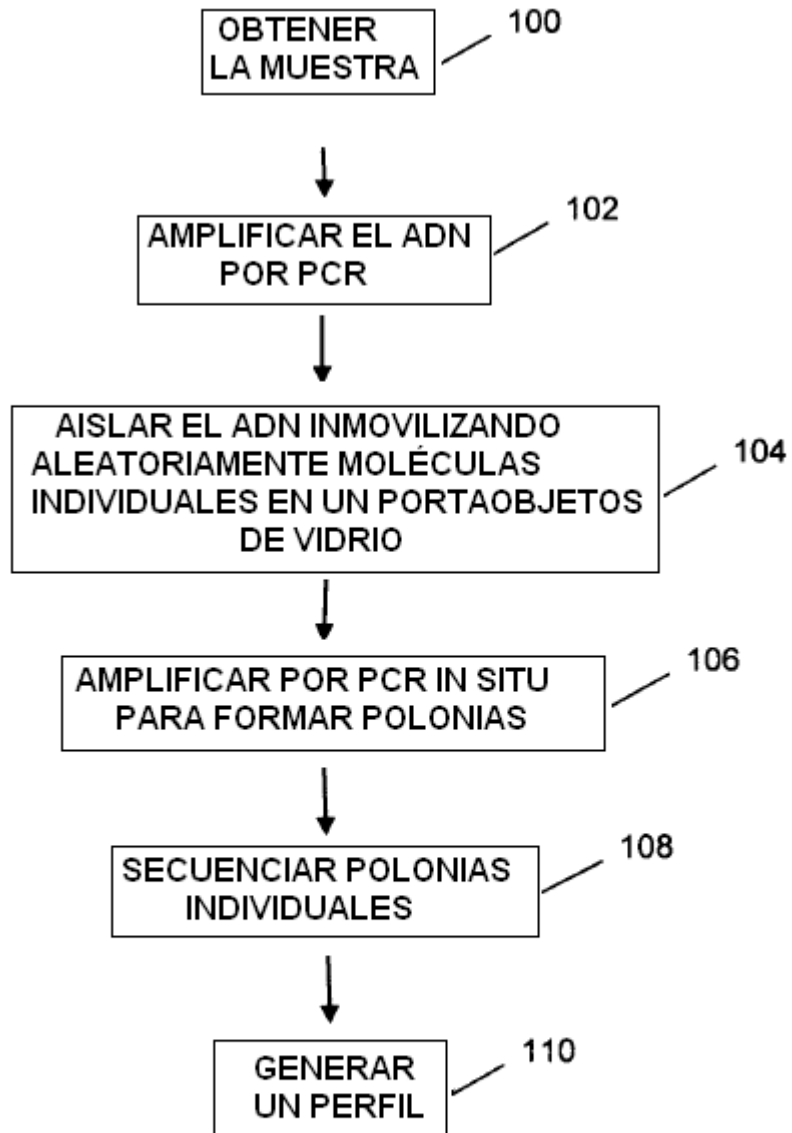
IgH, o semejantes. En un aspecto de los métodos de la invención, los segmentos de ácido nucleico que definen los clonotipos de un repertorio se seleccionan de manera que su diversidad (es decir, el número de secuencias de ácido nucleico distintas en el conjunto) es lo suficientemente grande como para que sustancialmente cada célula T o célula B o clon de éstas en un individuo porte una secuencia de ácido nucleico única de dicho repertorio. Esto es, según la invención, un médico puede seleccionar para definir los clonotipos un segmento o región particular de ácidos nucleicos recombinados que codifique TCR o BCR que no refleja la diversidad completa de una población de células T o células B; sin embargo, preferiblemente, los clonotipos se definen de manera que reflejan la diversidad de la población de células T y/o células B de la que derivan. Esto es, preferiblemente, cada clon diferente de una muestra tiene un clonotipo diferente. (Por supuesto, en algunas aplicaciones, habrá múltiples copias de uno o más clonotipos particulares en un perfil, tal como en el caso de muestras de pacientes con leucemia o linfoma). En otros aspectos de los métodos de la invención, la población de linfocitos correspondiente a un repertorio puede ser células B circulantes, o puede ser células T circulantes, o puede ser subpoblaciones de cualquiera de las poblaciones anteriores, incluyendo pero no limitado a, células T CD4+, o células T CD8+, u otras subpoblaciones definidas por marcadores de la superficie celular, o semejantes. Dichas subpoblaciones pueden adquirirse tomando muestras de tejidos particulares, por ejemplo, médula ósea, o nodos linfáticos, o semejantes, o separando o enriqueciendo células de una muestra (tal como sangre periférica) sobre la base de uno o más marcadores de la superficie celular, tamaño, morfología, o semejantes. En otros aspectos más de los métodos de la invención, la población de linfocitos correspondiente a un repertorio puede derivar de tejidos enfermos, tal como un tejido de tumor, un tejido infectado, o semejantes. En una realización, un repertorio que comprende cadenas  $\beta$  de TCR humano o fragmentos de éstas comprende un número de 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 repertorio que comprende cadenas de IgH humanas o fragmentos de éstas comprende un número de 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 repertorio comprende un conjunto de secuencias de nucleótido que codifican sustancialmente todos los segmentos de la región V(D)J de una cadena IgH. "Sustancialmente todo" tal y como se usa en la presente memoria puede significar cada segmento que tiene una abundancia relativa de 0,001 por ciento o mayor; o alternativamente, "sustancialmente todo" tal y como se usa en la presente memoria significa cada segmento que tiene una abundancia relativa de 0,0001 por ciento o mayor. En otra realización particular, un repertorio comprende un conjunto de secuencias de nucleótido que codifican sustancialmente todos los segmentos de la región V(D)J de una cadena  $\beta$  de TCR. En otra realización, un repertorio comprende un conjunto de secuencias de nucleótido que tienen longitudes en el intervalo de 25-200 nucleótidos e incluyen segmentos de las regiones V, D, y J de una cadena  $\beta$  de TCR. En otra realización, un repertorio comprende un conjunto de secuencias de nucleótido 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 repertorio comprende un número de secuencias de nucleótidos distintas que es sustancialmente equivalente al número de linfocitos que expresan una cadena de IgH distinta. En otra realización, un repertorio comprende un número de secuencias de nucleótidos distintas que es sustancialmente equivalente al número de linfocitos que expresan una cadena  $\beta$  de TCR distinta. 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  $\beta$  o parte de ésta portada 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  $\beta$  o parte de ésta portada o expresada por cada linfocito presente con una frecuencia de 0,0001 por ciento o mayor. Los conjuntos de clonotipos descritos en las dos frases anteriores se refieren algunas veces en la presente memoria como que representan el "repertorio completo" de secuencias de IgH y/o TCR $\beta$ . Como se ha mencionado anteriormente, cuando se mide o genera un perfil de clonotipos (o perfil de repertorio), se obtiene una muestra lo suficientemente grande de linfocitos de manera que dicho perfil proporciona una representación razonablemente exacta de un repertorio para una aplicación particular. En un aspecto de los métodos de la invención, se emplean muestras que comprenden de  $10^5$  a  $10^7$  linfocitos, especialmente cuando se obtiene de muestras de sangre periférica de 1-10 mL.

"Etiqueta de secuencia" (o "etiqueta") significa un oligonucleótido que está unido a un polinucleótido o molde y se usa para identificar y/o seguir al polinucleótido o molde en una reacción. Una etiqueta de oligonucleótido puede estar unida al extremo 3' ó 5' de un polinucleótido o molde o puede insertarse en el interior de dicho molde de polinucleótido para formar un conjugado lineal, referido algunas veces en la presente memoria como "polinucleótido etiquetado", o "molde etiquetado", o "conjugado etiqueta-polinucleótido", o semejantes. Las etiquetas de oligonucleótido pueden variar ampliamente en tamaño y composiciones; las referencias siguientes proporcionan una guía para seleccionar conjuntos de etiquetas de oligonucleótido apropiadas para realizaciones particulares: Brenner, patente U.S. 5.635.400; Brenner et al, Proc. Natl. Acad. Sci., 97: 1665-1670 (2000); Church et al, publicación de patente europea 0 303 459; Shoemaker et al, Nature Genetics, 14: 450-456 (1996); Morris et al, publicación de patente europea 0799897A1; Wallace, patente U.S. 5.981.179; y semejantes. Las longitudes y composiciones de las etiquetas de oligonucleótido pueden variar ampliamente, y la selección de longitudes y/o composiciones

particulares depende de varios factores incluyendo, sin limitación, cómo se usan las etiquetas para generar una lectura, por ejemplo, mediante una reacción de hibridación o mediante una reacción enzimática, tal como secuenciación; si están marcadas, por ejemplo, con un agente de tinción fluorescente o semejantes; el número de etiquetas de oligonucleótidos distinguibles que se requieren para identificar inequívocamente un conjunto de polinucleótidos, y semejantes, y cómo de diferentes deben ser las etiquetas de un conjunto con el fin de asegurar una identificación fiable, por ejemplo, sin hibridación cruzada o error en la identificación a partir de errores de secuenciación. Las etiquetas de oligonucleótido pueden tener cada una una longitud en el intervalo de de 2 a 36 nucleótidos, o de 4 a 30 nucleótidos, o de 8 a 20 nucleótidos, o de 6 a 10 nucleótidos, respectivamente. Pueden usarse conjuntos de etiquetas en los que cada etiqueta de oligonucleótido de un conjunto tiene una secuencia de nucleótidos única que se diferencia de la de cualquier otra etiqueta del mismo conjunto por al menos dos bases; alternativamente, se usan conjuntos de etiquetas en los que la secuencia de cada etiqueta de un conjunto se diferencia de la de cualquier otra etiqueta del mismo conjunto por al menos tres bases.

**REIVINDICACIONES**

1. Un método para detectar la recurrencia de un neoplasma linfoide en un paciente con clonotipos de cáncer definidos que comprende las etapas de:
- 5 (a) determinar un perfil de secuencias de ADN recombinadas en células T y/o células B que comprende las etapas de:
- (i) amplificar moléculas de ácido nucleico de células T y/o células B de una muestra obtenida del paciente;
- (ii) aislar espacialmente moléculas individuales del ácido nucleico amplificado;
- y
- (iii) secuenciar las moléculas individuales del ácido nucleico por secuenciación de nueva generación para
- 10 formar el perfil de secuencias de ADN recombinadas;
- (b) evaluar el perfil de secuencias de ADN recombinadas para determinar los niveles de los clonotipos de cáncer definidos y los clones que evolucionan a partir de éstos; y
- (c) usar los niveles para detectar la recurrencia del neoplasma linfoide.
2. El método de la reivindicación 1 que incluye además la etapa de repetir las etapas (a) y (b) para
- 15 monitorizar dicho neoplasma linfoide en el paciente.
3. El método de la reivindicación 1 o reivindicación 2 en el que los clones evolucionan a partir de los clonotipos de cáncer definidos por reemplazo de VH.
4. El método de la reivindicación 3 en el que el reemplazo de VH se identifica detectando la unión D-J.
5. El método de la reivindicación 1 o reivindicación 2 en el que dichos clones evolucionan a partir de los
- 20 clonotipos de cáncer definidos por hipermutación somática.
6. El método de la reivindicación 1 o reivindicación 2 en el que dichos clones evolucionan a partir de los clonotipos de cáncer definidos por una o más inserciones y/o deleciones de de 1-10 bases.
7. El método de la reivindicación 1 o reivindicación 2 en el que los clones que evolucionan a partir de los clonotipos de cáncer definidos tienen una región V y región J mutada de forma idéntica y una región NDN
- 25 diferente.
8. El método de la reivindicación 1 o reivindicación 2 en el que los clones que evolucionan a partir de los clonotipos de cáncer definidos se computan sobre la base de:
- a) el número de diferencias en la región NDN,
- b) la longitud de la región NDN, y
- 30 c) la presencia de otras hipermutaciones somáticas en los segmentos V y/o J.
9. El método de una cualquiera de las reivindicaciones 1 a 8 en el que dicho neoplasma linfoide es un neoplasma linfoide maligno.
10. El método de una cualquiera de las reivindicaciones 1 a 9 en el que el ácido nucleico es ADN genómico.
- 35 11. El método de una cualquiera de las reivindicaciones 1 a 10 en el que dichas secuencias recombinadas comprenden una reorganización genómica seleccionada del grupo que consiste en una reorganización VDJ de IgH, una reorganización DJ de IgH, una reorganización VJ de IgK, una reorganización VJ de IgL, una reorganización VDJ de TCR  $\beta$ , una reorganización DJ de TCR  $\beta$ , una reorganización VJ de TCR  $\alpha$ , una reorganización VJ de TCR  $\lambda$ , una reorganización VDJ de TCR  $\delta$ , y una reorganización VD de TCR  $\delta$ .
- 40 12. El método de una cualquiera de las reivindicaciones 1 a 11 en el que dicha etapa de secuenciación incluye generar lecturas de secuencia en un intervalo de 30 a 400 nucleótidos.



**Fig. 1A**



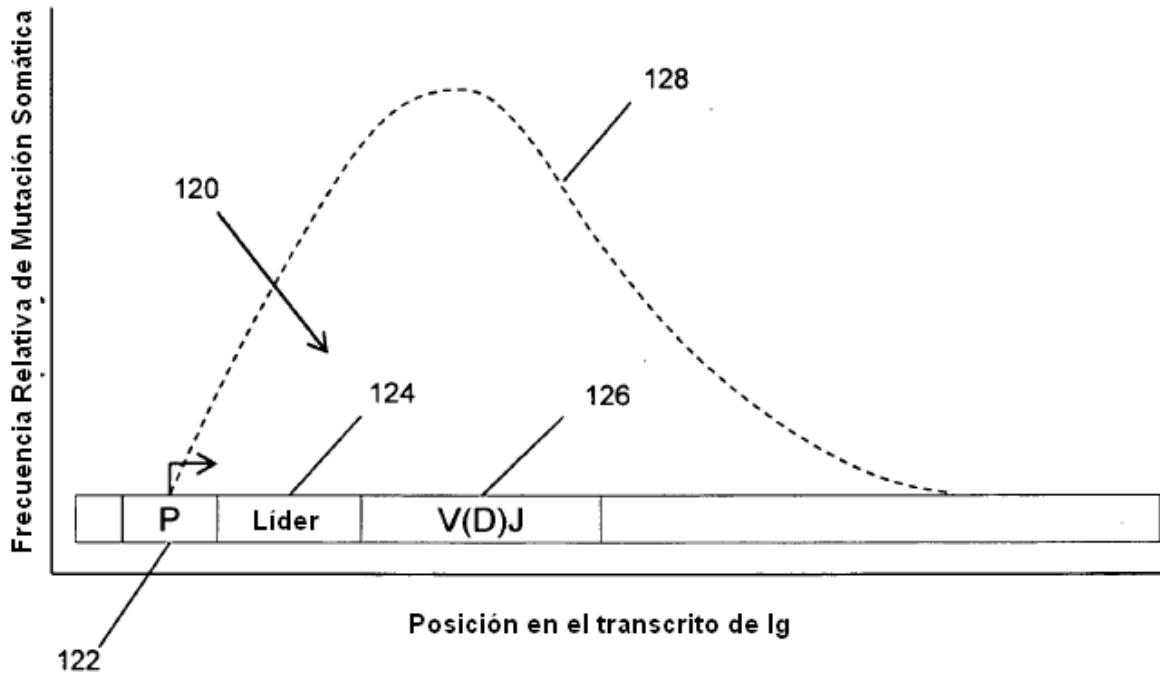


Fig. 1B

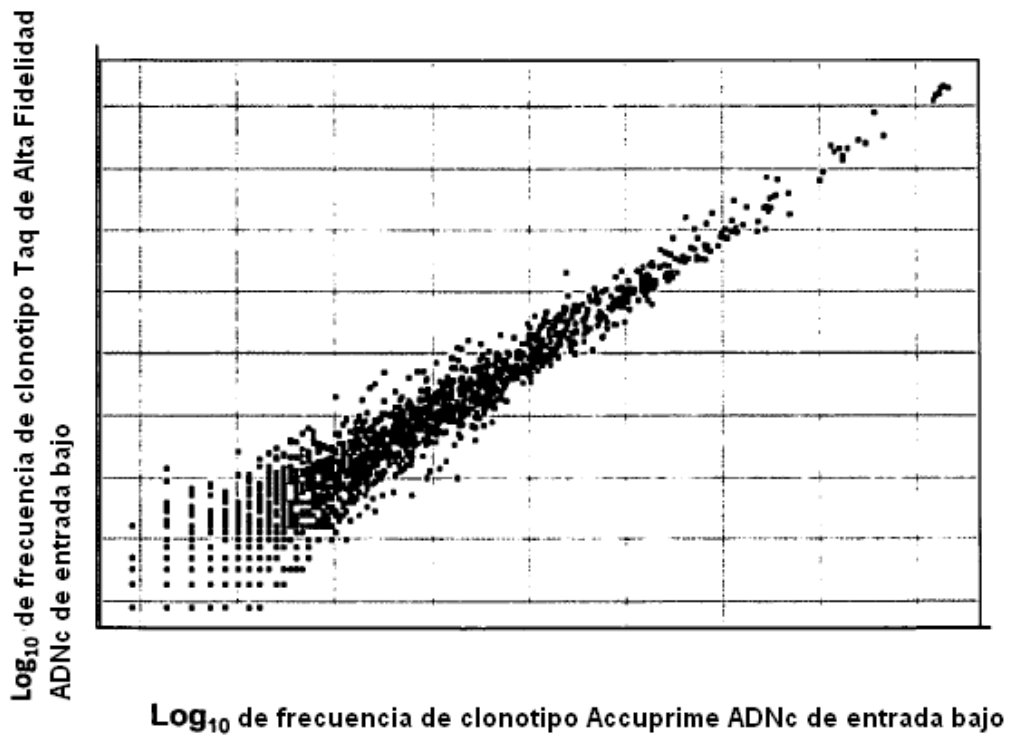


Fig. 7C

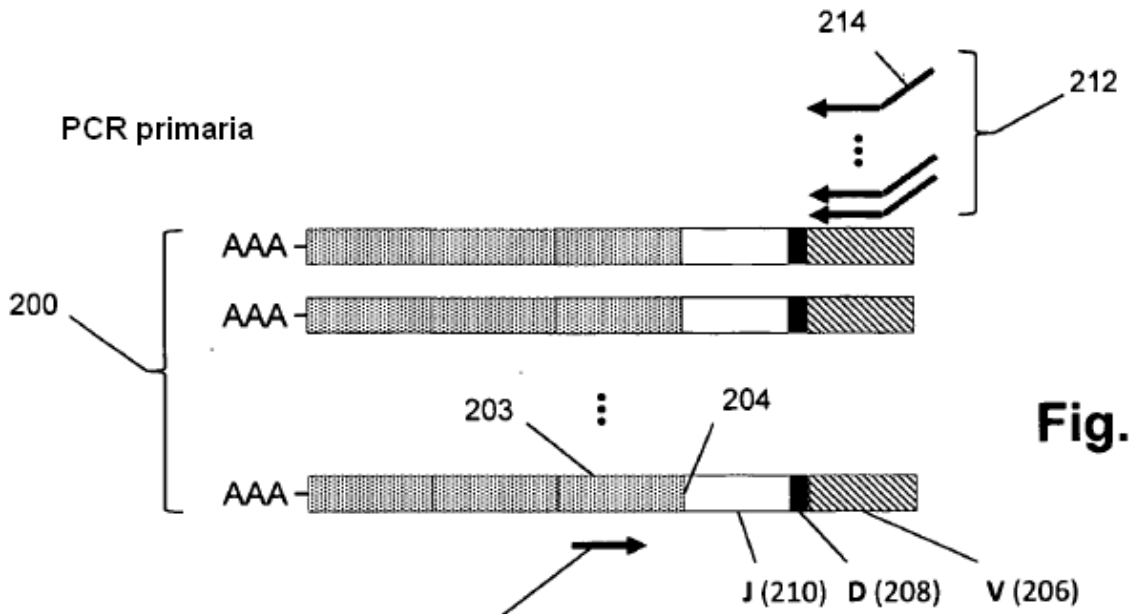


Fig. 2A

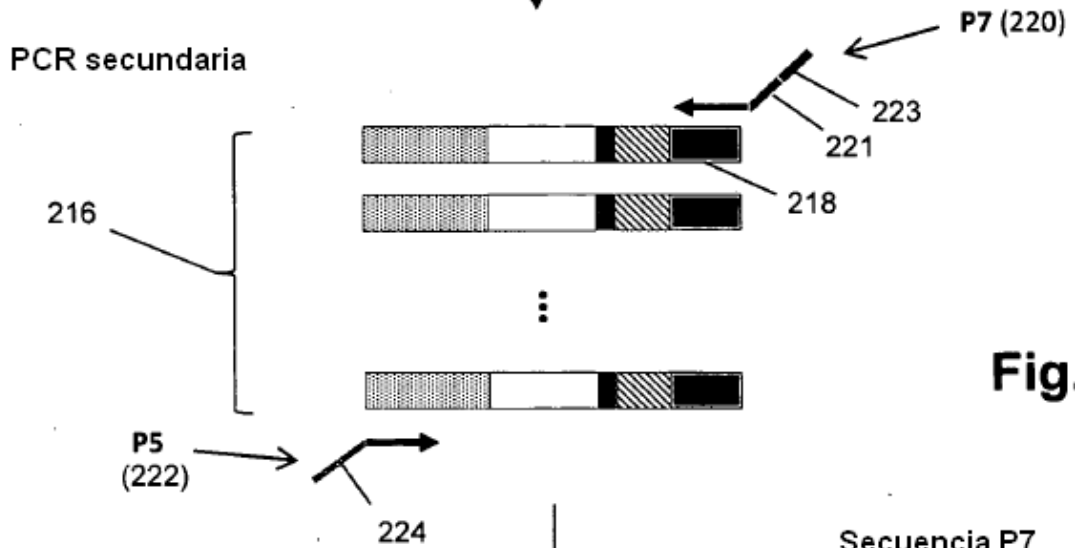


Fig. 2B

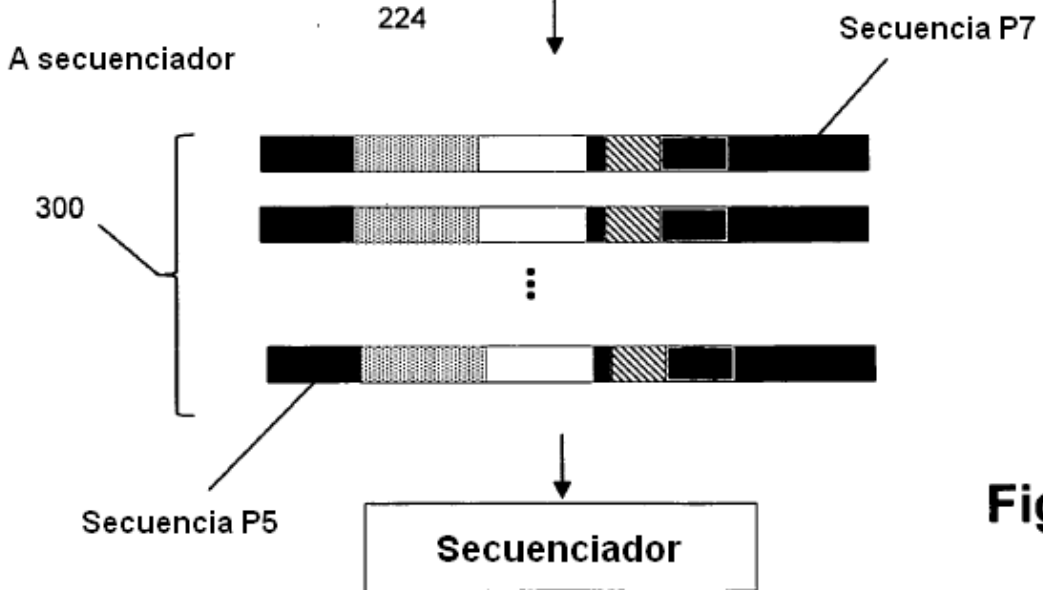
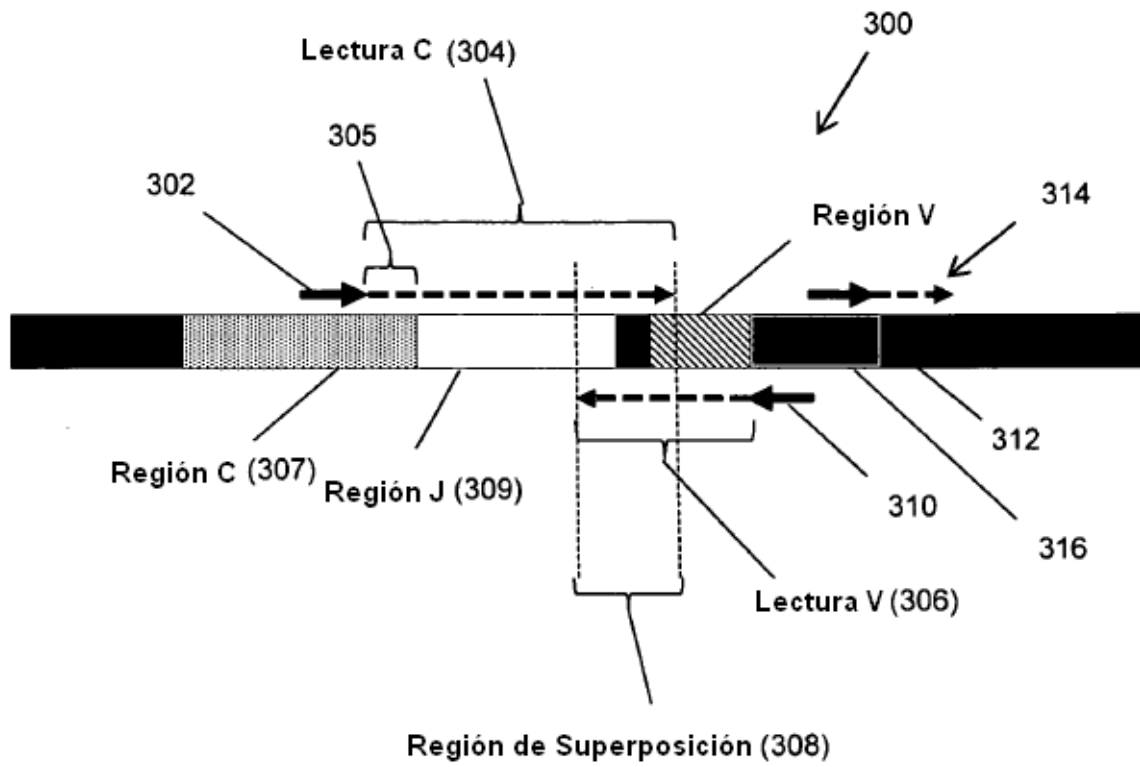
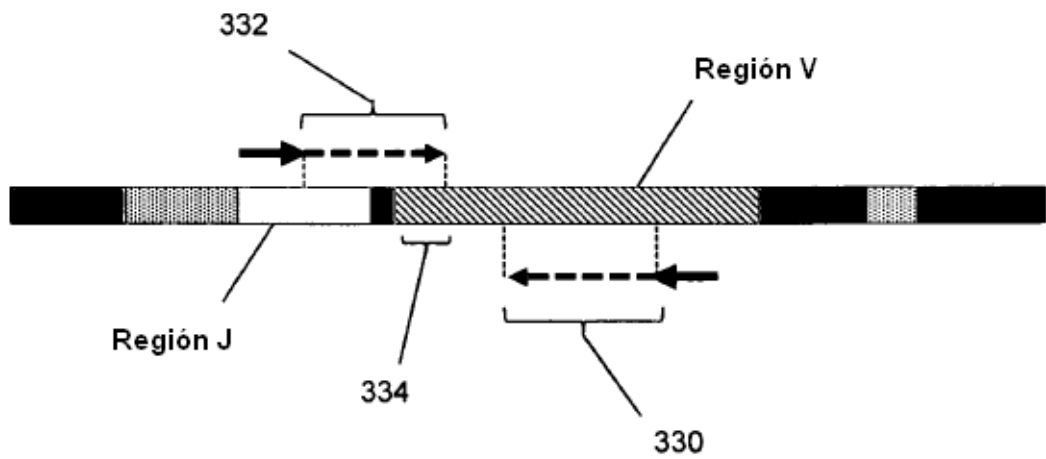


Fig. 3A



**Fig. 3B**



**Fig. 3C**

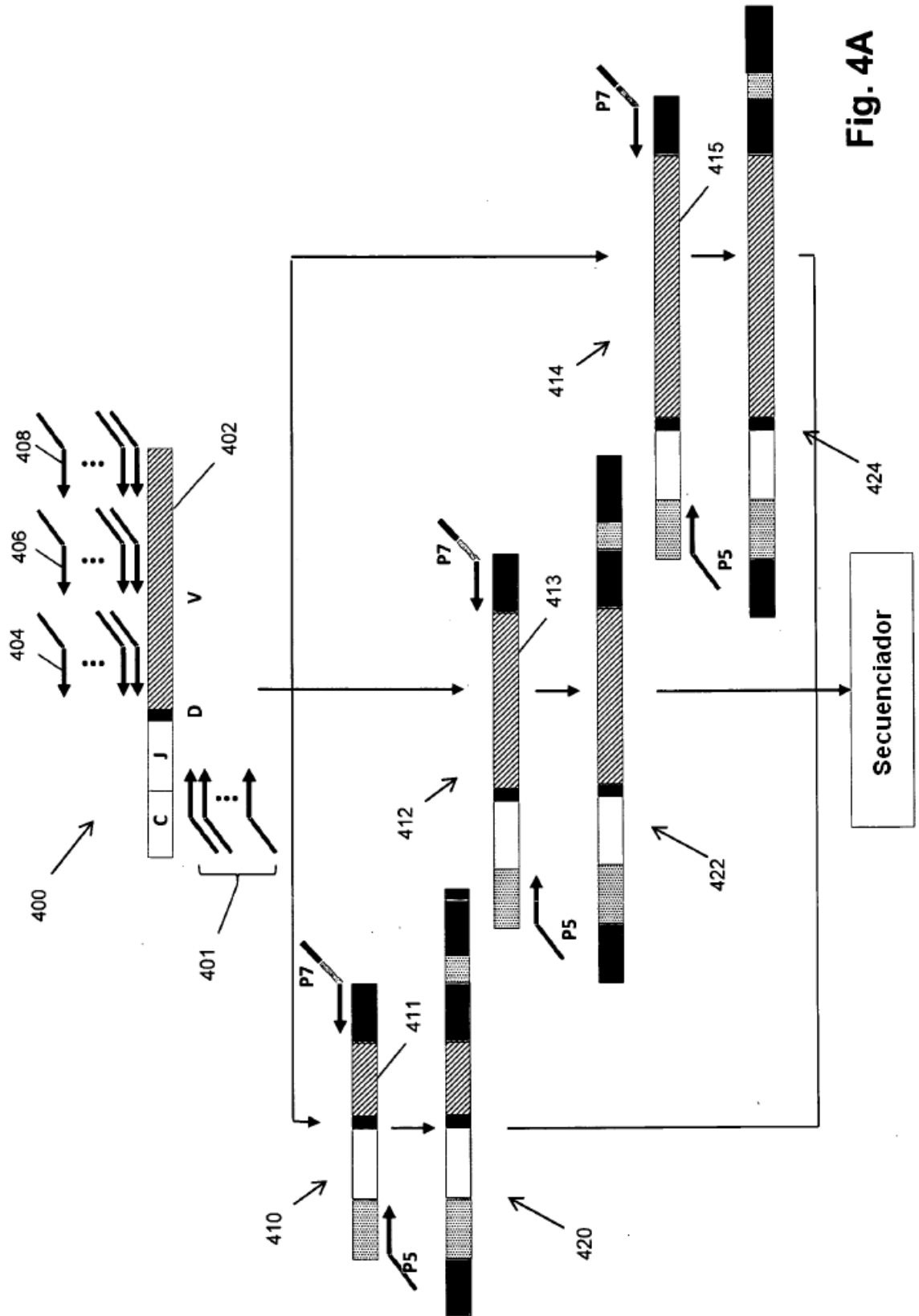


Fig. 4A

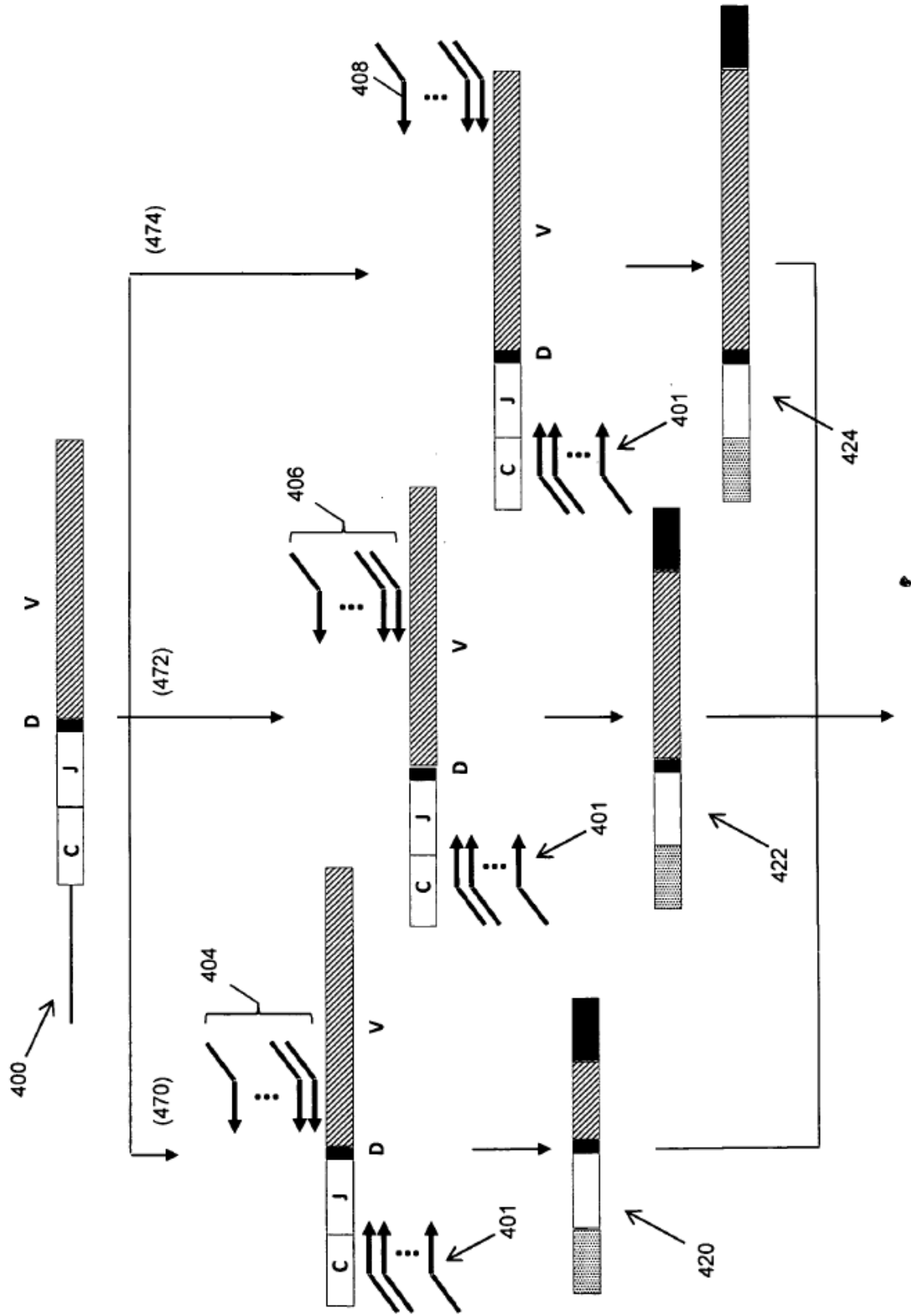


Fig. 4B

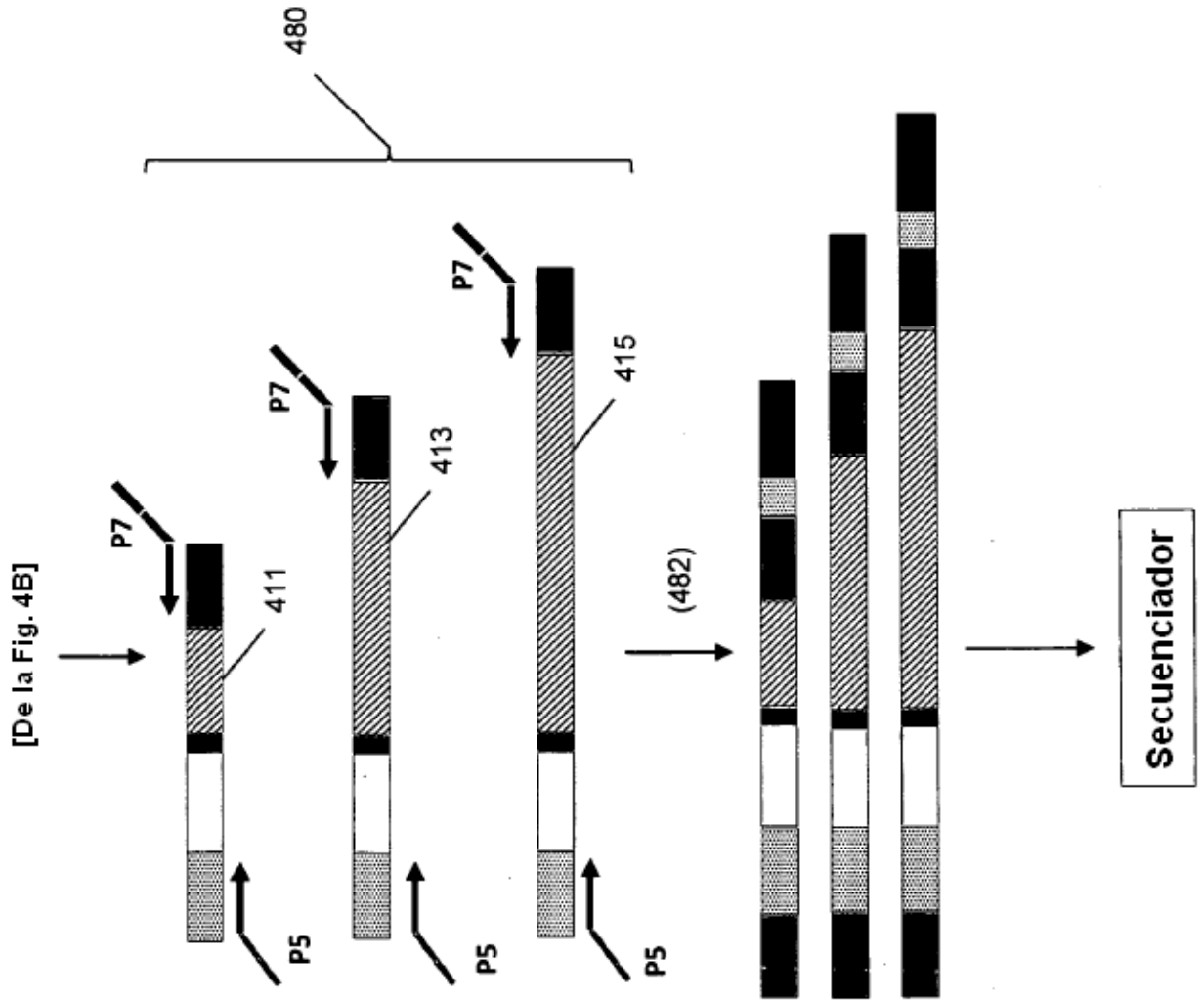


Fig. 4C

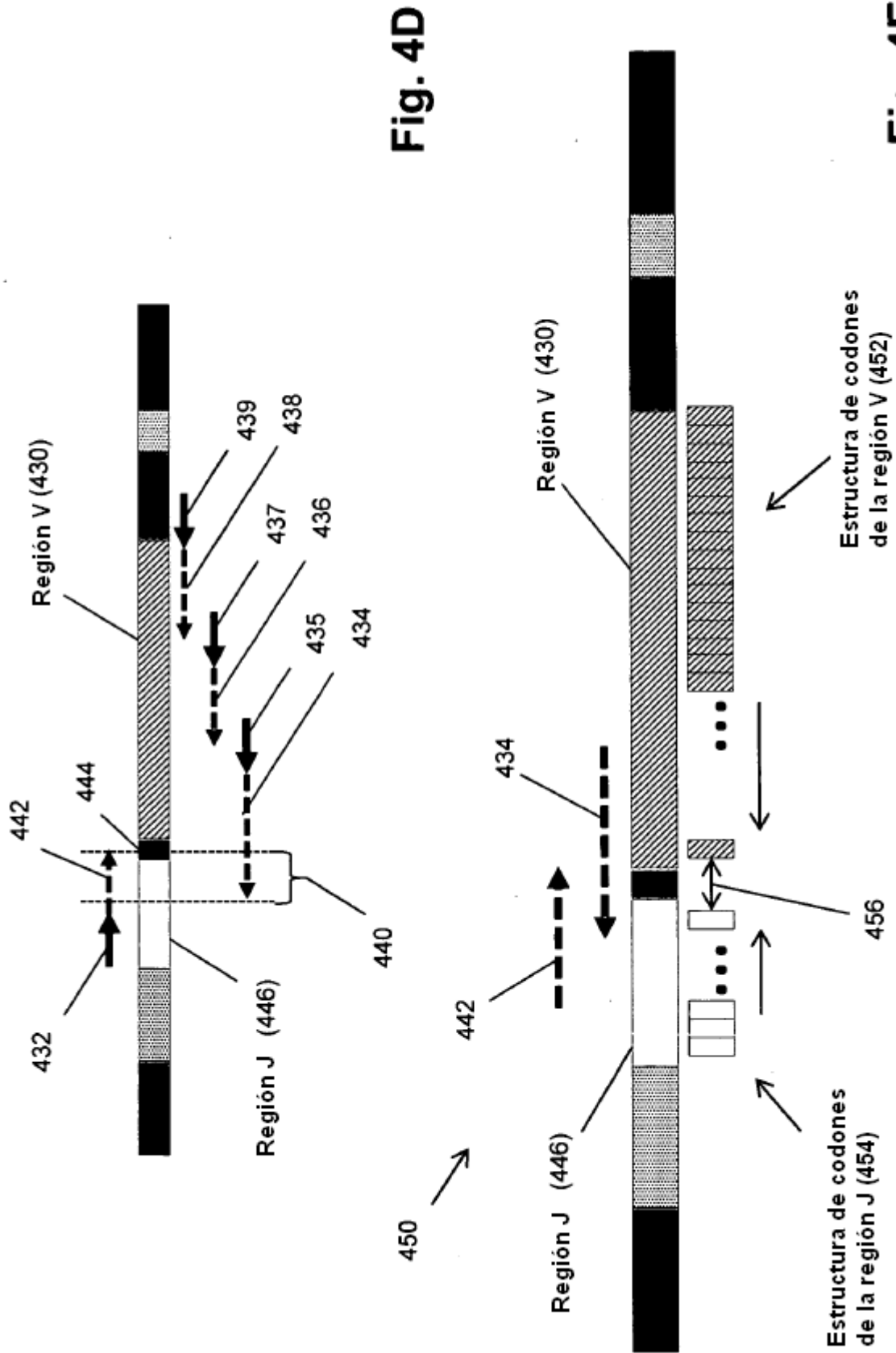


Fig. 4D

Fig. 4E

Número de ciclos para la muestra amplificada 2

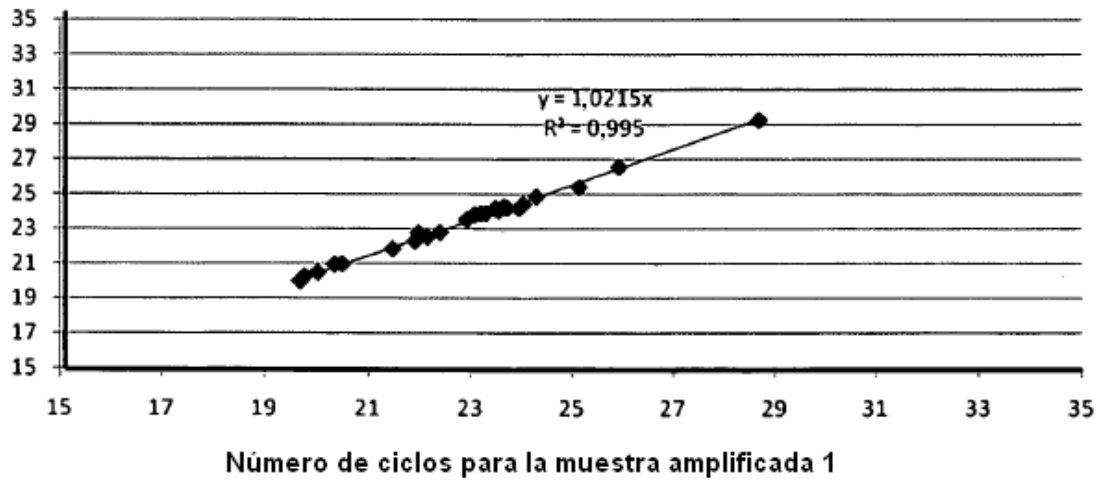


Fig. 5

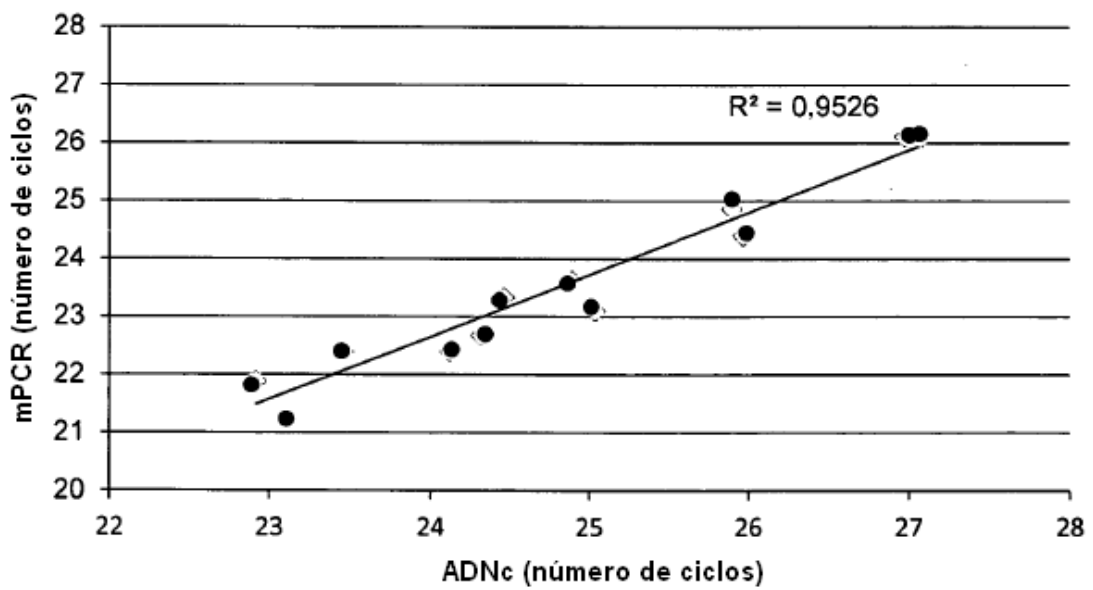
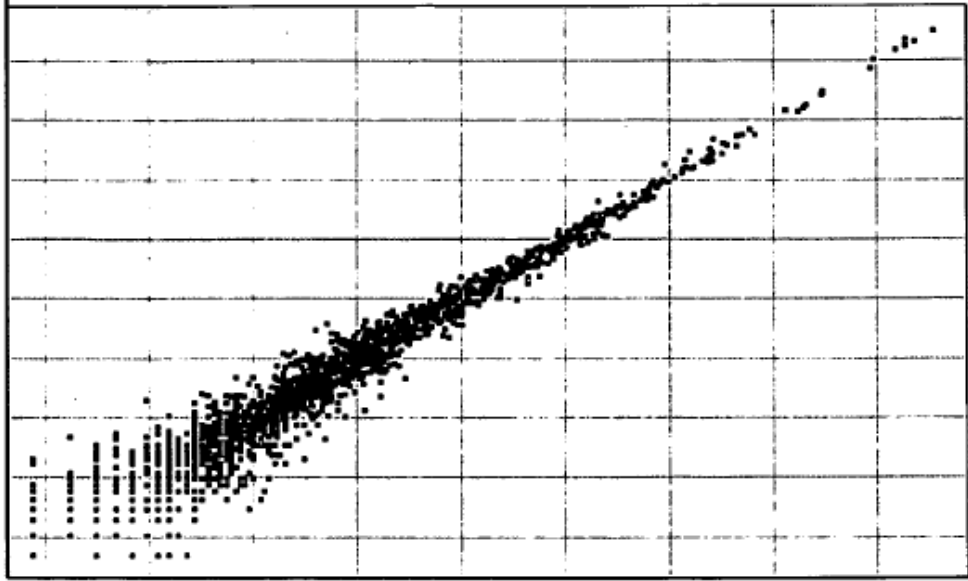


Fig. 6



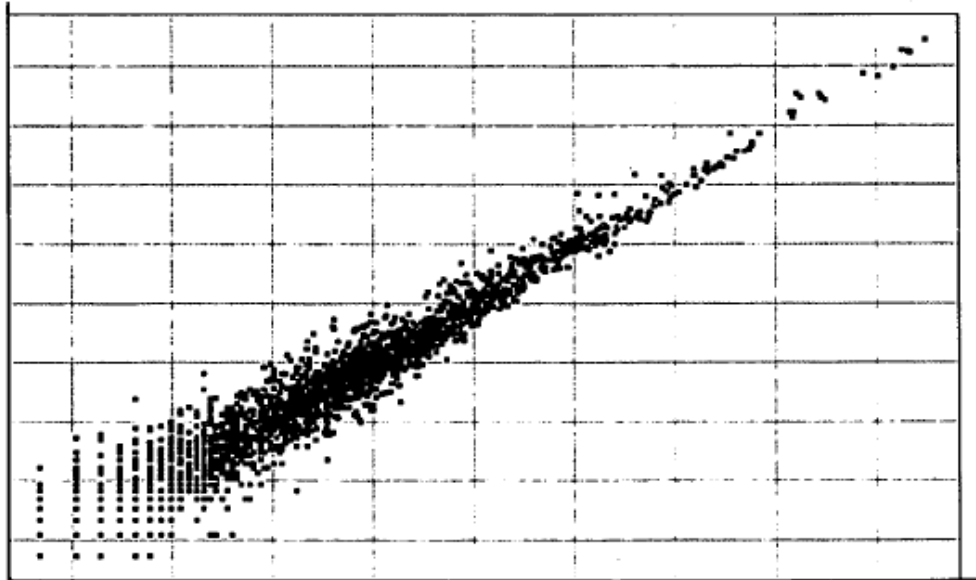
Log<sub>10</sub> frecuencia de clonotipo Accuprime muestra 2



Log<sub>10</sub> frecuencia de clonotipo Accuprime muestra 1

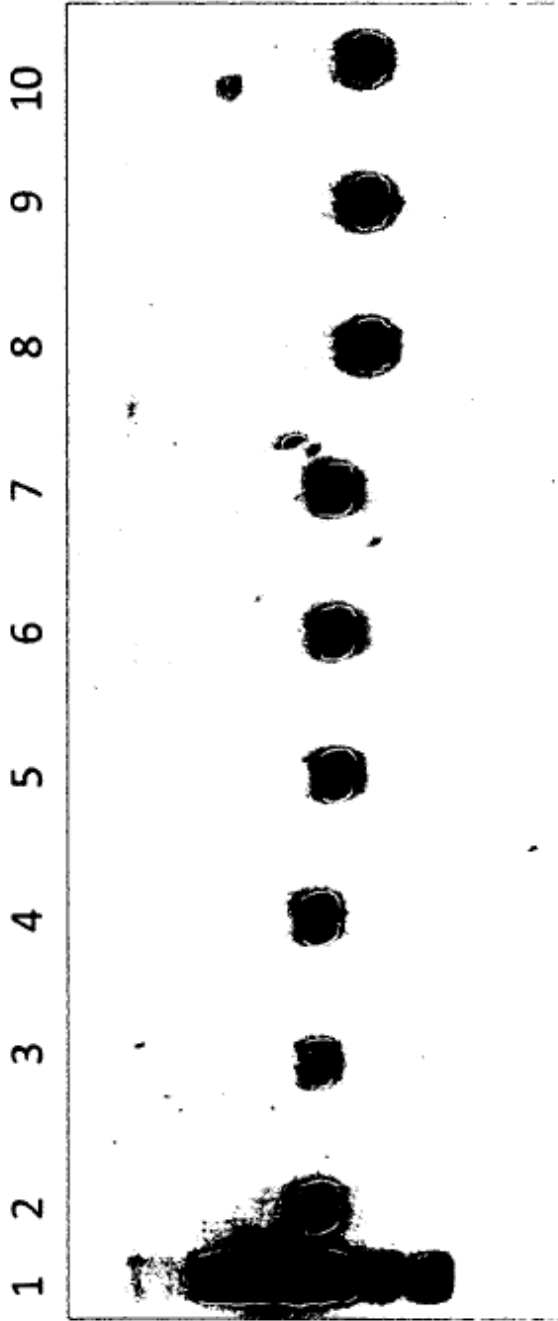
**Fig. 7A**

Log<sub>10</sub> frecuencia de clonotipo Taq alta fidelidad muestra 1



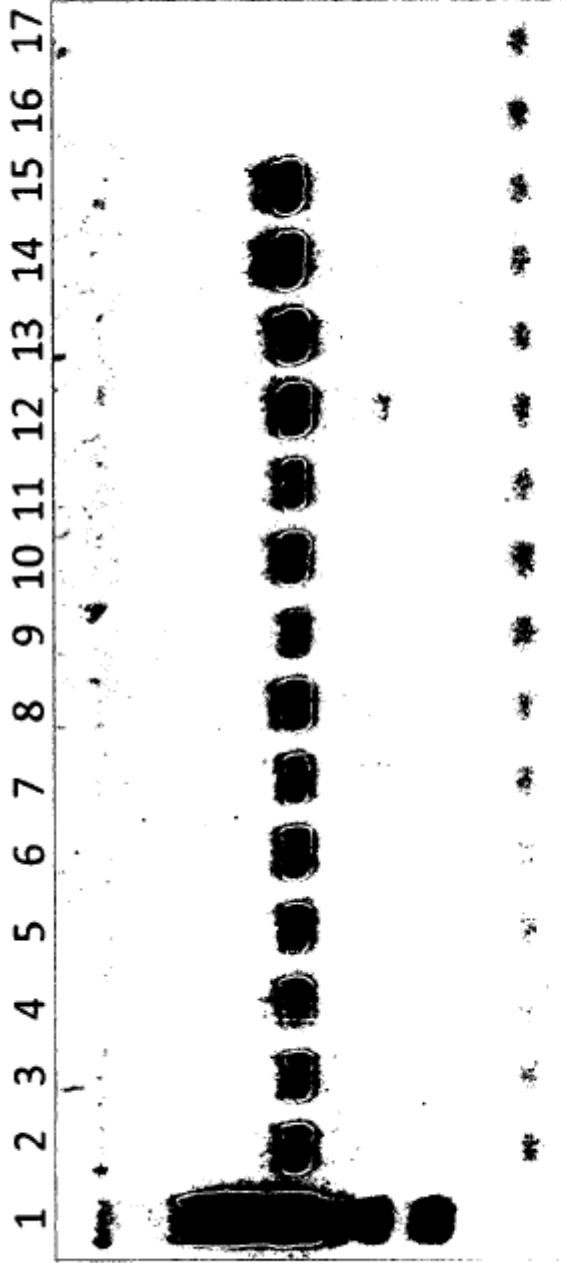
Log<sub>10</sub> frecuencia de clonotipo Accuprime muestra 2

**Fig. 7B**



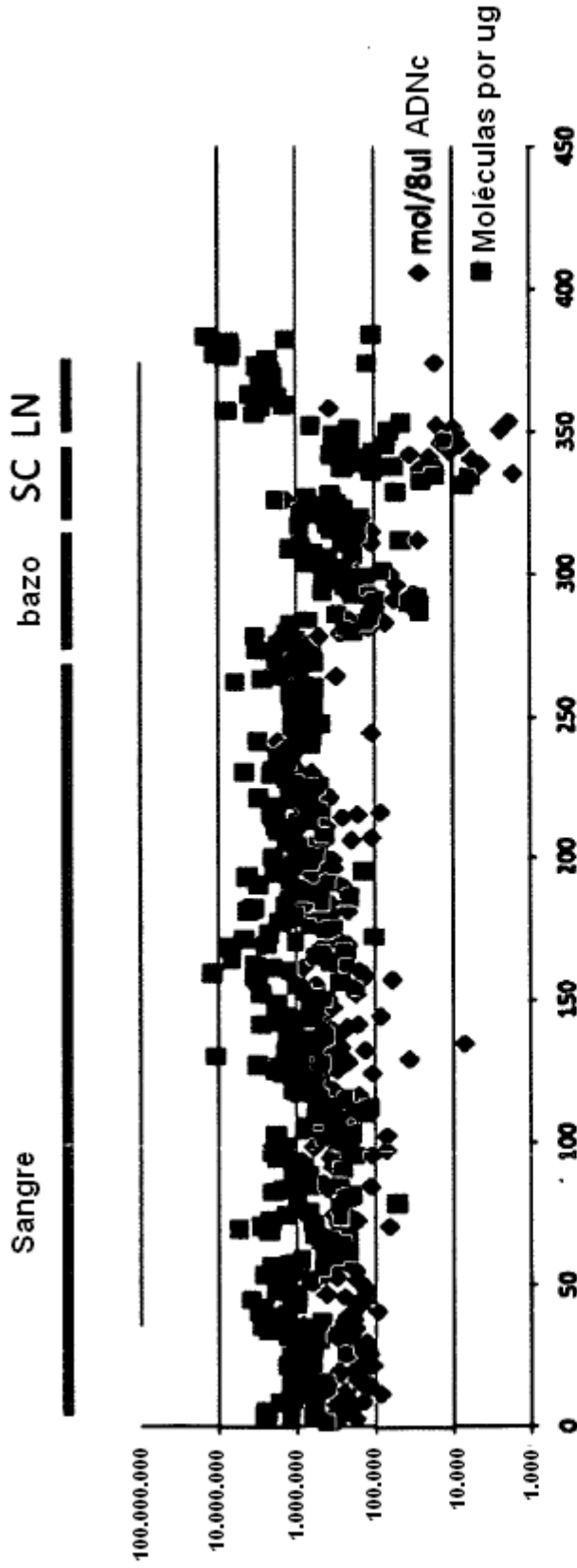
Se muestran los productos de amplificación después de 2 etapas de PCR con los 3 conjuntos de cebadores del segmento V en la etapa 1. Todos usan los cebadores de J en la primera etapa. El carril 1 es la escalera de 100 pb de tamaño (NEB). 2-4 amplificación con el conjunto de cebadores A, 5-7 con el conjunto de cebadores B, y 8-10 con el conjunto de cebadores C. Los 3 diferentes carriles para cada conjunto de cebadores es para 3 concentraciones diferentes de cebador del segmento V. Como los cebadores de la segunda etapa de PCR son los mismos para los tres combinados, también hemos mostrado que tres productos de la primera etapa de PCR pueden combinarse para ser amplificados en la segunda etapa de PCR.

**Fig. 8A**



Se muestran los productos de amplificación después de 2 etapas de PCR con los cebadores de J y el conjunto A de cebadores del segmento V en la etapa 1. El carril 1 es la escalera de 100 pb de tamaño (NEB), 2-9 usan una cantidad diferente de ADN genómico de sangre de entrada (carril 2-3: 250 ng, 4-5 y 10-11: 500 ng, 6-7: 1 ug, y 8-9: 2,5 ug). Los carriles 12-13 usan 250 ng de ADN genómico de bazo. Los carriles 14-15 usan 250 ng de la línea de células B Raji (control positivo), y los carriles 16-17 usan 250 ng de la línea de células T Jurkat (control negativo). Las dos reacciones para cada concentración de ADN genómico usan diferentes concentraciones de cebador.

**Fig. 8B**



Se usaron PCR en tiempo real usando estándares con números conocidos de moléculas para determinar el número de moléculas de  $TCR\beta$  en el ADN de entrada. Los puntos de color oscuro representan el número de moléculas que se amplificaron finalmente en la PCR múltiple, y los puntos de color claro muestran el número de moléculas por ug de ARN asumiendo una relación 1:1 perfecta entre ARN y ADNc. (SC es médula espinal y LN es nodo linfático).

**Fig. 9**

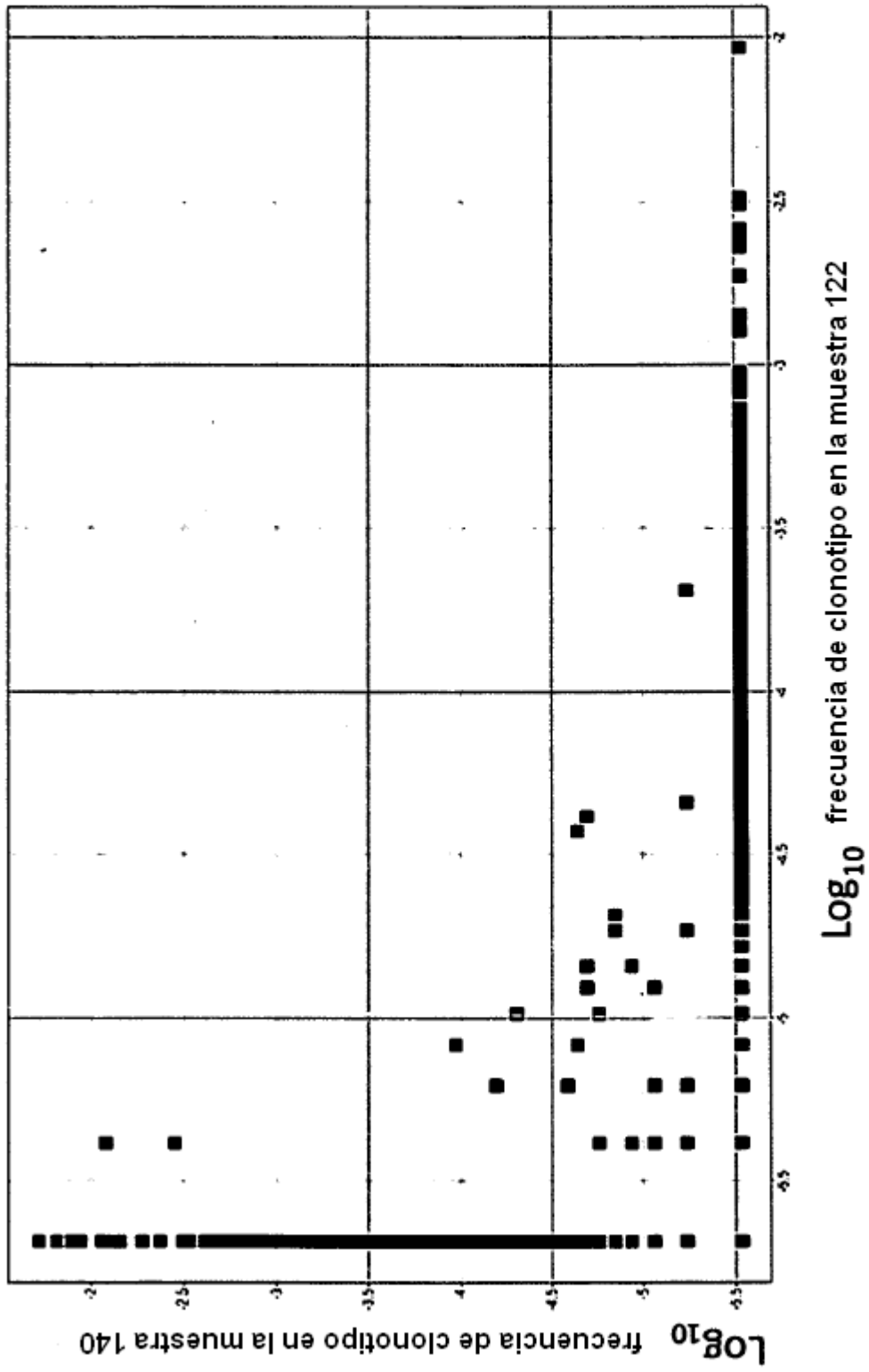


Fig. 10