



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS  
ESPAÑA



⑪ Número de publicación: **2 453 066**

⑮ Int. Cl.:  
**C12Q 1/68**  
(2006.01)

⑫

## TRADUCCIÓN DE PATENTE EUROPEA

T3

⑯ Fecha de presentación y número de la solicitud europea: **09.11.2009 E 09764927 (1)**  
⑯ Fecha y número de publicación de la concesión europea: **15.01.2014 EP 2364368**

⑮ Título: **Métodos para supervisar las condiciones por análisis de secuencia**

⑯ Prioridad:  
**07.11.2008 US 112693 P**

⑯ Fecha de publicación y mención en BOPI de la traducción de la patente:  
**03.04.2014**

⑮ Titular/es:  
**SEQUENTA, INC. (100.0%)  
409 Illinois Street  
San Francisco, CA 94158, US**

⑯ Inventor/es:  
**FAHAM, MALEK y  
WILLIS, THOMAS**

⑯ Agente/Representante:  
**UNGRÍA LÓPEZ, Javier**

**ES 2 453 066 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

Métodos para supervisar las condiciones por análisis de secuencia

5 **Antecedentes de la invención**

El sistema inmunitario comprende los sistemas inmunitarios innato y adaptativo. El sistema inmunitario innato comprende las células y mecanismos que utilizan métodos genéricos para reconocer patógenos ajenos. Las células implicadas en inmunidad innata incluyen neutrófilos, linfocitos citolíticos naturales, macrófagos, monocitos, basófilos, 10 eosinófilos, mastocitos y células dendríticas. Estas células llevan a cabo el acto de la fagocitosis así como la liberación de muchos productos químicos que destruyen patógenos invasores. Además, estas células están implicadas en mecanismos de defensa de inmunidad innata incluyendo la cascada del complemento e inflamación. Finalmente, algunas de estas células participan en el proceso de presentación de antígenos que desempeña un papel en el sistema inmunitario adaptativo.

15 El sistema inmunitario adaptativo ha evolucionado para atacar elementos específicos en sus dianas. La aparición de una respuesta a una diana específica proporciona al hospedador "memoria" de ella, provocando que monte una respuesta más fuerte si volviera a aparecer en otro momento. Habitualmente cualquier proteína o polisacárido puede actuar como la diana para algún subconjunto de las células de respuesta inmunitaria adaptativa o sus productos que 20 reconocen epítopos específicos en la diana. La respuesta inmunitaria adaptativa se divide en dos tipos: la respuesta inmunitaria humorla y la mediada por células, y los linfocitos B y linfocitos T desempeñan los papeles de especificidad en estas respuestas, respectivamente.

25 Puesto que la enfermedad autoinmunitaria implica el reconocimiento de algún elemento del sistema inmunitario adaptativo a dianas propias, se han examinado aspectos del sistema inmunitario adaptativo para ayudar en el diagnóstico o pronóstico. Usando técnicas inmunológicas convencionales, el sistema inmunitario humorla se ha investigado buscando autoanticuerpos en circulación. Se han identificado autoanticuerpos como antinucleares, anti ADNbc y factor reumatoide para varias enfermedades. Estos anticuerpos pueden no ser en sí mismos patológicos, ni es la diana que reconocen en el cuerpo necesariamente la misma que la ensayada *in vitro*, sin embargo, la medición 30 de sus niveles ayuda en el diagnóstico y en algunos casos tiene algunas implicaciones de pronóstico y tratamiento.

35 Otra metodología para estudiar el sistema inmunitario adaptativo en enfermedad autoinmunitaria se basa en el análisis de la diversidad de las células inmunitarias adaptativas. La activación de las células inmunitarias adaptativas conduce a su expansión clonal. Se consiguen habitualmente pruebas de esta expansión clonal mediante amplificación del ARN o ADN sanguíneo de parte de la secuencia de ácido nucleico que codifica la región de reconocimiento de antígenos. Por ejemplo, se usan cebadores de PCR para amplificar secuencias que tienen un segmento V específico de la cadena β en el receptor de linfocitos T (análogo a la cadena pesada de anticuerpo) para amplificar los segmentos J o segmentos J y D conectados con el segmento V específico. Cuando está presente una población celular diversa, se espera que amplifique fragmentos con una distribución de amplicones de tamaño 40 ligeramente diferente, pero la expansión clonal provoca que se enriquezcan tamaños específicos y por lo tanto sean más intensos al visualizarse 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.

45 Un problema del enfoque de espectrotipificación es que muchas secuencias distintas pueden tener la misma longitud y por lo tanto son indistinguibles. Por lo tanto solamente puede determinarse expansión clonal drástica por esta técnica. Experimental Hematology (2007) vol. 35, páginas 831-841 desvela la identificación del repertorio de TCR por hibridación con matrices de sondas específicas de J o V; sin embargo, dicha técnica no permite la identificación específica de secuencia de un perfil de secuencias J-V recombinadas y su cuantificación.

50 Existe la necesidad de mejorar métodos para diagnosticar y ayudar al pronóstico de enfermedad autoinmunitaria y patologías autoinmunitarias así como otras enfermedades para las que el sistema inmunitario desempeña un papel central.

55 Aunque la especificidad adicional en la realización del perfil del sistema inmunitario sería muy útil al permitir que se prediga mejor su impacto en la salud humana, se proporcionaría aún más utilidad si se desarrollaran métodos que permitieran que los linfocitos T y B específicos implicados en procesos de enfermedad se identifiquen incluso si esas secuencias particulares nunca se habían observado antes. La amplia diversidad del sistema inmunitario le proporciona una reserva inmensa de células potencialmente útiles pero también presenta un reto al investigador que intenta usar este repertorio para fines predictivos. Cualquier secuencia individual que se dirige a un antígeno es una de un amplio número que podría estar implicado con y/o correlacionarse con el proceso de enfermedad en un individuo dado. Los métodos que identificaran cuáles de las muchas células en un individuo dado están implicadas en procesos de enfermedad serían muy valiosos para la salud humana.

**Sumario de la invención**

- En un aspecto, se proporciona un método para determinar un perfil de secuencias de ADN recombinadas en linfocitos T y/o linfocitos B que comprende: obtener una muestra de un sujeto que comprende linfocitos T y/o linfocitos B, aislar espacialmente moléculas individuales de ADN genómico de dichas células; secuenciar dichas moléculas individuales aisladas espacialmente de ADN genómico, y determinar los niveles de secuencias diferentes de dicha muestra para generar dicho perfil de secuencias de ADN recombinado.
- 5 En otro aspecto, se proporciona un método para determinar un perfil de secuencias de ADN recombinado en linfocitos T y/o linfocitos B que comprende: obtener una muestra de un sujeto que comprende linfocitos T y/o linfocitos B, aislar espacialmente moléculas individuales de ADN genómico de dichas células, amplificar dichas moléculas individuales de ADN genómico, secuenciar dicho ADN amplificado y determinar los niveles de secuencias diferentes de dicha muestra para generar dicho perfil de secuencias de ADN recombinado.
- 10 En otro aspecto, se proporciona un método para determinar un perfil de secuencias de ADN recombinado en linfocitos T y/o linfocitos B que comprende: obtener una muestra de un sujeto que comprende linfocitos T y/o linfocitos B, amplificar 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 secuencias diferentes de dicha muestra para generar dicho perfil de secuencias de ADN recombinado.
- 15 En otro aspecto, se proporciona un método para determinar un perfil de secuencias de ADN recombinado en linfocitos T y/o linfocitos B que comprende: obtener una muestra de un sujeto que comprende linfocitos T y/o linfocitos B, amplificar ADN genómico de dichas células, aislar espacialmente moléculas individuales de dicho ADN amplificado, volver a amplificar dichas moléculas de ADN amplificado, secuenciar dichas moléculas de ADN que se ha vuelto a amplificar, y determinar los niveles de secuencias diferentes de dicha muestra para generar dicho perfil de secuencias de ADN recombinado.
- 20 En otro aspecto, se proporciona un método para determinar un perfil de secuencias de ADN recombinado en linfocitos T y/o linfocitos B que comprende: obtener una muestra de un sujeto que comprende linfocitos T y/o linfocitos B, amplificar ADN genómico de dichas células, aislar espacialmente moléculas individuales de dicho ADN amplificado, volver a amplificar opcionalmente dichas moléculas individuales aisladas espacialmente de ADNc, secuenciar dicho ADNc y/o ADNc reamplificado; y determinar los niveles de diferentes secuencias de dicha muestra para generar dicho perfil de secuencias de ADN recombinado.
- 25 En otro aspecto, se proporciona un método para determinar un perfil de secuencias de ADN recombinado en linfocitos T y/o linfocitos B que comprende: obtener una muestra de un sujeto que comprende linfocitos T y/o linfocitos B, transcribir de forma inversa ARN de dichas células para formar ADNc, aislar espacialmente moléculas individuales de dicho ADNc, volver a amplificar opcionalmente dichas moléculas individuales aisladas espacialmente de ADNc, secuenciar dicho ADNc y/o ADNc reamplificado; y determinar los niveles de diferentes secuencias de dicha muestra para generar dicho perfil de secuencias de ADN recombinado.
- 30 En otro aspecto, se proporciona un método para determinar un perfil de secuencias de ADN recombinado en linfocitos T y/o linfocitos B que comprende: obtener una muestra de un sujeto que comprende linfocitos T y/o linfocitos B, transcribir de forma inversa ARN de dichas células para formar ADNc, aislar espacialmente moléculas individuales de dicho ADNc, volver a amplificar opcionalmente dichas moléculas individuales aisladas espacialmente de ADNc, secuenciar dicho ADNc y/o ADNc reamplificado; y determinar los niveles de diferentes secuencias de dicha muestra para generar dicho perfil de secuencias de ADN recombinado.
- 35 En otro aspecto, se proporciona un método para determinar un perfil de secuencias de ADN recombinado en linfocitos T y/o linfocitos B que comprende: obtener una muestra de un sujeto que comprende linfocitos T y/o linfocitos B, aislar espacialmente células individuales en dicha muestra, secuenciar moléculas individuales de ácido nucleico de dichas células; y determinar los niveles de diferentes secuencias a partir de dicha muestra para generar dicho perfil de secuencias de ADN recombinado.
- 40 En una realización, dicha amplificación y/o reamplificación comprende PCR, PCR múltiple, TMA, NASBA o LAMP. En otra realización, dicho aislamiento espacial comprende subclonar dicho ADN o ADNc en vectores usados para transformar bacterias, separar dicho ADN o ADNc en dos dimensiones en un soporte sólido, separar dicho ADN o ADNc en tres dimensiones en una solución con micelas, o separar moléculas usando cámaras de microrreacción. En otra realización, dicha amplificación y/o reamplificación es mediante el crecimiento de bacterias que albergan ADN o ADNc subclonado, amplificación de ADN o ADNc en un portaobjetos, o amplificación de ADN o ADNc en una perla.
- 45 En otra realización, dicha secuenciación comprende secuenciación de didesoxi. En otra realización, dicha secuenciación comprende secuenciar por síntesis usando nucleótidos marcados en el extremo de forma reversible. En otra realización, dicha secuenciación comprende detección de liberación de pirofosfato en la incorporación de nucleótidos. En otra realización, dicha secuenciación comprende hibridación específica de alelos con una biblioteca de sondas oligonucleotídicas marcadas. En otra realización, dicha secuenciación comprende secuenciar por síntesis usando hibridación específica de alelo con una biblioteca de sondas oligonucleotídicas marcadas seguido de ligamiento de dichas sondas. En otra realización, dicha secuenciación comprende control en tiempo real de la incorporación de nucleótidos marcados durante una etapa de polimerización.
- 50 En otra realización, dichas secuencias de ADN recombinado comprenden genes de receptores de linfocitos T y/o genes de inmunoglobulina. En otra realización, dicha secuenciación comprende secuenciar un subconjunto de las secuencias clonales completas de inmunoglobulina y/o genes de receptores de linfocitos T.
- 55 En otra realización, dicho subconjunto de las secuencias clonales completas comprende el punto de unión V-D, el punto de unión D-J de una inmunoglobulina o gen del receptor de linfocitos T, la región variable completa de una inmunoglobulina o gen de receptor de linfocitos T, la región de reconocimiento de antígenos o la región determinante de complementariedad 3 (CDR3). En otra realización, dichos genes de receptor de linfocitos T comprenden genes  $\beta$  de receptor de linfocitos T. En otra realización, dichos genes de inmunoglobulina comprenden genes pesados de inmunoglobulina. En otra realización, dicha amplificación o reamplificación comprende una pluralidad de cebadores

complementarios de segmentos V y un cebador complementario de un segmento C. En otra realización, dicha amplificación o reamplificación comprende una pluralidad de cebadores complementarios de segmentos V y una pluralidad de cebadores complementarios de segmentos C.

- 5 En otra realización, dicha pluralidad de cebadores complementarios de segmentos V comprende al menos tres cebadores diferentes para cada segmento V y la pluralidad de cebadores complementarios de segmentos C comprende al menos 1, al menos 2, al menos 3, al menos 4, al menos 5 o al menos 6 cebadores.
- 10 En otras realizaciones, dichos linfocitos T o B son subconjuntos de los linfocitos T y B totales. En otra realización, dicho subconjunto de linfocitos T son linfocitos CD4+, CD8+, o linfocitos CD27<sup>alto</sup>. En otra realización, dicha muestra comprende al menos 100.000, al menos 500.000, al menos 750.000 o al menos 1.000.000 de linfocitos T.
- 15 En otra realización, dicha secuenciación comprende al menos 1000 lecturas por ciclo, al menos 10.000 lecturas por ciclo, al menos 100.000 lecturas por ciclo, o al menos 1.000.000 de lecturas por ciclo. En otra realización, 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.
- 20 En otra realización, dicha muestra se toma cuando el sujeto está en un estado de brote de una enfermedad autoinmunitaria. En otra realización, dicha muestra se toma de un sujeto que tiene o se sospecha que tiene lupus eritematoso sistémico.
- 25 En otro aspecto, se proporciona un método para determinar uno o más clonotipos correlacionados en un sujeto que comprende: generar uno o más perfiles de clonotipo por secuenciación de ácido nucleico de moléculas aisladas espacialmente, individuales, 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 correlacionados en el sujeto basándose en el o los perfiles de clonotipo.
- 30 En una realización, dicha al menos una muestra es de un tejido aquejado por la enfermedad. En otra realización, dicha determinación de uno o más clonotipos correlacionados comprende comparar perfiles de clonotipo de al menos dos muestras.
- 35 En otra realización, el primer estado de la enfermedad es un estado pico de la enfermedad. En otra realización, dichos uno o más clonotipos correlacionados están presentes en el estado pico de la enfermedad. En otra realización, dichos uno o más clonotipos correlacionados están ausentes en el estado pico de la enfermedad. En otra realización, dichos uno o más clonotipos correlacionados son altos en el estado pico de la enfermedad. En otra realización, dichos uno o más clonotipos correlacionados son bajos en el estado pico de la enfermedad.
- 40 En otra realización, dicha muestra comprende linfocitos T y/o linfocitos B. En otra realización, dichos linfocitos T y/o linfocitos B comprenden un subconjunto de linfocitos T y/o linfocitos B. En otra realización, dicho subconjunto de linfocitos T y/o linfocitos B está enriquecido por la interacción con un marcador. En otra realización, dicho marcador es un marcador de superficie celular en el subconjunto de linfocitos T y/o linfocitos B. En otra realización, dicho subconjunto de linfocitos T y/o linfocitos B interacciona con un antígeno específicamente presente en la enfermedad.
- 45 En otra realización, la enfermedad es lupus eritematoso sistémico o esclerosis múltiple.
- 50 En otro aspecto, se proporciona un método para desarrollar un algoritmo que puede predecir uno o más clonotipos correlacionados en cualquier muestra de un sujeto con una enfermedad que comprende: a) generar una pluralidad de perfiles de clonotipo de un conjunto de muestras, en el que las muestras son relevantes para la enfermedad, b) identificar uno o más clonotipos correlacionados del conjunto de muestras, c) usar parámetros de secuencia y/o datos funcionales de uno o más clonotipos correlacionados identificados en b) para desarrollar el algoritmo que puede predecir clonotipos correlacionados en cualquier muestra de un sujeto con la enfermedad.
- 55 En una realización, el conjunto de muestras se toma de uno o más tejidos aquejados por la enfermedad.
- 60 En otra realización, dicha identificación de uno o más clonotipos correlacionados comprende comparar perfiles de clonotipos de al menos dos muestras.
- 65 En otra realización, dichos datos funcionales incluyen la capacidad de unión de marcadores en la superficie de linfocitos T y/o linfocitos B o interacción con un antígeno por un linfocito T o linfocito B.
- 70 En otra realización, dichos parámetros de secuencia comprenden secuencia de ácido nucleico y secuencia de aminoácidos predicha.
- 75 En otra realización, las muestras son de uno o más individuos en un estado pico de la enfermedad. En otra realización, dichos uno o más clonotipos correlacionados están presentes en el estado pico de la enfermedad. En

otra realización, dichos uno o más clonotipos correlacionados están a un nivel alto en el estado pico de la enfermedad. En otra realización, dichos uno o más clonotipos correlacionados están a un nivel bajo en el estado pico de la enfermedad. En otra realización, el o los clonotipos correlacionados están ausentes en el estado pico de la enfermedad. En otra realización, la enfermedad es lupus eritematoso sistémico o esclerosis múltiple.

- 5 En otra realización, se proporciona un método para descubrir uno o más clonotipos correlacionados para un individuo, que comprende introducir un perfil de clonotipo de una muestra del individuo en un algoritmo, y usar el algoritmo para determinar uno o más clonotipos correlacionados para el individuo. En una realización, el algoritmo es un algoritmo que puede predecir uno o más clonotipos correlacionados en cualquier muestra de un sujeto con una enfermedad, desarrollándose dicho algoritmo: a) generando una pluralidad de perfiles de clonotipo de un conjunto de muestras, donde las muestras son relevantes para la enfermedad, b) identificando uno o más clonotipos correlacionados del conjunto de muestras, c) usando parámetros de secuencia y/o datos funcionales de uno o más clonotipos correlacionados identificados en b) para desarrollar un algoritmo que puede predecir clonotipos correlacionados en cualquier muestra de un sujeto con la enfermedad.
- 10 15 En una realización, dicha muestra se toma en un estado pico de la enfermedad. En otra realización, la muestra se toma de tejido aquejado por la enfermedad.
- 20 En otro aspecto, se proporciona un método para generar un algoritmo que calcula una puntuación de actividad de enfermedad que comprende: desarrollar un algoritmo que usa un conjunto de factores para combinar niveles de clonotipos correlacionados en una puntuación de actividad de la enfermedad, comparar la puntuación de la actividad de la enfermedad con datos clínicos con respecto a la patología, y optimizar los factores para maximizar la correlación entre datos clínicos y la puntuación de actividad de enfermedad.
- 25 En una realización, se proporciona un método para supervisar la patología de un individuo que comprende:
- 30 a) determinar un perfil de clonotipo de una muestra del individuo, b) introducir la información del perfil de clonotipo de a) en un algoritmo que calcula una puntuación de actividad de enfermedad, en el que el algoritmo se genera desarrollando un algoritmo que usa un conjunto de factores para combinar niveles de clonotipos correlacionados en una puntuación de actividad de enfermedad, comparar la puntuación de actividad de enfermedad con datos clínicos con respecto a la patología, y optimizar los factores para maximizar la correlación entre datos clínicos y la puntuación de actividad de enfermedad, y c) usar el algoritmo que calcula una puntuación de actividad de enfermedad para generar una puntuación predictiva de la patología del individuo.
- 35 40 En otra realización, el método para supervisar la patología del individuo comprende además determinar uno o más clonotipos correlacionados en el individuo, e introducir información del o los clonotipos correlacionados en el algoritmo. En otra realización, dicha determinación de uno o más clonotipos correlacionados en el individuo comprende a) generar uno o más perfiles de clonotipo por secuenciación de ácido nucleico de moléculas aisladas espacialmente, individuales, 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 correlacionados en el sujeto basándose en el o los perfiles de clonotipo.
- 45 50 En otra realización, dicha determinación de uno o más clonotipos correlacionados en el individuo comprende a) introducir un perfil de clonotipo de una muestra del individuo en un algoritmo que puede predecir uno o más clonotipos correlacionados, donde dicho algoritmo que puede predecir uno o más clonotipos correlacionados se desarrolla i) generando una pluralidad de perfiles de clonotipo de un conjunto de muestras, en el que las muestras son relevantes para la enfermedad, ii) identificando uno o más clonotipos correlacionados del conjunto de muestras, iii) usando parámetros de secuencia y/o datos funcionales de uno o más clonotipos correlacionados identificados en ii) para desarrollar el algoritmo que puede predecir clonotipos correlacionados en cualquier muestra de un sujeto con la enfermedad, y c) usar el algoritmo que puede predecir uno o más clonotipos correlacionados para determinar uno o más clonotipos correlacionados para el individuo.
- 55 En otra realización, la enfermedad es lupus eritematoso sistémico o esclerosis múltiple.

#### Breve descripción de los dibujos

- 60 Las nuevas características de la invención se exponen particularmente en las reivindicaciones adjuntas. Se obtendrá un mejor entendimiento de las características y ventajas de la presente invención por referencia a la siguiente descripción detallada que expone realizaciones ilustrativas, en las que se utilizan los principios de la invención, y los dibujos adjuntos de los que:

- 65 La **FIGURA 1** es un diagrama de flujo de una realización de un método de la invención proporcionada para determinar perfiles de clonotipo.
- La **FIGURA 2** muestra un esquema de PCR para amplificar genes de TCR $\beta$ .

La **FIGURA 3** ilustra un producto de PCR para secuenciar que se amplificó usando el esquema de la **FIGURA 2**.  
 Las **FIGURAS 4A y 4B** ilustran un esquema de PCR para amplificar secuencias de isotipo.  
 La **FIGURA 5** ilustra la reproducibilidad de amplificaciones múltiples.  
 La **FIGURA 6** ilustra que las amplificaciones múltiples tienen desviación de amplificación mínima.  
 5 La **FIGURA 7** ilustra electroforesis en gel de agarosa de amplificación múltiple de secuencias de IgH.  
 La **FIGURA 8A** muestra el  $\log_{10}$  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 **FIGURA 8B** 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 X) o Taq de Alta fidelidad (eje Y).  
 10 La **FIGURA 8C** 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 X) o Taq de Alta fidelidad (eje Y). Ilustra un diagrama de flujo para determinar el algoritmo por un estudio de descubrimiento.  
 La **FIGURA 9** ilustra una realización de un esquema para unir dos secuencias para formar un amplicón durante una reacción de amplificación. La información sobre la presencia de estas dos secuencias en la misma muestra (por ejemplo, célula) puede después conservarse incluso si se mezclan con un grupo de secuencias de otras muestras.  
 15 La **FIGURA 10** ilustra otra realización de un esquema de amplificación para unir dos secuencias.  
 La **FIGURA 11** ilustra otra realización de un esquema de amplificación para unir dos secuencias.  
 Las **FIGURAS 12A y 12B** ilustran un esquema para hacer múltiple la reacción que une dos secuencias por PCR.  
 20 Las **FIGURAS 13A - 13D** ilustran un esquema para unir tres secuencias entre sí.  
 La **FIGURA 14** ilustra un diagrama de flujo para descubrir clonotipos correlacionados usando un ensayo de calibración.  
 La **FIGURA 15** ilustra un diagrama de flujo para descubrir clonotipos correlacionados usando un estudio de población.  
 25 La **FIGURA 16** ilustra un diagrama de flujo para descubrir clonotipos correlacionados usando un estudio de población y un ensayo de calibración.  
 La **FIGURA 17** ilustra algoritmos que pueden predecir clonotipos correlacionados en una muestra.  
 La **FIGURA 18** ilustra un diagrama de flujo para generar un algoritmo de control para calcular la Carga Inmunitaria.  
 30 La **FIGURA 19** ilustra un diagrama de flujo para realizar un ensayo de control sin un ensayo de calibración.  
 La **FIGURA 20** ilustra un diagrama de flujo para realizar un ensayo de control usando un ensayo de calibración.

#### Descripción detallada de la invención

##### 35 Visión de conjunto

En general, la invención proporcionada incluye métodos para aplicar técnicas de secuenciación de ácido nucleico a la tarea de controlar el repertorio de células inmunitarias adaptativas para realizar el perfil del sistema inmunitario. Los perfiles del sistema inmunitario generados pueden usarse para el diagnóstico de enfermedades y trastornos, y 40 para el diagnóstico de estados de enfermedades y trastornos. Los métodos de realización de perfil inmunitario de la invención proporcionada pueden usarse para controlar enfermedades y trastornos y evaluar el tratamiento de enfermedades y trastornos. Estas enfermedades y trastornos a los que pueden aplicarse los métodos de la invención proporcionada incluyen enfermedad inmunitaria, incluyendo lupus eritematoso sistémico (SLE), esclerosis múltiple (MS), artritis reumatoide (RA) y espondilitis anquilosante. Los métodos de la invención proporcionada 45 pueden aplicarse al diagnóstico, control y tratamiento de rechazo de trasplantes y envejecimiento inmunitario. Además, los métodos de realización de perfil inmunitario de la invención proporcionada pueden usarse para diagnosticar, controlar y tratar otras enfermedades relacionadas con el sistema inmunitario, incluyendo cáncer y enfermedad infecciosa.

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

55 En un aspecto, se proporciona un método para determinar uno o más clonotipos correlacionados en un sujeto o individuo. En otro aspecto, se proporciona un método para desarrollar un algoritmo que puede predecir uno o más clonotipos correlacionados en cualquier muestra de un sujeto con una enfermedad. En otro aspecto, se proporciona un método para descubrir uno o más clonotipos correlacionados para un individuo usando un algoritmo que puede predecir uno o más clonotipos correlacionados en cualquier muestra de un sujeto. En otro aspecto, se proporciona un método para generar un algoritmo que calcula una puntuación de actividad de enfermedad. En otro aspecto, se proporciona un método para controlar la patología de un individuo.

**I. Métodos para determinar los perfiles de clonotipo****A. Visión de conjunto**

- 5 Los métodos de la presente invención pueden usarse para generar perfiles de secuencias de ADN recombinado, o clonotipos, en muestras de un sujeto.

En una realización, se proporciona un método para determinar un perfil de secuencias de ADN recombinado en linfocitos T y/o linfocitos B que incluye obtener una muestra de un sujeto que comprende linfocitos T y/o linfocitos B, aislar moléculas individuales de ADN genómico de dichas células, secuenciar las moléculas individuales aisladas de ADN genómico y determinar los niveles de secuencias diferentes de la muestra para generar dicho perfil de secuencias de ADN recombinado.

10 En otra realización, se proporciona un método para determinar un perfil de secuencias de ADN recombinado en linfocitos T y/o linfocitos B que incluye obtener una muestra de un sujeto que comprende linfocitos T y/o linfocitos 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 recombinado.

15 En otra realización, se proporciona un método para determinar un perfil de secuencias de ADN recombinado en linfocitos T y/o linfocitos B que incluye obtener una muestra de un sujeto que comprende linfocitos T y/o linfocitos B, amplificar 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 recombinado.

20 En otra realización, se proporciona un método para determinar un perfil de secuencias de ADN recombinado en linfocitos T y/o linfocitos B que incluye obtener una muestra de un sujeto que incluye linfocitos T y/o linfocitos B, amplificar ADN genómico de las células, aislar moléculas individuales del ADN amplificado, volver a amplificar las moléculas de ADN amplificado, secuenciar las moléculas de ADN reamplificadas, y determinar los niveles de diferentes secuencias de la muestra para generar el perfil de secuencias de ADN recombinado.

25 En otra realización, se proporciona un método para determinar un perfil de secuencias de ADN recombinado en linfocitos T y/o linfocitos B que incluye obtener una muestra de un sujeto que comprende linfocitos T y/o linfocitos B, aislar ARN de dicha muestra, transcribir de forma inversa el ARN de dichas células para formar ADNc, aislar moléculas individuales de dicho ADNc, volver a amplificar opcionalmente dicho ADNc, secuenciar dichas moléculas individuales aisladas de dicho ADNc o ADNc reamplificado, y determinar los niveles de secuencias diferentes de dicha muestra para generar dicho perfil de secuencias de ADN recombinado.

30 En otra realización, se proporciona un método para determinar un perfil de secuencias de ADN recombinado en linfocitos T y/o linfocitos B que incluye obtener una muestra de un sujeto que incluye linfocitos T y/o linfocitos 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 recombinado.

**B. Sujetos y muestras****1. Sujetos**

45 Los métodos de la invención proporcionada pueden usar muestras de sujetos e individuos (por ejemplo, pacientes). El sujeto puede ser un paciente, por ejemplo, un paciente con una enfermedad autoinmunitaria. El sujeto puede ser un paciente con una enfermedad infecciosa o cáncer. El sujeto puede ser un mamífero, por ejemplo, un ser humano. El sujeto puede ser hombre o mujer. El sujeto puede ser un bebé, un niño o un adulto.

**2. Muestras**

55 Las muestras usadas en los métodos de la invención proporcionada pueden incluir, por ejemplo, un fluido corporal de un sujeto, incluyendo líquido amniótico que rodea un feto, humor acuoso, bilis, sangre y plasma sanguíneo, cerumen (cera), fluido de Cowper o líquido preeyaculatorio, quilo, quimo, eyaculado femenino, líquido intersticial, linfa, menstruación, leche materna, mucus (incluyendo moco y flema), líquido pleural, pus, saliva, sebo (aceite cutáneo), semen, suero, sudor, lágrimas, orina, lubricación vaginal, vómito, agua, heces, líquidos corporales internos, incluyendo líquido cefalorraquídeo que rodea al cerebro y la médula espinal, líquido sinovial que rodea las articulaciones óseas, líquido intracelular que es líquido dentro de las células, y humor vítreo, los líquidos del globo ocular. En una realización, la muestra es una muestra sanguínea. La muestra sanguínea puede ser de aproximadamente 0,1, 0,2, 0,3, 0,4, 0,5, 0,6, 0,7, 0,8, 0,9, 1,0, 1,5, 2,0, 2,5, 3,0, 3,5, 4,0, 4,5 o 5,0 ml. La muestra puede ser Líquido Cefalorraquídeo (LCR) cuando el sujeto tiene esclerosis múltiple, líquido sinovial cuando el sujeto tiene artritis reumatoide y biopsia cutánea (o de otro órgano) cuando el sujeto tiene lupus sistémico. En una realización, el clonotipo puede identificarse a partir del líquido corporal/tejido disponible que más probablemente

refleje la patología seguido de control posterior de los niveles de los clonotipos a partir de un líquido corporal diferente, por ejemplo, sangre.

Las muestras pueden analizarse en un momento en el que la enfermedad está inactiva.

- 5 La muestra puede obtenerse por un asistente sanitario, por ejemplo, un médico, asistente médico, enfermero, veterinario, dermatólogo, reumatólogo, dentista, paramédico o cirujano. La muestra puede obtenerse por un técnico de investigación. Puede obtenerse más de una muestra de un sujeto.
- 10 La muestra puede ser una biopsia, por ejemplo, una biopsia cutánea. La biopsia puede ser de, por ejemplo, el cerebro, hígado, pulmón, corazón, colon, riñón o médula ósea. Puede usarse cualquier técnica de biopsia usada por los expertos en la materia 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 realiza un corte más pequeño que en una biopsia abierta. La biopsia puede ser una biopsia central o de incisión, en la que se retira parte del tejido. La biopsia puede ser una biopsia de escisión, en la que se intenta retirar una lesión completa. La biopsia puede ser una biopsia de aspiración de aguja fina, en la que se retira una muestra de tejido o fluido con una aguja.
- 15 La muestra puede incluir células inmunitarias. Las células inmunitarias pueden incluir linfocitos T y/o linfocitos B. Los linfocitos T incluyen, por ejemplo, células que expresan receptores de linfocitos T. Los linfocitos T incluyen linfocitos T auxiliares (linfocitos T efectores o linfocitos Tb), linfocitos T citotóxicos (CTL), linfocitos T de memoria y linfocitos T reguladores. La muestra puede incluir una única célula en algunas aplicaciones (por ejemplo, un ensayo de calibración para definir linfocitos 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 de linfocitos T.
- 20 La muestra puede incluir linfocitos B incluyen, por ejemplo linfocitos B del plasma, linfocitos B de memoria, linfocitos B1, linfocitos B2, linfocitos B de zona marginal y linfocitos B foliculares. Los linfocitos B pueden expresar inmunoglobulinas (anticuerpos, receptores de linfocitos B). La muestra puede incluir una única célula en algunas aplicaciones (por ejemplo, un ensayo de calibración para definir linfocitos 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 de linfocitos B.
- 25 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. 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 tanto como 10 millones de células o más lo que se traduce en un intervalo de ADN de 6 pg-60 µg, y ARN de aproximadamente 1 pg-10 µg.
- 30 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. 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 tanto como 10 millones de células o más lo que se traduce en un intervalo de ADN de 6 pg-60 µg, y ARN de aproximadamente 1 pg-10 µg.
- 35 40 **C. Medios para aislar, amplificar y volver a amplificar ácido nucleico**
- 1. Características de los genes de TCR y BCR**
- 45 Puesto que las recombinaciones de identificación están presentes en el ADN de cada célula inmunitaria adaptativa individual así como sus transcritos de ARN asociados, pueden secuenciarse ARN o ADN en los métodos de la invención proporcionada. Una secuencia recombinada de un linfocito T o linfocito B también puede denominarse clonotipo. El ADN o ARN puede corresponder a secuencias de genes del receptor de linfocitos T (TCR) o genes de inmunoglobulina (Ig) que codifican anticuerpos. Por ejemplo, el ADN y ARN puede corresponder a secuencias que codifican cadenas  $\alpha$ ,  $\beta$ ,  $\gamma$  o  $\delta$  de un TCR. En una mayoría de linfocitos T, el TCR es un heterodímero que consiste en una cadena  $\alpha$  y cadena  $\beta$ . La cadena de TCR $\alpha$  se genera por recombinación VJ, y después se genera receptor de cadena  $\beta$  por recombinación de V(D)J. Para la cadena de TCR $\beta$ , en seres humanos hay 48 segmentos V, 2 segmentos D y 13 segmentos J. Pueden suprimirse varias bases y añadirse otras (denominadas nucleótidos N y P) en cada uno de los dos puntos de unión. En una minoría de linfocitos T, los TCR consisten en cadenas  $\gamma$  y  $\delta$  delta. La cadena TCR  $\gamma$  se genera por recombinación de VJ, y la cadena de TCR  $\delta$  se genera por recombinación de V(D)J (Kenneth Murphy, Paul Travers, y Mark Walport, Janeway's Immunology 7<sup>a</sup> edición, Garland Science, 2007).
- 50 55 El ADN y ARN analizados en los métodos de la invención proporcionada pueden corresponder a secuencias que codifican inmunoglobulinas de cadena pesada (IgH) con regiones constantes ( $\alpha$ ,  $\delta$ ,  $\varepsilon$ ,  $\gamma$  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 de una región constante (C) y una variable. Para la cadena pesada, la región variable está compuesta de un segmento variable (V), de diversidad (D) y de unión (J). En el genoma están presentes varias secuencias distintas que codifican cada tipo de estos segmentos. Se produce un acontecimiento de recombinación de VDJ específico durante el desarrollo de un linfocito B, marcando esa célula para generar una cadena pesada específica. Se genera diversidad en la cadena ligera de una manera similar excepto que no hay región D de modo que solo haya recombinación de VJ. Con frecuencia se produce mutación
- 60 65

- 5 somática cerca del sitio de la recombinación, provocando la adición o delección de varios nucleótidos, aumentando adicionalmente la diversidad de cadenas pesadas y ligeras generadas por linfocitos B. La posible diversidad de los anticuerpos generados por linfocito B es entonces el producto de las diferentes cadenas pesadas y ligeras. Las regiones variables de las cadenas pesadas y ligeras contribuyen a formar la región o sitio de reconocimiento (o unión) de antígenos. A esta diversidad se añade un proceso de hipermutación somática que puede suceder después de montarse una respuesta específica contra algún epítopo. En este proceso aparecen mutaciones en los linfocitos B que son capaces de reconocer el epítopo específico lo que conduce a mayor diversidad en anticuerpos que pueden ser capaces de unirse con el epítopo específico con más fuerza. Todos estos factores contribuyen a una gran diversidad de anticuerpos generados por los linfocitos B. Pueden generarse muchos miles de millones y quizás 10 más de un billón de anticuerpos distintos. La premisa básica para generar diversidad de linfocitos T es similar a la de generar anticuerpos por linfocitos B. Un elemento de activación de linfocitos B y linfocitos B es su unión con epítopos ajenos. La activación de una célula específica conduce a la producción de más del mismo tipo de células lo que conduce a una expansión clonal.
- 15 10 Las regiones determinantes de complementariedad (CDR), o regiones hipervariables, son secuencias en los dominios variables de receptores de antígeno (por ejemplo, receptor de linfocitos T e inmunoglobulina) que pueden complementar un antígeno. La cadena de cada receptor de antígeno contiene tres CDR (CDR1, CDR2 y CDR3). Los dos polipéptidos que componen linfocitos T ( $\alpha$  y  $\beta$ ) e inmunoglobulina (IgH e IgK o IgL) contribuyen a la formación de las tres CDR.
- 20 15 La parte de CDR1 y CDR2 que se codifica por TCF $\beta$  queda dentro de uno de los 47 segmentos V funcionales. La mayor parte de la diversidad de las CDR se encuentra en CDR3, generándose la diversidad por acontecimientos de recombinación somática durante el desarrollo de linfocitos T.
- 25 20 Está presente una gran diversidad de BCR entre y dentro de individuos. El BCR está compuesto de dos genes IgH e IgK (o IgL) que codifican cadenas pesadas y ligeras de anticuerpo. Tres secuencias de Región Determinante de Complementariedad (CDR) que se unen a antígenos y moléculas de MHC tienen la mayor diversidad en IgH e IgK (o IgL). La parte de CDR1 y CDR2 que codifica IgH queda dentro de uno de 44 segmentos V funcionales. La mayor parte de la diversidad en linfocitos B sin tratamiento previo surge en la generación de CDR3 mediante 30 25 acontecimientos de recombinación somática durante el desarrollo de linfocitos B. La recombinación puede generar una molécula con uno de los segmentos V, D y J. En seres humanos hay 44 segmentos V, 27 D y 6 J; por lo tanto, existe una posibilidad teórica de más de 7.000 combinaciones. En una pequeña fracción de BCR (~5 %) se encuentran dos segmentos D. Además, pueden suprimirse varias bases y añadirse otras (denominados nucleótidos N y P) en cada uno de los dos puntos de unión generando un gran grado de diversidad. Después de la activación de 35 30 linfocitos B se produce un proceso de maduración de afinidad mediante hipermutación somática. En este proceso células descendientes de los linfocitos B activados acumulan distintas mutaciones somáticas en todo el gen con mayor concentración de mutación en las regiones CDR lo que conduce a generar anticuerpos con mayor afinidad por los antígenos. Por lo tanto pueden utilizarse múltiples cebadores para amplificar secuencias después de 40 35 hipermutación somática. Además de hipermutación somática, los linfocitos B activados experimentan el proceso de intercambio de isotipo. Los anticuerpos con los mismos segmentos variables pueden tener diferentes formas (isotipos) dependiendo del segmento constante. Mientras que todos los linfocitos B sin tratamiento previo expresan IgM (o IgD), los linfocitos B activados expresan principalmente IgG pero también IgM, IgA e IgE. Este cambio de expresión de IgM (y/o IgD) a IgG, IgA o IgE sucede mediante un acontecimiento de recombinación provocando que 45 40 una célula se especialice en la producción de un isotipo específico. Hay un segmento para cada IgM, IgD e IgE, dos segmentos para IgA y cuatro segmentos para IgG.

## **2. Reacciones de amplificación**

- 50 50 Puede usarse reacción en cadena de la polimerasa (PCR) para amplificar las regiones relevantes de una colección de células. Puede usarse Amplificación Mediada por Transcripción (TMA) para producir amplicones de ARN a partir de un ácido nucleico diana. El ácido nucleico de cada célula puede analizarse por separado, y cada célula portará su identificación única.
- 55 55 Pueden amplificarse secuencias de TCR $\beta$  o inmunoglobulina 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 pueden hibridar con uno o más segmentos V (**FIGURA 2** y **FIGURA 4**). El número de cebadores que hibridan con segmentos V en una reacción múltiple pueden 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 o 80. El número de cebadores que hibrida con segmentos V en una reacción múltiple puede ser, por ejemplo, 10-60, 20-50, 30-50, 40-50, 20-40, 30-40 o 35-40. Los cebadores pueden hibridar con diferentes segmentos V. Para genes 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 o 5 cebadores por segmento V. El número de cebadores que hibridan con segmentos C en una reacción múltiple pueden incluir, por ejemplo, al menos 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 o 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 o 3-6. La amplificación de genes de TCR o inmunoglobulina puede suceder como se describe en el Ejemplo

3 y/o Ejemplo 4.

- La región para amplificar puede incluir la secuencia clonal completa o un subconjunto de la secuencia clonal, incluyendo el punto de unión V-D, punto de unión D-J de una inmunoglobulina o gen del receptor de linfocitos T, la 5 región variable completa de un gen de inmunoglobulina o receptor de linfocitos T, la región de reconocimiento de antígenos o una CDR, por ejemplo, región determinante de 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 10 diferentes cebadores pueden introducir secuencia no original presente en la secuencia génica inmunitaria. Por ejemplo, el procedimiento de amplificación puede añadir uno o más marcadores al extremo 5' y/o 3' de la secuencia de TCR o inmunoglobulina amplificada (FIGURA 3). El marcador puede ser una secuencia que facilite la secuenciación posterior del ADN amplificado. El marcador puede ser una secuencia que facilite la unión de la 15 secuencia amplificada con un soporte sólido.
- Otros métodos para amplificación pueden no emplear ningún cebador en la región V. En su lugar, puede usarse un 20 cebador específico del segmento C y un cebador genérico puede ponerse en el otro lado (5'). El cebador genérico puede añadirse en la síntesis de ADNc mediante diferentes métodos incluyendo los métodos bien descritos de cambio de cadena. De forma similar, el cebador genérico puede añadirse después de la realización de ADNc mediante 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 25 digital (dPCR), PCR de emulsión digital (dePCR), PCR clonal, PCR de polimorfismos de longitud de fragmento amplificado (AFLP PCR), PCR específica de alelos, PCR de ensamblaje, PCR asimétrica, (en la que se un gran exceso de cebadores para una cadena dada), PCR de colonias, amplificación dependiente de helicasa (HDA), PCR de Arranque en Caliente, PCR inversa (IPCR), PCR *in situ*, PCR larga (extensión de ADN mayor de aproximadamente 5 kilobases), PCR múltiple, PCR anidada (usa más de un par de cebadores), PCR de célula individual, PCR touchdown, PCR isotérmica mediada por bucle (LAMP) y amplificación basada en secuencia de 30 ácidos nucleicos (NASBA). Otras esquemas de amplificación incluyen: Reacción en Cadena de la Ligasa, Amplificación de ADN Ramificado, Amplificación de Círculo Rodante, Amplificación de Círculo a Círculo, amplificación de SPIA, Amplificación de amplificación diana por Captura y Ligación (TACL) y amplificación RACE.
- La información en el ARN en una muestra puede convertirse a ADNc usando transcripción inversa. Pueden usarse 35 cebadores poliA, cebadores aleatorios y/o cebadores específicos de genes en reacciones de transcripción inversa.
- Después de amplificación de ADN del genoma (o amplificación del ácido nucleico en forma de ADNc por transcripción inversa de ARN), las moléculas de ácido nucleico individuales puede aislarse, opcionalmente volver a amplificarse, y después secuenciarse individualmente.
- 40 Las polimerasas que pueden usarse para amplificación en los métodos de la invención proporcionada incluyen, por ejemplo, Taq polimerasa, polimerasa AccuPrime o Pfu. La elección de polimerasa para usar puede basarse en si se prefiere fidelidad o eficacia.
- 45 En una realización, se aíslan células individuales en una muestra. Pueden unirse entre sí dos o más secuencias de cada célula aislada. Por ejemplo, pueden unirse secuencias de genes de TCR $\alpha$  y TCR $\beta$  o genes de IgH e IgK de una célula individual, por ejemplo por un esquema de amplificación (FIGURAS 9-13) o un esquema de ligación. Las secuencias de TCR $\alpha$  y TCR $\beta$  o IgH e IgK unidas para células aisladas pueden opcionalmente volver a amplificarse. Los 50 productos de amplificación unidos pueden opcionalmente volver a agruparse después de amplificación.
- 3. Medios para aislar ácidos nucleicos individuales**
- Los métodos para aislamiento de ácidos nucleicos a partir de un grupo incluyen subclonar ácido nucleico en 55 vectores de ADN y transformar bacterias (clonación bacteriana), 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 solución dentro de micelas (tal como puede conseguirse usando emulsiones en aceite con o sin inmovilización de las moléculas en una superficie sólida tal como perlas), o usar cámaras de microrreacción, por ejemplo, en microplacas microfluídicas o nanofluídicas. Puede usarse dilución para asegurar que esté presente de media una única molécula en un volumen dado, región espacial, perla o cámara de reacción.
- 60 Puede usarse PCR en tiempo real, tinción con pico green, electroforesis nanofluídica (por ejemplo LabChip) o mediciones de absorción de UV en una etapa inicial para considerar la cantidad funcional de material amplificable.
- Los métodos para volver a amplificar ácidos nucleicos incluyen cultivo bacteriano de colonias aisladas transformadas 65 con ácido nucleico, amplificación en un portaobjetos (por ejemplo, colonias de PCR (polonias)) y amplificación en una perla. Puede usarse el mismo método para amplificar y volver a amplificar el ácido nucleico o puede usarse un

método diferente para amplificar y volver a amplificar el ácido nucleico.

En ciertas realizaciones las etapas de subclonación incluyen una etapa en la que se une un cebador común con el ADN o ARN mediante una etapa de amplificación o ligación. Este cebador se usa después para amplificar los clones y como una secuencia de reconocimiento para hibridación de un cebador para secuenciación (**FIGURA 2**).

En otras realizaciones, se analizan ácidos nucleicos a partir de un subconjunto de células. Puede emplearse un método para separar células, por ejemplo usando un marcador de superficie celular. Por ejemplo, las células pueden aislarse por citometría de flujo de separación de células, separación por flujo, separación de células activadas por fluorescencia (FACS), separación basada en perlas tal como separación de células magnéticas (MACS; por ejemplo, usando partículas magnéticas recubiertas con anticuerpo), preparación basada en tamaño (por ejemplo, un tamiz, una serie de obstáculos o un filtro), separación en un dispositivo microfluídico, separación basada en anticuerpos, sedimentación, adsorción de afinidad, extracción de afinidad o centrifugación en gradiente de densidad. Las células pueden purificarse por microdissección de captura con láser. La separación puede basarse en el tamaño celular, morfología o marcadores intracelulares o extracelulares. Se describen métodos para aislar o separar células tumorales, por ejemplo, en Nagrath S. et al. (2007) *Nature* 450: 1235-1239; Patentes de Estados Unidos Nº 6008002, 7232653 y 7332288; Publicación de PCT Nº W02008157220A1; y Solicitud de Patente de Estados Unidos Nº US20080138805A1 y US20090186065; y Rosenberg R. et al. (2002) *Cytometry* 49: 150-158.

El subconjunto de células puede ser un subconjunto de linfocitos T y/o linfocitos B. El subconjunto de linfocitos T puede ser células CD4+, CD8+ o CD27<sup>alto</sup>.

La separación de células activadas por fluorescencia (FACS) usa dispersión de la luz y características fluorescentes para separar células. Puede transmitirse una propiedad fluorescente en una célula usando, por ejemplo, sondas de ácido nucleico o anticuerpos conjugados con un colorante fluorescente. Una suspensión celular puede formar una corriente de líquido fluido. 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 coloca una carga en un anillo de carga eléctrica antes de la medición de la intensidad de fluorescencia y la carga opuesta se porta en la gota a medida que se separa de la corriente. Las gotas cargadas pasan a través de dos placas de deflección de alta tensión que separan gotas en diferentes recipientes basándose en su carga. La carga puede aplicarse directamente a la corriente y la gota que se separa conserva la carga del mismo signo que la corriente. La corriente se devuelve después a neutro después de separarse la gota.

Puede usarse inmunofluorescencia directa o indirecta en FACS. En inmunofluorescencia directa, un anticuerpo se conjuga directamente con un colorante fluorescente. En inmunofluorescencia indirecta, el anticuerpo primario no está marcado, y un anticuerpo secundario está conjugado con un colorante fluorescente. En una realización, pueden aislarse células individuales de una muestra. Puede unirse entre sí información de secuencia de dos genes más en una célula. Por ejemplo, una muestra puede ser de un paciente con una enfermedad autoinmunitaria, y puede unirse físicamente información de secuencia de genes de TCR $\alpha$  y TCR $\beta$  de células aisladas espacialmente de la muestra, por ejemplo, por un esquema de amplificación o un esquema de ligación. Las secuencias de TCR $\alpha$  y TCR $\beta$  unidas pueden opcionalmente amplificarse y/o agruparse con secuencias unidas de otras células. Las secuencias unidas pueden como alternativa ser para IgH e IgK o para IgH e IgL.

### **C. Técnicas de secuenciación**

Puede usarse cualquier técnica para secuenciar ácido nucleico conocida por los expertos en la materia en los métodos de la invención proporcionada. Las técnicas de secuenciación de ADN incluyen reacciones de secuenciación de didesoxi clásicas (métodos de Sanger) usando terminadores marcados o cebadores y separación en gel en plancha o capilar, secuenciación por síntesis usando nucleótidos marcados en el extremo de forma reversible, pirosecuenciación, secuenciación 454, hibridación específica de alelo con una biblioteca de sondas oligonucleotídicas marcadas, secuenciación por síntesis usando hibridación específica de alelo con una biblioteca de clones marcados que se sigue de ligación, control en tiempo real de la incorporación de nucleótidos marcados durante una etapa de polimerización, secuenciación de polonia, y secuenciación de SOLiD. La secuenciación de las moléculas separadas se ha demostrado más recientemente por reacciones de extensión secuenciales o individuales usando polimerasas o ligasas así como por hibridaciones diferenciales individuales 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. Estos enfoques de secuenciación pueden por lo tanto usarse para estudiar el repertorio del receptor de linfocitos T (TCR) y/o receptor de linfocitos B (BCR).

La técnica de secuenciación usada en los métodos de la invención proporcionada puede generar al menos 1000 lecturas por ciclo, al menos 10.000 lecturas por ciclo, al menos 100.000 lecturas por ciclo, al menos 500.000 lecturas por ciclo o al menos 1.000.000 de lecturas por ciclo.

La técnica de secuenciación usada en los métodos de la invención proporcionada pueden 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, 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.

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

### **1. Secuenciación de molécula individual verdadera**

10 Una técnica de secuenciación que puede usarse en los métodos de la invención proporcionada incluye, por ejemplo Secuenciación de Molécula Individual Verdadera Helicos (tSMS) (Harris T. D. et al. (2008) *Science* 320: 106-109). En la técnica de tSMS, se escinde una muestra de ADN en cadenas de aproximadamente 100 a 200 nucleótidos y se añade una secuencia poli A al extremo 3' de cada cadena de ADN. Cada cadena se marca por la adición de un nucleótido adenosina marcado con fluorescencia. Las cadenas de ADN se hibridan después con una celda de flujo, que contiene millones de sitios de captura oligo T que se inmovilizan en la superficie de la celda de flujo. Los moldes pueden estar a una densidad de aproximadamente 100 millones de moldes/cm<sup>2</sup>. La celda de flujo se carga después en un instrumento, por ejemplo, secuenciador HeliScope™, y un láser ilumina la superficie de la celda de flujo, revelando la posición de cada molde. Una cámara CCD puede mapear la posición de los moldes en la superficie de la celda de flujo. El marcador fluorescente del molde se escinde después y se retira por lavado. La reacción de secuenciación comienza introduciendo ADN polimerasa y un nucleótido marcado con fluorescencia. El ácido nucleico oligo T actúa como un cebador. La polimerasa incorpora los nucleótidos marcados al cebador de una manera dirigida por molde. Se retira la polimerasa y los nucleótidos no incorporados. Los moldes que han dirigido la incorporación del nucleótido marcado con fluorescencia se detectan capturando imágenes de la superficie de la celda de flujo. Después de capturar las imágenes, una etapa de escisión retira el marcador fluorescente, y el procedimiento se repite con otros nucleótidos marcados con fluorescencia hasta que se consigue la longitud de lectura deseada. Se recoge información de secuencia con cada etapa de adición de nucleótidos.

### **2. Secuenciación 454**

30 Otro ejemplo de una técnica de secuenciación de ADN que puede usarse en los métodos de la invención proporcionada es secuenciación 454 (Roche) (Margulies, M et al. 2005, *Nature*, 437, 376-380). La secuenciación 454 implica dos etapas. En la primera etapa, el ADN se corta en fragmentos de aproximadamente 300-800 pares de bases, y los extremos de los fragmentos se hacen romos. Después se ligan adaptadores oligonucleotídicos a los extremos de los fragmentos. Los adaptadores actúan como cebadores para amplificación y secuenciación de los fragmentos. Los fragmentos pueden unirse a perlas de captura de ADN, por ejemplo, perlas recubiertas con estreptavidina usando, por ejemplo, adaptador B, que contiene un marcador de biotina 5'. Los fragmentos unidos con las perlas se amplifican por PCR dentro de gotas de una emulsión de aceite-agua. El resultado es múltiples copias de fragmentos de ADN amplificados de forma clonal en cada perla. En la segunda etapa, las perlas se capturan en pocillos (de tamaño de pico litros). Se realiza pirosecuenciación en cada fragmento de ADN en paralelo. La adición de uno o más nucleótidos genera una señal luminosa que se registra por una cámara CCD en un instrumento de secuenciación. La fuerza de la señal es proporcional al número de nucleótidos incorporado.

45 La pirosecuenciación hace uso de pirofosfato (PPi) que se libera tras la adición de nucleótidos. PPi se convierte a ATP por ATP sulfurylasa en presencia de adenosina 5' fosfatosulfato. La luciferasa usa ATP para convertir luciferina en oxiluciferina, y esta reacción genera luz que se detecta y se analiza.

### **3. Secuenciación SOLiD**

50 Otro ejemplo de una técnica de secuenciación de ADN que puede usarse en los métodos de la invención proporcionada es tecnología SOLiD (Applied Biosystems). En secuenciación SOLiD, el ADN genómico se corta en fragmentos, y se unen adaptadores a los extremos 5' y 3' de los fragmentos para generar una biblioteca de fragmentos. Como alternativa, pueden introducirse adaptadores internos ligando adaptadores a los extremos 5' y 3' de los fragmentos, circularizando los fragmentos, dirigiendo el fragmento circularizado para generar un adaptador interno, y uniendo adaptadores a los extremos 5' y 3' de los fragmentos resultantes para generar una biblioteca de compañeros emparejados. A continuación, se preparan poblaciones de perlas clonales en microrreactores que contienen perlas, cebadores, molde y componentes de PCR. Despues de la PCR, los moldes se desnaturizan y las perlas se enriquecen para separar las perlas con moldes extendidos. Los moldes en las perlas seleccionadas se someten a una modificación 3' que permite el enlace con un portaobjetos de vidrio.

60 La secuencia puede determinarse por hibridación secuencial y ligación de oligonucleótidos parcialmente aleatorios con una base determinada central (o par de bases) que se identifica por un fluoróforo específico. Despues de registrarse un color, el oligonucleótido ligado se escinde y se retira y despues se repite el procedimiento.

#### **4. Secuenciación SOLEXA**

Otro ejemplo de una tecnología de secuenciación que puede usarse en los métodos de la invención proporcionada es secuenciación SOLEXA (Illumina). La secuenciación SOLEXA se basa en la amplificación de ADN en una superficie sólida usando PCR regresiva y cebadores anclados. El ADN genómico se fragmenta, y los adaptadores se añaden a los extremos 5' y 3' de los fragmentos. Los fragmentos de ADN que se unen a la superficie de canales de la celda de flujo se extienden y se amplifica el puente. Los fragmentos se hacen bicatenarios, y se desnaturizan las moléculas bicatenarias. Múltiples ciclos de la amplificación de fase sólida seguida de desnaturización pueden crear varios millones de grupos de aproximadamente 1.000 copias de moléculas de ADN monocatenarias del mismo molde en cada canal de la celda de flujo. Se usan cebadores, ADN polimerasa y cuatro nucleótidos de terminación reversible, marcados con fluoróforo para realizar secuenciación secuencial. Después de la incorporación de nucleótidos, se usa un láser para excitar los fluoróforos, y se captura una imagen y se registra la identidad de la primera base. Los terminadores 3' y fluoróforos de cada base incorporada se retiran y se repiten las etapas de incorporación, detección e identificación.

15

#### **5. Secuenciación de SMRT**

Otro ejemplo de una tecnología de secuenciación que puede usarse en los métodos de la invención proporcionada incluye la tecnología de molécula individual, en tiempo real (SMRT™) de Pacific Biosciences. En SMRT, cada una de las cuatro bases de ADN se une a uno de cuatro colorantes fluorescentes diferentes. Estos colorantes están fosfoligados. Se inmoviliza una única ADN polimerasa con una molécula individual de ADN monocatenario molde en el fondo de una guía de ondas en modo cero (ZMW). Una ZMW es una estructura de confinamiento que permite la observación de la incorporación de un único nucleótido por ADN polimerasa frente al fondo de nucleótidos fluorescentes que rápidamente se difunden dentro y fuera de la ZMW (en microsegundos). Se tardan varios milisegundos en incorporar un nucleótido en una cadena creciente. Durante ese tiempo, el marcador fluorescente se excita y produce una señal fluorescente y el marcador fluorescente se escinde. La detección de la fluorescencia correspondiente del colorante indica que base se incorporó. Se repite el procedimiento.

20

#### **6. Secuenciación de nanoporos**

25

Otro ejemplo de una técnica de secuenciación que puede usarse en los métodos de la invención proporcionada es secuenciación de nanoporos (Soni GV y Meller A. (2007) Clin Chem 53: 1996-2001). Un nanoporo es un orificio pequeño, del orden de un nanómetro de diámetro. La inmersión de un nanoporo en un líquido conductor y aplicación de un potencial a través de él da como resultado una corriente eléctrica suave debido a la conducción de iones a través del nanoporo. La cantidad de corriente que fluye es sensible al tamaño del nanoporo. A medida que una molécula de ADN pasa a través de un nanoporo, cada nucleótido en la molécula de ADN obstruye el nanoporo en un grado diferente. Por lo tanto, el cambio de la corriente que pasa a través del nanoporo a medida que la molécula de ADN pasa a través del nanoporo representa una lectura de la secuencia de ADN.

30

#### **7. Secuenciación de serie de transistores de efecto de campo sensible a productos químicos**

Otro ejemplo de una técnica de secuenciación que puede usarse en los métodos de la invención proporcionada implica usar una serie de transistores de efecto de campo sensible a productos químicos (chemFET) para secuenciar ADN (por ejemplo, como se describe en la Publicación de Solicitud de Patente de Estados Unidos Nº 20090026082). En un ejemplo de la técnica, pueden colocarse moléculas de ADN en cámaras de reacción, y las moléculas molde pueden hibridarse con un cebador de secuenciación unido a una polimerasa. La incorporación de uno o más trifosfatos en una nueva cadena de ácido nucleico en el extremo 3' del cebador de secuenciación puede detectarse por un cambio de la corriente por un chemFET. Una serie puede tener múltiples sensores chemFET. En otro ejemplo, puede unirse ácidos nucleicos individuales a perlas, y los ácidos nucleicos pueden amplificarse en la perla, y las perlas individuales pueden transferirse a cámaras de reacción individuales en una serie de chemFET, teniendo cada cámara un sensor chemFET, y pueden secuenciarse los ácidos nucleicos.

35

#### **8. Secuenciación con un microscopio electrónico**

40

Otro ejemplo de una técnica de secuenciación que puede usarse en los métodos de la invención proporcionada implica usar un microscopio electrónico (Moudrianakis E. N. y Beer M. Proc Natl Acad Sci USA. 1965 Mar; 53: 564-71). En un ejemplo de la técnica, se marcan moléculas de ADN individuales usando perlas metálicas que son distinguibles usando un microscopio electrónico. Estas moléculas se estiran después en una superficie plana y se capturan imágenes usando un microscopio electrónico para medir las secuencias.

45

Una cualquiera de las técnicas de secuenciación descritas en el presente documento puede usarse en los métodos de la invención proporcionada.

**D. Métodos para secuenciar el repertorio de TCR y BCR**

- Pueden leerse secuencias que se originan a partir de una molécula individual o que se origina a partir de amplificaciones de una molécula individual. Pueden realizarse millones de amplificaciones independientes de moléculas individuales en paralelo bien en una superficie sólida o bien en compartimentos minúsculos en emulsión de agua/aceite. La muestra de ADN para secuenciar puede diluirse y/o dispersarse lo suficiente para obtener una molécula en cada compartimento. Esta dilución puede seguirse de amplificación de ADN para generar copias de las secuencias de ADN originales y crear "grupos" de moléculas que tienen todas la misma secuencia. Estos grupos pueden después secuenciarse. Pueden generarse muchos millones de lecturas en un ciclo. La secuencia puede generarse comenzando en el extremo 5' de una cadena dada de una secuencia amplificada y/o la secuencia puede generarse partiendo del extremo 5' de la secuencia complementaria. En una realización preferida, se genera una secuencia a partir de cadenas, es decir lecturas de extremos emparejados.
- La prevalencia de una secuencia particular en la secuencia de ADN original puede medirse después contando cuántos grupos portan esa secuencia. Secuencias más prevalentes en la muestra original conducen a más compartimentos y más grupos que contienen las secuencias específicas.
- Pueden usarse métodos en los esquemas de amplificación para asegurar que la frecuencia de las secuencias de ADN medida coincide con la frecuencia de la secuencia de ADN en la muestra original. Los métodos pueden incluir asegurar que la concentración de cebadores de PCR sea lo suficientemente alta para conducir cada reacción de hibridación hasta saturación en cada ciclo, ajustar concentraciones de cebadores individuales para minimizar la amplificación diferencial de diferentes secuencias, etc.
- Pueden usarse algoritmos para determinar qué secuencias generadas por el secuenciador se originan de la secuencia de ADN. Las secuencias medidas individualmente (lecturas) pueden compensarse entre sí, contener errores introducidos por amplificación y/o por secuenciación. Puede usarse un algoritmo para combinar lecturas entre sí para determinar con más precisión la frecuencia de una secuencia de ADN en el material de partida.
- Puede obtenerse un millón de lecturas de secuenciación o más para IgH y/o TCR $\beta$  originalmente amplificado a partir de una muestra sanguínea que comprende ADN o ARN. El número de lecturas para una secuencia de IgH o TCR $\beta$  específica está relacionado con la frecuencia del clonotipo específico en la muestra sanguínea. Por lo tanto, puede determinarse la cantidad de cada uno de los clonotipos. Si se conocen los clonotipos patógenos para un paciente particular, su nivel puede determinarse de forma precisa mediante este enfoque de secuenciación.
- En ciertas realizaciones de la invención proporcionada, se extrae una colección de moléculas de ADN incluyendo una representación del ADN genómico o ARN transcripto inverso de las regiones de TCR y BCR de células inmunitarias de uno o más sujetos y opcionalmente se amplifica de tal modo que cada molécula pueda secuenciarse usando una o más de las técnicas de secuenciación descritas anteriormente para poder detectar la presencia y frecuencia de las secuencias en un sujeto.
- Pueden secuenciarse diferentes regiones de genes de inmunoglobulina o receptores de linfocitos T. En algunas realizaciones, la secuencia completa de las regiones variables puede secuenciarse para identificar y cuantificar un clonotipo.
- Puede secuenciarse un subconjunto único de las secuencias clonales completas. En algunas realizaciones, se secuencian nucleótidos que comprenden los puntos de unión VD y DJ para identificar de forma única y cuantificar 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 antígenos o la región determinante de complementariedad 3 (CDR3). Puede amplificarse un fragmento que contiene la CDR3 completa o la región variable completa para permitir la secuenciación de las partes que comprenden CDR3 de los segmentos V, D y J.
- Uno o más marcadores en productos amplificados puede usarse para secuenciar genes del receptor de linfocitos T o inmunoglobulina. Pueden usarse uno o más cebadores que hibridan con los marcadores en las reacciones de secuenciación. Pueden secuenciarse diferentes secciones de una molécula amplificada en reacciones separadas, y los resultados de secuenciación pueden reconstituirse para generar una secuencia parcial o una completa de la molécula.
- En una realización, solamente se amplifica y secuencia la CDR3. La amplificación y secuenciación de la CDR3 puede conseguirse usando cebadores específicos para 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 lo que conduce a la amplificación del repertorio completo de secuencias. Este repertorio de secuencias puede después mezclarse y someterse a separación, con o sin amplificación, y secuenciarse usando cualquiera de las técnicas de secuenciación descritas. Cuando la amplificación con los diversos cebadores V se realiza en tubos separados, el número de moléculas que porta los diferentes segmentos V puede "normalizarse" debido a la saturación de PCR. Por ejemplo, si un segmento V particular tuviera una o varias expansiones clonales que condujeran a su representación más que otros segmentos esta información

podría suprimirse o reducirse puesto que la reacción de PCR para cada segmento puede conducirse a saturación o cerca de ella. Puede usarse PCR en tiempo real para cuantificar cuánto de cada segmento V está presente. Puede secuenciarse la CDR completa, o puede secuenciarse un subconjunto de la secuencia de CDR3.

- 5 En una realización, solamente se analiza un subconjunto de clonotipos. Esto puede conseguirse amplificando con un cebador específico del subconjunto de clonotipos, por ejemplo, un cebador que es específico del segmento V. Pueden identificarse clonotipos únicos secuenciando con lecturas contiguas largas que proporcionan conectividad completa. En algunas realizaciones, cuando están presentes varias secuencias de interés, una longitud de lectura corta a través de solamente uno de los puntos de unión puede generar marcadores degradados que no son únicos de un clonotipo específico sino que se comparten entre múltiples clonotipos. Por ejemplo, la secuenciación a través del punto de unión V/J puede agrupar todas las secuencias con el mismo 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 están conectadas a diferentes segmentos D, por ejemplo.
- 10 15 Puede realizarse el mismo análisis cuando solamente está presente V y D (por ejemplo, la cadena ligera de un anticuerpo o la subunidad  $\alpha$  en TCR). La diversidad completa de TCR o BCR incorpora ambas subunidades. Sin embargo, es posible realizar el análisis en las secuencias de ambas subunidades.
- 20 Pueden tenerse en cuenta errores generados por secuenciación y/o por amplificación cuando se genera el perfil de clonotipo. Por ejemplo, véase Ejemplo 5.

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

25 **II. Métodos para determinar clonotipos correlacionados, puntuaciones de actividad de enfermedad y algoritmos para determinar uno o ambos**

**A. Clonotipos correlacionados frente a no correlacionados**

- 30 35 40 45 El amplio repertorio de secuencias de receptor de linfocitos T y B crea un reto para hallar células individuales que se correlacionen con resultados de salud humana específicos. En muchos casos las secuencias de clonotipos que serán de interés serán únicas del individuo que se estudie. Los métodos de la presente invención proporcionan medios para distinguir a) clonotipos correlacionados (que pueden ser los clonotipos cuyo nivel se correlaciona con la enfermedad) de b) clonotipos no correlacionados (que pueden ser los clonotipos cuyos niveles no se correlacionan con la enfermedad). En una realización, un clonotipo correlacionado puede presentar correlación positiva o negativa con la enfermedad. En otra realización, un clonotipo presente en un estado pico de una enfermedad pero no presente en un estado no pico de una enfermedad puede ser un clonotipo correlacionado (correlación positiva con la enfermedad). En otra realización, un clonotipo que es más abundante (es decir está presente a un mayor nivel de moléculas) en un estado (o estadio) pico de una enfermedad que en un estado no pico de la enfermedad puede ser un clonotipo correlacionado (correlación positiva con la enfermedad). En otra realización, un clonotipo ausente en un estado pico de una enfermedad pero presente durante un estado no pico de la enfermedad puede ser un clonotipo correlacionado (correlación negativa con la enfermedad). En otra realización, un clonotipo que es menos abundante en un estado pico de una enfermedad que en un estado no pico de una enfermedad puede ser un clonotipo correlacionado (correlación negativa con la enfermedad). En otra realización, un clonotipo correlacionado para un individuo se determina por un algoritmo.

**B. Descubrir los clonotipos correlacionados y no correlacionados usando un ensayo de calibración sin un estudio de población**

- 50 55 60 En esta realización de la invención, se identifican clonotipos correlacionados mirando los clonotipos presentes en algunas muestras que tienen relevancia para una patología (por ejemplo, véase **FIGURA 14**). Este estado puede ser sangre de una muestra en un estado pico de la enfermedad (por ejemplo una muestra sanguínea de un paciente con MS o lupus durante un brote agudo) o tejido aquejado que se supone que está enriquecido con respecto a linfocitos T y B implicados en la enfermedad para ese individuo. Los ejemplos de estos tejidos podrían ser biopsias de riñón de pacientes con lupus con inflamación de los riñones, LCR en pacientes con MS durante un brote, líquido sinovial para pacientes con artritis reumatoide o muestras tumorales de pacientes con cáncer. En todos estos ejemplos, es probable que los tejidos contengan linfocitos T y B relevantes que están relacionados con la enfermedad (aunque no necesariamente los agentes causantes). Es notable que si este método se usa para identificar los clonotipos que son relevantes para la enfermedad, solamente serán relevantes para el individuo en cuya muestra se detectaron. Como resultado, será necesario un ensayo de calibración específico para usar este método para identificar clonotipos correlacionados en cualquier individuo dado con una enfermedad.

- 65 En una realización, se proporciona un método para determinar uno o más clonotipos correlacionados en un sujeto. El método puede incluir etapas para a) generar uno o más perfiles de clonotipo por secuenciación de ácido nucleico de moléculas aisladas espacialmente, individuales, 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

correlacionados en el sujeto basándose en el o los perfiles de clonotipo.

En una realización, la al menos una muestra es de un tejido aquejado por la enfermedad. En otra realización, dicha determinación de uno o más clonotipos correlacionados comprende comparar perfiles de clonotipo de al menos dos muestras.

- 5 En otra realización, el primer estado de la enfermedad es un estado pico de la enfermedad. En otra realización, uno o más clonotipos correlacionados están presentes en el estado pico de la enfermedad. En otra realización, el o los clonotipos correlacionados están ausentes en el estado pico de la enfermedad. En otra realización, uno o más clonotipos correlacionados son altos en el estado pico de la enfermedad. En otra realización, uno o más clonotipos correlacionados son bajos en el estado pico de la enfermedad. En otra realización, la muestra comprende linfocitos T y/o linfocitos B. En otra realización, los linfocitos T y/o linfocitos B comprenden un subconjunto de linfocitos T y/o linfocitos B. En otra realización, el subconjunto de linfocitos T y/o linfocitos B se enriquece por la interacción con un marcador. En otra realización, el marcador es un marcador de superficie celular en el subconjunto de linfocitos T y/o linfocitos B. En otra realización, el subconjunto de linfocitos T y/o linfocitos B interacciona con un antígeno específicamente presente en la enfermedad.

- 10 15 En una realización, la enfermedad es una enfermedad autoinmunitaria. En otra realización, la enfermedad autoinmunitaria es lupus eritematoso sistémico, esclerosis múltiple, artritis reumatoide o Espondilitis Anquilosante.

### **C. Descubrir clonotipos correlacionados y no correlacionados usando un estudio de población**

- 20 25 En una realización, se proporciona un método para identificar clonotipos correlacionados usando un estudio de población (por ejemplo, véase **FIGURA 15**). La utilidad del estudio de población es que permite que la información específica acerca de clonotipos correlacionados que se han determinado en individuos con resultados de patología conocidos se generalice para permitir identificar dichos clonotipos correlacionados en todos los sujetos futuros sin la necesidad de un ensayo de calibración. Puede usarse el conocimiento de un conjunto específico de clonotipos correlacionados para extraer normas acerca de los atributos probables (parámetros) de clonotipos que se correlacionarán en sujetos futuros.

- 30 35 En una realización, la invención proporcionada abarca métodos que incluyen identificar clonotipos correlacionados y no correlacionados secuenciando el repertorio celular inmunitario en un estudio de muestras de pacientes con enfermedad o enfermedades y opcionalmente controles de salud en diferentes momentos y, en el caso de los pacientes con una enfermedad, en diferentes estados (y conocidos) de la evolución de la enfermedad caracterizada por datos clínicos. La enfermedad puede ser, por ejemplo, una enfermedad autoinmunitaria. Los clonotipos cuyo nivel se correlacionada con medidas de la enfermedad en estos estados diferentes pueden usarse para desarrollar un algoritmo que prediga la identidad de un conjunto mayor de secuencias que se correlacionarán con la enfermedad distinto de las que no se correlacionarán con la enfermedad en todos los individuos. A diferencia del caso del ensayo de calibración, no es necesario que estén presentes secuencias correlacionadas en el estudio de descubrimiento pero puede predecirse basándose en estas secuencias. Por ejemplo, una secuencia correlacionada puede ser 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 correlacionados puede usarse para identificar clonotipos en una muestra de cualquier individuo y no es de ningún modo único de un individuo dado, permitiendo de este modo predecir los clonotipos correlacionados en una nueva muestra sin conocimiento previo de los clonotipos presentes en ese individuo.

- 40 45 50 En un aspecto, se proporciona un método para desarrollar un algoritmo que prediga uno o más clonotipos correlacionados en cualquier muestra de un sujeto con una enfermedad que comprende: a) generar una pluralidad de perfiles de clonotipo de un conjunto de muestras, donde las muestras son relevantes para la enfermedad, b) identificar uno o más clonotipos correlacionados del conjunto de muestras, c) usar parámetros de secuencia y/o datos funcionales de uno o más clonotipos correlacionados identificados en b) para desarrollar un algoritmo que pueda predecir clonotipos correlacionados en cualquier muestra de un sujeto con la enfermedad.

En una realización, el conjunto de muestras se toma de uno o más tejidos aquejados por la enfermedad.

- 55 En otra realización, la identificación de uno o más clonotipos correlacionados comprende comparar perfiles de clonotipo de al menos dos muestras. En otra realización, los datos funcionales incluyen la capacidad de unión de marcadores en linfocitos T y/o linfocitos B o interacción con un antígeno por un linfocito T o linfocito B. En otras realizaciones, dichos parámetros de secuencia comprenden secuencia de ácido nucleico y secuencia de aminoácidos predicha. En otra realización, las muestras son de uno o más individuos en un estadio pico de la enfermedad. En otra realización dichos uno o más clonotipos correlacionados están presentes en el estadio pico de la enfermedad. En otra realización dichos uno o más clonotipos correlacionados están a un nivel alto en el estadio pico de la enfermedad. En otra realización, uno o más clonotipos correlacionados están a un nivel bajo en el estadio pico de la enfermedad. En otra realización, uno o más clonotipos correlacionados están ausentes en el estadio pico de la enfermedad.

- 60 65 En una realización, la enfermedad es una enfermedad autoinmunitaria. En otra realización, la enfermedad autoinmunitaria es lupus eritematoso sistémico, esclerosis múltiple, artritis reumatoide o Espondilitis Anquilosante.

En otro aspecto, se proporciona un método para descubrir uno o más clonotipos correlacionados para un individuo, que comprende a) introducir un perfil de clonotipo de una muestra del individuo en un algoritmo, y b) usar el algoritmo para determinar uno o más clonotipos correlacionados para el individuo. El algoritmo puede ser un algoritmo desarrollado: a) generando una pluralidad de perfiles de clonotipo de un conjunto de muestras, donde las

5 muestras son relevantes para la enfermedad, b) identificando uno o más clonotipos correlacionados del conjunto de muestras y c) usando parámetros de secuencia y/o datos funcionales de uno o más clonotipos correlacionados identificados en b) para desarrollar el algoritmo que puede predecir clonotipos correlacionados en cualquier muestra de un sujeto con la enfermedad.

10 **D. Descubrir clonotipos correlacionados y no correlacionados usando un ensayo de calibración combinado con un estudio de población**

En una realización de la invención los clonotipos correlacionados se identifican usando un ensayo de calibración combinado con un estudio de población (por ejemplo, véase **FIGURA 17**). En esta realización el estudio de población no da como resultado un algoritmo que permita predecir clonotipos en cualquier muestra sino que permite desarrollar un algoritmo para predecir los clonotipos correlacionados en cualquier muestra de un sujeto para el que se haya generado un perfil de clonotipo de calibración particular (por ejemplo, véase **FIGURA 16**). Un ejemplo de esto podría ser el desarrollo de un algoritmo que predijera los clonotipos correlacionados en un paciente con lupus basándose en el perfil del clonotipo medido de una muestra sanguínea en cualquier estadio de la enfermedad 20 después de haberse tomado en primer lugar un análisis de sangre durante un estado de brote clínico que se usó para calibrar el algoritmo.

En esta realización, la invención proporcionada abarca métodos para identificar clonotipos correlacionados y no correlacionados secuenciando el repertorio de células inmunitarias en un estudio de muestras de pacientes de la enfermedad o enfermedades y opcionalmente controles sanitarios en momentos diferentes y, en el caso de los pacientes con una enfermedad, en estados diferentes (y conocidos) de la evolución de la enfermedad caracterizada por los datos clínicos. Los clonotipos que se encuentran a diferente frecuencia (o nivel) en el primer estado que en el segundo estado se usan después para desarrollar un algoritmo que predice cuales de las secuencias halladas en los repertorios de cada individuo en el primer estado de enfermedad se correlacionarán con enfermedad en el estado 30 posterior en cada individuo distinto de los que no se correlacionan con la enfermedad en ese individuo. A diferencia del caso del ensayo de calibración solamente, las secuencias correlacionadas pueden ser un subconjunto de todas las secuencias que se ha descubierto que son diferentes entre patologías. También es posible que no se encuentren clonotipos correlacionados en la muestra de calibración pero se predice basándose en el algoritmo que están 35 correlacionados si aparecen en una muestra futura. Como un ejemplo, puede predecirse que un clonotipo que codifica la misma secuencia de aminoácidos que un clonotipo hallado en una muestra de calibración es un clonotipo correlacionado basándose en el algoritmo que resulta del estudio de población. A diferencia de las realizaciones anteriores, el algoritmo se desarrolla para predecir los clonotipos correlacionados basándose en un perfil de clonotipo de calibración que es un perfil de clonotipo generado en el individuo para el que van a predecirse los 40 clonotipos correlacionados en un estado específico de la enfermedad. En esta realización el algoritmo no puede usarse para generar clonotipos correlacionados en un individuo particular hasta que se haya medido un perfil de clonotipo de calibración específico. Después de haberse medido este perfil de calibración en un sujeto particular, todos los clonotipos correlacionados posteriores pueden predecirse basándose en la medición de los perfiles de clonotipo en ese individuo.

45 En otro aspecto, se proporciona un método para descubrir uno o más clonotipos correlacionados para un individuo, que comprende a) introducir un perfil de clonotipo de una muestra del individuo en un algoritmo, y b) usar el algoritmo para determinar uno o más clonotipos correlacionados para el individuo. El algoritmo puede ser un algoritmo desarrollado: a) generando una pluralidad de perfiles de clonotipo de un conjunto de muestras, en el que las muestras son relevantes para la enfermedad, b) identificando uno o más clonotipos correlacionados del conjunto 50 de muestras y c) usando parámetros de secuencia y/o datos funcionales de uno o más clonotipos correlacionados identificados en b) para desarrollar un algoritmo que pueda predecir clonotipos correlacionados en cualquier muestra de un sujeto con la enfermedad. En una realización, la muestra se toma en un estado pico de la enfermedad. En otra realización, la muestra se toma de tejido aquejado por la enfermedad.

55 **E. Parámetros relacionados con la secuencia que pueden usarse para predecir clonotipos correlacionados**

Para realizar un estudio de población puede usarse un conjunto de entrenamiento para entender las características de clonotipos correlacionados ensayando diversos parámetros que pueden distinguir los clonotipos correlacionados de los que no. Estos parámetros incluyen la secuencia para los segmentos V, D y J específicos usados. En una 60 realización se muestra que es más probable que los segmentos V específicos se correlacionen con algunas enfermedades como sucede si es probable que los clonotipos para una enfermedad específica reconozcan epítopos relacionados y por lo tanto puedan tener similitud de secuencia. Otros parámetros incluidos en realizaciones adicionales incluyen el alcance de mutación hipersomática identificada y el nivel de un clonotipo en el pico de un episodio y su nivel cuando la enfermedad está relativamente inactiva. Otros parámetros que pueden predecir clonotipos correlacionados incluyen sin limitación: 1) motivos de secuencia incluyendo región V o J, una combinación VJ, secuencias cortas en región DJ; 2) Longitud de secuencia del clonotipo; 3) Nivel del clonotipo incluyendo nivel

absoluto (número de clones por millón de moléculas) o nivel de rango; 4) Similitud de secuencia de aminoácidos y ácido nucleico con otros clonotipos: la frecuencia de otros clonotipos altamente relacionados, incluyendo los que tienen cambios silenciosos (diferencias de nucleótidos que codifican los mismos aminoácidos) o los que tienen 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 difieren por mutaciones somáticas de algunos clonotipos de línea germinal; 6) clonotipos cuyas proteínas asociadas tienen estructuras tridimensionales similares.

#### **F. Datos funcionales para refinar la determinación de clonotipos correlacionados**

10 Realizaciones adicionales usarán datos funcionales para ayudar a identificar clonotipos correlacionados. Por ejemplo, los linfocitos T y/o linfocitos B que contienen ciertos marcadores que están enriquecidos en células que contienen clonotipos correlacionados pueden capturarse mediante métodos convencionales como FACS o MACS. En otra realización el marcador es un marcador de superficie celular. En otra realización la reactividad de linfocitos T y/o linfocitos B a un antígeno relevante para la patología o para el tejido afectado serán buenas pruebas de la relevancia patológica de un clonotipo.

15 En otra realización, la secuencia de los clonotipos candidatos puede sintetizarse y ponerse en el contexto del TCR o BCR completo y evaluarse con respecto a la reactividad relevante. Como alternativa, los fragmentos amplificados de las secuencias diferentes pueden usarse como una aportación a las técnicas de presentación de fagos, ribosomas o ARN. 20 Estas técnicas pueden seleccionar las secuencias con la reactividad relevante. La comparación de los resultados de secuenciación para los de antes y después de la selección puede identificar los clones que tienen la reactividad y por lo tanto probablemente sean patológicos. En otra realización, las técnicas de presentación específicas (por ejemplo presentación de fagos, ribosomas o ARN) pueden usarse en un formato de matriz. Las 25 moléculas individuales (o amplificaciones de estas moléculas individuales) que portan secuencias individuales del TCR o BCR (por ejemplo secuencias de CDR3) pueden disponerse como fagos, ribosomas o ARN. Después pueden estudiarse antígenos específicos para identificar la secuencia o las secuencias que codifican péptidos que se unen a ellos. Los péptidos que se unen a antígenos relevantes para la enfermedad probablemente sean patológicos.

#### **G. Generar un algoritmo de carga inmunitaria**

30 Puede usarse un algoritmo para calcular una Carga Inmunitaria (por ejemplo, véase **FIGURA 18**). La Carga Inmunitaria puede usarse para tomar una decisión clínica. Usando 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 combine la información acerca de los 35 niveles de los clonotipos correlacionados y no correlacionados en una única puntuación (Carga Inmunitaria). Los parámetros de este algoritmo pueden después ajustarse para maximizar la correlación entre Carga Inmunitaria y los datos clínicos. Por ejemplo, los datos clínicos pueden ser una medida clínica de gravedad de la enfermedad (por ejemplo, el alcance de lesiones en un MRI para un paciente con esclerosis múltiple).

40 Los clonotipos correlacionados usados en la generación de un algoritmo de carga inmunitaria pueden generarse usando un ensayo de calibración, un estudio de población, o un ensayo de calibración y un estudio de población como se ha descrito anteriormente.

45 Algunos de los factores que pueden considerarse al combinar los clonotipos correlacionados son el número de clonotipos correlacionados, su nivel, su tasa de cambio (velocidad), y la tasa de cambio en la velocidad (aceleración). Otros factores para evaluar incluyen el nivel de los clonotipos en el pico del episodio y en el estado de enfermedad inactivo.

50 En una realización, la Carga Inmunitaria generada se relaciona con una enfermedad autoinmunitaria. Dicha Carga puede denominarse Carga Autoinmunitaria.

55 En un aspecto, se proporciona un método para generar un algoritmo que calcula una puntuación de actividad de enfermedad, que comprende: a) desarrollar un algoritmo que usa un conjunto de factores para combinar niveles de clonotipos correlacionados en una puntuación de actividad de la enfermedad, b) comparar la puntuación de actividad de la enfermedad con datos clínicos con respecto a la patología, y c) optimizar los factores para maximizar la correlación entre los datos clínicos y la puntuación de actividad de la enfermedad.

#### **H. Control de la enfermedad usando el algoritmo de carga**

##### **1. Control de la enfermedad sin un ensayo de calibración**

60 En una realización de la invención los clonotipos y el algoritmo de carga inmunitaria se determinan usando un estudio de población (por ejemplo, véase **FIGURA 19**). Puede usarse Carga Inmunitaria directamente sin tener que calibrar en primer lugar al paciente individual. Este ensayo puede realizarse cuando el paciente está en cualquier estado de la enfermedad. Este ensayo puede usarse para generar clonotipos correlacionados y no correlacionados específicos basándose en el algoritmo desarrollado anteriormente. La carga inmunitaria puede calcularse después

usando el segundo algoritmo generado en un estudio de población. Esta puntuación puede expresarse después clínicamente.

- 5 En otro aspecto, se proporciona un método para controlar el estado de la enfermedad de un individuo que comprende: a) determinar un perfil de clonotipo de una muestra de un sujeto, b) introducir la información del perfil de clonotipo de a) en un algoritmo, y c) usar el algoritmo para generar una puntuación predictiva de la patología del individuo. El algoritmo puede ser un algoritmo generado a) desarrollando un algoritmo que use un conjunto de factores para combinar niveles de clonotipos correlacionados en una puntuación de actividad de enfermedad, b) comparar la puntuación de actividad de enfermedad con datos clínicos con respecto al estado de la enfermedad, y c) optimizar los factores para maximizar la correlación entre los datos clínicos y la puntuación de actividad de la enfermedad.
- 10

## **2. Controlar la enfermedad usando un ensayo de calibración**

- 15 En una realización de la invención proporcionada los clonotipos correlacionados y el algoritmo de Carga Inmunitaria se determinan usando un ensayo de calibración o un ensayo de calibración y un estudio de población (por ejemplo, véase **FIGURA 20**). Puede usarse carga inmunitaria en la clínica realizando en primer lugar un ensayo de calibración. Este ensayo puede realizarse cuando el paciente está en un estado que es similar al primer estado usando en el estudio que generó los clonotipos correlacionados y no correlacionados que se usan en el algoritmo de carga inmunitaria. Por ejemplo, este estado puede ser un estado de brote de una enfermedad autoinmunitaria si así es como se derivó el algoritmo de carga inmunitaria. Este ensayo de calibración puede usarse después para generar los clonotipos correlacionados y no correlacionados específicos para usar en los ensayos de control de enfermedad posteriores. En un momento posterior en el tratamiento de este paciente, se realizó otro ensayo en el paciente y se puede calcular la carga inmunitaria usando el algoritmo generado en el estudio de descubrimiento, y la lista de niveles de clonotipo generados en el ensayo de calibración específico de este paciente. Esta puntuación de Carga Inmunitaria puede usarse después clínicamente.

- 20 En otro aspecto, se proporciona un método para controlar el estado de la enfermedad de un individuo que comprende: a) determinar un perfil de clonotipo de una muestra de un sujeto, b) introducir la información del perfil de clonotipo de a) en un algoritmo, y c) usar el algoritmo para generar una puntuación predictiva de la patología del individuo. El algoritmo puede ser un algoritmo generado a) desarrollando un algoritmo que use un conjunto de factores para combinar niveles de clonotipos correlacionados en una puntuación de actividad de enfermedad, b) comparar la puntuación de actividad de enfermedad con datos clínicos con respecto a la patología, y c) optimizar los factores para maximizar la correlación entre datos clínicos y la puntuación de actividad de enfermedad. En otra 25 realización, el método puede comprender además determinar uno o más clonotipos correlacionados en el individuo por cualquiera de los métodos de la invención proporcionada, e introducir información del o los clonotipos correlacionados en el algoritmo.

- 30 En una realización, la enfermedad es una enfermedad autoinmunitaria. En otra realización, la enfermedad autoinmunitaria es lupus eritematoso sistémico, esclerosis múltiple, artritis reumatoide o Espondilitis Anquilosante.

## **3. Otros factores relacionados con el uso de la carga inmunitaria**

- 35 La misma Carga Inmunitaria puede significar diferentes cosas para pacientes diferentes. Para empezar, debe considerarse la imagen clínica completa de un paciente. Desde una perspectiva de ensayo, se puede considerar la velocidad (tasa de cambio de Carga Inmunitaria a lo largo del tiempo) y aceleración (tasa de cambio de la velocidad a lo largo del tiempo) además del nivel de Carga Inmunitaria al tomar decisiones clínicas. Por ejemplo, si la puntuación de Carga Autoinmunitaria está aumentando (alta velocidad) puede ser predictivo de un brote incipiente en una enfermedad autoinmunitaria.

- 40 Pueden integrarse ensayos adicionales en la puntuación de Carga, por ejemplo, una puntuación de Carga Autoinmunitaria incluye, por ejemplo, la tasa de sedimentación de eritrocitos (ESR), los niveles de proteína C reactiva (CRP), anti ADNbc, otros títulos de autoanticuerpos, niveles del complemento, niveles de proteína en orina, relación de creatinina/proteína en orina, niveles de creatinina, niveles de nitrógeno de urea en sangre (BUN), niveles 45 de plaquetas, recuentos de WBC, hematocrito (Hct), Hb, resultados de análisis de orina. Otros ensayos que se relacionan con SLE que pueden integrarse incluyen, por ejemplo, el nivel de CD27, el nivel de células CD27++, genes sensibles a INF (Baechler, EC *et al.* (2003) Proc. Natl. Acad. Sci. 100: 2610-2615), y puntuación de quimiocinas (Bauer JW *et al.* (2009) Arthritis Rheum. 60: 3098-3107). Otros ensayos no relacionados con lupus incluyen, por ejemplo, ensayo de hormona estimulante del tiroides (TSH), ensayo de triyodotironina (T3), ensayo de tiroxina (T4), ensayos de la función hepática (LFT), otros antíanticuerpos, ensayo de calprotectina, ensayo de lactoferrina y análisis del líquido sinovial. Los ensayos adicionales incluyen ensayo de formación de imágenes, incluyendo, por ejemplo, MRI, exploración de CT, rayos X y ultrasonidos.

### III. Determinación de las patologías

Debido a que el sistema inmunitario es tan importante para la salud humana, la capacidad de medir las respuestas inmunitarias tiene amplias aplicaciones en medicina. La presente invención enseña la capacidad de usar el sistema inmunitario para entender el estado de enfermedad subyacente cuando está mediado por el sistema inmunitario. Esto permite que un conjunto muy potente de aplicaciones de diagnóstico y pronóstico que usan los perfiles inmunitarios para proporcionar información sobre los riesgos de una amplia diversidad de resultados clínicos y permita a los médicos intervenir más eficazmente.

10 **A. Utilidad de los perfiles inmunitarios en el tratamiento de enfermedad autoinmunitaria**

Los métodos de la invención proporcionada pueden usarse para diagnosticar y tratar enfermedad autoinmunitaria en un sujeto. La enfermedad autoinmunitaria implica células inmunitarias adaptativas que escapan del proceso habitual que confiere autoinmunidad y que atacan a alguna diana o algunas dianas en el tejido corporal. Las enfermedades autoinmunitarias incluyen, por ejemplo, encefalomielitis diseminada aguda, enfermedad de Addison, espondilitis anquilosante, síndrome de anticuerpos antifosfolípidos, anemia hemolítica autoinmunitaria, hepatitis autoinmunitaria, enfermedad del oído interno autoinmunitaria, enfermedad de Behcet, penfigoide ampolloso, enfermedad Celiaca, enfermedad de Chagas, enfermedad pulmonar obstructiva Crónica, dermatomiositis, diabetes mellitus de tipo 1, síndrome de Goodpasture, enfermedad de Graves, síndrome de Guillain-Barré, tiroiditis de Hashimoto, Hidradenitis supurante, púrpura trombocitopénica idiopática, cistitis intersticial, esclerosis múltiple, miastenia grave, neuromiotonía, pénfigo vulgar, anemia perniciosa, polimiositis, cirrosis biliar primaria, artritis reumatoide, esclerodermia, lupus eritematoso sistémico, síndrome de Sjögren y síndromes de vasculitis. Los estadios de estas enfermedades autoinmunitarias pueden diagnosticarse usando los métodos de la invención proporcionada. Los tratamientos pueden sugerirse a un sujeto basándose en el estadio de la enfermedad autoinmunitaria.

25 Puede usarse información clínica con respecto a un sujeto con una enfermedad autoinmunitaria o que se sospecha que tiene una enfermedad autoinmunitaria, para determinar la patología (o Carga Autoinmunitaria). Puede usarse la información clínica para identificar patrones de un perfil de clonotipo que se correlaciona con una patología. 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 de marcadores moleculares.

30 La información clínica puede incluir síntomas de una o más enfermedades autoinmunitarias. Para hepatitis autoinmunitaria los síntomas pueden incluir fatiga, hepatomegalia, ictericia, prurito, erupción cutánea, artralgia, incomodidad abdominal, angiomas de araña, náuseas, vómitos, anorexia, orina oscura, heces pálidas o grises. Para dermatomiositis (DM) los síntomas pueden incluir erupción (en parche, descoloraciones azuladas-púrpura de la cara, cuello, hombros, parte superior del pecho, codos, rodillas, nudillos y espalda) que acompaña o precede a debilidad muscular, disfagia, mialgia, fatiga, pérdida de peso y fiebre de grado bajo. Para enfermedad de Graves, los síntomas pueden incluir pérdida de peso debido a mayor gasto energético, aumento del apetito, el ritmo cardíaco y la presión sanguínea, y temblores, nerviosismo y sudores. Para tiroiditis de Hashimoto, los síntomas pueden incluir ralentización mental y física, mayor sensibilidad al frío, aumento de peso, aumento de la aspereza de la piel, bocio. Para enfermedad del tejido conectivo mixta (MCTD), los síntomas pueden incluir características de lupus eritematoso sistémico (SLE), esclerodermia y polimiositis. Para el Penfigoide ampolloso (BP) los síntomas pueden incluir ronchas levemente pruríticas hasta ampollas graves e infección, ampollas orales o esofágicas. Para el pénfigo, los síntomas pueden incluir formación de ampollas de las membranas cutánea y mucosa. Para anemia perniciosa, los síntomas pueden incluir apnea, fatiga, palidez, taquicardia, inapetencia, diarrea, hormigueo y endurecimiento de manos y pies, llagas bucales y paso inestable. Para polimiositis (PM), los síntomas pueden incluir debilidad muscular, disfagia y mialgia. Para cirrosis biliar primaria (PBC), los síntomas puede incluir fatiga y prurito. Para esclerodermia (esclerosis sistémica), los síntomas pueden incluir hinchamiento e hinchazón de los dedos o las manos, engrosamiento de la piel, úlceras cutáneas en los dedos, rigidez de las articulaciones de las manos, dolor, úlceras bucales y diarrea. Para el síndrome de Sjögren, los síntomas pueden incluir sequedad de los ojos y la boca, glándulas del cuello hinchadas, dificultad al tragiar o al hablar, sabores u olores inusuales, sed y úlceras de la lengua. Para lupus eritematoso sistémico (SLE), los síntomas pueden incluir fiebre, pérdida de peso, pérdida de pelo, úlceras bucales y nasales, malestar, fatigas, ataques y síntomas de enfermedad mental, inflamación de las articulaciones similar a RA, erupción en forma de mariposa en nariz y mejillas, sensibilidad extrema al frío en las manos y los pies. Para 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 temporales.

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

65 La utilidad clínica de la Carga Autoinmunitaria se analiza con respecto a las enfermedades autoinmunitarias específicas posteriormente. Otra realización de la presente invención contempla la combinación de ensayos de perfiles inmunitarios con otros marcadores que ya se usan para la detección de la actividad de enfermedad en estas

enfermedades para permitir ensayos con mayor sensibilidad y especificidad. Pueden usarse otros identificadores o marcadores moleculares al calcular la Carga Autoinmunitaria o para determinar la patología. 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,

- 5 bacteriano). Los identificadores o marcadores pueden determinarse por técnicas que incluyen, por ejemplo, hibridación genómica comparativa (CGH), análisis de micromatrizes cromosómicas (CMA), perfiles de expresión, micromatriz de ADN, micromatriz de oligonucleótidos de alta densidad, matriz de expresión de ARN de genoma completo, micromatriz peptídica, ensayo inmunoabsorbente ligado a enzimas (ELISA), secuenciación de genoma, análisis del número de copias (CNV), análisis de polimorfismo de nucleótidos pequeños (SNP), inmunohistoquímica, hibridación *in situ*, hibridación *in situ* fluorescente (FISH), PCR, transferencia de Western, transferencia de Southern, SDS-PAGE, electroforesis en gel y transferencia de Northern.
- 10

Para lupus eritematoso sistémico, los marcadores pueden incluir niveles de tasa de sedimentación de eritrocitos (ESR), niveles de proteína C reactiva (CRP), Anti ADNbc, otros títulos de autoanticuerpos, niveles del complemento, 15 niveles de proteína en orina, relación de creatinina/proteína en orina, niveles de creatinina, niveles de nitrógeno de urea en sangre (BUN), niveles de plaquetas, recuentos de WBC, hematocrito (Hct), Hb y resultados de análisis de orina. Otros ensayos que se relacionan por ejemplo con SLE que pueden integrarse incluyen, por ejemplo, el nivel de CD27, el nivel de células CD27++, genes sensibles a INF y puntuación de quimiocinas.

20 **1. Lupus eritematoso sistémico (SLE)**

Los métodos de la invención proporcionada pueden usarse para determinar estados o estadios de lupus eritematoso sistémico (SLE o lupus). El SLE es una afección autoinmunitaria grave que con frecuencia aqueja a adultos jóvenes (principalmente mujeres). Se caracteriza por procesos inflamatorios que pueden afectar a muchos órganos 25 incluyendo la piel, articulaciones, riñones, pulmones, corazón y sistema nervioso central lo que conduce a discapacidades frecuentes y en ocasiones la muerte. La enfermedad sigue una evolución muy impredecible marcada por periodos de brote seguidos de periodos quiescentes de remisión. No obstante, los pacientes a los que se ha diagnosticado SLE son vistos regularmente por un reumatólogo y se tratan con una diversidad de medicaciones 30 importantes. Estas medicaciones incluyen esteroides tales como prednisona y otros inmunosupresores tales como Cellcept (mofetil micofenolato). Aunque estos fármacos pueden reducir el daño a los órganos, contienen efectos secundarios significativos incluyendo riesgo de infección e infertilidad. La poca fiabilidad para algunos de los 35 síntomas (por ejemplo, dolor y fatiga) y la evolución impredecible de la enfermedad hacen difícil adaptar las dosis de medicación, dando como resultado un sobretratamiento de algunos pacientes e infratratamiento de otros. Como resultado, el tratamiento del SLE presenta retos terapéuticos significativos al especialista clínico.

35 Hay varios métodos convencionales que puede usar un especialista clínico para evaluar la actividad del SLE. El estado de la enfermedad puede medirse observando los síntomas clínicos de la enfermedad. Estos métodos incluyen evaluación de señales (por ejemplo, erupción cutánea) y síntomas (por ejemplo, dolor de las articulaciones y fatiga) así como resultados de laboratorio (por ejemplo, relación de creatinina/proteína en orina, anticuerpo anti 40 ADNbc y recuentos sanguíneos). Estos marcadores clínicos, sin embargo, pueden ser indicadores retardados del estado de la enfermedad y como tal los pacientes pueden responder solamente 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 inflamación, por ejemplo anticuerpo anti ADNbc, nivel de complemento (por ejemplo, C3), proteína C reactiva (CRP) y tasa de sedimentación de eritrocitos (ESR) habitualmente carecen de especificidad 45 y/o sensibilidad. Los métodos invasivos tales como biopsia renal son poco prácticos para su uso rutinario. Como resultado los especialistas clínicos realizan ensayos bastante frecuentes de sus pacientes sin una medida perfecta de la patología. Los síntomas clínicos y la evaluación de laboratorio se integran en medidas tales como el Índice de la Actividad de Enfermedad del Lupus Eritematoso Sistémico (SLEDAI) y Evaluación Global Médica (PGA). Estas 50 medidas no se realizan rutinariamente en la práctica clínica y con frecuencia se quedan cortas en varias situaciones clínicas.

55 Se realizan ejemplos específicos de la utilidad de la Carga Autoinmunitaria en la realización de intervenciones terapéuticas en SLE con más detalle en la sección de ejemplos junto con estudios habilitantes específicos que determinan la Carga Autoinmunitaria.

**2. Esclerosis múltiple (MS)**

Los métodos de la invención proporcionada también pueden usarse para determinar los estados o estadios de Esclerosis Múltiple (MS). La MS es una enfermedad autoinmunitaria que afecta al cerebro y la médula espinal (sistema nervioso central). Los síntomas varían, porque la localización y gravedad de cada ataque puede ser diferente. Los episodios pueden durar días, semanas o meses. Estos episodios alternan con periodos sin síntomas o con síntomas reducidos (remisiones). Es habitual que la enfermedad vuelva (recaída). Sin embargo, la enfermedad puede continuar empeorando sin periodos de remisión.

65 Debido a que los nervios de cualquier parte del cerebro o la 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 anómala en cualquier área, dolor debido a espasmos musculares, dolor en los brazos o piernas, problemas al mover los brazos o piernas, problemas al caminar, problemas con la coordinación y realización de movimientos pequeños, habla mal articulada o difícil de entender, temblor en uno o más brazos y piernas, espasmo incontrolable de los grupos musculares (espasticidad muscular) y debilidad en uno o más brazos o piernas.
- 5 Los síntomas oculares incluyen, por ejemplo, visión doble, incomodidad ocular, movimientos rápidos del ojo incontrolables y pérdida de visión (habitualmente afectan a un ojo cada vez).
- 10 Otros síntomas cerebrales y nerviosos incluyen, por ejemplo, reducción del lapso de atención, reducción de discernimiento, reducción de la memoria, depresión o sentimientos de tristeza, mareos y problemas de equilibrio, dolor facial, pérdida de audición y fatiga.
- 15 Los síntomas del intestino y la vejiga incluyen, por ejemplo, estreñimiento, dificultad al comenzar a orinar, necesidad frecuente de orinar, filtración de heces, fuerte necesidad de orinar y filtración de orina (incontinencia).
- No existe cura conocida para esclerosis múltiple en este momento. Sin embargo, hay terapias que puede ralentizar la enfermedad. El objetivo del tratamiento es controlar los síntomas y ayudar al paciente a mantener una calidad de vida normal.
- 20 25 Las medicaciones usadas para ralentizar la progresión de esclerosis múltiple pueden incluir, por ejemplo, moduladores inmunitarios para ayudar a controlar el sistema inmunitario, incluyendo interferones (Avonex, Betaseron o Rebif), anticuerpos monoclonales (Tysabri), acetato de glatiramer (Copaxone), mitoxantrona (Novantrone), metotrexato, azatioprina (Imuran), ciclofosfamida (Cytoxan) y natalizumab (Tysabri). Pueden usarse esteroides para reducir la gravedad de los ataques.
- 30 Las medicaciones para controlar los síntomas pueden incluir, por ejemplo, medicinas para reducir los espasmos musculares tales como Lioresal (Baclofen), tizanidina (Zanaflex) o una benzodiacepina, medicaciones colinérgicas para reducir los problemas urinarios, antidepresivos para síntomas anímicos o conductuales y amantadina para fatiga.
- 35 40 La MS afecta a las mujeres más que a los hombres. El trastorno comienza más habitualmente entre las edades de 20 y 40, pero puede verse a cualquier edad. La MS está provocada por daño a la vaina de mielina, la cobertura protectora que rodea a las células nerviosas. Cuando esta cobertura nerviosa está dañada, los impulsos nerviosos se ralentizan o se detienen. La MS es una enfermedad progresiva, lo que significa que el daño nervioso (neurodegeneración) empeora a lo largo del tiempo. Lo rápido que la MS empeora varía de una persona a otra. El daño nervioso está provocado por inflamación. La inflamación se produce cuando las células inmunitarias del propio cuerpo atacan al sistema nervioso. Pueden producirse episodios repetidos de inflamación a lo largo de cualquier área del cerebro y la médula espinal. Los investigadores no están seguros de qué desencadena la inflamación. Las teorías más habituales apuntan a un virus o defecto genético, o una combinación de ambos. Es más probable que la MS suceda en el norte de Europa, el norte de los Estados Unidos, el sur de Australia y Nueva Zelanda que en otras áreas. Los estudios geográficos indican que puede haber un factor ambiental implicado. Las personas con un historial familiar de MS y los que viven en un área geográfica con una mayor tasa de incidencia para MS tienen un mayor riesgo de la enfermedad.
- 45 50 55 60 65 Los síntomas de MS pueden imitar a los de muchos otros trastornos del sistema nervioso. La enfermedad se diagnostica descartando otras afecciones. Las personas que tienen una forma de MS denominada recidivante-remitente tienen un historial de al menos dos ataques, separados por un periodo sin síntomas o de síntomas reducidos. El proveedor de cuidados sanitarios puede sospechar MS si hay reducciones de la función de dos partes diferentes del sistema nervioso central (tales como reflejos anómalos) en dos momentos diferentes. Un examen neurológico puede mostrar reducción de la función nerviosa en un área del cuerpo, o extendida en muchas partes del cuerpo.
- Los ensayos para diagnosticar esclerosis múltiple incluyen, por ejemplo, ensayos del líquido cefalorraquídeo, incluyendo formación de bandas oligoclonales del LCR, exploración craneal por MRI, punción lumbar (punción raquídea), estudio de la función nerviosa (ensayo de potencial evocado) y MRI raquídeo.
- Como otras enfermedades autoinmunitarias, la MS sigue una evolución impredecible con brotes agudos y periodos de remisión. Hay un número creciente de terapias, cada una con efectos secundarios que varían de graves (aumento de peso y depresión) a con peligro para la vida (pancitopenia e infecciones por PML), eficacia variable en diferentes pacientes y altos costes. Al mismo tiempo, la falta de ensayos rutinarios altamente precisos y específicos de la actividad de la enfermedad MS hacen al reto de administrar eficazmente terapia complicado. Pueden separarse episodios clínicos por periodos de tiempo largos (hasta años en enfermedad de estadio temprano) incluso sin tratamiento. Además, las medicaciones disponibles reducen la probabilidad de recaída pero no la evitan completamente. Por lo tanto, la actividad de la enfermedad es difícil de evaluar y, por lo tanto, hay una medida a corto plazo inadecuada de la actividad de enfermedad que podría usarse para medir si una terapia específica

muestra eficacia en un paciente dado midiendo la reducción en número o gravedad de recaídas. El único otro ensayo disponible para controlar la actividad de enfermedad es MRI cerebral para seguir el estado de las lesiones como se revela con la ayuda de agentes potenciadores de contrastes tales como gadolinio. Sin embargo, dicha formación de imágenes solo ofrece una vista integrada del daño cerebral y carece de especificidad y resolución en el tiempo. Intentar usar formación de imágenes de MRI para seguir la evolución de la enfermedad en escalas de tiempo más cortas de un año es poco práctico dados los costes, la falta de especificidad y los peligros de exposición excesiva a contraste. Como resultado, los pacientes se tratan con frecuencia con gran gasto durante períodos prolongados de tiempo sin ninguna retroalimentación eficaz que permita al médico modificar la dosificación y/o cambiar o añadir terapias.

5 10 **3. Artritis reumatoide (RA)**

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

20 25 Aproximadamente el 1 % de la población mundial está aquejada de artritis reumatoide, las mujeres con tres veces más frecuencia que los hombres. La aparición es más frecuente entre las edades de 40 y 50, pero pueden verse afectadas personas de cualquier edad. Puede ser una afección discapacitante y dolorosa, que puede conducir a pérdida sustancial de operatividad y movilidad. La RA se diagnostica principalmente basándose en síntomas y señales, pero también puede diagnosticarse con análisis de sangre (especialmente un ensayo llamado factor reumatoide) y rayos X. El diagnóstico y tratamiento a largo plazo típicamente se realizan por un reumatólogo, un experto en las enfermedades de las articulaciones y los tejidos conectivos.

30 35 Están disponibles diversos tratamientos. El tratamiento no farmacológico incluye terapia física, ortosis y terapia ocupacional. Puede usarse analgesia (analgésicos) y fármacos antiinflamatorios, incluyendo esteroides, para suprimir los síntomas, mientras que pueden usarse fármacos antirreumáticos modificadores de enfermedad (DMARD) para inhibir o detener el proceso inmunitario subyacente y evitar daño a largo plazo. En tiempos recientes, el grupo más nuevo de productos biológicos ha aumentado las opciones de tratamiento.

40 45 Cuando se sospecha clínicamente RA, pueden realizarse estudios inmunológicos, tales como ensayos con respecto a la presencia de factor reumatoide (RF, un anticuerpo específico). Un RF negativo no descarta RA; en su lugar, la artritis se llama seronegativa. Este es el caso en aproximadamente el 15 % de los pacientes. Durante el primer año de enfermedad, es más probable que el factor reumatoide sea negativo convirtiéndose algunos individuos a estado seropositivo a lo largo del tiempo. También se ve RF en otras enfermedades, por ejemplo síndrome de Sjögren y en aproximadamente 10 % de la población sana, por lo tanto el ensayo no es muy específico.

50 55 Debido a esta baja especificidad, se han desarrollado nuevos ensayos serológicos, que se ensayan con respecto a la presencia de los llamados anticuerpos anti proteína citrulinada (ACPA). Como RF, estos ensayos son positivos solamente en una proporción (67 %) de todos los casos de RA, pero en pocas ocasiones son positivos si no está presente RA, proporcionándole una especificidad de aproximadamente el 95 %. Como con RF, hay pruebas de que los ACPA están presentes en muchos casos incluso antes de la aparición de la enfermedad clínica.

60 65 Los ensayos más comunes para ACPA son el ensayo anti CCP (péptido citrulinado cíclico) y el ensayo anti MCV (anticuerpos contra Vimentina citrulinada mutada). Recientemente, se ha desarrollado un ensayo serológico en el punto de cuidado (POCT) para la detección temprana de RA. Este ensayo combina la detección de factor reumatoide y anti MCV para el diagnóstico de artritis reumatoide y muestra una sensibilidad del 72 % y especificidad del 99,7 %.

Además, pueden realizarse varios otros análisis de sangre para tener en cuenta otras causas de artritis, tales como lupus eritematoso. La tasa de sedimentación de eritrocitos (ESR), proteína C reactiva, recuento de sangre completo, función renal, enzimas hepáticas y otros ensayos inmunológicos (por ejemplo, anticuerpo antinuclear/ANA), se realizan todos en este estadio. Los niveles de ferritina elevados pueden revelar hemocromatosis, una RA mimética, o ser una señal de enfermedad de Still, una variante seronegativa, habitualmente juvenil, de reumatoide.

La expresión fármaco antirreumático modificador de enfermedad (DMARD) significaba originalmente un fármaco que afecta a las mediciones biológicas tales como ESR y niveles de hemoglobina y autoanticuerpo, pero se usa ahora habitualmente para indicar un fármaco que reduce la tasa de daño al hueso y cartílago. Se ha descubierto que los DMARD producen remisiones sintomáticas duraderas y retardan o detienen la progresión. Esto es significativo, ya que dicho daño es habitualmente irreversible. Los antiinflamatorios y analgésicos mejoran el dolor y la rigidez pero no evitan el daño a las articulaciones ni ralentizan la progresión de enfermedad.

- Hay un reconocimiento creciente entre los reumatólogos de que el daño permanente a las articulaciones se produce en un estadio muy temprano de la enfermedad. En el pasado era habitual comenzar la terapia solamente con un fármaco antiinflamatorio, y evaluar la progresión clínicamente y usando rayos X. Si hubo pruebas de que comenzaba a producirse daño a las articulaciones, entonces se prescribiría un DMARD más potente. Los ultrasonidos y MRI son métodos más sensibles de formar imágenes de las articulaciones y han demostrado que el daño de las articulaciones se produce mucho más pronto y en más enfermos de lo que se creía previamente. Las personas con rayos X normales con frecuencia tendrán erosiones detectables por ultrasonidos que los rayos X no podrían demostrar. El objetivo ahora es tratar antes de que se produzca el daño.
- 5
- 10 Puede haber otras razones por las que comenzar los DMARD temprano es beneficioso para evitar el daño estructural a las articulaciones. A partir de los estadios más tempranos de la enfermedad, las articulaciones se infiltran por células del sistema inmunitario que se señalan entre sí de maneras que pueden implicar una diversidad de bucles de retroalimentación positiva (se ha observado durante mucho tiempo que una única inyección de corticosteroides puede evitar la sinovitis en una articulación particular durante períodos largos). La interrupción de este proceso tan pronto como sea posible con un DMARD eficaz (tal como metotrexato) parece mejorar el resultado de la RA durante años posteriores. El retardo de la terapia hasta solamente unos pocos meses después de la aparición de los síntomas puede dar como resultado peores consecuencias a largo plazo. Hay por lo tanto interés considerable en establecer la terapia más eficaz con artritis temprana, cuando el paciente es más sensible a la terapia y puede obtener mayor beneficio.
- 15
- 20 Los fármacos de masa molecular pequeña tradicionales usados para tratar artritis incluyen, por ejemplo, DMARD sintetizados químicamente: azatioprina, ciclosporina (ciclosporina A), D-penicilamina, sales de oro, hidroxicloroquina, leflunomida, metotrexato (MTX), minociclina, y sulfasalacina (SSZ). Los fármacos citotóxicos incluyen Ciclofosfamida.
- 25
- 30 Los acontecimientos adversos más habituales se relacionan con toxicidad del hígado y 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 provocar toxicidad ocular, aunque esta es poco común, y debido a que la hidroxicloroquina no afecta a la médula ósea o el hígado se considera con frecuencia el DMARD con la menor toxicidad. Desafortunadamente la hidroxicloroquina no es muy potente, y habitualmente es insuficiente para controlar los síntomas por sí misma.
- 35
- 35 Pueden producirse agentes biológicos (productos biológicos) mediante ingeniería genética, e incluyen, por ejemplo, bloqueadores del factor de necrosis tumoral alfa (TNF $\alpha$ ) - etanercept (Enbrel), infliximab (Remicade), adalimumab (Humira), bloqueadores de Interleucina 1 (IL-1) - anaquinra (Kineret), anticuerpos monoclonales contra linfocitos B - rituximab (Rituxan), bloqueador de la coestimulación de linfocitos T - abatacept (Orencia), bloqueadores de Interleucina 6 (IL-6) - tocilizumab (un anticuerpo anti receptor de IL-6) (RoActemra, Actemra)
- 40
- 40 Los agentes anti inflamatorios incluyen, por ejemplo, glucocorticoides, fármacos antiinflamatorios no esteroideos (AINE, la mayoría también actúan como analgésicos). Los analgésicos incluyen, por ejemplo, paracetamol (acetaminofén en Estados Unidos y Canadá), opiáceos, diprocualona y lidocaína tópica.
- 45
- 45 El reto de tratar la RA queda en el hecho de que la enfermedad es una enfermedad crónica a largo plazo que puede dar como resultado una discapacidad problemática para la que existe una amplia serie de tratamientos cada uno de los cuales tiene inconvenientes significativos. Muchos de los DMARD someten a los pacientes a efectos secundarios significativos incluyendo mayor riesgo de infecciones graves, cáncer o incluso enfermedad autoinmunitaria. Además, los fármacos derivados biológicamente son muy caros, y el paciente se somete a inyecciones frecuentes.
- 50
- 50 Un doctor que inicie la terapia para un paciente se enfrenta con muchas opciones posibles. Sería deseable obtener retroalimentación rápida una vez que el paciente comience la terapia para entender si el paciente está respondiendo a la terapia que se ha seleccionado antes de que se presente la manifestación clínica en sí misma. La formación de imágenes no es sensible y es cara y muchos marcadores sanguíneos tales como CRP carecen de suficiente sensibilidad. Un ensayo que permitiría al médico determinar rápidamente el estado de la enfermedad le permitiría ajustar la terapia rápidamente a una terapia más eficaz, ahorrando al paciente daño adicional a las articulaciones y usando más eficazmente las caras terapias disponibles.
- 55
- 60 Un paciente que no ha experimentado ningún brote agudo desde el comienzo del tratamiento puede de hecho estar experimentando aún daño inflamatorio continuado a las articulaciones que no se ha manifestado clínicamente. Un ensayo que permitiría al doctor diferenciar este estado del fondo permitiría ajustar la terapia para intentar trae al paciente más cerca de un estado en el que no se experimente ningún daño a las articulaciones continuado. Se describen ejemplos específicos de cómo puede usarse la Carga Autoinmunitaria para tratar pacientes con MS en más detalle en la sección de ejemplos del presente documento.

#### **4. Espondilitis anquilosante**

Los métodos pueden usarse para detectar actividad de enfermedad para espondilitis Anquilosante. La espondilitis Anquilosante (AS, del griego ankylos, doblado; spondylos, vértebra), previamente conocida como enfermedad de

5 Bechterew, síndrome de Bechterew y enfermedad de Marie Strümpell, una forma de Espondiloartritis, es una artritis inflamatoria, crónica, y enfermedad autoinmunitaria. Principalmente afecta a las articulaciones en la columna vertebral y el sacroiliion en la pelvis, provocando posterior fusión de la columna vertebral. Es un miembro del grupo de las espondiloartropatías con una fuerte predisposición genética. La fusión completa da como resultado una rigidez completa de la columna vertebral, una afección conocida como columna de bambú.

10 El paciente típico es un hombre joven, de edad entre 18 y 30 años, cuando aparecen por primera vez los síntomas de la enfermedad, con dolor crónico y rigidez en la parte inferior de la columna vertebral o en ocasiones en la columna completa, con frecuencia con dolor indicado en una o la otra nalga o la parte posterior del muslo de la articulación sacroiliia. Los hombres se ven afectados más que las mujeres en una relación de aproximadamente 3:1,

15 habitualmente tomando 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 (iritociclitis y uveítis), provocando enrojecimiento, dolor ocular, pérdida de visión, partículas flotantes y fotofobia. Otro síntoma habitual es fatiga generalizada y en ocasiones náuseas. Menos habitualmente pueden aparecer aortitis, fibrosis pulmonar apical y ectasia de las vainas de las raíces nerviosas sacras. Como con todas las espondiloartropatías seronegativas, puede producirse levantamiento de las uñas (onicolisis).

20 No hay ensayo directo para diagnosticar AS. Un examen crónico y estudios de rayos X de la columna vertebral, que muestran cambios raquídeos característicos y sacroilitis, son las principales herramientas de diagnóstico. Un inconveniente del diagnóstico por rayos X es que se han establecido habitualmente señales y síntomas de AS hasta 8-10 años antes de que aparecieran cambios evidentes en los rayos X en una película sencilla de rayos X, lo que significa un retardo de hasta 10 años antes de poder introducir terapias adecuadas. Las opciones para diagnóstico más temprano son tomografía y formación de imágenes por resonancia magnética de las articulaciones sacroiliacas, pero la fiabilidad de estos ensayos aún no está clara. El ensayo de Schober es una medida clínica útil de la flexión de la columna lumbar realizada durante el examen.

25 30 Durante periodos inflamatorios agudos, los pacientes con AS en ocasiones mostrarán un aumento de la concentración en sangre de proteína C reactiva (CRP) y un aumento de la tasa de sedimentación de eritrocitos (ESR), pero hay muchos con AS cuyas tasas de CRP y ESR no aumentan de modo que los resultados de CRP y ESR normales no siempre se corresponden con la cantidad de inflamación que de hecho tiene una persona. En ocasiones, las personas con AS tienen resultados de niveles normales, pero experimentan una cantidad significativa de inflamación en sus cuerpos.

35 40 La espondilitis anquilosante (AS, del griego ankylos, doblado; spondylos, vértebra), previamente conocida como enfermedad de Bechterew, síndrome de Bechterew y enfermedad de Marie Strümpell, una forma de Espondiloartritis, es una artritis crónica, inflamatoria y enfermedad autoinmunitaria. Principalmente afecta a las articulaciones de la médula espinal y el sacroiliion en la pelvis, provocando con el tiempo fusión de la columna vertebral.

45 Es un miembro del grupo de las espondiloartropatías con una fuerte predisposición genética. La fusión completa da como resultado una rigidez completa de la columna vertebral, una afección conocida como columna de bambú.

50 55 Hay tres tipos principales de medicaciones usadas para tratar la espondilitis anquilosante: 1) fármacos antiinflamatorios, que incluyen AINE tales como ibuprofeno, fenilbutazona, indometacina, naproxeno e inhibidores de COX-2, que reduce la inflamación y el dolor. También se ha demostrado por pruebas clínicas que los analgésicos opioideos son muy eficaces en el alivio del tipo de dolor crónico habitualmente experimentado por los que padecen AS, especialmente en formulaciones de liberación temporalizada. 2) DMARD tales como ciclosporina, metotrexato, sulfasalacina y corticosteroides, usados para reducir la respuesta del sistema inmunitario mediante inmunosupresión; 3) bloqueadores de TNF $\alpha$  (antagonistas) tales como etanercept, infliximab y adalimumab (también conocidos como productos biológicos), se indican para el tratamiento de y son inmunosupresores eficaces en AS como en otras enfermedades autoinmunitarias.

60 65 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 eficaces en el tratamiento no solamente de la artritis de las articulaciones sino también de la artritis raquídea asociada con AS. Un inconveniente, además del frecuentemente alto coste, es el hecho de que estos fármacos aumentan 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 comenzar el tratamiento. En el caso de infecciones recurrentes, incluso dolor de garganta recurrente, la terapia puede suspenderse debido a la inmunosupresión implicada. Se recomienda a los pacientes que toman las mediciones de TNF que limiten su exposición a otros que portan o pueden portar un virus (tal como resfriado o gripe) o que pueden tener una infección bacteriana o fúngica.

La AS produce síntomas que son muy comunes en las poblaciones sanas. Por ejemplo, no es necesario que un paciente que presente dolor de espalda grave experimente un brote de AS sino que podía tener simplemente dolor de espalda habitual. El médico se ve forzado a tomar una decisión acerca de si tratar estos síntomas con fármacos caros con efectos secundarios potencialmente graves sin una visión muy precisa del estado de la enfermedad. CRP

5 y ESR no proporcionan una visión muy precisa del estado de la enfermedad. Al mismo tiempo la evolución de la enfermedad no tratada puede dar como resultado daño raquídeo a largo plazo debilitante. Este estado de la situación conduce a un reto clínico difícil y se usa sobretratamiento significativo. La disponibilidad de una medida objetiva que refleje la actividad de la enfermedad puede ser de gran ayuda en el tratamiento de pacientes con AS.

10 **B. Utilidad del perfil inmunitario en la detección de cáncer**

Estos métodos pueden usarse para medir el riesgo de cáncer. El cáncer se ha convertido en la causa principal de muerte en el mundo industrializado. Por lo tanto, hay gran necesidad de métodos de tratamiento del cáncer. Se 15 están intentando muchos enfoques para el tratamiento del cáncer incluyendo el desarrollo de nuevos fármacos de moléculas pequeñas así como anticuerpos que se dirigen al tumor.

Un conjunto de métodos que se ha propuesto es la inmunoterapia. La vigilancia de los tumores es una de las 20 funciones de las células del sistema inmunitario. Hay varias categorías de antígenos tumorales que se reconocen por el sistema inmunitario. La primera categoría está comprendida por antígenos que se generan de nuevo por mutación somática (mutación puntuación o una traslocación) en el tumor. Otra categoría consiste en antígenos de 25 proteínas que se expresan solamente en células germinales masculinas que no expresan moléculas de MHC. La desregulación de la expresión génica en muchos tumores puede permitir que se expresen algunos de esos antígenos. Una tercera categoría incluye antígenos de proteínas expresadas solamente en tejidos particulares. La cuarta categoría comprende antígenos que se sobreexpresan significativamente en el tejido tumoral. Finalmente la 25 quinta categoría incluye antígenos que resultan de la modificación postraduccional anómala.

Una de las propiedades de los tumores es su capacidad para escapar a la eliminación eficaz por el sistema 30 inmunitario. Se cree que nuevas mutaciones adquiridas en el tumor permiten que vaya de la fase de equilibrio (en la que el tumor no está eliminado completamente pero su crecimiento se mantiene controlado) a la fase de escape en la que el tumor crece sin control eficaz por el sistema inmunitario. Hay muchos mecanismos que los tumores emplean para escapar al sistema inmunitario. Estos mecanismos incluyen la falta de péptidos antigenicos 35 específicos o las moléculas coestimuladoras que pueden activar los linfocitos T. Otros mecanismos incluyen la secreción tumoral del factor que inhibe los linfocitos T y la creación de un sitio privilegiado inducido por tumor creando una barrera física que separa el tumor de los linfocitos. La inducción del sistema inmunitario para luchar 40 mejor contra el tumor como una estrategia para tratar el cáncer se está estudiando y ensayando de múltiples maneras. Un enfoque es la terapia de linfocitos T adoptiva. Este enfoque se centra en la identificación de linfocitos T que se dirigen a antígenos tumorales mediante el aislamiento de células que se infiltran en el tumor y/o reaccionan a 45 un antígeno tumoral específico. Estos linfocitos T pueden cultivarse *in vitro* en condiciones que potencien su eficacia, como el uso de IL-2 y/o células presentadoras de antígenos. Las células expandidas se vuelven a infundir a la sangre del paciente. Otro enfoque es usar retrovirus que contienen TCR específico de tumor. Estos retrovirus pueden infundirse en el paciente en células especiales que posteriormente secretan el retrovirus permitiendo que infecte linfocitos T que después comienzan a expresar el TCR específico de tumor. Finalmente un enfoque habitual es el uso de vacunación. La premisa de este enfoque de terapia es que la inmunización del paciente con uno o más de los antígenos tumorales estimulará la capacidad del sistema inmunitario para luchar contra el tumor. Con 50 frecuencia se realiza inmunización con el uso de un adyuvante como Bacilo de Calmette-Guerin (BCG). Este enfoque ha sido exitoso en la prevención de cáncer inducido por virus como se demuestra por la capacidad para evitar cánceres cervicales inducidos por HPV-16 y HPV-18. Sin embargo, este ha sido menos exitoso en el tratamiento de otros tumores.

55 Mucha de la mejora en la mortalidad debido al cáncer se ha producido debido a la disponibilidad de mejores métodos de detección temprana que conducen por ejemplo a tasas reducidas de mortalidad en cáncer de mama y cánceres cervicales. La mutabilidad de los tumores hace su tratamiento temprano mucho más eficaz que cuando se detectan tarde. Tradicionalmente, la búsqueda de biomarcadores de detección de cáncer habitualmente implicaba buscar marcadores que se expresaban a alto nivel en el cáncer y estaban a nivel bajo o ausentes en el tejido normal. Esto ha conducido a la identificación de varios marcadores tumorales, como PSA. Un problema con la detección temprana del cáncer es que el mayor valor para detección de cáncer sucede cuando la detección del biomarcador es más difícil, es decir, el tumor es muy pequeño. Por lo tanto, para tener un biomarcador de detección de cáncer eficaz que pueda distinguir a pacientes con tumores pequeños de los que no, es necesario que haya una enorme diferencia en la expresión entre el tejido tumoral y el normal debido a la gran diferencia de tamaño entre el tejido tumoral y el normal. Adicionalmente, es necesario que el marcador se "vientre" eficazmente a la sangre u otro fluido corporal para permitir la detección usando una técnica no invasiva.

60 La presente invención enseña un mecanismo nuevo para detección de cáncer usando la respuesta celular inmunitaria. Visto así, la detección de cáncer no se consigue por la detección de un marcador producido por el tumor en sí mismo sino por la respuesta del sistema inmunitario al tumor. Específicamente el perfil de TCR y/o BCR puede proporcionar información acerca de si el cuerpo está montando una respuesta a un tumor o no. Esto puede aliviar

algunos de los problemas con los biomarcadores actuales. En primer lugar la respuesta inmunitaria es una señal de amplificación que puede ser más fácil de detectar. En segundo lugar los linfocitos pasan a través de la sangre regularmente y por lo tanto es una señal de amplificación que puede ser más fácil de detectar. En segundo lugar los linfocitos pasan a través de la sangre regularmente y por lo tanto el biomarcador relevante puede ya estar presente y ser detectable en la sangre periférica frente al biomarcador tumoral tradicional. Finalmente, el problema del material de biomarcador "de fondo" generado por el tejido normal se reduce en gran medida. La gran diversidad de linfocitos T y/o B proporciona un modo de detectar el biomarcador relevante con alta sensibilidad y especificidad, particularmente con la reciente disponibilidad de métodos de alto rendimiento para secuenciación de ADN. El enfoque de usar la respuesta del sistema inmunitario al cáncer para detectarlo impulsa los cimientos plantados para este campo por la promesa de immunoterapia. Sin embargo el riesgo para las dos aplicaciones es probablemente muy diferente. Usar la respuesta inmunitaria al cáncer para su detección no requiere que el clonotipo específico sea eficaz en el tratamiento del tumor sino más bien que esté asociado con la respuesta inmunitaria al tumor.

Otra realización de la presente invención contempla la combinación de los ensayos de perfiles inmunitarios con otros marcadores que ya se usan para la detección de cáncer para permitir ensayos con mayor sensibilidad y especificidad. Otros identificadores o marcadores moleculares pueden usarse en el cálculo del 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 micromatriz cromosómica (CMA), perfiles de expresión, micromatriz de ADN, micromatriz de oligonucleótidos de alta densidad, matriz de expresión de ARN de genoma completo, micromatriz peptídica, ensayo inmunoabsorbente ligado a enzimas (ELISA), secuenciación de genoma, análisis del número de copias (CNV), análisis de polimorfismos de nucleótidos pequeños (SNP), inmunohistoquímica, hibridación *in situ*, hibridación *in situ* fluorescente (FISH), PCR, transferencia de Western, transferencia de Southern, SDS-PAGE, electroforesis en gel y transferencia de Northern.

### **C. Utilidad de los perfiles inmunitarios en medicina de trasplantes**

Estos métodos pueden usarse para detectar rechazo inmunitario de órganos transplantados. El trasplante de órganos se ha convertido en una parte integral de la medicina realizándose 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 Estados Unidos cada año. Estos son procedimientos generalmente complicados realizados en centros de cuidados terciarios. Para minimizar el riesgo del rechazo de trasplantes, los pacientes se someten con frecuencia a inmunosupresión durante períodos prolongados de tiempo sometiéndolos al riesgo de cáncer e infecciones. Además, muchos trasplantes se rechazan de forma aguda o años después del trasplante. A pesar de estos problemas el trasplante de órganos sigue siendo una modalidad de tratamiento esencial ya que los pacientes con insuficiencias orgánicas tienen pocas otras alternativas.

El rechazo de trasplante de órganos sólidos se produce principalmente debido a la respuesta del sistema inmunitario adaptativo al órgano transplantado. Esto se debe a la presencia de aloantígenos en el injerto que se reconocen por el sistema inmunitario del hospedador. El rechazo puede producirse en tres fases diferentes. La primera es la fase hiperaguda en un periodo de minutos desde el trasplante en la que los anticuerpos preformados montan una respuesta 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 producirse años después del trasplante. Dados estos riesgos se ha tenido cuidado en 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 tienen subtipos ABO coincidentes así como ensayados con respecto a coincidencia cruzada (determinación de si el receptor tiene anticuerpos que reaccionan con los leucocitos del donante). Se realiza emparejamiento cuidadoso de forma similar para el complejo Mayor de Histocompatibilidad (MHC) para reducir el rechazo agudo. Sin embargo, dado que las moléculas del MHC son muy polimórficas es muy difícil identificar una coincidencia perfecta. Los gemelos monocigóticos tienen una coincidencia de MHC perfecta. De forma similar se espera que un cuarto de hermanos tengan una coincidencia de MHC perfecta. Los individuos no relacionados que tienen los mismos alelos detectados según el ensayo clínico con frecuencia tienen diferencias debido a otros sitios polimórficos que no se ensayan en la práctica clínica habitual. Sin embargo, incluso con coincidencia de MHC perfecta de hermanos, hay aún un riesgo significativo de rechazo debido a la existencia de antígenos de histocompatibilidad menores, y de hecho es muy común el rechazo agudo que sucede a más de la mitad de los injertos.

Se podría imaginar que ensayos más agresivos del locus del MHC así como la identificación y emparejamiento de los antígenos menores de histocompatibilidad mejorarían significativamente el rechazo de injertos y posiblemente las tasas de supervivencia. Aunque eso podría ser cierto, el número limitado de órganos de donantes disponibles hace esta tarea poco práctica ya que los ensayos más agresivos podrían retardar significativamente la identificación de un injerto apropiado para usar para cada paciente. Por lo tanto, mucho del progreso que se ha producido en el campo de los trasplantes estaría en el uso de agentes inmunosupresores para evitar y tratar el rechazo. En la actualidad se utilizan muchos fármacos para este fin incluyendo: Azatioprina, corticosteroides, Ciclosporina, Tacrolimus, Ácido de Micofenolato, Sirolimus, Muromonab-CD3, Anticuerpo Monoclonal Anti CD25, Anticuerpo Monoclonal Anti CD20 e

inhibidores de Calcineurina.

El trasplante de médula ósea se usa más frecuentemente para el tratamiento de leucemia y linfoma. Típicamente el receptor experimenta un régimen agresivo de radiación y/o quimioterapia para reducir la carga del tumor antes del trasplante.

Los linfocitos T maduros del donante pueden atacar a algunos de los tejidos hospedadores en el rechazo inverso que se denomina Enfermedad de Injerto contra Hospedador (GVHD). Esta se manifiesta con frecuencia por erupción, diarrea y enfermedad hepática. El emparejamiento cuidadoso del MHC puede aliviar pero no eliminar este problema. Una solución es el agotamiento en la médula ósea del donante *in vitro* de linfocitos T maduros que son en última instancia responsables de la GVHD. Un problema con esto es que el mismo fenómeno que provoca la GVHD puede ser responsable de algunos de los efectos terapéuticos del trasplante de médula ósea mediante el efecto de injerto frente a leucemia en el que los linfocitos T donantes atacan a las células cancerosas restantes. Además, el agotamiento de linfocitos T donantes puede exponer al paciente al riesgo de ser inmunodeficiente. Por lo tanto el riesgo y los beneficios tienen que equilibrarse cuando se consideran estos enfoques. Los pacientes se tratan por lo tanto con frecuencia con inmunosupresores para prevenir así como tratar la GVHD.

El tratamiento actual de la médula ósea pero aún más para trasplante de órganos sólidos se basa en gran medida en el tratamiento con agentes inmunosupresores fuertes. Sin embargo, dado que estos fármacos tienen riesgos significativos se usan de una manera para equilibrar 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 a la que el riesgo y los beneficios están equilibrados para el paciente medio. Los ensayos que pueden predecir acontecimientos de rechazo futuros pueden potencialmente ser útiles en la adaptación del tratamiento a los pacientes en los momentos apropiados en que los necesiten. Esto puede dar como resultado la reducción de las dosis inmunosupresoras para algunos de los pacientes mientras que se mejora la tasa de rechazo y se espera que la supervivencia de injerto.

Otra realización de la presente invención contempla la combinación de los ensayos de perfiles inmunitarios con otros marcadores que ya se usan para la detección de rechazo de trasplantes para permitir ensayos con mayor sensibilidad y especificidad. Otros identificadores o marcadores moleculares pueden usarse en el cálculo del 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 micromatrizes cromosómicas (CMA), perfiles de expresión, micromatriz de ADN, micromatriz de oligonucleótidos de alta densidad, matriz de expresión de ARN de genoma completo, micromatriz peptídica, ensayo inmunoabsorbente ligado a enzimas (ELISA), secuenciación genómica, análisis del número de copias (CNV), análisis de polimorfismo de nucleótido pequeño (SNP), inmunohistoquímica, hibridación *in situ*, hibridación *in situ* fluorescente (FISH), PCR, transferencia de Western, transferencia de Southern, SDS-PAGE, electroforesis en gel y transferencia de Northern.

#### **D. Utilidad de los perfiles inmunitarios en el tratamiento de infección**

Estos métodos tienen utilidad al guiar el tratamiento de infecciones particularmente cuando estas infecciones pueden existir en estados activos y latentes. La aparición de antibióticos para el tratamiento de enfermedad infecciosa durante el pasado siglo ha tenido un gran impacto en la esperanza de vida. Durante la pasada década las técnicas de diagnóstico molecular han tomado un papel rápidamente creciente en el diagnóstico y tratamiento de enfermedad infecciosa. La excelente sensibilidad y especificidad proporcionada por 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 realiza con frecuencia por un ensayo molecular que emplea técnicas de amplificación de ácidos nucleicos. Otro conjunto de aplicaciones implican 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 a los que ya se ha diagnosticado SIDA. Este ensayo ayuda al médico a determinar el estado de la enfermedad del paciente y por lo tanto puede proporcionar directrices acerca de la eficacia del régimen de tratamiento que se use.

En ocasiones es útil considerar no solamente el nivel del agente infeccioso sino también la respuesta inmunitaria al agente infeccioso. Un ejemplo en el que la respuesta inmunitaria a la infección se usa habitualmente en la práctica médica es en la hepatitis B. Un aspecto del ensayo de la hepatitis B se basa en detectar el agente infeccioso mediante detección de antígenos de la hepatitis B o por un ensayo de amplificación de ácido nucleico. Además es habitual en la práctica clínica rutinaria ensayar con respecto a la presencia de diferentes anticuerpos que se dirigen al virus de la hepatitis B. La presencia de IgM anti HBc habitualmente sucede en una situación de infección aguda, la aparición de IgG anti HBc indica que la infección es crónica. De forma similar la aparición de anticuerpo anti HBs señala la eliminación de la infección.

En una realización de la presente invención el valor de elaboración de respuesta inmunitaria a una infección se utiliza junto con la sensibilidad y especificidad del ensayo molecular. Esto puede ser particularmente útil para

enfermedades infecciosas que son crónicas cuando el agente infeccioso permanece latente en el cuerpo. El perfil de TCR y/o BCR puede usarse para evaluar la respuesta inmunitaria a una infección. La secuenciación puede usarse para obtener un perfil del TCR y/o BCR 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 varios enfoques.

Otra realización de la presente invención contempla la combinación de los ensayos de perfiles inmunitarios con otros marcadores que ya se usan para la detección de agentes infecciosos para permitir ensayos con mayor sensibilidad y especificidad. Pueden usarse otros identificadores o marcadores moleculares en el cálculo del algoritmo de carga o para determinar la patología. 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), perfiles de expresión, micromatriz de ADN, micromatriz oligonucleotídica de alta densidad, matriz de expresión de ARN de genoma completo, micromatriz peptídica, ensayo inmunoabsorbente ligado a enzimas (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 de Western, transferencia de Southern, SDS-PAGE, electroforesis en gel y transferencia de Northern.

#### **E. Utilidad del perfil inmunitario en el tratamiento de pacientes ancianos**

Estos métodos tienen utilidad en el control del estado del sistema inmunitario en los ancianos. Las personas mayores padecen un deterioro del sistema inmunitario denominado inmunosenescencia que afecta a su capacidad para responder a las infecciones y para inducir respuestas eficaces a las vacunas (Weinberger *et al.*, 2008). Esto resulta evidente a partir de las altas tasas de mortalidad debido 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 cree que el deterioro de la capacidad del sistema inmunitario explica el aumento de la tasa de cánceres en los ancianos. Además la inmunosenescencia puede contribuir a otras enfermedades importantes de los ancianos con un componente significativo de procesos inflamatorios, como Alzheimer y enfermedad cardiaca. La capacidad de predecir qué individuos tienen mayor riesgo para estos resultados letales sería útil para médicos geriatras ya que toman decisiones clínicas acerca de vacunación, tratamiento agresivo de infecciones y hospitalización.

Muchos aspectos del sistema inmunitario innato y adaptativo están alterados en la inmunosenescencia. Los linfocitos T pierden sensibilidad, los macrófagos tienen una capacidad presentadora de antígenos reducida y secreción de citocinas alterada, los linfocitos citolíticos naturales tienen toxicidad reducida, las células dendríticas foliculares no pueden presentar antígenos tan eficazmente, y los neutrófilos pierden su capacidad fagocítica. Hay un grupo menor de linfocitos T y B sin tratamiento previo y un aumento del grupo de memoria y efecto que conduce a una diversidad reducida de repertorios de linfocitos T y B lo que conduce a la reducción de la capacidad del sistema inmunitario adaptativo para responder a nuevos antígenos. En particular, los repertorios de linfocitos T que están asociados con citomegalovirus (CMV) están aumentados en gran medida y hasta el 45% del repertorio de linfocitos T total puede estar dedicado a él. Se ha observado que estas expansiones son menos pronunciadas en centenarios.

Los estudios han sugerido que los marcadores inmunitarios pueden predecir la supervivencia en los ancianos. Se ha mostrado que el grado de diversidad del repertorio de linfocitos B predice la supervivencia en los ancianos al menos en una población. Incluso aunque se ha mostrado 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 del repertorio puede proporcionar significativamente más precisión en la predicción. Por ejemplo, las expansiones sensibles a CMV pueden tener una importancia diferente que otras expansiones.

En una realización de la presente invención, puede recogerse ARN de los linfocitos T y B hallados en sangre periférica a partir 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 en varios puntos temporales en sus historiales clínicos. Los perfiles de pacientes con supervivencia larga se compararán con pacientes con supervivencia corta. En primer lugar, puede obtenerse una medida global de la diversidad. Esto no incluirá solamente el número de clonotipos diferentes identificados sino también su diversidad. Por ejemplo, ¿es el uso del segmento V, D, J el mismo en los dos grupos o está un grupo más restringido en 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 sea menos versátil al responder a un nuevo antígeno en comparación con la otra muestra cuyos clonotipos se distribuyen entre todos los segmentos V.

Además de la diversidad global, se determinará si los clonotipos expandidos en pacientes que han tenido una larga supervivencia pueden distinguirse basándose en algún parámetro de secuencia en comparación con clonotipos en pacientes que han tenido una supervivencia corta. Este enfoque puede complementarse mirando los clonotipos que

responden a antígenos específicos. Por ejemplo dadas las pruebas disponibles la identificación de clonotipos sensibles a CMV puede tener potencial predictivo. Puede realizarse captura de clonotipos de linfocitos T que son sensibles a CMV en un estudio de descubrimiento a partir de un conjunto de pacientes ancianos así como sanos.

5 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 añadir esta información puede aumentar la capacidad para predecir el paciente que sobrevive durante largo tiempo frente al que no.

10 Otra realización de la presente invención contempla la combinación de los ensayos de perfiles inmunitarios con otros marcadores que ya se usan para la detección de la salud en la población anciana para permitir ensayos con mayor sensibilidad y especificidad. Otros identificadores o marcadores moleculares pueden usarse en el cálculo del algoritmo de carga o para determinar la patología. 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

15 pueden determinarse por técnicas que incluyen, por ejemplo, hibridación genómica comparativa (CGH), análisis de micromatrices cromosómicas (CMA), perfiles de expresión, micromatriz de ADN, micromatriz oligonucleotídica de alta densidad, matriz de expresión de ARN de genoma completo, micromatriz peptídica, ensayo inmunoabsorbente ligado a enzimas (ELISA), secuenciación genómica, análisis del número de copias (CNV), análisis de polimorfismo de nucleótido pequeño (SNP), inmunohistoquímica, hibridación *in situ*, hibridación *in situ* fluorescente (FISH), PCR, transferencia de Western, transferencia de Southern, SDS-PAGE, electroforesis en gel y transferencia de Northern.

20

#### **F. Utilidad de los perfiles inmunitarios en la administración de vacunas**

25 Estos métodos tienen utilidad en la administración de vacunas. El uso de la vacunación ha conducido a una gran reducción en la tasa de infecciones de múltiples organismos. Una enfermedad infecciosa que continúa teniendo un impacto sanitario significativo con más de 30.000 muertes al año en los Estados Unidos es la gripe. La vacunación para la gripe debe realizarse anualmente ya que la cepa muta rápidamente. La mayoría de las secuelas graves de la enfermedad se producen en los ancianos. Desafortunadamente los ancianos experimentan con frecuencia inmunosenescencia haciéndolos insuficientemente sensibles a la vacunación.

30 Para distinguir pacientes que son sensibles a la vacunación de los que no, es necesario realizar un estudio de descubrimiento. En esta población están disponibles muestras sanguíneas antes y después (en uno o más momentos establecidos) de vacunación para una cohorte de pacientes vacunados para gripe con resultado de gripe conocido (es decir, si se protegieron posteriormente de la infección o no). Puede obtenerse secuencia de TCR y/o BCR de estas muestras. Se determinan los clonotipos que se enriquecen después de la vacunación en cada paciente. Los clonotipos enriquecidos en pacientes que respondieron a la vacunación se compara después con un conjunto de control de clonotipos (por ejemplo, el resto de los clonotipos en el mismo conjunto de pacientes) para distinguir los clonotipos correlacionados de otros clonotipos. El algoritmo para predecir estos clonotipos se usa después para predecir clonotipos correlacionados entre pacientes que no respondieron a la vacunación. Los pacientes que no respondieron pueden generar el mismo tipo de clonotipos que los que respondieron pero a niveles menores. Como alternativa podría ser que los que no responden generen una clase distinta de clonotipos. El número de clonotipos correlacionados identificados en el que no responde puede distinguir estas dos posibilidades.

45 Con los clonotipos correlacionados identificados, se construye después un algoritmo para generar una puntuación para predecir la probabilidad de inmunización. Para generar este algoritmo se utilizan datos de los perfiles de los que responden a la vacuna y los que no responden. Este algoritmo puede usarse después para predecir la probabilidad de inmunización en el siguiente paciente usando los clonotipos correlacionados predichos de una muestra obtenida después de la inmunización. La predicción se realiza mediante la aplicación de otro algoritmo que también se ha generado en el estudio de descubrimiento. Puede opcionalmente ayudarse de (o sustituirse por) datos de la precalibración para limitar la búsqueda de clonotipos correlacionados a los que se enriquecieron después de la inmunización.

55 Otra realización de la presente invención contempla la combinación de los ensayos de perfiles inmunitarios con otros marcadores que ya se usan para la detección de respuesta a vacunación para permitir ensayos con mayor sensibilidad y especificidad. Otros identificadores o marcadores moleculares pueden usarse en el cálculo del 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), perfiles de expresión, micromatriz de ADN, micromatriz oligonucleotídica de alta densidad, matriz de expresión de ARN de genoma completo, micromatriz peptídica, ensayo inmunoabsorbente ligado a enzimas (ELISA), secuenciación genómica, análisis del número de copias (CNV), análisis de polimorfismo de nucleótido pequeño (SNP), inmunohistoquímica, hibridación *in situ*, hibridación *in situ* fluorescente (FISH), PCR, transferencia de Western, transferencia de Southern, SDS-PAGE, electroforesis en gel y transferencia de Northern.

60

65

### **G. Utilidad de los perfiles inmunitarios en el control de la hipersensibilidad inmunitaria (alergia)**

El sistema inmunitario adaptativo ha evolucionado para responder a antígenos que se asocian con patógenos. Como sucede en el caso de las enfermedades autoinmunitarias, el sistema inmunitario puede en ocasiones tener la diana errónea. Mientras que en enfermedades autoinmunitarias el sistema inmunitario se dirige a autoantígenos, en reacciones de hipersensibilidad monta una respuesta a estímulos inocuos como medicaciones, polvo y alimentos. La hipersensibilidad es muy común, teniendo hasta el 50% de la población de Estados Unidos alergia a un estímulo ambiental, y está provocada por mecanismos. La hipersensibilidad se divide en cuatro tipos. La hipersensibilidad de

- 5 Tipo I es la hipersensibilidad de tipo inmediato y está mediada por IgE. La de Tipo II con frecuencia se debe a unión de anticuerpo IgG con antígeno asociado con superficie celular. Por ejemplo, un fármaco inocuo que se une a la superficie de la célula puede hacer a la célula una diana para IgG anti fármaco en pacientes que resultaban tener estos anticuerpos. La de Tipo III está provocada por deposición de complejos de antígeno-anticuerpo en tejidos. Esto sucede por ejemplo cuando la cantidad de antígeno es grande dando como resultado complejos inmunitarios pequeños que no pueden eliminarse eficazmente y en su lugar se depositan en paredes de vasos sanguíneos. La sensibilidad de Tipo IV es una hipersensibilidad de tipo retardado mediada por linfocitos T. Las de Tipo I y Tipo IV tienen el mayor impacto en la salud humana.

En la reacción de hipersensibilidad de Tipo I el paciente se hace sensible a un antígeno inocuo (alérgeno) produciendo anticuerpo IgE contra él. La exposición posterior al alérgeno induce la activación de células que se

- 20 unen a IgE, tales como mastocitos y basófilos. Una vez activadas estas células provocan la reacción alérgica mediante inducción de un proceso inflamatorio secretando productos químicos almacenados y sintetizando citocinas, 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 a colapso circulatorio con peligro para la vida en anafilaxis. Con frecuencia, la reacción de Tipo I se sigue después de otra fase tardía que desempeña un papel en muchos de los procesos patológicos resultantes. La fase tardía de reclutamiento de linfocitos T auxiliares y otras células inflamatorias es esencialmente una reacción de hipersensibilidad de Tipo IV. Algunas reacciones alérgicas de Tipo I incluyen rinoconjuntivitis de temporada (fiebre del heno), alergia alimentaria, anafilaxis inducida por fármacos, dermatitis atópica (eccema) y asma. Estas son afecciones muy comunes con prevalencia creciente que provocan costes significativos así como morbilidad y mortalidad. Por ejemplo, el asma es una enfermedad crónica que afecta a
- 25 30 ~7% de la población de Estados Unidos provocando ~4.000 muertes al año. Algunas de estas enfermedades tienen algunos aspectos relacionados. Por ejemplo, los pacientes con dermatitis atópica tienen un riesgo significativamente aumentado de tener asma. La alergia alimentaria puede provocar vómitos y diarrea pero también puede dar como resultado anafilaxis en un número significativo de pacientes, 30,000 casos que dan como resultado ~200 muertes al año en Estados Unidos. Algunos de los mismos alérgenos que activan mastocitos submucosos en la nariz provocando síntomas de rinitis alérgica también pueden activar mastocitos en las vías respiratorias inferiores provocando constricción bronquial, un síntoma típico del asma. Algunas reacciones de hipersensibilidad de Tipo IV son dermatitis de contacto (por ejemplo, hiedra venenosa), rinitis crónica, asma crónico, y enfermedad celiaca. La enfermedad celiaca es una enfermedad crónica provocada por una alergia alimentaria no mediada por IgE. Es una enfermedad del intestino delgado provocada por la respuesta alérgica frente al gluten, un componente presente en el
- 35 40 trigo y otros alimentos. Más del 95% de los pacientes celiacos tienen un alelo del MHC de clase II específico, el HLA-DQ2.

El tratamiento de las reacciones de hipersensibilidad difiere, pero con frecuencia tienen dos aspectos: el tratamiento agudo y el tratamiento crónico o prevención. Algunas de estas afecciones pueden presentar peligro para la vida

- 45 (anafilaxis y asma agudo) e implican atención médica inmediata. El tratamiento crónico en general implica tratar de evitar el alérgeno específico. Eso puede ser eficaz 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 ambiente, como el polen o el polvo. Por lo tanto se utiliza con frecuencia tratamiento crónico con medicaciones para algunas de estas enfermedades (por ejemplo, asma y rinitis alérgica). El nivel de eficacia del control del tratamiento se ensaya en última instancia cuando el paciente se vuelve a exponer al alérgeno o los alérgenos. Por lo tanto algunos pacientes pueden someterse a sobre o infratratamiento. Idealmente estaría disponible un ensayo que evalúe la actividad de la enfermedad y el grado en que el paciente es propenso a montar una respuesta de hipersensibilidad. Dicho ensayo permitiría la adaptación del tratamiento a las necesidades del paciente individual.

#### **55 Ejemplos**

##### **EJEMPLO 1**

###### **Determinar la secuencia de ADN recombinado en un sujeto con una enfermedad autoinmunitaria**

- 60 Se toma una muestra sanguínea de una paciente con una enfermedad autoinmunitaria. Se aislan células CD4+ y CD8+ de la muestra sanguínea usando perlas magnéticas recubiertas con anticuerpo. Se usa PCR para amplificar la región variable completa del gen  $\beta$  del receptor de linfocitos T. Los fragmentos amplificados se subclonian en vectores y se transforman en bacterias para aislar los fragmentos de ADN. Las bacterias se cultivan para amplificar el ADN, y se usa secuenciación de didesoxi para secuenciar las regiones variables del gen  $\beta$  del receptor de linfocitos T para identificar los clonotipos. La información de secuenciación se usa para generar un perfil de clonotipo

para el paciente. Se muestra un método similar en la **FIGURA 1**.

#### EJEMPLO 2

5 Determinar el estado de una enfermedad autoinmunitaria

Se toma una muestra de líquido cefalorraquídeo (LCR) y sangre de un paciente con un pico episódico de esclerosis múltiple. Se aíslan células CD4+ del LCR y la sangre, y se amplifica la CDR3 del gen  $\beta$  del receptor de linfocitos T mediante PCR. Los fragmentos amplificados se subclonian en vectores y se transforman en bacterias para aislar los fragmentos de ADN. Las bacterias se cultivan para amplificar el ADN, y se usa la secuenciación de didesoxi para secuenciar las regiones variables del gen  $\beta$  del receptor de linfocitos T para identificar los clonotipos. La información de secuenciación se usa para generar un perfil de clonotipo para el paciente.

10 Se toma otra muestra de sangre cuando el paciente está en un estado relativamente inactivo de esclerosis múltiple. Se repite el mismo procedimiento anterior para generar un perfil de clonotipo. Se identifican clonotipos patológicos como los que son altos en el episodio pico y bajaron significativamente en el estado inactivo. Se toma otra muestra sanguínea del paciente en un estado posterior. En este momento se amplifica solamente una fracción de las regiones CDR3 del gen  $\beta$  del receptor de linfocitos T y después se secuencia. Este subconjunto contiene los clonotipos patológicos. El nivel de los diversos clonotipos se determina para evaluar la patología del paciente.

15 20 Se toma otra muestra de sangre cuando el paciente está en un estado relativamente inactivo de esclerosis múltiple. Se repite el mismo procedimiento anterior para generar un perfil de clonotipo. Se identifican clonotipos patológicos como los que son altos en el episodio pico y bajaron significativamente en el estado inactivo. Se toma otra muestra sanguínea del paciente en un estado posterior. En este momento se amplifica solamente una fracción de las regiones CDR3 del gen  $\beta$  del receptor de linfocitos T y después se secuencia. Este subconjunto contiene los clonotipos patológicos. El nivel de los diversos clonotipos se determina para evaluar la patología del paciente.

#### EJEMPLO 3

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

25 Para estudiar la amplificación del repertorio de TCR, se analizarán cadenas TCR $\beta$ . El análisis incluirá amplificación, secuenciación, y análisis de las secuencias de TCR $\beta$ . Un cebador AGCGACCTCGGGTGGGAACA es complementario de una secuencia común en C $\beta$ 1 y C $\beta$ 2, y hay 34 cebadores V (**Tabla 1**) capaces de amplificar los 48 segmentos V. C $\beta$ 1 o C $\beta$ 2 difieren entre sí en la posición 10 y 14 del punto de unión J/C. El cebador para C $\beta$ 1 y C $\beta$ 2 terminará en la posición 16 pb y no debería tener preferencia por C $\beta$ 1 o C $\beta$ 2.

30 35 Los 34 cebadores V se modifican a partir de un conjunto original de cebadores publicado por el grupo BIOMED-2 para amplificar los 48 segmentos V y todos sus alelos publicados como se definen por el sistema de información internacional ImMunoGeneTics (<http://imgt.cines.fr/>).

35 Los cebadores de BIOMED-2 se han usado en cadena múltiple para identificar la clonalidad en enfermedades linfoproliferativas.

**Tabla 1.** Secuencias de cebadores complementarias de las diferentes familias V.

| Familia de segmento V               | Secuencia de cebador   |
|-------------------------------------|------------------------|
| <b>V20-1</b>                        | AACTATGTTTGGTATCGTCAGT |
| <b>V29-1</b>                        | TTCTGGTACCGTCAGCAAC    |
| <b>V9, 5-1, 5-6, 5-5, 5-8, 5-4A</b> | AGTGTATCCTGGTACCAACAG  |
| <b>V9, 5-1, 5-6, 5-5, 5-8, 5-4B</b> | AGTGTGTACTGGTACCAACAG  |
| <b>V9, 5-1, 5-6, 5-5, 5-8, 5-4C</b> | ACTGTGTCCTGGTACCAACAG  |
| <b>V9, 5-1, 5-6, 5-5, 5-8, 5-4D</b> | AGTGTGTCCTGGTACCAACAG  |
| <b>V9, 5-1, 5-6, 5-5, 5-8, 5-4E</b> | TCTGTGTACTGGTACCAACAG  |
| <b>V7-3, 7-6, 7-9, 7-2, 7-4A</b>    | CCCTTTACTGGTACCGACAG   |
| <b>V7-3, 7-6, 7-9, 7-2, 7-4B</b>    | GCCTTTACTGGTACCGACAG   |
| <b>V7-3, 7-6, 7-9, 7-2, 7-4C</b>    | CCCTTTACTGGTACCGACAAA  |
| <b>V7-8, 16A</b>                    | TTTTGGTACCAACAGGTCC    |
| <b>V7-8, 16B</b>                    | TTTTGGTACCAACAGGCC     |
| <b>V 7-7</b>                        | AACCCTTATTGGTATCACAG   |
| <b>V4-1, 4-3, 4-2A</b>              | CGCTATGTATTGGTACAAGCA  |
| <b>V4-1, 4-3, 4-2B</b>              | GGCAATGTATTGGTACAAGCA  |

| Familia de segmento V                | Secuencia de cebador        |
|--------------------------------------|-----------------------------|
| V12-3, 12-4, 12-5                    | TTTCTGGTACAGACAGACCATGA     |
| V3-1                                 | TACTATGTATTGGTATAAACAGGACTC |
| V25-1                                | CAAAATGTAUTGGTATCAACAA      |
| V28, 10-3, 6-2, 6-3, 6-1, 6-6, 24-1A | ATGTTCTGGTATCGACAAGACC      |
| V28, 10-3, 6-2, 6-3, 6-1, 6-6, 24-1B | ATGTACTGGTATCGACAAGACC      |
| V6-4, 6-9A                           | TGCCATGTAUTGGTATAGACAAG     |
| V6-4, 6-9B                           | ATACTTGCCTGGTATCGACAAG      |
| V10-1, 10-2, 6-5, 6-9, 6-8, 27A      | ATATGTTCTGGTATCGACAAGA      |
| V10-1, 10-2, 6-5, 6-9, 6-8, 27B      | ATATGTCCTGGTATCGACAAGA      |
| V10-1, 10-2, 6-5, 6-9, 6-8, 27C      | ACATGTCCTGGTATCGACAAGA      |
| V14                                  | TAATCTTATTGGTATCGACGTGT     |
| V19                                  | GCCATGTAUTGGTACCGACA        |
| V18                                  | TCATGTTACTGGTATCGGCAG       |
| V30                                  | CAACCTATACTGGTACCGACA       |
| V11-1, 11-3, 11-2A                   | CATGCTACCCCTTACTGGTACC      |
| V11-1, 11-3, 11-2B                   | CACAATACCCCTTACTGGTACC      |
| V2                                   | ATACTTCTATTGGTACAGACAAATCT  |
| V13                                  | CACTGTACTGGTACCGAGCA        |
| V15                                  | CGTCATGTAUTGGTACCGAGCA      |

El uso de los cebadores para amplificación se ensayó con 34 secuencias sintéticas. Las secuencias sintéticas contenían en un lado la secuencia de uno de los oligonucleótidos y en el otro lado el complemento del cebador C.

5 Entre los dos cebadores había 6 pb correspondientes al sitio de la enzima de restricción Cla I. Todas las secuencias sintéticas se amplificaron con los cebadores apropiados, y se demostró mediante digestión con Cla I que los productos de amplificación eran el resultado de amplificar las secuencias sintéticas y no mediante formación de dímeros de cebadores.

10 El Analizador de Genoma Illumina es la plataforma de secuenciación seleccionada. En cada carril, pueden realizarse ~15 millones de lecturas. Se procesarán doce muestras humanas y 96 de ratón por carril, y se usarán marcadores

15 de secuencia para distinguir lecturas de una muestra de las de otra. Se puede realizar una exploración de amplificación de dos etapas, como se ilustra en la **FIGURA 2**. Como se muestra en la **FIGURA 2A**, la PCR primaria usará en un lado un cebador de 20 pb cuyo extremo 3' está a 16 bases del punto de unión J/C y es perfectamente complementario de C $\beta$ 1 y los dos alelos de C $\beta$ 2. En la PCR secundaria, en el mismo lado del molde, se usa un

20 cebador que tiene en su extremo 3' la secuencia de las 10 bases más cercanas al punto de unión J/C, seguido de 17 pb con la secuencia de las posiciones 15-31 desde el punto de unión J/C, seguido de la secuencia P5. Este cebador se denomina C10-17-P5. P5 desempeña un papel en la formación de grupos. Cuando el cebador C10-17-P5 hibrida con el molde generado de la primera PCR, se crea un bucle de 4 pb (posición 11-14) en el molde, a media que el cebador hibrida con la secuencia de las 10 bases más cercanas al punto de unión J/C y las bases en las posiciones

25 15-31 desde el punto de unión J/C. La formación de bucle de las posiciones 11-14 elimina la amplificación diferencial de moldes que portan C $\beta$ 1 o C $\beta$ 2. En última instancia, se realiza secuenciación con un cebador complementario de la secuencia de las 10 bases más cercanas al punto de unión J/C y bases en las posiciones 15-31 desde el punto de unión J/C (este cebador se denominará C'). El cebador C10-17-P5 puede purificarse por HPLC para asegurar que todo el material amplificado tenga extremos intactos que puedan utilizarse eficazmente en la formación de grupos.

25 En la **FIGURA 2B**, se muestra que la longitud del saliente en los cebadores V es de 14 pb. La primera PCR puede ayudarse de un saliente más corto. Por otro lado para la segunda PCR, puede ser ventajoso usar el saliente en el cebador V en la primera PCR tanto tiempo como sea posible debido a que la segunda PCR usará el cebador a partir de esta secuencia. Un uso de cebadores muy eficaz en la segunda PCR puede provocar limitación en la

30 representación en los datos finales.

- Se investigó un tamaño mínimo del saliente que soporta una segunda PCR eficaz. Se realizaron dos series de cebadores V (para dos segmentos V diferentes) con tamaños de salientes de 10 a 30 con etapas de 2 pb. Usando las secuencias sintéticas apropiadas, se realizó la primera PCR con cada uno de los cebadores en la serie y se realizó electroforesis en gel para mostrar que todos se amplificaron. Para medir la eficacia de la segunda amplificación por PCR se realizó PCR en tiempo real con verde SYBR usando como un molde los productos de PCR de las diferentes primeras reacciones de PCR y como cebadores Read2-tag1-P7 y Read2-tag2-P7. Usando las 4 series de datos en tiempo real (2 PCR primarias con dos segmentos V diferentes y dos PCR secundarias con diferentes cebadores que contienen dos marcadores diferentes) apareció una imagen uniforme. Hubo una mejora en la eficacia entre los tamaños de salientes de 10 y 14 pb. Sin embargo hubo poca o ninguna mejora en la eficacia con un saliente de más de 14 pb. La eficacia permaneció alta a medida que el saliente se hizo tan pequeño como 14 pb debido a la alta concentración de cebadores que permiten que los 14 pb sean un molde de cebador suficiente a una temperatura mucho mayor que su temperatura de fusión. Al mismo tiempo, se mantuvo la especificidad debido a que el molde no era todo el ADNc sino más bien un producto de PCR de baja complejidad en el que todas las moléculas tenían el saliente de 14 pb.
- Como se ilustra en la **FIGURA 2B**, la PCR primaria usará 34 cebadores V diferentes que hibridan con el otro lado del molde y contienen un saliente de 14 pb común en la cola 5'. La de 14 pb es la secuencia parcial de uno de los cebadores de secuenciación Illumina (denominado el cebador Read 2). El segundo cebador de amplificación en el mismo lado incluye la secuencia P7, un marcador, y la secuencia de cebador Read 2 (este cebador se denomina Read2\_tagX\_P7). La secuencia P7 se usa para formación de grupos. Read 2 y su complemento se usan para secuenciar el segmento V y el marcador respectivamente. Se creó un conjunto de 96 de estos cebadores con marcadores numerados del 1 al 96 (véase posteriormente). Estos cebadores pueden purificarse por HPLC para asegurar que todo el material amplificado tiene extremos intactos que pueden utilizarse eficazmente en la formación de grupos.
- Como se ha mencionado anteriormente, el cebador de segundo estadio, C-10-17-P5 (**FIGURA 2A**) tiene homología interrumpida con el molde generado en la PCR del primer estadio. Se ha validado la eficacia de amplificación usando este cebador. Un cebador alternativo a C-10-17-P5, terminado en CsegP5, tiene perfecta homología con el cebador C de primer estadio y una cola 5' que porta P5. La eficacia de usar C-10-17-P5 y CsegP5 en la amplificación de moldes de PCR de primer estadio se comparó realizando PCR en tiempo real. En varias repeticiones, se descubrió que la PCR que usaba el cebador C-10-17-P5 tenía pocas o ninguna diferencia en eficacia en comparación con PCR usando el cebador CsegP5.
- La molécula resultante de la amplificación de 2 estadios ilustrada en la **FIGURA 2** tendrá la estructura típicamente usada con el secuenciador Illumina como se muestra en la **FIGURA 3**. Dos cebadores que hibridan con la parte más externa de la molécula, los cebadores de Illumina P5 (AATGATACGGCGACCACCGAG) y P7 (CAAGCAGAAGACGGCATACGAGAT) se usarán para amplificación en fase sólida de la molécula (formación de grupos). Se realizan tres lecturas de secuencia por molécula. La primera lectura de 100 pb se realiza con el cebador C', que tiene una temperatura de fusión que es apropiada para el procedimiento de secuenciación Illumina. La segunda lectura es de solamente 6 pb de longitud y será únicamente para el fin de identificar el marcador de la muestra. Se genera usando el cebador Tag Illumina (AGATCGGAAGAGCACACGTCTGAACCCAGTCAC). La lectura final es el cebador Read 2, un cebador Illumina con la secuencia GTGACTGGAGTTCAGACGTGTCTCCGATCT. Usando este cebador, se generará una lectura de 100 pb en el segmento V comenzando con la secuencia del cebador V de la primera PCR.
- Se diseñó un conjunto de marcadores de secuencia de 6 pb para distinguir diferentes muestras procesadas en el mismo carril de secuenciación, en el que cada marcador es diferente de todos los otros marcadores en el conjunto en al menos 2 diferencias. Las 2 diferencias evitan la asignación errónea de una lectura a la muestra errónea si hay un error de secuenciación. El alineamiento realizado para comparar los marcadores permitió huecos y por lo tanto un error de delección o inserción por secuenciación tampoco asignará la lectura a la muestra errónea. Características adicionales al seleccionar los marcadores fueron limitar los procesamientos de una única base (4 A o T y 3 G o C) así como ninguna similitud con los cebadores Illumina. En total se generaron 143 marcadores con la premisa de que se usarán 96 de ellos.
- Se ha realizado PCR en tiempo real con los 34 cebadores diferentes usando ADNc obtenido de una muestra sanguínea. Se obtuvieron valores de Ct diferentes para los diferentes cebadores. Cada uno de los productos de PCR se procesó por electroforesis en gel y demostró una única banda. Además los 34 cebadores se agruparon y se realizó una PCR y de nuevo se obtuvo una única banda de PCR.
- Optimización de amplificación**
- La amplificación múltiple puede usar todos los segmentos V. Un problema en la amplificación de diferentes secuencias es la eficacia de amplificación relativa de las diferentes secuencias y la conservación de la reactividad relativa inicial de las diferentes secuencias en el material amplificado final. La eficacia de amplificación relativa puede subdividirse en diferentes eficacias de las distintas secuencias de cebadores así como las diferentes eficacias de amplificación de diferentes secuencias usando el mismo cebador. La diferencia de eficacia puede deberse a

distintas secuencias de cebadores. La reacción se optimizará para intentar conseguir amplificación que mantenga la abundancia relativa de las secuencias amplificadas por diferentes cebadores de segmento V. Algunos de los cebadores están relacionados, y por lo tanto muchos de los cebadores pueden “intercomunicarse”, amplificando moldes que no coinciden perfectamente con ellos. Las condiciones pueden optimizarse de modo que cada molde 5 puede amplificarse de una manera similar independientemente de qué cebador lo amplificó. En otras palabras, si hay dos moldes, entonces después de amplificación 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 portaran un cebador diferente debido a la intercomunicación. En análisis posteriores de los datos de secuenciación 10 la secuencia de cebadores se eliminará del análisis y por lo tanto no importa qué cebador se usó en la amplificación siempre que los moldes se amplifiquen igualmente.

Puesto que la cantidad de cada molde no se conoce en ADNc, se han generado conjuntos de patrones usando la reacción de PCR de 34 cadenas individuales de ADNc. El producto en cada una estas reacciones comprendía una pluralidad de secuencias con un cebador V. Los productos diferentes se cuantificaron cuidadosamente para crear un 15 conjunto de patrones a la misma concentración. Se usó un grupo de los 34 cebadores y se realizaron 34 PCR en tiempo real usando el grupo de cebadores y cada una de las secuencias convencionales como molde. Idealmente los 34 patrones mostrarán la misma eficacia de amplificación por PCR en tiempo real. Esto sugiere que cada secuencia se amplifica igualmente incluso aunque la presencia de intercomunicación no deje claro qué cebadores 20 están llevando a cabo la amplificación. Esta optimización es coherente con el objetivo de tener amplificación igual independientemente de los cebadores reales que se incorporan en el producto de amplificación. El aumento de la concentración del grupo de cebadores total redujo significativamente el intervalo dinámico como se esperaba a partir del aumento de la eficacia de la amplificación. Además para moldes que parecían amplificarse más eficazmente que la media, se redujo la concentración de su cebador perfectamente coincidente en el grupo. Por el contrario para 25 moldes que se amplificaban ineficazmente aumentó la concentración de su cebador perfectamente coincidente. Esta optimización demostró que todos los moldes se amplifican dentro del doble de la amplificación media.

Idealmente la PCR primaria tendrá un número pequeño de ciclos para minimizar la amplificación diferencial por los diferentes cebadores. La amplificación secundaria se realiza con un par de cebadores y por lo tanto el problema de la amplificación diferencial es mínimo. El 1% de la PCR primaria se lleva directamente a la PCR secundaria. Treinta 30 y cinco ciclos (equivalentes a ~28 ciclos sin la etapa de dilución 100 veces) usados entre las dos amplificaciones fueron suficientes para mostrar una amplificación robusta independientemente de si la distribución de ciclos fue: un ciclo primario y 34 secundarios o 25 primarios y 10 secundarios. Incluso aunque idealmente la realización de solamente un ciclo en la PCR primaria puede reducir la desviación de amplificación, hay otras consideraciones. Un 35 aspecto de esto es la representación. Esta desempeña un papel cuando la cantidad de entrada de partida no excede el número de lecturas que se obtienen en última instancia. Por ejemplo, si se obtienen 1.000.000 de lecturas y se comienza con 1.000.000 millón de moléculas de entrada entonces llevar solamente representación de 100.000 moléculas a la amplificación secundaria degradaría la precisión de la estimación de la abundancia relativa de las diferentes especies en la muestra original. La dilución 100 veces entre las dos etapas significa que la representación se reducirá a no ser que la amplificación por PCR primaria generara significativamente más de 100 moléculas. Esto 40 se traduce a un mínimo de 8 ciclos (256 veces), pero más cómodamente 10 ciclos (~1.000 veces). La alternativa a esto es llevar más del 1% de la PCR primaria a la secundaria pero debido a la alta concentración de cebadores usada en la PCR primaria, puede usarse un gran factor de dilución para asegurar que estos cebadores no interfieran 45 con la amplificación y empeoren la desviación de amplificación entre secuencias. Otra alternativa es añadir una etapa de purificación o enzimática para eliminar los cebadores de la PCR primaria para permitir una menor dilución de ellos. En este ejemplo, la PCR primaria fue de 10 ciclos y la secundaria de 25 ciclos.

### Ensayos de enzimas de alta fidelidad

50 Pueden usarse enzimas con mayor fidelidad para minimizar los errores. El ensayo se ha optimizado usando Taq polimerasa. Para validar el uso de Accuprime así como Taq de alta fidelidad se amplificó un ADNc con el grupo de cebadores usando Taq polimerasa, Accuprime, o Taq de alta fidelidad. Cada uno del material amplificado se usó después como un molde para 37 PCR en tiempo real con cada uno de los 34 cebadores V y el cebador C. Se cuantificó la cantidad relativa de los moldes. Se encontró una alta correlación ( $r^2 > 0,95$ ) entre la concentración de cada molde en el Accuprime, Taq de alta fidelidad y Taq, validando el uso de estas enzimas de alta fidelidad.

### 55 Validación de las condiciones de amplificación

La optimización se realizó usando un grupo de cebadores en el molde convencional no con el fondo de ADNc. El 60 objetivo fue obtener validación para estos resultados en la mezcla de ADNc. Para mostrar reproducibilidad, se usaron grupos de oligos para amplificar el ADNc por duplicado. Cada uno de los 34 productos se cuantificó en cada una de las dos amplificaciones. Como se muestra en la FIGURA 5, la reproducibilidad fue excelente.

Para la FIGURA 5, se realizaron dos reacciones de PCR primarias usando los cebadores de TCR $\beta$  agrupados y el cebador de C y una muestra de ADNc como un molde. La abundancia relativa en cada uno del material amplificado del molde que es amplificable con cada uno de los 34 cebadores V (y el cebador C) se evaluó usando PCR en tiempo real. Usando cada uno de los dos productos amplificados como un molde, se realizaron treinta reacciones de 65

PCR en tiempo real diferentes usando el cebador C y uno de los cebadores V en cada reacción. La abundancia relativa determinada por PCR en tiempo real fue altamente reproducible usando todos los cebadores V entre las dos muestras, lo que indica que la amplificación múltiple es altamente reproducible. El número de ciclo (valor de 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 X y usando el segundo producto de amplificación múltiple como un molde se representa en el eje Y.

Para evaluar la desviación de amplificación puede emplearse una técnica similar. El grupo de oligos puede usarse para amplificar usando ADNc como un molde. Después puede cuantificarse la cantidad de molde amplificado por cada uno de los 34 cebadores diferentes (junto con el cebador de segmento C) usando PCR en tiempo real y esa cantidad puede compararse con la cantidad amplificada usando el mismo cebador del ADNc. Sin embargo, puesto que hay intercomunicación incluso si la abundancia relativa entre las secuencias internas en el producto amplificado y el ADNc es la misma, pueden detectarse grandes diferencias usando esta lectura. Para aliviar este problema se diseñaron 12 oligos que pueden, cuando se usan con el cebador de segmento C, amplificar secuencia internas de los cebadores del segmento V. Si se realiza la optimización de forma apropiada, entonces la concentración de estas secuencias internas debería cambiar poco entre el ADNc y los productos amplificados. Esto se muestra en la **FIGURA 6**.

Para la **FIGURA 6**, se usó una muestra de ADNc como un molde para una amplificación múltiple usando los cebadores de TCR $\beta$  agrupados y el cebador C. Se usaron el cebador C y cebadores que están cadena abajo (denominados cebadores internos) de los cebadores V usados para la amplificación inicial y el material del material de amplificación múltiple como un molde para evaluar la abundancia relativa de las secuencias diferentes. De forma similar se usó PCR en tiempo real para evaluar la abundancia relativa de estas mismas secuencias en el ADNc. Si la amplificación múltiple tuvo mucha desviación, la abundancia relativa en el material amplificado puede ser muy diferente de la del ADNc. Como puede verse en la **FIGURA 6**, se vio una alta correlación que demostraba desviación de amplificación mínima en la amplificación múltiple. El número de ciclo (valor Ct) para cada una de las amplificaciones por PCR en tiempo real usando cebadores internos, y ADNc y el producto de amplificación múltiple como molde se muestran en el eje X y Y, respectivamente.

### 30 Secuenciación de TCR $\beta$

Se usaron seis amplificaciones múltiples con los oligos agrupados y una muestra de ADNc como un molde. Se realizaron tres de cada una de las amplificaciones con Accuprime y otras 3 con Taq de alta fidelidad. Dos amplificaciones con cada enzima usaron ADNc que correspondía a 500 ng de ARN inicial, y una amplificación con cada enzima usaba 10 veces menos ADNc. Para cada una de las seis reacciones se realizó una PCR de 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 realizó usando los mismos conceptos descritos posteriormente.

40 Para evaluar la reproducibilidad del ensayo se determinó si los niveles de clonotipo eran uniformes en los experimentos por duplicado. Como se muestra en las **FIGURAS 8A-C**, se obtiene alta correlación cuando se usa la misma cantidad de enzima y ADNc de entrada de partida (cada una de las 2 comparaciones tuvo  $r^2=0,944$ ). Cuando se usaron diferentes enzimas la correlación empeora (mediana de la correlación para las 4 posibles combinaciones  $r^2=0,931$ ) y solo se reduce moderadamente ( $r^2=0,924$ ) cuando las dos enzimas se usaron para amplificar ADNc de entrada más pequeño (correspondiente a solamente 50 ng de ARN).

45 Para la **FIGURA 8**, se identificaron secuencias idénticas en cada muestra. Después para ocuparse de errores de secuenciación se fusionaron algunos clonotipos para formar clonotipos mayores usando los enfoques generales descritos en la sección de análisis primario de secuencia. Los recuentos de clonotipos se calcularon después en cada muestra. Una fracción de los clonotipos (no mostrada en la figura) estaba presente en una muestra pero no en otra, probablemente debido al algoritmo que las fusiona con otro clonotipo en una muestra pero no en la otra. La frecuencia de clonotipos en una muestra se calcula después como su número de recuentos dividido por el número total de lecturas obtenidas para esa muestra. Por ejemplo si se observaron 1.000 recuentos para un clonotipo en una muestra con un millón de lecturas, su frecuencia se calcula como 0,1%. La **FIGURA 8A** muestra el  $\log_{10}$  de la frecuencia de cada clonotipo en las dos muestras duplicadas usando Accuprime y ADNc correspondiente a 500 ng de ARN como un molde de entrada. La correlación ( $r^2$ ) entre estos duplicados es de 0,944. La **FIGURA 8B** 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 X) o Taq de Alta fidelidad (eje Y). Hay 4 comparaciones con esta combinación con una mediana de correlación  $r^2=0,931$ . La mostrada en la figura tiene  $r^2=0,929$ . La **FIGURA 8C** 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 X) o Taq de Alta fidelidad (eje Y). La correlación observada  $r^2=0,924$ .

50 Estos resultados validan la reproducibilidad del ensayo, y se ajusta a la expectativa de que la reproducibilidad empeora cuando se comparan diferentes enzimas. Se ve reducción adicional cuando se usa una cantidad menor de ADNc de entrada lo que refleja que la representación menor en el material de entrada conduce a peor precisión al reflejar la abundancia relativa de los diferentes clonotipos en la sangre. Adicionalmente es posible que parte de la

reducción en la correlación se deba a la amplificación adicional (10 veces) necesaria para la menor entrada, pero esto probablemente sea el efecto menor dadas las pruebas de la alta reproducibilidad de amplificación.

## EJEMPLO 4

5 Análisis del repertorio de IgH: estrategia de amplificación y secuenciación

10 Una diferencia entre la amplificación de CDR3 en TCR $\beta$  e IgH es que se usarán múltiples cebadores para cada secuencia V en la investigación de IgH debido a la posibilidad de mutaciones somáticas en IgH. Se usarán tres cebadores diferentes para cada segmento V. Los cebadores están en regiones que evitan las CDR, que tienen las mutaciones somáticas mayores. Se realizarán tres reacciones de amplificación diferentes. En cada reacción, cada uno de los segmentos V se amplificará en uno de los tres cebadores y todos usarán los mismos cebadores del segmento C. Los cebadores en cada reacción estarán aproximadamente a la misma distancia del punto de unión V-D. Suponiendo que la última posición del segmento V es 0, entonces el primer conjunto de cebadores (A) tiene el extremo 3' a aproximadamente -255, el segundo (B) tiene el extremo 3' a aproximadamente -160, y el tercero (C) tiene el extremo 3' a aproximadamente -30. Dada la homología entre varios segmentos V, para amplificar los 48 segmentos V y los muchos alelos conocidos (como se define por el sistema de información internacional ImMunoGeneTics <http://imgt.cines.fr/>) serán necesarios 23, 33 y 32 cebadores en las fases A, B y C, respectivamente. La lista de cebadores se muestra en las **Tablas 2, 3 y 4**.

15

20

**Tabla 2** Cebadores de fase A**Fase de secuencias A**

|                |                               |
|----------------|-------------------------------|
| <b>IGHV1_1</b> | CCTCAGTGAAGGTCTCCTGCAAGG      |
| <b>IGHV1_2</b> | CCTCGGTGAAGGTCTCCTGCAAGG      |
| <b>IGHV1_3</b> | CCTCAGTGAAGGTTCTGCAAGG        |
| <b>IGHV1_4</b> | GGGCTACAGTAAAATCTCCTGCAAGG    |
| <b>IGHV2_1</b> | AAACCCACACAGACCCCTACGCTGAC    |
| <b>IGHV2_2</b> | AAACCCACAGAGACCCCTACGCTGAC    |
| <b>IGHV2_3</b> | AAACCCACACAGACCCCTACACTGAC    |
| <b>IGHV3_1</b> | CTGGGCGGTCCCTGAGACTCTCCTG     |
| <b>IGHV3_2</b> | CTGGGGGGTCCCTAGACTCTCCTG      |
| <b>IGHV3_3</b> | CAGGGCGGTCCCTGAGACTCTCCTG     |
| <b>IGHV3_4</b> | CAGGGCGTCCCTGAGACTCTCCTG      |
| <b>IGHV3_5</b> | CTCCCCCCTCCCTGAAACTCTCCTG     |
| <b>IGHV3_6</b> | CTGGCAGGTCCCTGAGACTCTCCTG     |
| <b>IGHV3_7</b> | CTGGAGGGTCCCTGAGACTCTCCTG     |
| <b>IGHV3_8</b> | CTGGGAGGTCCCTGAGACTCTCCTG     |
| <b>IGHV3_9</b> | TGGGGGGGGCCCTGAGACTCTCCT      |
| <b>IGHV4_1</b> | CTTCGGAGACCCCTGTCCCTCACCTG    |
| <b>IGHV4_2</b> | CTTCGGACACCCCTGTCCCTCACCTG    |
| <b>IGHV4_3</b> | CTTCACAGACCCCTGTCCCTCACCTG    |
| <b>IGHV4_4</b> | CTTCGGAGACCCCGTCCCTCACCTG     |
| <b>IGHV4_5</b> | CGGGGACCTGTCCCTCACCTG         |
| <b>IGHV5_1</b> | GATCTCCTGTAAGGGTTCTGGATACAGCT |
| <b>IGHV6</b>   | TCGCAGACCCCTCTCACTCACCTGT     |

**Tabla 3** Cebadores para fase B**Fase de secuencias B**

|                |                           |
|----------------|---------------------------|
| <b>IGHV6</b>   | TGGATCAGGCAGTCCCCATCGAGAG |
| <b>IGHV5_1</b> | GCTGGGTGCGCCAGATGCC       |
| <b>IGHV2_1</b> | TGGATCCGTAGCCCCCAGG       |
| <b>IGHV2_2</b> | TGGATCCGTAGCCCCCGGG       |
| <b>IGHV1_1</b> | GTGCGACAGGCCCTGGACAA      |
| <b>IGHV1_2</b> | GGGTGCGACAGGCCACTGGACCA   |
| <b>IGHV1_3</b> | GTGCGCCAGGCCCGGACAA       |
| <b>IGHV1_4</b> | GGGTGCGACAGGCCCTGGACAA    |
| <b>IGHV1_5</b> | GGGTGCAACAGGCCCTGGAAAA    |
| <b>IGHV1_6</b> | GGGTGCGACAGGCCCTGGAAAA    |
| <b>IGHV1_7</b> | GTGCGACAGGCCCGGACAA       |
| <b>IGHV1_8</b> | GTGCGACAGGCCCGAGACAA      |
| <b>IGHV4_1</b> | TCCGCCAGCCCCCAGGGAAGG     |
| <b>IGHV4_2</b> | TCCGGCAGCCCCCAGGGAAGG     |
| <b>IGHV4_3</b> | TCCGGCAGCCACCAGGGAAGG     |
| <b>IGHV4_4</b> | TCCGCCAGCACCCAGGGAAGG     |

**Fase de secuencias B**

**IGHV4\_5** TCCGGCAGCCGCCGGGAA  
**IGHV4\_6** TCCGGCAGCCGCCGGGAA  
**IGHV4\_7** TCCGGCAGCCGCTGGGAAGG  
**IGHV4\_8** TCCGCCAGCCCTAGGGAAAGG  
**IGHV3\_1** GGTCCGCCAGGCTCCAGGGAA  
**IGHV3\_2** GTTCCGCCAGGCTCCAGGGAA  
**IGHV3\_3** GGTCCGCCAGGCTCCAGGGAA  
**IGHV3\_4** GGTCCGTCAAGCTCCGGGGAA  
**IGHV3\_5** GATCCGCCAGGCTCCAGGGAA  
**IGHV3\_6** GGTCCGCCAGGCTCCAGGGAA  
**IGHV3\_7** GGTCCGCCAGGCTCCAGGGCAA  
**IGHV3\_8** GGTCCGCCAGGCCAGGCAA  
**IGHV3\_9** GGTCCGCCAGGCTCCGGGGCAA  
**IGHV3\_10** GGGTCCGTCAAGCTCCAGGGAAAGG  
**IGHV3\_11** CTGGGTCCGCCAGCTACAGGAAA  
**IGHV3\_12** GGTCCGCCAGCCTCCAGGGAA  
**IGHV3\_13** GGTCCGGCAAGCTCCAGGGAA

**Tabla 4** Cebadores para fase C  
**Fase de secuencias C**

**IGHV7** CTAAGGCTGAGGACACTGCCGTGT  
**IGHV6** CTCTGTGACTCCCGAGGACACGGCT  
**IGHV5\_1** AGTGGAGCAGCCTGAAGGCCTC  
**IGHV2\_1** TGACCCAACATGGACCCTGTGGACAC  
**IGHV1\_1** ACATGGAGCTGAGCAGCCTGAGATC  
**IGHV1\_2** ACATGGAGCTGAGCAGGCTGAGATC  
**IGHV1\_3** ACATGGAGCTGAGGAGCCTGAGATC  
**IGHV1\_4** ACATGGAGCTGAGGAGCCTAAGATCTGA  
**IGHV4\_1** GAGCTCTGTGACCGCCGCGAC  
**IGHV4\_2** GAGCTCTGTGACCGCCGTGGACA  
**IGHV4\_3** GAGCTCTGTGACCGCTGCAGACACG  
**IGHV4\_4** GAGCTCTGTGACCGCTGCAGGACA  
**IGHV4\_5** GAGCTCTGTGACTGCCGAGACACG  
**IGHV4\_6** GAGCTCTGTGACTGCAGCAGACACG  
**IGHV4\_7** GAGCTCTGTGACTGCCGCGACA  
**IGHV4\_8** GAGCTCTGTGACCGCGGACGCG  
**IGHV4\_9** GGCTCTGTGACCGCCGCGAC  
**IGHV4\_10** GAGCTCTGTGACCGCCGAGACA  
**IGHV4\_11** GAGCTCTGTGACCGCTGACACGG  
**IGHV3\_1** CAAATGAACAGCCTGAGAGCCGAGGACA  
**IGHV3\_2** CAAATGAACAGCCTGAAAACCGAGGACA  
**IGHV3\_3** CAAATGAACAGTCTGAAAACCGAGGACA  
**IGHV3\_4** CAAATGATCAGCCTGAAAACCGAGGACA  
**IGHV3\_5** CAAATGAACAGTCTGAGAACTGAGGACACC  
**IGHV3\_6** CAAATGAACAGTCTGAGAGCCGAGGACA  
**IGHV3\_7** CAAATGAACAGCCTGAGAGCTGAGGACA  
**IGHV3\_8** CAAATGAGCAGCCTGAGAGCTGAGGACA  
**IGHV3\_9** CAAATGAACAGCCTGAGAGACGAGGACA  
**IGHV3\_10** CAAATGGGCAGCCTGAGAGCTGAGGACA  
**IGHV3\_11** CAAATGAACAGCCTGAGAGCCGGGAA  
**IGHV3\_12** CAAATGAACAGTCTGAGAGCTGAGGACA  
**IGHV3\_13** CAAATGAGCAGTCTGAGAGCTGAGGACA

En el lado del segmento C, dos secuencias con una diferencia de una base entre ellas (GCCAGGGGGAAAGACCGATGG y GCCAGGGGGAAAGACGGATGG) abarcan los cuatro segmentos y los múltiples alelos conocidos de IgG. Se usará un esquema similar para los dos estadios de PCR para genes de TCR $\beta$ . En el lado V, se usará el mismo saliente de 14 pb 5' en cada uno de los cebadores V. En la PCR secundaria, se emplea el mismo cebador Read2-tagX-P7 en el lado V. En el lado C se usará 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 (AATGATAACGGCGACCACCGAGATCTGGG AAG ACGAT GG GCC CTT GGT GGA) comprende la secuencia del segmento C de las posiciones 3-19 y 21-28 y se salta la posición 20 que tiene una base diferente en al menos uno de los alelos de IgG diferentes y la secuencia para P5 que puede usarse para la formación de los grupos como se muestra en la **FIGURA 4**.

Todos los cebadores en las 3 fases tuvieron éxito en la amplificación de una única banda de ADNc. De forma similar la estrategia de PCR primaria y secundaria usando los tres grupos de cebadores en la PCR primaria mostró una única banda como se muestra en la **FIGURA 7**.

Para la **FIGURA 7**, se realizó PCR múltiple usando 3 grupos de cebadores correspondientes a las 3 fases usando ADNc como un molde. Después de la PCR primaria y secundaria los productos se procesaron en un gel de agarosa. A indica el producto de PCR del grupo de oligos de la fase A. De forma similar B y C indican los productos de los grupos de B y C. M es un carril de marcadores. Se obtuvieron bandas individuales con los tamaños apropiados usando los 3 grupos.

En última instancia, las 3 reacciones diferentes de una única muestra se mezclarán después a relación equimolar y se someterán a secuenciación. La secuenciación se realizará desde ambas direcciones usando los dos cebadores Illumina. Se secuenciarán 100 pb de cada lado. Las secuencias de línea germinal máxima que abarcan los segmentos D+J son ~30 pb más largas para BCR que TCR. Por lo tanto si el resultado neto de la retirada y adición de nucleótidos en los puntos de unión (nucleótidos N y P) generan una distribución similar para IgH y TCR $\beta$ , se espera que de media 90 pb y como máximo 120 pb de secuencia después del segmento C sea suficiente para alcanzar el extremo 3' del segmento V. Por lo tanto, en la mayoría de los casos, la secuencia del cebador C será suficiente para alcanzar el segmento V. La secuenciación de uno de los adaptadores Illumina debería identificar el segmento V usado así como identificar hipermutaciones somáticas en los segmentos V. Se secuenciarán diferentes trozos de los segmentos V dependiendo de cuál de las tres reacciones de amplificación se origine la secuencia. La secuencia completa del BCR puede alinearse a partir de tres lecturas diferentes que se originaron de diferentes reacciones de amplificación. La reacción de secuenciación de un extremo que muestra la secuencia de CDR3 completa facilitará en gran medida el alineamiento preciso de diferentes lecturas.

#### 35 EJEMPLO 5

##### Análisis primario de datos de secuencia humana

40 Para cada muestra de paciente, se obtendrán aproximadamente 1 millón de lecturas de alta calidad de lecturas de extremos emparejados de 100 pb. Se supone que este millón de lecturas son independientes, originándose a partir de 1 millón o más moléculas de ARN obtenidas de 1 millón de células o más. Se eliminarán lecturas con baja calidad.

45 Se anticipará una tasa de error de ~1%. El error puede surgir de la transcriptasa inversa, amplificación durante PCR o durante la secuenciación. Se usarán enzimas a prueba de errores para las etapas de PCR; por lo tanto los errores de secuenciación (~1%) será la principal fuente de error ya que el error de la PCR es de menos de 0,1%. La relevancia de la PCR y error de transcriptasa inversa se ampliará en gran medida en situaciones en las que haya un cuello de botella. Puede producirse un cuello de botella, por ejemplo, si se comienza con más de 100.000 moléculas de ARN o una de las etapas de manipulación molecular diferentes es ineficaz para hacer a la población efectiva de moléculas 100.000. En estas situaciones, el mismo error que apareció debido a PCR o transcriptasa inversa puede aparecer en muchos grupos.

50 55 Se obtendrán datos para TCR $\beta$  e IgG. Dada la hipermutación somática en IgH y la diferencia en la estrategia de amplificación, el análisis primario de TCR $\beta$  e IgH será algo diferente.

#### TCR $\beta$

60 En un extremo, se usará el cebador del segmento C' para secuenciación. La longitud del segmento V que se secuenciará dependerá de la longitud de los nucleótidos N+P añadidos. Dado el número medio de nucleótidos añadidos, se secuenciarán aproximadamente 40 nucleótidos del segmento V. En el otro extremo se usará el cebador de secuenciación Illumina P7. Se obtendrá la secuencia de 20 pb de la secuencia del cebador V seguida de 80 pb de la secuencia del segmento V.

- Las lecturas se alinearán con los segmentos V de línea germinal (incluyendo los alelos conocidos diferentes) para asignar un segmento V a cada lectura. Las lecturas que no coincidan sustancialmente con ningún segmento V se descartarán de lecturas de análisis adicionales. Una coincidencia sustancial se definirá como una en la que no hay más de 5 errores. Dada una tasa de error aleatoria del 1% se espera que este esquema descarte <1% de las lecturas debido a error. El resto de la secuencia se asignará al segmento V que tenga la mayor coincidencia. Los cebadores que amplifican segmentos V de la misma familia con frecuencia estarán altamente relacionados con una o pocas bases de diferencia entre ellos. Estos cebadores pueden "intercomunicarse", amplificando otros miembros de la familia. Por lo tanto el comienzo de la lectura de secuencia (el cebador) puede ser para un segmento V que es diferente del segmento V al que pertenece el resto de la lectura. Se permitirán dichos casos y no se contarán como errores.
- Se evalúan las secuencias del extremo del segmento V al comienzo del segmento C. Esta secuencia se denominará la región DJ. Esta secuencia es de media de 60 bases con una fracción muy pequeña de hasta 90 bases. Es probable que se pueda asignar el segmento J en la mayoría de los casos ya que probablemente el extremo cerca del segmento C esté conservado. Por otro lado, el segmento D puede ser difícil de asignar dado su pequeño tamaño (15-16 pb) y el corte y adición que se produce en ambos de sus extremos. Debe observarse que cuando se asigna un segmento J, es posible predecir el segmento D y C. Despues de asignar el segmento, se definirá la secuencia de la región DJ. El error complica este análisis más de lo que lo hace para el alineamiento de los segmentos C y V por dos razones. En primer lugar puede producirse error en el alineamiento DJ en una de las dos lecturas que se alinean eficazmente duplicando la tasa de error de alineamiento de una única lectura con la secuencia de la base de datos.
- Adicionalmente, puede atribuirse fácilmente una diferencia de una base en el alineamiento de C o V al error (excepto por el caso poco común de un alelo de línea germinal V no descrito previamente), y las secuencias que tienen una diferencia de una base con un segmento V pueden asignarse para tener la secuencia del segmento V. Por otro lado no estará claro si una diferencia de una base entre dos lecturas de región DJ se debe a un error o una diferencia de secuencia genuina entre dos clonotipos. Hay dos posibilidades: las lecturas pertenecen al mismo clonotipo pero tienen errores, o hay dos o más clonotipos distintos. Los clonotipos se designarán como distintos cuando la probabilidad de su aparición por error de secuenciación sea baja bien debido a sus observaciones frecuentes o bien porque divergen en demasiadas bases.
- Se espera que el error de PCR se concentre en algunas bases que se mutaron en los ciclos tempranos de PCR. Se espera que el error de secuenciación se distribuya en muchas bases incluso aunque sea totalmente aleatorio ya que el error probablemente tenga algunas desviaciones sistemáticas. Se supondrá que algunas bases tendrán error de secuenciación a una mayor tasa, por ejemplo 5% (5 veces la media). Dadas estas suposiciones, el error de secuenciación se convierte en el tipo de error dominante. Distinguir los errores de PCR de la aparición de clonotipos altamente relacionados desempeñará un papel en el análisis. Dada la importancia biológica de determinar que hay dos o más clonotipos altamente relacionados, se tomará un enfoque conservativo para tomar dichas decisiones. Se considerará la detección de suficientes de los clonotipos menores para estar seguro con alta confianza (por ejemplo 99,9%) de que hay más de un clonotipo. Para ejemplos de clonotipos que están presentes en 100 copias/1.000.000, la variante menor se detectará 14 o más veces para designarla como un clonotipo independiente. De forma similar, para clonotipos presentes en 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 obtendrá con cada base secuenciada. Si la relación entre la puntuación de calidad y la tasa de error está validada anteriormente, entonces en lugar de emplear la tasa de error del 5% conservativa para toda las bases, puede usarse la puntuación de calidad para decidir el número de lecturas que es necesario que estén presentes para determinar un clonotipo independiente. Puede usarse la mediana de la puntuación de calidad de la base específica en todas las lecturas, o de forma más rigurosa, la probabilidad de ser un error puede calcularse dada la puntuación de calidad de la base específica en cada lectura, y después pueden combinarse las probabilidades (suponiendo la independencia) para estimar el número probable de error de secuenciación para esa base. Como resultado, habrá diferentes umbrales de rechazo de la hipótesis del error de secuenciación para diferentes bases con diferentes puntuaciones de calidad. Por ejemplo para un clonotipo presente en 1.000 copias/1.000.000 la variante menor se designa independiente cuando se detecta 22 y 74 veces si la probabilidad de error fue de 0,01 y 0,05, respectivamente.
- Después de designar clonotipos que aparecen con demasiada frecuencia debido a error como distintos o independientes, se considerarán criterios que permite la designación de clonotipos como independientes debido a sus diferencias en demasiadas bases. Se espera que menos del 0,1% de las veces, dos lecturas tengan más de 4 errores en 60 pb entre ellos. Por lo tanto esto se usará como un punto de corte para considerar dos clones como independientes o distintos. El algoritmo que se empleará será el siguiente. Se observará el clonotipo con el mayor número de recuentos (clonotipo 1), y se determinará si hay algún otro clonotipo que tenga el mismo segmento V y tenga 4 o menos diferencias de base con respecto a él en la región DJ. Si se identifica más de uno de dichos clonotipos, se evaluará en primer lugar el mayor de estos clonotipos. La regla descrita anteriormente se aplicará para decir si designar el clonotipo como un clonotipo independiente o el mismo como el clonotipo principal. Si no se designa como un clonotipo independiente, se contará después como si tuviera la secuencia de clonotipo 1. Al final de este ejercicio se obtendrá la secuencia y los recuentos para todos los clonotipos. Este enfoque asegura que los clonotipos no se designarán independientes cuando no lo sean. Sin embargo, algunos clonotipos verdaderamente

independientes pueden clasificarse erróneamente (no es frecuente con número bajo de diferencias con respecto al clonotipo principal) como iguales. Este tipo de error será mucho menos perjudicial que considerar dos clonotipos independientes cuando no lo son.

5 **IgH**

- Para secuenciar el extremo del segmento C, se usará el cebador Illumina. Las primeras bases para secuenciar serán el cebador del segmento C seguido de 0-2 bases del segmento C y después la región DJ. La secuencia del cebador identificará a qué isotipo pertenece la lectura específica. Todas o la mayoría de las secuencias de DJ pueden obtenerse mediante la lectura del lado del segmento C. Se espera que la región DJ tenga una media de 80 pb. Por lo tanto, una lectura de 100 pb abarcará el cebador del segmento C y la región DJ media. Algunas regiones DJ pueden ser de hasta 120 pb, y su lectura completa puede incluir datos de secuenciación de la región V (sin contar los casos en los que se encuentran 2 segmentos D en la misma IgH).
- 10 La secuencia de la región DJ obtenida del segmento C se considerará inicialmente. Se contará el número de cada secuencia única. Como se analiza para TCR $\beta$ , algunas de las secuencias relacionadas se originan probablemente del mismo clonotipo pero tienen algún error de secuenciación y PCR. Para designar dos clonotipos como distintos, se determinará si es muy probable que la diferencia haya surgido mediante error de PCR. Se usará el mismo esquema que antes para demandar un número mínimo de observaciones independientes de los clonotipos menores o un número mínimo de diferencias entre los dos clonotipos. Se emplearán las mismas normas descritas para TCR $\beta$  para asegurar que se clasifican erróneamente menos del 0,1% de los clonotipos como distintos.
- 15

20 Se realizará secuenciación de la región V usando un cebador Illumina. Habrá 3 cebadores diferentes para cada segmento V. Los cebadores se colocarán a aproximadamente -200, aproximadamente -100 y aproximadamente -30. Las primeras bases secuenciadas serán la secuencia de cebador V seguido de más de las bases del segmento V. Se asignará la lectura específica para una de las tres fases de cebadores de un segmento V específico. La asignación se realizará en primer lugar investigando la secuencia de cebador puesto que los cebadores tienen una secuencia conocida que no se puede cambiar. Los cebadores de la misma familia pueden en ocasiones tener algo de "intercomunicación" que amplifica secuencias altamente relacionadas en la misma familia. Los cebadores se usarán para asignar la familia. El segmento V específico entre los segmentos en la familia se determinará identificando la secuencia V de la familia que es más similar a la de la lectura de secuenciación. Los segmentos V en IgH pueden tener mutaciones somáticas en el transcurso del proceso de maduración de afinidad de anticuerpos y por lo tanto se permitirá una mayor proporción de diferencias de las secuencias de línea germinal que para TCR $\beta$ . Se han observado anticuerpos con más de 25 mutaciones (~10%) en la región VDJ. Se asignará una lectura al armazón del segmento V con la secuencia más cercana siempre que tenga >85% de homología con ella. Los clonotipos relacionados se evaluarán con respecto a independencia usando el mismo esquema que se ha descrito anteriormente. Específicamente para que los clonotipos se consideren distintos y debido a mutaciones somáticas, no es necesario que sean suficientemente frecuentes o tengan variación amplia entre ellos para asegurar que menos del 0,1% de los clonotipos se clasifiquen erróneamente como distintos cuando no lo son.

25

30 40 Para lecturas que se determina que tienen el tercer armazón (más cercano al punto de unión VD), se determinará la secuencia solapante entre las lecturas de extremos emparejados. Las bases no alineadas con el segmento V se alinearán con el complemento de la primera lectura para determinar el solapamiento. Los cebadores para el tercer armazón estarán a ~30 pb de distancia del punto de unión. Por lo tanto, si el V está intacto pueden usarse aproximadamente 50 pb de secuencia para alcanzar el punto de unión de VD (cebador de 20 pb + 30 pb) y las lecturas de 100 pb permitirán leer 50 pb después del punto de unión. Este es el número mínimo esperado de bases que se leería después del punto de unión VD ya que la deleción de algunas bases en V permitirá una lectura más larga después del punto de unión. Incluso para la región DJ más larga, se espera que haya un solapamiento de 10 pb entre las secuencias de las lecturas emparejadas. Se espera que la región DJ más larga sea de 120 pb, y se espera que 80 pb de ella se lean del segmento C y 50 pb se lean de la otra dirección de la región V lo que conduce a un solapamiento de 10 pb entre las lecturas emparejadas.

45

50 55 Los clonotipos que leen la misma secuencia del cebador C pero tienen diferentes armazones del mismo segmento V se convierten en candidatos para la consolidación en los mismos clonotipos. Se hay un solapamiento entre las secuencias obtenidas de los diferentes armazones del segmento V entonces la determinación de si los clonotipos son independientes o no se realiza con las mismas normas que se han descrito anteriormente.

60 Como resultado del análisis anterior, puede contarse el número de lecturas para cada clonotipo. Se identifican clonotipos que son de la misma familia que difieren entre sí solamente por mutaciones somáticas. Estas mutaciones somáticas pueden estar restringidas a secuencias en la lectura del segmento V de solamente un armazón, más de un armazón, o en la región DJ.

## EJEMPLO 6

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

- 5 Se ensayan en primer lugar si hay clonotipos que se correlacionan con la actividad de enfermedad en pacientes. En segundo lugar, se definirá un conjunto de características de secuencia y/o marcadores de superficie celular que distingan clonotipos que se correlacionan con la enfermedad de los que no. En tercer lugar, se medirá el grado en el que el análisis de clonotipo proporciona información clínicamente útil, tal como la correlación con el resultado a corto plazo (por ejemplo, 3 meses).

10 **1. Presencia de clonotipos correlacionados con la enfermedad**

Habrá dos tareas principales: identificar clonotipos correlacionados y medir la actividad de enfermedad de su nivel. Estas tareas pueden realizarse en una situación clínica en dos etapas para cada paciente:

- 15 1) Puede realizarse un ensayo de Calibración para determinar la identidad de los clonotipos correlacionados para el paciente específico. Esto puede hacerse secuenciando el ARN de IgH y TCR $\beta$  (o secuencia TCR $\alpha$ -TCR $\beta$  unida de una única célula) para cada paciente en un momento de un pico de un episodio, en cuyo momento el nivel del clonotipo correlacionado puede alcanzar sus mayores niveles.

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

25 Se usará amplificación, secuenciación y desarrollo de análisis primario como se ha descrito anteriormente para evaluar las muestras del paciente. Específicamente, se evaluará un conjunto de pacientes con lupus eritematoso sistémico (SLE) que tienen un periodo de seguimiento de un año y muestras de sangre en serie durante este periodo. Estos pacientes fueron vistos por el Dr. Michele Petri en la Escuela Médica Johns Hopkins cada tres meses durante un año, y están disponibles medidas clínicas de la actividad de enfermedad incluyendo Índice de Actividad de Enfermedad del Lupus Eritematoso Sistémico (SLEDAI), Evaluación Global del Médico (PGA), así como múltiples ensayos de laboratorio incluyendo C3 (Complemento 3) y niveles anti ADNbc para todas las visitas de todos los pacientes. Los fármacos que se administran a los pacientes incluyen prednisona, plaquenilo, AINE, Tipo AINE, dosis de ácido acetilsalicílico (ASA), plavix, diurético, inhibidores de ACE o bloqueadores del receptor de angiotensina (ARB), bloqueador del canal de calcio, Triam y solumedrol. Se estudiarán los pacientes que han tenido al menos una vez durante el seguimiento un cambio significativo en la actividad de enfermedad como se define por un cambio de 3 puntos en el SLEDAI o un cambio de 1 punto en la PGA. En general hubo 181 pacientes (con un total de 815 muestras de sangre) que cumplían estos criterios. Se sometió a ARN de todas estas muestras sanguíneas a PCR

- 40 múltiple usando cebadores descritos anteriormente para amplificar las secuencias que abarcaban CDR3 en IgH y TCR $\beta$ . Todos los materiales amplificados se secuenciarán (hasta un millón de lecturas) y se determinará la abundancia de diferentes clonotipos.

45 Usando los datos clínicos, secuenciación, se identificarán las características que distinguen clonotipos cuyo nivel se correlaciona con actividad de enfermedad de los que no. En segundo lugar, se desarrollarán un algoritmo para determinar la actividad de enfermedad usando el perfil de IgH y TCR $\beta$  en sangre.

**2. Identificación de características de clonotipos correlacionados**

- 50 Se anticipa que los clonotipos que son relevantes para la enfermedad aumentarán 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 necesariamente se correlacionan con la enfermedad. Por ejemplo, en un paciente particular podría haber 10 clonotipos enriquecidos en el momento de alta actividad de la enfermedad, pero solamente 5 se correlacionan con la enfermedad. Para identificar estos clonotipos relevantes, se estudiará un subconjunto de clonotipos que se correlacionan claramente con la enfermedad y otro conjunto que claramente no se correlaciona con la enfermedad. Se investigarán las características que distinguen esas dos clases de clonotipos.

55 Todos los pacientes tendrán al menos un cambio significativo en la actividad de enfermedad durante el seguimiento de un año en este diseño experimental. Se analizarán los clonotipos de IgH y TCR obtenidos en el pico de la actividad de enfermedad en cada paciente. Se seleccionarán conjuntos de clonotipos correlacionados y no correlacionados entre los de los clonotipos de mayor nivel. Por lo tanto la primera etapa es definir clonotipos que están a alto nivel. Los criterios específicos para seleccionar los clonotipos que entrarán en el análisis incluirán una combinación de intervalo de frecuencias del clonotipo y el nivel de clonotipo (número de lecturas de clonotipo por millón), así como pruebas de que el clonotipo no pertenece a la distribución de clonotipos de baja frecuencia.

Este conjunto de clonotipos de cada muestra de paciente, denominados Clonotipos Altamente Prevalentes (HPC) se analizará adicionalmente. Se evaluará la correlación del nivel de cada uno de estos clonotipos con medidas clínicas. Se calculará la correlación de la puntuación de SLEDAI con el nivel de clonotipo. Para cada paciente habrá 4-5 puntos de estudio que pueden usarse para evaluar la correlación de SLEDAI con el nivel de cada HPC. Se investigará la distribución de estas correlaciones obtenidas. Se anticipa que la mayoría de los HPC tendrán una baja correlación con SLEDAI. Se investigará 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  $>0,9$  aleatoriamente. Una mayor proporción de HPC con  $r^2>0,9$  indica la presencia de un clonotipo que se correlaciona con la enfermedad. Además de comparar el número de clonotipos correlacionados con expectativa aleatoria, se realizará un análisis de permutación en el que se calculará la correlación de las puntuaciones de SLEDAI de un paciente y el nivel de HPC individuales de otro. La distribución de correlaciones generadas a partir de esta permutación puede usarse como la relación "de fondo". (Para asegurar su validez, se confirmará que hay poca correlación entre SLEDAI entre diferentes pacientes). La correlación excesiva en el extremo de alta correlación, por ejemplo,  $r^2>0,9$  indicará la presencia de clonotipos que se correlacionan con la enfermedad. Se seleccionarán los clonotipos más correlacionados como el conjunto de clonotipos correlacionados. Debido a que se conoce el número de HPC que tiene una correlación aleatoria mayor que un umbral determinado (a partir del cálculo usando suposición aleatoria o mediante el análisis de permutación descrito anteriormente), el umbral para definir el clonotipo correlacionado puede establecerse de tal modo que tenga una tasa de falso descubrimiento del 10%, es decir el 10% del conjunto de clonotipos correlacionados se correlacionarán aleatoriamente. Se seleccionará un conjunto de HPC que tengan muy poca correlación con la puntuación de SLEDAI. Estos actuarán como el conjunto de clonotipos no correlacionados. Estos 2 conjuntos de clonotipos pueden analizarse adicionalmente para identificar características que pueden distinguirlos. Estas características pueden después buscarse en nuevas muestras para identificar los clonotipos que probablemente se correlacionen con la actividad de enfermedad en estas muestras. Los niveles en sangre de estos clonotipos pueden después seguirse para determinar la actividad de enfermedad.

Surge una complicación a partir de la premisa de que el nivel de clonotipo puede cambiar antes de que lo haga la actividad de enfermedad.

Por lo tanto es posible que al intentar estudiar solamente HPC altamente correlacionados con SLEDAI, puedan eliminarse clonotipos clínicamente útiles que cambian antes que SLEDAI. Puede seleccionarse otro conjunto de clonotipos que se correlaciona con una puntuación de SLEDAI Modificado (MSLEDAI). El MSLEDAI es el mismo que SLEDAI en todos los puntos de estudio excepto los que están justo antes de un cambio significativo. Para estos puntos de datos la puntuación de MSLEDAI será la media entre la puntuación de SLEDAI en ese punto y el siguiente punto de estudio. Los clonotipos que cambian antes de SLEDAI probablemente muestren mejor correlación con MSLEDAI que SLEDAI. Proporcionará información el cálculo del número en exceso de HPC que tienen mayor correlación con MSLEDAI de lo esperado por expectativas generadas por permutación o aleatorias.

Después se identificarán características que distinguen clonotipos correlacionados de los que no se correlacionan. El análisis se realizará exactamente de la misma manera para los clonotipos que se correlacionan con SLEDAI o MSLEDAI. En cualquier caso el objetivo sería que este conjunto de características recapitulen correctamente esta clasificación que permite la identificación de clonotipos correlacionados en el siguiente conjunto de muestras. Se espera que cada paciente tenga un conjunto único de clonotipos correlacionados, pero el estudio de entrenamiento se diseñará para generar las normas que predicen los clonotipos correlacionados a partir de una muestra de calibración (a alta actividad de enfermedad). Puede ensayarse dos tipos de parámetros generales: los que se obtienen de los datos de secuenciación en sí mismos, y los que pueden usar experimentación extra. La experimentación extra puede incluir la evaluación de diferentes células con diferentes marcadores de superficie celular u otros. Aquí hay algunos tipos de parámetros que se investigarán:

- 50 1) Motivo de secuencia: el motivo puede ser una región V o J específica, una combinación VJ, o secuencias cortas en la región DJ que está asociada con que un clonotipo esté correlacionado.
- 51 2) Tamaño del clonotipo.
- 52 3) Nivel: nivel absoluto (número de lecturas por millón) o nivel de rangos.
- 53 4) Similitud con otros clonotipos: la presencia de otros clonotipos altamente relacionados, como los que tienen cambios silenciosos (diferencias de nucleótidos que codifican los mismos aminoácidos) o los que tienen cambios de aminoácidos conservativos.
- 54 5) Para los BCR el nivel de mutaciones somáticas en el clonotipo y/o el número de clonotipos distintos que difieren en mutaciones somáticas de algunos clonotipos de línea germinal.

Cada uno de estos parámetros se estudiará individualmente con respecto a asociación con clonotipos correlacionados. Se atravesará un umbral de 0,05 (no corregido para múltiples ensayos) para eliminar factores que probablemente no contribuyan a la predicción de clonotipos correlacionados. Dados los múltiples parámetros, se

realizarán muchos ensayos para generar múltiples resultados positivos aleatoriamente. Sin embargo el objetivo principal de esta etapa es filtrar los parámetros para un conjunto más pequeño. El conjunto de parámetros positivos se usará después para crear un algoritmo para clasificar los dos conjuntos de clonotipos. Se empleará un algoritmo de aprendizaje mecanizado que usa los diferentes parámetros para clasificar los dos conjuntos de clonotipos. Para 5 minimizar el riesgo de sobreajuste, se usará la técnica de validación cruzada. Usando este algoritmo cada clonotipo obtendrá una puntuación que se corresponde con la probabilidad de que sea un clonotipo correlacionado. Se colocará después un umbral para clasificar los clonotipos por encima de él como correlacionados y los que están por debajo de él como no correlacionados. La precisión de la clasificación puede estimarse por la técnica de validación 10 cruzada; por ejemplo, los clonotipos se ponen en grupos iguales y el algoritmo que usa todos los clonotipos excepto un grupo. Después los clonotipos se clasifican en el último grupo (grupo de ensayo) usando el algoritmo que se obtuvo usando el resto de los clonotipos. Esto se repite tantas veces como el número de grupos, y en cada iteración todos los grupos excepto uno se usan para entrenamiento y se clasifica un grupo. La precisión del algoritmo puede 15 estimarse a partir de la precisión media de las diferentes clasificaciones en las diferentes iteraciones. Debe observarse que en todas estas iteraciones el algoritmo exacto sería ligeramente diferente. La precisión de la clasificación es entonces una estimación ya que no está en el algoritmo final sino más bien en un conjunto de algoritmos relacionados generados con datos de entrenamiento de todos los clonotipos excepto uno.

En última instancia, se generarán dos algoritmos entrenados en dos conjuntos de clonotipos correlacionados diferentes: uno correlacionado con SLEDAI y el otro correlacionado con MSLEDAI. Incluso si los clonotipos en el 20 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. Se compararán los algoritmos. Adicionalmente estos algoritmos se usarán para identificar clonotipos correlacionados que no estaban inicialmente en el conjunto de 25 entrenamiento. Se compararán los clonotipos identificados en los dos algoritmos, y si los clonotipos iniciales en los dos conjuntos de entrenamiento eran de la misma población, probablemente los clonotipos identificados sean muy similares. A no ser que los resultados del algoritmo sean bastante similares, ambos algoritmos se llevarán a cabo para identificar clonotipos correlacionados para medir la actividad de enfermedad de lupus.

Otros enfoques experimentales pueden añadir a la potencia de secuenciación en la identificación de clonotipos que 30 se correlacionan con enfermedades. Los clonotipos correlacionados pueden enriquecerse en células con algún marcador de superficie u otros. Por ejemplo, se conocen linfocitos B con altos niveles de CD27 en pacientes con lupus activo, y por lo tanto podría ser que los clonotipos correlacionados pudieran estar enriquecidos en la población de células CD27. Si esto resulta ser cierto, puede mejorarse la predicción de clonotipos correlacionados realizando 35 un enriquecimiento con respecto a células con altos niveles de CD27. Específicamente, puede realizarse una reacción de secuenciación de las secuencias de IgH de todos los linfocitos B en la muestra de sangre así como de los linfocitos B con alto CD27. Se espera que los clonotipos correlacionados estén presentes a mayor frecuencia en la población de CD27 alto que en toda la muestra sanguínea.

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

40 La sección anterior ha descrito análisis basados en clonotipo para identificar características de los clonotipos correlacionados. Además, para ese análisis solamente se usó una fracción de todos los HPC para designar claramente clonotipos como correlacionados o no correlacionados. Esta sección describe el análisis que se dirige al nivel del paciente para calcular una medida de la actividad de enfermedad, que se denominará puntuación Autoinmunitaria (AI). El algoritmo desarrollado según la sección anterior se aplicará para identificar clonotipos 45 correlacionados entre todos los HPC. Se determinará el nivel de estos HPC correlacionados. El nivel de los clonotipos correlacionados puede normalizarse con el número total de clonotipos de TCR así como con HPC que se ha predicho que no se correlacionarán con enfermedad. El nivel de estos clonotipos correlacionados en diferentes puntos temporales se usará para calcular la puntuación de AI en estos diferentes puntos.

50 En paciente con más de un clonotipo correlacionado, se combinará la información con respecto al nivel de estos clonotipos diferentes. Además se integrarán los datos de los clonotipos de IgH y TCR $\beta$ . Se intentarán diferentes algoritmos para realizar la combinación. Por ejemplo, se estudiará la media, mediana, suma y mayor nivel de 55 clonotipos correlacionados. El nivel de clonotipos puede ser sus recuentos de lectura lineal sencilla, el logaritmo de esos o alguna otra conversión. Potencialmente puede ser la diferencia entre los clonotipos correlacionados y no correlacionados. Pueden utilizarse además métodos para media ponderada. La ponderación puede basarse en la probabilidad de que un clonotipo esté correlacionado.

60 Para evaluar cuál de los modelos es óptimo, todos los modelos se evaluarán para identificar el que genera la mayor correlación entre la puntuación de AI y la puntuación de SLEDAI. Para este análisis la correlación de las puntuaciones de SLEDAI y AI se realiza entre todos los datos obtenidos de todos los puntos de estudio de todos los pacientes. Para estimar y mejorar el grado de sobreajuste, se usará la técnica de validación cruzada. El nivel de correlación medida refleja la relación "de sección transversal" entre las puntuaciones de AI y SLEDAI. Además de SLEDAI, se estudiará la correlación con otras medidas clínicas como los niveles de C3 y anticuerpo anti ADNbc así como proteína en la 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 enfermedad alta y baja, y no es necesariamente un reflejo de que la puntuación de AI se correlacione con SLEDAI.

dentro de un paciente. Para demostrar eso, se realizará evaluación “longitudinal”.

## 4. Análisis longitudinal

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

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

15 20 Otro modo de demostrar la capacidad de la puntuación de AI para medir los cambios en la actividad de enfermedad en pacientes individuales es determinando su precisión en la distinción de estados de actividad de enfermedad alta y baja en los mismos pacientes. Para cada uno de los 181 pacientes, se seleccionarán los dos puntos de estudio cuando el SLEDAI era mayor (que se denominará HDAP para el punto de actividad de enfermedad alto) y niveles menores (que se denominará LDAP para punto de actividad de enfermedad bajo). Se comparará la distribución del AI de todos los HDAP con la del AI de todos los LDAP, y se calculará el p valor de que sean diferentes. Además, se evaluará la frecuencia de que el AI en HDAP sea mayor que LDAP en cada paciente. Si el AI no cambia con la actividad de enfermedad en un paciente individual entonces se espera que el AI en HDAP sea mayor que el de LDAP solamente el 50% de las veces. Se realizará otro análisis en el que se determina la fracción de momentos en que AI en HDAP es mayor que en LDAP por una diferencia significativa (es decir, por encima de la variación de AI probable). Para medir la fluctuación de AI, se usarán todos los puntos de estudio de todos los pacientes, y se puede calcular la desviación típica (y desviación típica relativa) de AI en los diferentes grupos de valores de SLEDAI. Esto generará una desviación típica relativa entre todos los pacientes (AI-RSDall) y este valor puede depender o no de SLEDAI (es decir el AI-RSDall puede ser diferente en diferentes valores de SLEDAI). Puede calcularse la proporción de pacientes en los que el AI en HDAP es mayor que el AI en LDAP por un número específico (por ejemplo, 2) de AI-RSDall. Puede haber alguna desviación sistemática en la que el AI calculado en algunos pacientes sea uniformemente más alto (o más bajo) de lo que se esperaba de la puntuación de SLEDAI. Por lo tanto AI-RSDall es una combinación de la fluctuación intrínseca de AI dentro de un paciente así como la diferencia sistemática de AI para pacientes con SLEDAI similares. La fluctuación intrínseca de AI puede calcularse dentro de un paciente calculando la desviación típica (y desviación típica relativa) de las puntuaciones de AI entre puntos de estudio con valores de SLEDAI similares (<2 puntos de diferencia) dentro de un paciente. Puede calcularse la mediana entre todos los pacientes de la desviación típica relativa (AI-RSDpt-med). Después puede evaluarse la proporción de pacientes en los que AI en HDAP es mayor que AI en LDAP por un número específico (por ejemplo, 2) de AI-RSDpt-med.

25 30 35 40 45 50 55 60 65

45 Después de la demostración de que AI de hecho fluctúa con SLEDAI dentro de pacientes individuales se evaluará si AI puede predecir SLEDAI en el siguiente punto de estudio, 3 meses después. Para evaluar esto puede cuantificarse el nivel de correlación entre la puntuación de AI en el momento 0 y la puntuación de SLEDAI en el momento + 3 meses. La correlación puede calcularse en un nivel de paciente y después puede obtenerse la mediana de la correlación del paciente. Otro modo de demostrar la capacidad de AI para predecir actividad de enfermedad en el futuro cercano es evaluar la sensibilidad y especificidad de AI en la predicción de la actividad de enfermedad 3 meses en el futuro. Clínicamente, los pacientes que tienen buena respuesta a su tratamiento actual pueden distinguirse de los que no. Un estado del paciente en un momento particular se clasificará en una de dos clases: Control Bajo (PC) e incluye pacientes que en 3 meses tendrán alta actividad de enfermedad (SLEDAI >6 puntos) y/o un brote (aumento de SLEDAI en 3 puntos), y Control Bueno (GC) e incluye pacientes que a los 3 meses tendrán actividad de enfermedad baja o moderada (SLEDAI <6) y/o una reducción significativa en la actividad de enfermedad (reducción de SLEDAI en 3 puntos). La sensibilidad de clasificación puede después evaluarse y obtenerse especificidad usando diferentes umbrales de AI. Puede generarse 3 meses antes una curva de ROC que describa el rendimiento de AI en la predicción del estado del paciente (PC o GC). El rendimiento obtenido por este ensayo se comparará con el de medidas clínicas convencionales incluyendo SLEDAI, niveles de anti ADNbc y C3.

65

60 También se realizará un análisis para evaluar la capacidad de AI para producir cambios en las puntuaciones de SLEDAI 3 meses después. Usando datos de todos los puntos de estudio de todos los pacientes, puede representarse la relación entre las puntuaciones de AI y SLEDAI para identificar el nivel de correlación "de sección transversal" como se ha analizado anteriormente. Esto determina la relación entre SLEDAI y AI en el mismo punto de estudio. Esta relación se ajustará con una ecuación que permite la predicción de la puntuación de SLEDAI dada una puntuación de AI (o viceversa). Si el AI predice brotes entonces los cambios en SLEDAI en algún punto de

- estudio 1 se precederán por cambios en el AI en el punto 0. Por lo tanto, si se produce un brote entre el punto 0 y 1, la puntuación de AI en el punto 0 (que se denominará Almeas) será mayor de lo que se esperaba (que se denominará Alexp) dado el SLEDAI en el punto de estudio 0. Por otro lado, sin cambios en la actividad de enfermedad entre el punto de estudio 0 y el punto de estudio 1, la puntuación de AI en el punto 0 será muy similar a la que se espera dado el SLEDAI en el punto de estudio 0. El cambio de AI relativo (Rel-AI-diff) puede calcularse dividiendo la diferencia de Almeas y Alexp por Almeas. La sensibilidad y especificidad de AI en la predicción de un cambio significativo en SLEDAI 3 meses después puede evaluarse usando diferentes umbrales de Rel-AI-diff. Los umbrales pueden ser bidireccionales de modo que si el Rel-AI-diff en un punto de estudio específico es mayor que un umbral específico se predice un brote, y de forma similar si es menor que el negativo del umbral específico se espera una reducción significativa en SLEDAI. Por otro lado cuando el Rel-AI-diff en un punto de estudio está entre el umbral y su negativo, no se esperan cambios significativos en la actividad de enfermedad. Puede generarse una curva de ROC que muestra la relación de la sensibilidad y falsos positivos usando muchos umbrales diferentes de Rel-AI-diff. Pueden generarse curvas de ROC similares usando medidas clínicas convencionales incluyendo SLEDAI, niveles de anti ADNbc y C3.
- Si la fluctuación de AI varía a diferentes valores de SLEDAI, se refinará el análisis anterior. Una sección anterior ha descrito el cálculo de AI-RSDall y AI-RSDpt-med y mencionó evaluar si cambian a diferentes valores de SLEDAI. Si lo hacen entonces el análisis de ROC puede realizarse como se ha descrito anteriormente pero en lugar de usar diferentes umbrales de Rel-AI-diff, se usarán diferentes umbrales de AI-RSDall y AI-RSDpt-med. Se comparará el rendimiento obtenido por el ensayo con el de medidas clínicas convencionales incluyendo SLEDAI, niveles anti ADNbc y C3.
- En el análisis anterior, se realizaron intentos de predecir el SLEDAI en el punto 1 a partir de la puntuación de AI en el punto 0. Es probable que además del nivel absoluto en el punto 0, el cambio de AI del punto -1 al 0 sea informativo para predecir SLEDAI en el punto 1. Por ejemplo se considera un paciente que tiene en el punto de estudio -1 una puntuación de AI de X-1, y en el punto 0 la puntuación de AI aumenta a un nuevo valor X0 que es apreciablemente mayor que X-1. Este paciente puede tener mayor probabilidad de un brote en el punto 1 que un paciente cuyo AI ha sido estable en X0 en los puntos de estudio -1 y 0. Este concepto de cambio o velocidad de AI se incorporará para generar una puntuación de AI modificado (MAI). Para generar un MAI en el punto 0 serán necesarias la puntuación de AI en el punto -1 y en el punto 0, y por lo tanto un punto de datos por paciente no tendrá un MAI asociado con él. La fórmula específica para incorporar la velocidad en el cálculo de AI para obtener MAI se optimizará. Esta optimización puede realizarse mediante maximización de la correlación de MAI y SLEDAI 3 meses después. Se usará el diseño de validación cruzada para evaluar y controlar el grado de sobreajuste. Puede realizarse correlación con respecto a puntos de datos de todas las muestras, pero también puede realizarse a un nivel de paciente y puede evaluarse la mediana de la correlación entre todos los pacientes. Este último enfoque mejora el problema de que algunos pacientes tienen una desviación sistemática de puntuación de AI demasiado baja o demasiado alta. Usando MAI, puede realizarse el mismo tipo de análisis de ROC que se mencionó para AI para evaluar su capacidad para predecir SLEDAI 3 meses después. En primer lugar, de forma análoga a lo que se ha descrito para AI, puede realizarse un análisis para mostrar la capacidad de MAI en el punto 0 para distinguir los estados de PC y GC en el punto 1. Adicionalmente, puede realizarse un análisis similar al que se describió para AI para evaluar la capacidad de MAI en el punto 0 para predecir cambio de actividad de enfermedad significativo (cambio de 3 puntos en SLEDAI) entre los puntos 0 y 1. Para este último análisis se pueden usar diferentes umbrales de Rel-AI-diff, AI-RSDall o AI-RSDpt-med. El rendimiento de MAI se comparará con el de AI para determinar si la adición del factor de velocidad es útil.
- Una complicación del estudio descrito es que se realizan cambios de tratamiento para diferentes pacientes durante el periodo de seguimiento del estudio. Esto probablemente complique la predicción de la actividad de enfermedad. Por ejemplo, considérense dos pacientes con la misma puntuación de AI en el punto 0 y uno de esos pacientes tuvo una reducción en la medicación al mismo tiempo. La probabilidad de que este paciente tenga un aumento de la actividad de enfermedad en el punto 1 es por lo tanto probable que sea más alta que para el paciente que no cambió de medicación en el punto 0. Esto conduzca probablemente a una subestimación del rendimiento de AI. Un modo de aliviar esto es eliminar todos los puntos con cambios de medicación significativos del estudio. Otro es modificar la puntuación de AI para incluir si un paciente tiene un cambio de medicación y crear un AI con medicación modificada. Por lo tanto en el ejemplo anterior con los dos pacientes, el que tiene el cambio de medicación tendrá un AI de medicación modificada mayor.
- 5. Integración con otros marcadores predictivos**
- La capacidad predictiva del marcador de actividad de enfermedad puede maximizarse. Por lo tanto se ensayarán la capacidad predictiva de la información del repertorio de TCR/BCR integrada con otros marcadores. Estos marcadores incluyen marcadores convencionales usados en la clínica como los niveles de anti ADNbc y C3. También se incluirán otros marcadores que están publicados. Por ejemplo se ha mostrado ya que un panel de quimiocinas tiene cierta capacidad predictiva usando el mismo conjunto de pacientes que se van a usar. Se evaluará si este panel aumenta la capacidad predictiva del repertorio de TCR y BCR. La primera etapa es integrar la puntuación de AI con la medida adicional para generar una puntuación de AI expandido (EAI). Pueden evaluarse diferentes modos de realizar la integración, y estos pueden optimizarse mediante maximización de la correlación de

EAI y SLEDAI tres meses después. Se usará el diseño de validación cruzada para evaluar y controlar el grado de sobreajuste. Usando EAI puede evaluarse la capacidad para predecir la actividad de enfermedad 3 meses después por su capacidad para distinguir GC de PC y predecir cambios en la actividad de enfermedad. El rendimiento en la medición de la actividad de enfermedad y cambio en la actividad de enfermedad puede describirse mediante análisis de ROC como se ha descrito anteriormente.

## 6. Validación

El número de variables que se ensaya es alto en comparación con el número de muestras. Esto puede prestarse a sobreajuste, no pudiendo validarse en estudios posteriores resultados inicialmente prometedores. Se usará un enfoque de validación cruzada en el entrenamiento para obtener una medida del alcance del sobreajuste. Sin embargo, una validación de un conjunto independiente de muestras estará implicada en trabajos posteriores. 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, debería validarse y el modo específico de calcular la medida de interés. Se tomará un algoritmo específico para validación. Además se especificarán uno o más criterios de valoración específicos. La sensibilidad y especificidad de AI puede evaluarse en la capacidad para distinguir GC de PC 3 meses después para evaluar la capacidad de AI para predecir la actividad de enfermedad. En otro ejemplo, puede evaluarse la sensibilidad y especificidad de AI para predecir el cambio de actividad de enfermedad significativo en 3 meses usando un umbral específico de Rel-AI-diff.

## EJEMPLO 7

### Medir la respuesta de un paciente con SLE a la terapia farmacológica

Los métodos de la invención proporcionada se usarán para medir la respuesta de un paciente con SLE a la terapia farmacológica. La determinación de si un paciente con SLE al que se proporciona un fármaco caro con efectos secundarios graves responde al fármaco desempeña un papel tanto en el cuidado del paciente como en hacer la administración de dicho cuidado rentable. Muchos indicadores clínicos de actividad de enfermedad responden al tratamiento de forma imprecisa y después de un retardo temporal de hasta varios meses. Durante este tiempo, la enfermedad puede progresar y los efectos secundarios pueden añadir complicaciones a la terapia. Un entendimiento rápido de la respuesta farmacológica permitirá a los pacientes cambiar a terapias más eficaces más rápidamente.

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 del paciente se evalúa trimestralmente mediante una evaluación clínica exhaustiva además de ensayos de laboratorio incluyendo medición de los niveles de C3, anticuerpo anti ADNbc, recuentos sanguíneos y análisis de orina. Durante una visita el paciente se queja de lesiones cutáneas y fatiga, y el análisis de orina muestra pruebas de proteinuria y/o cilindros celulares. El reumatólogo refiere al paciente a un nefrólogo para una biopsia renal para evaluar el estado inflamatorio del riñón y pide creatinina en suero y relación de proteína y creatinina en orina de 24 horas para evaluar el grado de alteración de la función renal. Una biopsia renal muestra pruebas de nefritis lúpica difusa, mientras que el ensayo de creatinina y proteína en la orina revela pruebas de síndrome nefrótico (relación de proteína en la orina y creatinina de 3,6). Basándose en esta información se proporciona un diagnóstico de nefritis lúpica aguda y el paciente comienza un ciclo de terapia farmacológica. Hay varios fármacos posibles que pueden elegirse en este punto. Con frecuencia se usan inmunomoduladores tales como mofetil micofenolato (Cellcept) aunque algunas veces en casos graves se prescriben fármacos tales como Metotrexato, Azatiopurina (Imuran) Ciclofosfamida (cytoxan). En ocasiones también se usa Rituximab (Rituxan) como una segunda o tercera opción. Uno de estos fármacos se usa con frecuencia en combinación con esteroide sistémico tal como Prednisona o metilprednisolona para suprimir los síntomas agudos. Aquí, se prescribe mofetil micofenolato a 150 mg por día junto con 60 mg de prednisona. Dados los muchos efectos secundarios de los esteroides, incluyendo el riesgo de osteoporosis, hiperglucemia, aumento de peso y otros síntomas Cushingoides a largo plazo, la dosis de prednisona del paciente se disminuye a lo largo de ~6 semanas si la imagen clínica lo permite.

La primera cuestión que se determina es si el paciente responde a la terapia, y como resultado, si puede reducirse apropiadamente la dosis de esteroide. Por lo tanto, durante este periodo se sigue la creatinina en suero del paciente así como la creatinina y proteína en orina para asegurar que el paciente responde a las medicaciones. Se pueden realizar biopsias renales frecuentes para detectar si se está invirtiendo el daño inflamatorio; sin embargo, el uso rutinario de biopsia de riñón porta un riesgo demasiado grande y es demasiado invasiva para ser práctica. Los marcadores basados en sangre actuales que se usan para evaluar el estado inflamatorio son de uso limitado al tomar esta decisión porque no están suficientemente bien correlacionados con la enfermedad subyacente para basarse en ellos para arriesgarse a los efectos secundarios aumentados que acompañan las dosis altas de esteroides. Los marcadores de función de suero y orina pueden tener algún retardo en la detección de la mejora en el estado inflamatorio y por lo tanto los esteroides pueden reducirse antes de que estos marcadores muestren un cambio definitivo y por lo tanto extender el periodo del brote renal. Una reducción más lenta, con información de marcadores más sensibles, en estos casos podría haber acortado el periodo de brote evitando daño adicional al tejido renal. Después de la reducción del esteroide hasta una dosis de mantenimiento de aproximadamente 10 mg el paciente puede mostrar niveles persistentemente elevados de proteína en la orina y la alta relación de proteína y

creatinina en orina de 2, y el médico debe decidir ahora si cambiar de Cellcept a otra medicación. A favor de esto se argumentan las pruebas continuadas de pérdida de función renal pero sin una medida precisa del estado inflamatorio del riñón, puede ser difícil conocer si la enfermedad en sí misma está en remisión habiendo hecho no obstante algún nivel de daño renal irreversible que dé como resultado estos niveles persistentes de proteinuria. Aquí

5 de nuevo los marcadores basados en sangre existentes proporcionan información imperfecta y no son prácticas las biopsias renales adicionales. Esta decisión se vería ayudada en gran medida por una medida basada en sangre precisa del estado de la enfermedad.

10 La Carga Autoinmunitaria sería muy útil en esta situación para evaluar la respuesta a terapia midiendo la actividad de enfermedad sola o en combinación con otros marcadores de la actividad de enfermedad. Se desarrolló un algoritmo para Carga Autoinmunitaria usando el estudio descrito anteriormente. Los clonotipos correlacionados que se usarán para calcular la Carga Autoinmunitaria se medirán usando un ensayo de calibración. Este ensayo de calibración se realizará usando sangre de un paciente en un momento de actividad de enfermedad pico, por ejemplo al comienzo de la terapia. El ensayo de calibración se realizará usando sangre o como alternativa usando el tejido que está afectado (por ejemplo, biopsia renal o biopsia cutánea). En un momento posterior en el que la respuesta a terapia va a evaluarse, se tomará una muestra sanguínea y se usará junto con el ensayo de calibración para medir la carga autoinmunitaria. Esta se usará para tomar una decisión de tratamiento. Si los clonotipos correlacionados derivan de un estudio de población, no hay necesidad del ensayo de calibración y un análisis de sangre en el momento en que va a evaluarse la respuesta a la terapia es suficiente para medir la Carga Autoinmunitaria para proporcionar información para la decisión del tratamiento.

#### EJEMPLO 8

##### Determinación del momento apropiado para reducir o detener la terapia para un paciente con SLE

25 Los métodos de la invención proporcionada pueden usarse para determinar el momento apropiado para reducir o detener la terapia para un paciente con SLE. Además del retardo temporal que puede mostrarse por las medidas clínicas de la actividad de enfermedad, una dificultad adicional está en la falta de sensibilidad de estas mediciones. La enfermedad subclínica puede no obstante dar como resultado un rebrote de la enfermedad si la terapia se reduce demasiado pronto. Como resultado de esto, típicamente se administran ciclos de terapia inmunosupresora durante un periodo de tiempo que es mucho más largo de lo que es necesario para el paciente medio para asegurar que el riesgo de rebrote sea bajo para el paciente medio pero aún sea suficientemente largo para el tramo final de la distribución. Por lo tanto se produce sobretratamiento significativo, provocando efectos secundarios y costes en la mayoría de los pacientes, mientras que se produce infratratamiento de algunos pacientes provocando rebrotos potencialmente prevenibles. Un método que podría medir la actividad subclínica que era predictivo del riesgo de rebrote sería permitir que la terapia se redujera basándose en dichas medidas en lugar de basándose en el sobretratamiento por diseño.

40 En este ejemplo, el paciente del Ejemplo 7 toma prednisona y mofetil micofenolato durante un periodo de 6 meses y la relación de proteína y creatinina en orina vuelve a un nivel de 0,5. Este nivel permanece por encima del nivel de la línea basal 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 el paciente no presenta ningún otro síntoma. Al mismo tiempo el paciente experimenta niveles moderados de náuseas y aumento de peso como posibles efectos secundarios de las medicaciones que adicionalmente tienen efectos secundarios a largo plazo graves. El médico se enfrenta a una decisión difícil: equilibrar el miedo de reducir el Cellcept y/o los esteroides demasiado rápido, lo que podría dar como resultado inflamación renal renovada y probablemente daño renal irreversible a más largo plazo adicional y las reacciones adversas que puedan producirse debido a las medicaciones. Aquí de nuevo desempeñaría un papel una evaluación no ambigua del estado de la enfermedad sin tener que realizar una biopsia renal al tomar esta decisión. Se recomienda intentar reducir los esteroides mediante ensayos repetidos de esteroides que conducen a la recurrencia del mismo dilema clínico. De hecho esta cuestión surge cada vez que el paciente está en remisión y el paciente toma esteroides o inmunomoduladores.

45 La Carga Autoinmunitaria sería muy útil en esta situación para evaluar si reducir o no la terapia midiendo la actividad de la enfermedad sola o en combinación con otros marcadores de la actividad de enfermedad. Se desarrollará un algoritmo para la Carga Autoinmunitaria usando el estudio descrito anteriormente. Los clonotipos correlacionados que se usarán para calcular la Carga Autoinmunitaria se medirán usando un ensayo de calibración. Este ensayo de calibración se realizará usando sangre de un paciente en un momento de actividad de enfermedad pico, por ejemplo al comienzo de la terapia. El ensayo de calibración podría realizarse usando sangre o como alternativa usando el tejido que esté afectado (biopsia renal o biopsia cutánea). En un momento posterior en el que se va a evaluar el nivel de actividad de enfermedad, puede tomarse una muestra sanguínea y usarse junto con ensayo de calibración para medir la Carga Autoinmunitaria. Esta se evaluará para tomar una decisión de tratamiento y evaluar si el paciente tiene alguna actividad de enfermedad detectable. Si los clonotipos correlacionados derivan de un estudio de poblaciones, entonces no es necesario el ensayo de calibración y un análisis de sangre en el momento en el que va a evaluarse la respuesta a la terapia es suficiente para medir la Carga Autoinmunitaria para proporcionar información para la decisión de tratamiento.

## EJEMPLO 9

Predicción de brotes en un paciente con SLE

- 5 Un reto en el tratamiento de pacientes con SLE es el representado por el hecho de que surgen brotes sin advertencia previa, frustrando así los intentos de los médicos de tratar la enfermedad de forma preventiva. Esperar a que se produzcan brotes antes de comenzar el tratamiento somete a los pacientes a síntomas clínicos potencialmente destructivos, puede implicar hospitalización cara e inconveniente, y puede provocar que se realicen daños orgánicos a largo plazo necesitando además intervenciones terapéuticas agresivas que tienen en sí mismas
- 10 múltiples efectos secundarios. Un paradigma mucho más deseable sería un paradigma terapéutico en el que los brotes se detectaron en una fase subclínica momento en el cual podría administrarse terapia de forma proactiva salvando sufrimiento significativo al paciente, dando como resultado hospitalizaciones menos caras y en última instancia permitiendo un mejor pronóstico a largo plazo para los pacientes.
- 15 El paciente en el Ejemplo 7 se recupera del brote agudo descrito anteriormente y se reducen todas las terapias del paciente excepto Plaquenil y una dosis baja de 5 mg de Prednisona. No obstante este paciente permanece con alto riesgo de tener otro episodio inflamatorio. Como resultado, este paciente permanecerá bajo los cuidados de un reumatólogo que continuará siguiendo los síntomas clínicos del paciente y ensayos de laboratorio. Desafortunadamente estos síntomas y ensayos no proporcionan advertencias tempranas para un brote inminente
- 20 hasta que los pacientes han mostrado de hecho síntomas clínicos de un brote y la secuencia se repite. Podría incluirse un marcador altamente específico del aumento de la actividad subclínica en la evaluación clínica rutinaria del paciente para detectar señales no ambiguas de un brote que podrían alcanzar un estadio clínicamente detectable en los siguientes 1-3 meses. Comenzar las terapias antes podría hacer al brote menos grave y podría permitir conseguir tratamiento con menos daño orgánico a largo plazo o menos esteroides usados de lo que sucede
- 25 actualmente.

La Carga Autoinmunitaria sería muy útil en esta situación para evaluar la probabilidad de un brote incipiente midiendo la actividad de enfermedad sola o en combinación con otros marcadores de la actividad de enfermedad. Esta puntuación por sí sola o la tasa de aumento (velocidad) o aceleración de esta puntuación pueden usarse para evaluar la probabilidad de progresión hasta un brote. Podría desarrollarse un algoritmo para Carga Autoinmunitaria usando el estudio descrito anteriormente. Los clonotipos correlacionados que se usarán para calcular la Carga Autoinmunitaria podrían medirse usando un ensayo de calibración. Este ensayo de calibración podría realizarse usando sangre de un paciente en un momento de actividad de enfermedad pico, por ejemplo al comienzo de la terapia. El ensayo de calibración podría realizarse usando sangre o como alternativa usando el tejido que está afectado (por ejemplo biopsia renal o biopsia cutánea). En un momento posterior en el que va a evaluarse la respuesta a la terapia, puede tomarse una muestra sanguínea y usarse junto con el ensayo de calibración para medir la Carga Autoinmunitaria. Esta puede usarse para tomar una decisión de tratamiento. Si los clonotipos correlacionados derivan de un estudio de poblaciones, no es necesario el ensayo de calibración y un análisis de sangre en el momento en el que va a evaluarse el riesgo de brote es suficiente para medir la Carga Autoinmunitaria para proporcionar información para la decisión del tratamiento.

## EJEMPLO 10

Medición objetiva para evaluar síntomas subjetivos de pacientes con SLE

- 45 El SLE afecta a muchos órganos y produce muchos síntomas potenciales incluyendo algunos que son muy comunes en las poblaciones sanas. Por ejemplo, si un paciente con SLE se queja de una cefalea, la cefalea puede ser una señal de lupus del SNC o puede deberse a la cefalea normal. De forma similar, si pacientes con SLE se quejan de fatiga que empeora durante un periodo de tiempo, la fatiga que empeora puede deberse a deterioro de su enfermedad o puede deberse a depresión u otras causas. La disponibilidad de una medida objetiva que refleja actividad de enfermedad puede ser de gran ayuda en el tratamiento de pacientes con SLE.

55 La paciente en el Ejemplo 7 se presenta al traumatólogo con quejas principales de cefalea, fatiga y dificultad con la concentración. La cefalea de la paciente es recurrente y solamente mejora transitoriamente con tratamiento con Motrin. El SLE de la paciente está en lo demás bien controlado. Los tensores psicosociales relevantes en la vida de la paciente incluyen que está pasando por un divorcio. Los médicos tienen un dilema cuando se enfrenta a pacientes con SLE con síntomas que son no específicos del SLE y son habituales en la población general. ¿Padece el paciente lupus del SNC? ¿O podría estar padeciendo otras causas comunes de sus síntomas, como depresión? Los ensayos de laboratorio actuales actualmente carecen de la sensibilidad y especificidad para basarse en ellos para distinguir estas posibilidades. Un ensayo fiable para medir la actividad de enfermedad del SLE puede utilizarse de forma rutinaria para ayudar a distinguir las dos posibilidades.

- 60 La Carga Autoinmunitaria sería muy útil en esta situación para evaluar objetivamente la actividad de enfermedad bien sola o bien en combinación con otros marcadores de actividad de enfermedad. Se desarrollará un algoritmo para la Carga Autoinmunitaria usando el estudio descrito anteriormente. Los clonotipos correlacionados que se usarán para calcular la carga autoinmunitaria se medirán usando un ensayo de calibración. Este ensayo de

calibración se realizará usando sangre de un paciente en un momento de actividad de enfermedad pico, por ejemplo al comienzo de la terapia. El ensayo de calibración se realizará usando sangre o como alternativa usando el tejido que está afectado (por ejemplo biopsia renal o biopsia cutánea). En un momento posterior en el que va a evaluarse la actividad de enfermedad objetiva, puede tomarse una muestra de sangre y usarse junto con el ensayo de

5 calibración para medir la Carga Autoinmunitaria. Esto se usará para tomar una decisión de tratamiento. Si los clonotipos correlacionados derivan de un estudio de poblaciones, no hay necesidad del ensayo de calibración y un análisis de sangre en el momento en el que va a evaluarse la actividad de enfermedad objetiva es suficiente para medir la Carga Autoinmunitaria para proporcionar información para la decisión del tratamiento.

## 10 EJEMPLO 11:

## Medir la respuesta a terapia farmacológica de un paciente con MS

15 Como se ha indicado anteriormente, uno de los principales retos en la terapia de MS es medir si un paciente responde a una terapia farmacológica y cuánto. Durante enfermedad de estadio progresivo y tardío hay evaluaciones clínicas tales como la puntuación del estado de discapacidad expandido (EDSS) que miden el grado de deterioro físico que ha resultado de la enfermedad. Sin embargo, estas evaluaciones no son útiles en el estadio temprano de enfermedad recidivante/remitente. Pueden usarse parámetros clínicos en torno a recaídas para evaluar la progresión de la enfermedad, pero estos son indicadores aproximados y retardados, ya que los pacientes pueden 20 pasar varios años entre recaídas, durante los que pueden vislumbrarse pocas pruebas a partir de las evaluaciones clínicas. En último lugar, puede usarse formación de imágenes cerebrales tales como MRI potenciada por gadolinio para examinar lesiones cerebrales. Los pacientes con MS se someten típicamente a dicha MRI cada año. Sin embargo, dichas imágenes carecen de especificidad. Además, como una medida del daño cerebral integrado, no 25 son buenas medidas de actividad de enfermedad actual sino que más bien reflejan el historial de la enfermedad y su impacto en el cerebro.

30 Aunque es verdad que el paradigma de tratamiento clínico actual para MS es que los pacientes diagnosticados con enfermedad recidivante remitente deberían tener terapia continua para retardar la aparición de la enfermedad progresiva, el repertorio creciente de fármacos aprobados para tratar MS hace la falta de retroalimentación biológica crecientemente problemática. La lista mostrada anteriormente de fármacos aprobados para tratar MS continúa haciéndose más larga a medida que la importante inversión en terapias de MS comienza a dar sus frutos. Cada uno de estos fármacos tiene efectos secundarios graves y es muy caro de administrar, con costes de 30.000-100.000 \$ al año de tratamiento. Los pacientes que no se tratan bien pasarán antes a enfermedad progresiva que es debilitante y provoca intervenciones de cuidados sanitarios caras incluyendo hospitalizaciones y cuidados a largo plazo. Por lo 35 tanto, puede permitirse que el paciente reciba terapia óptima pronto en el tratamiento.

## **Ejemplo de utilidad clínica**

40 Perfil del paciente: una mujer de 30 años de edad viene al hospital con deterioro visual monocular con dolor. Se le realiza una evaluación neurológica y una punción lumbar para obtener líquido cefalorraquídeo que se usa para evaluar si están presentes linfocitos T clonales. También se le encarga una MRI cerebral. Basándose en estos ensayos, se realiza un diagnóstico de MS. Se le prescribe Betaseron 250 µg por inyección para autoadministrar por vía subcutánea cada dos días. En una vista de seguimiento seis meses después, la paciente se queja de depresión y aumento de peso. No se han indicado acontecimientos neurológicos adicionales al médico. El médico se enfrenta 45 ahora con un dilema clínico. ¿Debería el médico mantener la terapia como se ha administrado? ¿Debería usarse una nueva terapia? ¿Debería el médico pedir un MRI incurriendo en coste y sometiendo al paciente a exposición a contraste adicional? ¿Debería el médico esperar hasta que el siguiente MRI programado muestre nuevas lesiones? ¿Debería el médico esperar para ver si reaparecen brotes? Todas estas decisiones se beneficiarían de una medida no ambigua de si la enfermedad está activa o no.

50 La Carga Autoinmunitaria sería muy útil en esta situación para evaluar la respuesta a terapia midiendo la actividad de enfermedad bien sola o bien en combinación con otros marcadores de la actividad de enfermedad. Se desarrollará un algoritmo para la Carga Autoinmunitaria usando los estudios descritos en el presente documento.  
55 Los clonotipos correlacionados que se usen para calcular la Carga Autoinmunitaria se medirán usando un ensayo de calibración. Este ensayo de calibración se realizará usando sangre de un paciente en un momento de actividad de enfermedad pico, por ejemplo al comienzo de la terapia. El ensayo de calibración podría realizarse usando sangre o como alternativa usando el tejido que está afectado (por ejemplo LCR). En un momento posterior en el que va a evaluarse la respuesta a terapia, puede tomarse una muestra sanguínea y usarse junto con el ensayo de calibración para medir la Carga Autoinmunitaria. Esto puede usarse para tomar una decisión de tratamiento. Si los clonotipos correlacionados derivan de un estudio de población, no hay necesidad del ensayo de calibración y un análisis de sangre en el momento en el que va a evaluarse la respuesta a terapia es suficiente para medir la Carga Autoinmunitaria para proporcionar información para la decisión del tratamiento.  
60

## EJEMPLO 12:

Predicción de brotes de MS

- 5 Como en todas las enfermedades autoinmunitarias, el alivio de brotes es el principal objetivo de la terapia. No solamente son los brotes debilitantes para el paciente y caros de tratar, sino que se cree crecientemente que cada brote contribuye a progresión de enfermedad a más largo plazo no reversible. Pueden usarse varias terapias para controlar los brotes incipientes tales como metilprednisolona IV o prednisona oral. Dichas medicaciones tienen efectos secundarios significativos y como tal no se prescriben sin pruebas de un brote activo. Podría usarse una medida de actividad subclínica creciente que se correlacionaba con brotes clínicos posteriores para proporcionar información para este tipo de tratamiento de brote proactivo que podría dar como resultado brotes más cortos y menos perjudiciales. Además hay terapias que demuestran alta eficacia clínica para la reducción de brotes que portan riesgos de efectos secundarios muy significativos y letales. Uno de dichos fármacos es Tysabri, un fármaco que se ha visto que da como resultado tanto resultados clínicos mejorados como aumento del riesgo de infecciones cerebrales letales tales como PML. Estos riesgos han reducido el valor de dichos fármacos a terapia de última línea cuando se ha demostrado que otros fármacos ya no controlan la progresión y ha limitado el valor de estos fármacos como tratamientos crónicos. Un ensayo que pudiera predecir cuándo el estado de brote es incipiente podría aumentar la utilidad de dichos fármacos ya que podría usarse de una manera similar a esteroides para controlar los períodos de brote agudos minimizando a la vez los riesgos de efectos secundarios letales.

20 Ejemplo de utilidad clínica

El paciente del Ejemplo 11 toma Betaseron durante 3 años y presenta un brote clínico que dura una semana. La MIR del paciente al final del año muestra nuevas lesiones significativas (múltiples lesiones hiperintensas FLAIR y T2W dirigidas perpendicularmente ovoides de tamaño variable discretas (placas), que aparecen isohipointensas en imágenes T1W e hiperintensas en imágenes T2W que implican regiones de materia blanca subcortical y periventricular bilateral, incluyendo la interfaz calloso septal). El médico se preocupa porque el paciente tiene un alto riesgo de brotes durante el transcurso de los siguientes 12 meses. Se presenta un dilema clínico. ¿Espera el médico a síntomas clínicos adicionales para intervenir con terapia adicional? ¿Debería el doctor cambiar terapias? Si es así, ¿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 controlar la actividad de enfermedad subclínica y mostrar cuándo aumenta la enfermedad y cuando se produce probablemente un brote podría usarse para ayudar a tomar estas decisiones clínicas.

35 La Carga Autoinmunitaria sería muy útil en esta situación para evaluar el riesgo de brote midiendo la actividad de enfermedad bien sola o bien en combinación con otros marcadores de actividad de enfermedad. Podría desarrollarse un algoritmo para Carga Autoinmunitaria usando los estudios descritos en la presente invención. Los clonotipos correlacionados que se usarán para calcular la Carga Autoinmunitaria podrían medirse usando un ensayo de calibración. Este ensayo de calibración podría realizarse usando sangre de un paciente en un momento de actividad de enfermedad pico, por ejemplo al comienzo de la terapia. El ensayo de calibración podría realizarse usando sangre o como alternativa usando el tejido que está afectado (por ejemplo LCR). En un momento posterior en el que va a evaluarse el riesgo de brote, puede tomarse una muestra de sangre y usarse junto con el ensayo de calibración para medir la Carga Autoinmunitaria. Esta puede usarse para tomar una decisión de tratamiento. Si los clonotipos correlacionados derivan de un estudio de población, no hay necesidad del ensayo de calibración y un análisis de sangre en el momento en el que va a evaluarse el riesgo de brote es suficiente para medir la Carga Autoinmunitaria para proporcionar información para la decisión del tratamiento.

## EJEMPLO 13

50 Control de la conformidad con la terapia para MS

Debido a la relativa infrecuencia de los 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 se prescriben son tanto caras como inconvenientes para el paciente, lo que implica autoinyecciones que pueden provocar una reacción dolorosa y efectos secundarios. Hay como resultado un grado significativo de falta de conformidad con los regímenes terapéuticos que son difíciles de controlar para un 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 doctor como al paciente ver de forma rutinaria lo bien controlada que está la enfermedad subyacente. Se ha mostrado que dichos métodos son muy eficaces en pacientes con VIH para motivarlos a seguir la terapia eficazmente. Un análisis de sangre que se realizará trimestralmente permitiría al médico ver al paciente y medir el estado de la enfermedad.

65 La Carga Autoinmunitaria sería muy útil en esta situación para evaluar la conformidad con la terapia midiendo la actividad de enfermedad sola o en combinación con otros marcadores de la actividad de enfermedad. Se desarrollará un algoritmo para Carga Autoinmunitaria usando los estudios descritos en el presente documento. Los clonotipos correlacionados que se usarán para calcular la Carga Autoinmunitaria se medirán usando un ensayo de

calibración. Este ensayo de calibración se realizará usando sangre de un paciente en un momento de actividad de enfermedad pico, por ejemplo al comienzo de la terapia. El ensayo de calibración podría realizarse usando sangre o como alternativa usando el tejido que está afectado (por ejemplo LCR). En un momento posterior, en el que va a evaluarse la conformidad con la terapia, se tomará una muestra de sangre y se usará junto con el ensayo de calibración para medir la Carga Autoinmunitaria. Esta se usará para tomar una decisión de tratamiento y guiar mejor al paciente a una mejor conformidad. Si los clonotipos correlacionados derivan de un estudio de población, no hay necesidad de un ensayo de calibración y un análisis de sangre en el momento en el que va a evaluarse la conformidad con la terapia es suficiente para medir la Carga Autoinmunitaria para proporcionar información para la decisión del tratamiento.

10

#### EJEMPLO 14

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

15 Se desarrollará un esquema de amplificación y secuenciación para TCR $\beta$  e IgH de ratón que es similar al desarrollado para seres humanos. Se aplicarán métodos similares para minimizar las diferencias en la eficacia de amplificación de diferentes secuencias y técnicas de validación similares usando adiciones y la técnica de 5' RACE descrita anteriormente. La cantidad de aportación mínima de ADNC se determinará en una metodología similar como se ha descrito para muestras humanas. Una diferencia en el esquema de amplificación entre ratón y seres humanos 20 es que los dos segmentos T para TCR $\beta$  en ratón no tienen ningún polimorfismo en los 50 pb más cercanos al punto de unión J/C. Por lo tanto, en el esquema el cebador para la amplificación de primer estadio se colocará en las posiciones 25-50, y para la amplificación del segundo estadio el cebador se colocará en las posiciones 1-25, y el cebador tendrá una cola 5' para el último cebador que contiene la secuencia P5. Las diferentes secuencias 25 mejorarán la especificidad y son similares a la estrategia usada en seres humanos excepto que no hay necesidad de "formar bucle" con ninguna base para polimorfismos.

#### EJEMPLO 15

##### Ánálisis primario de datos de secuencias de ratón

30 El armazón de análisis que se usará para análisis de datos de ratón es similar al descrito anteriormente para los datos humanos. Una diferencia es que las muestras de ratón se secuenciarán en menos profundidad que las muestras humanas. Se anticipa que las muestras sanguíneas del ratón serán de 100  $\mu$ l. En 100  $\mu$ l de sangre hay ~100.000 linfocitos y por lo tanto la secuenciación a una profundidad mucho mayor de 100.000 no mejora 35 significativamente la precisión. Por lo tanto, solo se obtendrán 100.000 lecturas para cada muestra de ratón. Incluso aunque el número de lecturas será menor para ratón que para seres humanos, se tomarán muestras de una fracción mayor de linfocitos totales y de sangre de ratón. Se espera que el número de linfocitos de ratón totales sea más de 3 órdenes de magnitud menor que el de seres humanos. De forma similar 100  $\mu$ l de sangre proporcionarán una mejor muestra (~10%) de los linfocitos en la sangre de ratón en el momento en que se comparan con las muestras 40 obtenidas usando 10 ml de sangre humana (0,2%).

#### EJEMPLO 16

##### Ánálisis de repertorio de IgH y TCR en modelo de SLE de ratón

45 Se usará un modelo de ratón de SLE para estudiar la relación entre el repertorio de TCR/BCR y la actividad de enfermedad. El modelo de ratón será el B6 con los loci sle1 y sle3 de NZM2410. Estos ratones B6.sle1.sle3 (BSS) desarrollan nefritis de tipo SLE de una manera espontánea. Se estudiarán tres tipos de cohortes. Para todos los puntos de estudio, se obtendrán BUN de sangre, creatinina y autoanticuerpos anti nucleares, proteína en orina y el nivel de creatinina. Se determinará si una puntuación generada a partir del repertorio de TCR/BCR de sangre se correlaciona bien con estos índices medidos de enfermedad renal. La primera cohorte será similar a la cohorte humana descrita cuando se recojan muestras de sangre longitudinales junto con la evaluación de la función renal. Específicamente, se seguirán 7 ratones BSS mensualmente hasta el mes 8. Al final, estos ratones se sacrificarán y además de la sangre, se analizan tejidos del bazo y del riñón. Como control, se evaluarán 5 ratones B6 de una manera similar. Las segundas cohortes serán de corte transversal en las que se sacrificarán diferentes cohortes de animales en momentos específicos y se analizarán en ese momento muestras de bazo, riñón y sangre. Específicamente, se sacrificarán 5 ratones BSS cada mes y se analizará la sangre, el bazo y el riñón. Como control, se evaluarán dos ratones de control B6 de la misma manera. Finalmente, se tratará una tercera cohorte con esteroides después de la aparición de la enfermedad y evaluación de nefritis y muestras de sangre obtenidas 50 regularmente después de eso. Específicamente a los 4 meses de edad, se tratarán 20 ratones que tienen la enfermedad con esteroides y después semanalmente durante los siguientes 4 meses se toma sangre para análisis del repertorio de TCR/BCR y evaluación de la función renal. Como control se tratarán 5 ratones BSS con placebo y se seguirán de una manera similar. Se realizará análisis del repertorio de TCR y BCR a partir de todos los puntos de estudio (es decir diferentes puntos temporales y diferentes tejidos para el mismo punto temporal). El análisis 55 implicará PCR de 2 estadios, procesamiento de secuenciación y análisis de datos primarios como se ha descrito

anteriormente.

EJEMPLO 17

5 Identificación y dinámica de los clonotipos que se correlacionan con SLE de ratón

En primer lugar, se identificará un conjunto de clonotipos que se correlacionan con la función renal. Como medida de la función renal, puede usarse la relación de creatinina/proteína en orina, creatinina en suero o niveles de BUN. En la primera y tercera cohortes, puede evaluarse la correlación del nivel en sangre de cada clonotipo de HPC con cada una de las tres medidas. De una manera similar a lo que se ha descrito en seres humanos, puede evaluarse si hay un gran aumento en el número de clonotipos con alta correlación con 1, 2 o las 3 medidas de función renal frente a la expectativa aleatoria (o ensayos de permutación). Dada la expectación aleatoria, el umbral de correlación se seleccionará donde se espere que solamente el 10% de los clonotipos con un nivel de correlación por encima de ese umbral tengan el nivel de correlación observado al azar (10% de falso descubrimiento). Se ha hecho énfasis en estos clonotipos, y este conjunto se definirá como “clonotipos correlacionados”.

Además de este método estadístico para identificar clonotipos correlacionados, los clonotipos podrían identificarse como relevantes para la enfermedad por un método “funcional” de enriquecimiento de clonotipos específicos en tejido renal. Por el método funcional puede identificarse un conjunto de clonotipos en la cohorte 2 que pueden ser relevantes para la enfermedad, y estos se denominarán clonotipos correlacionados identificados funcionalmente. Puede evaluarse el alcance del solapamiento entre la definición “estadística” y la definición “funcional” de clonotipos correlacionados. Las cohortes 1 y 3 tienen muestras renales recogidas en el último punto temporal. Puede evaluarse si los clonotipos enriquecidos en estas muestras renales están presentes en la sangre y están entre los clonotipos con mayor correlación con la función renal.

25 La dinámica de los clonotipos correlacionados (identificados estadística y funcionalmente) puede evaluarse después. Por ejemplo, usando datos de la cohorte 2, se evaluará el ciclo temporal de la subida y bajada (si la hubiera) de sus niveles en los tres compartimentos: riñón, sangre y bazo.

30 En los clonotipos correlacionados identificados estadísticamente, se identificaría un subconjunto de los clonotipos correlacionados según su correlación con la función renal. Los clonotipos correlacionados pueden identificarse sin conocer los datos de la función renal. En otras palabras, pueden entenderse las características que distinguen los clonotipos correlacionados, de los que son irrelevantes para enfermedad. Para hacerlo se identificará un conjunto de clonotipos con baja correlación con la función renal como clonotipos no correlacionados de control.

35 Características de los clonotipos que se correlacionan con la enfermedad

40 Después de la identificación de los dos conjuntos de clonotipos, correlacionados y no correlacionados, se buscarán características que distingan estos dos conjuntos. Se realizará análisis separado y combinado usando los clonotipos correlacionados identificados estadística y funcionalmente. Se evaluará el mismo tipo de características estudiado en seres humanos, por ejemplo el nivel del clonotipo, la presencia de motivos de secuencia particulares, y la secuencia de otros clonotipos relacionados. Como se ha descrito para el estudio humano, hay un riesgo significativo de sobreajuste y por lo tanto es necesario emplear técnica de validación cruzada o conjuntos de ensayo y entrenamiento separados.

45 Una utilidad para el experimento de ratón es la disponibilidad de células que permiten la evaluación de si los clonotipos correlacionados están enriquecidos en un subtipo específico de células. Se estudiará si los clonotipos correlacionados están enriquecidos en algunos subtipos celulares; puede realizarse secuenciación a partir del conjunto completo de linfocitos y del subtipo específico donde los clonotipos correlacionados están enriquecidos, y este criterio de enriquecimiento puede usarse como una característica extra para distinguir clonotipos correlacionados de otros clonotipos irrelevantres para la enfermedad. Para saber qué clonotipos de subtipos celulares están enriquecidos pueden usarse un par de enfoques: conducido por hipótesis y sin hipótesis. El primer es probar una docena de marcadores de superficie candidatos en linfocitos T o B en un conjunto de muestras. Por ejemplo, un candidato es CD69 en linfocitos T para seleccionar linfocitos T activados. Para linfocitos B los estudios han mostrado el aumento de células CD27<sup>alto</sup> en SLE activo, y por lo tanto ese es un buen candidato para un marcador de células que puede tener enriquecimiento de los clonotipos correlacionados. En cada uno de estos experimentos, los subtipos celulares específicos se purifican mediante FACS. Después se realiza una reacción de secuenciación para ADNc del complemento completo de los linfocitos así como para ADNc de los linfocitos que se purificaron por FACS de una colección de muestras diferentes. Se evaluará si los dos conjuntos de clonotipos correlacionados y no correlacionados están presentes en diferentes proporciones en el complemento completo de linfocitos en comparación con el subconjunto purificado por FACS. Los marcadores que tienen una gran diferencia pueden ser útiles en la identificación de clonotipos correlacionados. El enriquecimiento de clonotipos en subtipos de células con estos marcadores se usará además de los parámetros de secuencia para detectar clonotipos correlacionados.

En el enfoque sin hipótesis, se buscarán marcadores que se expresen diferencialmente en células con un clonotipo correlacionado de otras células. Se seleccionarán algunos casos en los que un clonotipo de TCR específico está claramente correlacionado con la enfermedad, y se seleccionarán casos en los que esté altamente enriquecido el clonotipo que represente la mayoría de los clonotipos con el mismo segmento V. Se realizará FACS usando anticuerpos para el segmento V específico (están disponibles en el mercado anticuerpos contra todos los segmentos V) para seleccionar una población que está altamente enriquecida con respecto a células que portan el clonotipo correlacionado. El ARN puede prepararse a partir de estas células y puede estudiarse la expresión de todos los genes realizando un experimento de matrices. Como control, puede usarse ARN total de linfocitos y/o ARN de células purificadas por FACS que portan otro segmento V irrelevante. Pueden buscarse marcadores que distingan de forma máxima la muestra obtenida del segmento V purificado con FACS con el clonotipo correlacionado de los controles. Pueden encontrarse marcadores, incluyendo marcadores de superficie (puesto que es mucho más fácil realizar FACS con proteínas de superficie) que distinguir las dos poblaciones. Si se observa un marcador de ARN uniforme de muestras de varios ratones, este se validará al nivel proteico. Usando las mismas muestras, se usarán anticuerpos contra la proteína marcadora en un ensayo de FACS para purificar células que portan la proteína marcadora. Puede ensayarse más de un marcador para aumentar la probabilidad de validar uno de ellos. Se secuenciará el TCR y/o BCR de las células purificadas. Si los resultados de ARN se mantienen al nivel proteico entonces los clonotipos correlacionados deberían estar enriquecidos en el subconjunto purificado de células. Despues de validar que los resultados de ARN aún se mantienen a nivel proteico los resultados se validarán en otras muestras. Las muestras que no se sometieron al análisis de matrices, se someterán a análisis de FACS usando el anticuerpo para la proteína o las proteínas marcadoras. Se secuenciará el TCR y/o BCR de las células purificadas. Se evaluará si los clonotipos correlacionados están enriquecidos en las células purificadas usando anticuerpo para el marcador o los marcadores específicos. Esto validará la utilidad del marcador o los marcadores en la identificación de clonotipos correlacionados.

## 25 EJEMPLO 18

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

El algoritmo para clonotipos correlacionados anterior puede aplicarse para identificar en todas las muestras de cohortes 1 y 3 clonotipos correlacionados según su secuencia y/o marcadores. Usando el nivel de los clonotipos correlacionados en cada paciente, puede generarse una puntuación de 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 de entrenamiento y ensayo separados. La correlación de AI y las medidas de función renal puede evaluarse de una manera de sección transversal (todos los puntos de estudio de todos los ratones). También puede evaluarse la cuestión de si la puntuación de AI cambia en un ratón individual cuando cambia la función renal. Esto puede evaluarse comparando el AI de función renal alta y baja en el mismo animal de una manera similar a lo que se ha descrito en seres humanos.

## 40 EJEMPLO 19

### Unión de secuencias de la misma célula

Pueden amplificarse dos secuencias de la misma célula, y durante la amplificación pueden unirse para formar un amplicón. La información de la presencia de estas dos secuencias en la misma célula puede conservarse después incluso si las secuencias unidas están mezcladas con un grupo de secuencias de otras muestras.

Un ejemplo de la utilidad de este esquema de unión es para evaluación de la diversidad de TCR. La diversidad de TCR se genera a partir de la diversidad de cada TCR $\alpha$  y TCR $\beta$ . Además, la combinación de un TCR $\alpha$  y TCR $\beta$  en una célula aumenta significativamente la diversidad. Sin embargo, cuando se extraen ácidos nucleicos de una muestra con una pluralidad de linfocitos T, se pierde la información de qué TCR $\alpha$  está presente en la misma célula que TCR $\beta$ . Se presenta aquí un método que permite la conservación de esta información. Este método comprende separar las células en compartimentos distintos, amplificar la secuencia deseada de una manera que una covalentemente amplicones inicialmente separados y opcionalmente mezclar todas las secuencias amplificadas para análisis posterior. Pueden concebirse varios métodos para colocar cada célula en un compartimento. Por ejemplo, un método es poner células en una microgota o una emulsión de micelas que puede usarse en PCR. Estas gotas pueden cargarse de una manera directa o cargarse aleatoriamente de tal manera que la mayoría de las gotas contengan como máximo una única célula. Además, puede usarse separación celular para colocar una única célula en un recipiente de PCR. Despues puede realizarse amplificación de ácido nucleico en cada gota.

## 60 Esquema 1

Como se ilustra en la **FIGURA 9**, la secuencia 1 puede amplificarse usando cebador 1 y cebador 2. El cebador 2 porta una secuencia saliente 5' que no es complementaria de la secuencia genómica (**FIGURA 9A**, línea fina). De forma similar la secuencia 2 puede amplificarse usando el cebador 3 y cebador 4 (**FIGURA 9A**). El cebador 3 porta una secuencia saliente 5' que es complementaria de la secuencia saliente del cebador 2 (**FIGURA 9A**, línea discontinua). En esta figura los dos salientes (o las dos secuencias de enlace) que representan dos secuencias

complementarias se dibujan en líneas finas; una secuencia se muestra como una línea continua y su complemento se muestra como una línea discontinua. Se dibujan otras secuencias complementarias que tienen los mismos colores sólidos: negro y gris para la secuencia 1 y 2, respectivamente.

- 5 Despues de la amplificación con los cebadores 1-4, cada uno de los dos productos de amplificación tiene una secuencia de enlace en un extremo y los dos productos pueden hibridar entre sí y las cadenas pueden extenderse para formar una molécula bicatenaria completa (**FIGURA 9B**). Esta molécula tiene ahora las secuencias 1 y 2 enlazadas entre sí y puede despues amplificarse con los cebadores 1 y 4 (**FIGURA 9C**).
- 10 Los 4 cebadores pueden ponerse en la reacción en el mismo momento para conseguir enlace y amplificación de secuencia. Puede ser beneficioso añadir concentración baja de los cebadores 2 y 3. La baja concentración de los cebadores 2 y 3 asegurará que los dos amplicones de secuencia individuales alcancen la saturación pronto en la reacción permitiendo que el amplicón enlazado domine la reacción de PCR en los estadios posteriores de la reacción. Esto conducirá a que la reacción final tenga una alta concentración del amplicón enlazado en relación con 15 los amplicones de secuencia individuales.

#### Esquema 1(a)

- 20 El esquema 1(a) es una variante del Esquema 1 en el que la secuencia de enlace es idéntica a la secuencia del cebador 2 (**FIGURA 10**). La secuencia 1 puede amplificarse usando el cebador 1 y cebador 2 sin secuencias salientes de los cebadores (**FIGURA 10A**). El cebador 3 porta una secuencia 5' que es complementaria del cebador 2 (**FIGURA 10A**). La secuencia 2 puede amplificarse usando el cebador 3 y el cebador 4 creando una secuencia de enlace que es complementaria de la Secuencia 1 (**FIGURA 10A**). Otras secuencias complementarias se dibujan con los mismos colores: negro y gris para secuencia 1 y 2, respectivamente.
- 25 Despues de amplificación con los cebadores 1-4, los dos productos pueden hibridar entre sí mediante la secuencia del Cebador 2 y pueden extenderse cadenas para formar una molécula bicatenaria completa (**FIGURA 10B**). Esta molécula tiene ahora las secuencias 1 y 2 enlazadas entre sí y puede despues amplificarse con el cebador 1 y 4 (**FIGURA 10C**).

30 Esquema 2

- 35 El esquema 2, mostrado en la **FIGURA 11** es un esquema similar al Esquema 1 excepto que se consigue la amplificación última con secuencias que no son complementarias del genoma. Una ventaja de este enfoque es que las secuencias cebadoras pueden seleccionarse para ser ideales para amplificación sin amplificación fuera de la diana. Esto puede ser útil en casos en los que los cebadores complementarios a la secuencia genómica para amplificar no son ideales. Usando cebadores no complementarios para el genoma para amplificación, pueden usarse concentraciones bajas de los cebadores 1-4, minimizando la amplificación fuera de la diana. Además, el esquema 2 puede adaptarse para un esquema múltiple en el que se usa más de un par de cebadores sin provocar 40 tantas interacciones de cebadores. Cada par de secuencias para enlazar tendrá sus propios 4 cebadores únicos que no es necesario que estén a alta concentración. Un par de cebadores de amplificación puede amplificar todos los pares de secuencias enlazadas (**FIGURA 11C**).

- 45 La secuencia 1 puede amplificarse usando el cebador 1 y cebador 2 (**FIGURA 11A**). Los cebadores 1 y 2 portan en sus extremos 5' secuencias salientes distintas que no son complementarias de la secuencia genómica (**FIGURA 11A**, líneas discontinuas y finas, respectivamente). De forma similar, la secuencia 2 puede amplificarse usando el cebador 3 y el cebador 4 (**FIGURA 11A**). Los cebadores 3 y 4 portan en sus extremos 5' secuencias salientes distintas que no son complementarias de la secuencia genómica (**FIGURA 11A**). El saliente en los cebadores 1 y 4 se marca como "Amp 1" (punteado) y "Amp 2" (ondulado) y son secuencias no complementarias del genoma que se 50 usa en última instancia para amplificación (**FIGURA 11A**). De forma análoga al esquema 1, los salientes del cebador 2 (fino) y 3 (fino/discontinuo) son las secuencias de enlace que son complementarias entre sí. Se dibujan otras secuencias complementarias que tienen los mismos colores: negro y gris para la secuencia 1 y la secuencia 2, respectivamente.

- 55 Despues de amplificación con los cebadores 1-4, cada uno de los dos productos de amplificación tiene una secuencia de enlace en un extremo y los dos productos pueden hibridar entre sí y las cadenas pueden extenderse para formar una molécula bicatenaria completa (**FIGURA 11B**). Esta molécula tiene ahora las secuencias 1 y 2 enlazadas entre sí y puede despues amplificarse con el cebador 5 y 6 (**FIGURA 11C**).
- 60 Opcionalmente, pueden usarse inicialmente los cebadores 1-4, y despues del enlace de las dos secuencias, pueden añadirse los cebadores 5 y 6. Una realización más preferida tendrá todos los cebadores añadidos en la primera etapa. Otra realización más preferida tendrá todos los cebadores presentes inicialmente con la concentración de cebadores 1-4 menor que la de 5 y 6. Esto permite que se produzca enlace y amplificación completo en una etapa. La baja concentración de los cebadores 2 y 3 asegurará que los dos amplicones de secuencia individuales alcancen 65 la saturación pronto en la reacción permitiendo que el amplicón enlazado domine la PCR en los estadios posteriores de la reacción. Esto conducirá a que la reacción final tenga una alta concentración del amplicón enlazado en relación

con los amplicones de secuencia individuales. Además, la baja concentración de los cebadores 1-4 minimiza cualquier amplificación fuera de la diana que pueda producirse si estos cebadores fueran de menor calidad que los cebadores 5 y 6.

- 5 El uso de los cebadores 5 y 6 para amplificación permite una multiplicación más eficaz (**FIGURA 12**). Un par de cebadores (cebadores 5 y 6) puede usarse para amplificar todas las secuencias enlazadas. Las secuencias enlazadas pueden diseñarse de diferentes maneras para aplicaciones diferentes. El ejemplo ilustrado en la **FIGURA 12** es para enlazar dos pares de secuencias, pero este esquema puede extenderse adicionalmente a decenas, centenares o miles de secuencias. Si hay un conjunto de pares de genes para enlazar (por ejemplo, TCR $\alpha$  con 10 TCR $\beta$  e IgH con IgK) entonces las secuencias de enlace para cada par puede ser diferente. En este ejemplo la secuenciación de enlace para TCR $\alpha$  y  $\beta$  será diferente de la de IgH e IgK como se representa por líneas discontinuas gruesas (TCR $\alpha$  y TCR $\beta$ ) o líneas discontinuas finas (IgH con IgK) (**FIGURA 12A**). Todas las secuencias amplificadas de este ejemplo se muestran en el mismo color. Los cebadores de amplificación para todas las secuencias enlazadas serán los mismos cebadores: 5 y 6 como se representa en la **FIGURA 11**. En otras 15 aplicaciones pueden usarse las mismas secuencias de enlace si no hay emparejamiento específico.

También se concibe que pueden enlazarse más de 2 secuencias. Por ejemplo pueden enlazarse 3 o más secuencias entre sí (**FIGURAS 13A-13D**). Para crear una molécula que enlaza 3 secuencias, uno de los productos puede tener dos secuencias de enlace diferentes en sus extremos, enlazándose cada uno con un producto (**FIGURA 20 13A**). En el ejemplo representado, la secuencia 2 tiene dos secuencias de enlace. La secuencia de enlace del cebador 3 permite el enlace con la secuencia 1 mediante la secuencia de enlace del cebador 2 (par complementario de la secuencia de enlace LS 1). De forma similar, la secuencia de enlace del cebador 4 permite el enlace con la secuencia 3 mediante la secuencia de enlace del cebador 5 (par complementario de la secuencia de enlace 2, LS 2) (**FIGURA 13A**). En otro ciclo la totalidad de la secuencia 2 se convierte en una secuencia de enlace para enlazar la secuencia 1 y la secuencia 3. Las secuencias Amp 1 y Amp 2 complementarias de los cebadores 1 y 6 permiten la amplificación después de formación de una molécula con secuencias enlazadas 1-3.

#### EJEMPLO 20

30 Control de la recurrencia metastásica en paciente con cáncer de colon

Muchos cánceres que se detectan en un estadio tratable aún portan un riesgo continuo para el paciente de recurrencia de tumor metastásico. Dichas recurrencias con frecuencia se detectan tarde y en estadios intratables y pueden ser letales para los pacientes. Un ejemplo de dicha situación es la del cáncer de colon recurrente. A pesar de 35 programas de exploración para cáncer de colon crecientemente agresivo, el cáncer de colon representa uno de los tumores malignos más habituales en los Estados Unidos. Se diagnostica aproximadamente a 150.000 pacientes cada año cáncer de colon en estadios graves pero tratables (Estadio II y Estadio III). Estos pacientes se tratan por resección tumoral seguida de un ciclo de quimioterapia. Aunque estos tratamientos son generalmente eficaces, hay no obstante una probabilidad significativa de que estos pacientes tengan recurrencias metastásicas del tumor 40 primario en los años después del tratamiento. El 50% de los pacientes del Estadio III, por ejemplo, tendrán una recurrencia en un periodo de 5 años de cirugía. Estas recurrencias pueden ser aisladas (por ejemplo en el colon o el hígado) o multifocales. En cualquiera de los casos, pero particularmente si están aisladas, detectarlas en un estadio temprano puede desempeñar un papel en la maximización de las probabilidades de terapia exitosa (cirugía y/o quimioterapia).

45 Hay actualmente dos ensayos usados en vigilancia post tratamiento. Se usa exploración de CT del abdomen y el pecho para identificar tumores visibles en estas imágenes. Típicamente estas exploraciones se realizan en intervalos de 6-12 meses durante los primeros 5 años después de la terapia. Aunque estas exploraciones pueden revelar tumores malignos en estadio temprano, su eficacia clínica está en entredicho. Las desventajas de estas 50 exploraciones incluyen el hecho de que someten a los pacientes a cantidades significativas de radiación lo que en sí mismo puede provocar tumores adicionales y el gasto significativo. Otro ensayo basado en sangre que se ha mostrado que tiene algún valor es el ensayo de CEA. Este ensayo de anticuerpos mide el nivel de una proteína en suero que es específica para algunos tumores de colon. La desventaja del ensayo de CEA es su falta de sensibilidad (<60% de pacientes con exploraciones de CT positivas tienen un ensayo de CEA positivo).

55 En esta realización de la invención, se usan linfocitos obtenidos del tumor primario resecado para desarrollar un perfil inmunitario que pueda usarse para añadir sensibilidad a un ensayo basado en sangre para recurrencia de cáncer temprana. Los TCR (y/o BCR) de los linfocitos hallados en el tumor resecado pueden amplificarse y secuenciarse. Los clonotipos que están enriquecidos en la muestra tumoral probablemente sean relevantes para la respuesta inmunitaria al tumor. Pueden usarse extracciones de sangre posteriores del paciente para evaluar el nivel de estos clonotipos. Un aumento en el nivel de estos clonotipos puede señalizar una respuesta inmunitaria a una recurrencia tumoral. En este caso la detección de la respuesta inmunitaria puede ser más sensible que la detección 60 del marcador tumoral en sí mismo.

**Estudio de descubrimiento para la detección de recurrencia del cáncer usando un ensayo de calibración**

Se concibe que puede realizarse un estudio de descubrimiento para determinar la probabilidad de detección de la recurrencia dado el perfil de TCR (y/o BCR) en sangre. Pueden usarse muestras de tumor resecado así como

5 muestras sanguíneas de seguimiento de pacientes con resultado conocido para este estudio. Pueden secuenciarse TCR (y/o BCR) de todas estas muestras. Son candidatos para los clonotipos correlacionados los que están presentes en los datos de TCR (y/o BCR) de las muestras tumorales. Dados los resultados conocidos en este estudio de entrenamiento se puede idear usando las técnicas de validación cruzada convencionales, un modelo que 10 genere una puntuación (Riesgo de Recurrencia) dado el nivel de los clonotipos diferentes. Esta puntuación de recurrencia puede después calcularse en un nuevo paciente midiendo los clonotipos en el tumor resecado (puntos de calibración) y los datos de los clonotipos hallados en la sangre del mismo paciente en un momento posterior durante la vigilancia con respecto a recurrencia. El uso de los datos tumorales permite una reducción en el número 15 de clonotipos presentes en la sangre que se consideran en este análisis.

**15 Estudio de descubrimiento para detección de recurrencia de cáncer usando un ensayo de calibración y un estudio de población**

Es probable que no todos los clonotipos que están enriquecidos en la muestra de ensayo tumoral sean relevantes para la respuesta inmunitaria al tumor. Podría haber algún linfocito que se expandiera localmente debido a una

20 condición inflamatoria favorable. En otra realización de la presente invención, el estudio de descubrimiento puede realizarse usando las mismas muestras pero el estudio se usa para identificar parámetros que distinguen clonotipos "correlacionados" de "no correlacionados". Estos parámetros pueden incluir 1) Motivo de secuencia: el motivo puede ser una región V o J específica, una combinación VJ, o secuencias cortas en la región DJ que se asocian con que un 25 clonotipo sea correlacionado; 2) Tamaño del clonotipo; 3) Nivel: nivel absoluto (número de lecturas por millón) o nivel de rangos; 4) Similitud con otros clonotipos: la presencia de otros clonotipos altamente relacionados, como los que tienen cambios silenciosos (diferencias de nucleótidos que codifican los mismos aminoácidos) o los que tienen cambios de aminoácidos conservativos; 5) Para los BCR el nivel de mutaciones somáticas en el clonotipo y/o el 30 número de clonotipos distintos que difieren en mutaciones somáticas de algún clonotipo de línea germinal. 6) Presencia en una célula que porta un marcador específico. Este estudio dará como resultado después un algoritmo que puede predecir qué clonotipos probablemente se correlacionen con la recurrencia del cáncer en sangre tomada de un conjunto específico de clonotipos presentes en una muestra tumoral dada. Estos clonotipos pueden usarse después para desarrollar una puntuación de Riesgo de Recurrencia del mismo modo que se ha descrito anteriormente.

**35 Estudio de descubrimiento para la detección de recurrencia de cáncer usando un estudio de población**

En otra realización de la presente invención, los clonotipos medidos en el tumor resecado se usan para generar un modelo que predice clonotipos correlacionados en muestras aún no vistas. Este modelo también puede usarse para 40 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 de cáncer resecado en un nuevo paciente que se somete a vigilancia de recurrencia sino que más bien el riesgo de recurrencia podría evaluarse simplemente

midendo los clonotipos en una muestra de sangre dada.

**Estudio de descubrimiento para la detección de cáncer de colon primario usando un estudio de población**

45 Como extensión se concibe que la detección de cánceres primarios pueda conseguirse usando la misma metodología. Con los cánceres primarios no hay tumor resecado que pueda usarse para enriquecer con respecto a clonotipos relevantes. Sin embargo, incluso en presencia de datos de resección de tumor, se concibe que es necesario usar secuencias adicionales y otros parámetros para identificar clonotipos relevantes y en última instancia 50 generar una puntuación con respecto a probabilidad de detección de cáncer. Por lo tanto, por extensión si el algoritmo es suficientemente predictivo se puede detectar el cáncer de sangre (u otro fluido corporal) sin los datos del tumor resecado. En esta realización de la invención, es necesario que esté disponible un estudio de descubrimiento con muestras de sangre de pacientes que preceden a su diagnóstico de cáncer primario. De una 55 manera análoga a la descrita anteriormente, pueden identificarse parámetros (secuencia y otros) para predecir los clonotipos que se correlacionan con la respuesta del sistema inmunitario al tumor. Después puede usarse un modelo para generar una puntuación de riesgo de cáncer que predice el riesgo de progresión a cáncer de colon. Este algoritmo puede después aplicarse a una nueva muestra de sangre del paciente para medir el riesgo de cáncer de colon primario.

**60 EJEMPLO 21****Control con respecto a rechazo en pacientes de trasplante de corazón**

65 Los trasplantes de corazón son un procedimiento relativamente poco común ya que el suministro de órganos es muy limitado. Se realizan cada año en todo el mundo 3.500 trasplantes de corazón. Cada procedimiento es muy caro y

los órganos que se usan tienen valor incalculable. Como resultado los pacientes que reciben estos órganos se tratan

de forma extremadamente proactiva. Para medir el estado de la reacción inmunitaria al órgano donado en un momento en el que pueden ser eficaces intervenciones con inmunosupresores, se realiza a los pacientes biopsias cardíacas periódicas para medir la inflamación del órgano. Basándose en estos ensayos, pueden proporcionarse ciclos agresivos de inmunosupresores. Estos procedimientos tienen varias limitaciones. Como procedimiento quirúrgico invasivo tienen riesgos para el paciente. Además son caros y pueden realizarse normalmente a intervalos infrecuentes. Se ha mostrado que un ensayo basado en sangre basado en la realización del perfil de expresión de un panel de 11 genes de ensayo (Alomapa) es bastante sensible para detectar rechazo de órganos pero carece de suficiente sensibilidad para usar como un reemplazo de biopsia y se usa en su lugar para decidir cuándo realizar una biopsia. En una realización de la presente invención se usan perfiles de TCR (y/o BCR) para evaluar el estado de "rechazo" y generar una puntuación de Riesgo de Rechazo que prediga la probabilidad de rechazo en un marco temporal específico. Se concibe que puede realizarse un estudio de descubrimiento para determinar la probabilidad de rechazo dado el perfil de TCR en sangre (y/o BCR). Este puede usarse en la clínica para proporcionar información para las terapias inmunosupresoras que se usan.

15 **Descubrimiento de clonotipos correlacionados usando un estudio de población**

En esta realización de la invención puede usarse una población de pacientes post trasplante con muestras sanguíneas con resultado clínico conocido. Puede secuenciarse el TCR (y/o BCR) de todas estas muestras y puede usarse la correlación de clonotipos individuales con resultado de rechazo para distinguir los clonotipos correlacionados de no correlacionados. Posteriormente, pueden derivarse parámetros que distinguen estas dos clases de clonotipos. Estos parámetros pueden incluir 1) Motivo de secuencia: el motivo puede ser una región V o J específica, una combinación VJ, o secuencias cortas en la región DJ que se asocian con que un clonotipo esté correlacionado; 2) Tamaño del clonotipo; 3) Nivel: nivel absoluto (número de lecturas por millón) o nivel de rangos; 4) Similitud con otros clonotipos: la presencia de otros clonotipos altamente relacionados, como los que tienen cambios silenciosos (diferencias de nucleótidos que codifican los mismos aminoácidos) o los que tienen cambios de aminoácidos conservativos; 5) Para los BCR el nivel de mutaciones somáticas en el clonotipo y/o el número de distintos clonotipos que difieren en mutaciones somáticas de algún clonotipo de línea germinal. 6) Presencia en una célula que porta un marcador específico. Aparecería un método alternativo o complementario para definir el clonotipo correlacionado y no correlacionado si las muestras de estudio tienen muestras de biopsia del injerto, particularmente si estuviera en rechazo activo. Se espera que en ese momento haya un gran enriquecimiento de los clonotipos correlacionados. Pueden identificarse parámetros para distinguir estos de los otros clonotipos como se ha analizado anteriormente.

35 Los datos del perfil de las muestras sanguíneas se usan después para predecir la probabilidad de rechazo. Dados los resultados conocidos en este estudio de entrenamiento se puede idear un modelo usando las técnicas de validación cruzada convencionales que generen 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 relacionada con la probabilidad de rechazo.

40 **Descubrimiento de clonotipos correlacionados usando un ensayo de calibración**

En otra realización puede implementarse un método para identificar clonotipos correlacionados usando un ensayo de calibración para cada paciente. Este método implica tomar una primera muestra de biopsia después del trasplante. La presencia del material de biopsia del injerto después del trasplante ofrece la posibilidad de analizar los TCR de la muestra de biopsia para identificar los clonotipos correlacionados como se define por los que son prevalentes en esta muestra. Este conjunto de clonotipos puede después seguirse en sangre y se genera una puntuación para la probabilidad de rechazo. El algoritmo para generar la puntuación de riesgo de rechazo deriva mediante un estudio de descubrimiento que es similar al descrito anteriormente que utiliza los datos clínicos disponibles y los niveles de los clonotipos correlacionados para generar una puntuación de Riesgo de Rechazo que se aproxima a la probabilidad de rechazo.

55 En esta realización se realizará un ensayo de calibración específico usando material de una primera biopsia después del trasplante pero podrían reemplazarse biopsias adicionales por el uso de muestras sanguíneas cuyos clonotipos podrían usarse junto con este ensayo de calibración para medir una puntuación de riesgo de rechazo.

60 Además de la biopsia de injerto, se pueden usar las muestras sanguíneas antes del trasplante como otro punto de calibración. Los clonotipos que son prevalentes en esta muestra probablemente no estén relacionados con el rechazo, representando más bien el historial de抗ígenos previos que ha visto el paciente. Por lo tanto, cuando se consideran las muestras sanguíneas después del trasplante se pueden restar los clonotipos que estaban presentes antes del trasplante al determinar los clonotipos correlacionados. Estos clonotipos pueden usarse después para generar un modelo de Riesgo de Rechazo.

65 En esta realización, podrían usarse dos ensayos de calibración: uno antes del trasplante y uno de una biopsia después del trasplante. Estas calibraciones podrían usarse después junto con clonotipos derivados de un análisis de sangre para medir el Riesgo de Rechazo.

**Descubrimiento de clonotipos correlacionados usando un ensayo de calibración y un estudio de población**

En otra realización, la identificación de los clonotipos correlacionados puede conseguirse mediante una combinación de los enfoques anteriores. Específicamente, esto puede conseguirse usando el estudio de población para generar

- 5 un algoritmo para predecir clonotipos correlacionados. Además puede conseguirse mediante datos de calibración del mismo paciente usando biopsia de injerto y/o muestras sanguíneas pretrasplante. Una realización más preferida empleará ambos enfoques: algoritmo construido en población y calibración individual para identificar de forma más precisa los clonotipos correlacionados. Se genera después una puntuación de Riesgo de Rechazo usando el nivel de estos clonotipos para predecir la probabilidad de rechazo mediante el uso del estudio de población como un

10 conjunto de entrenamiento.

En esta realización, pueden usarse dos ensayos de calibración: uno antes del trasplante y uno de una biopsia después del trasplante. Estas calibraciones podrían después usarse junto con clonotipos derivados de una muestra sanguínea para medir el Riesgo de Rechazo.

- 15 La predicción de GVHD puede realizarse de una manera muy similar con el mismo concepto del estudio de población para generar un algoritmo que prediga clonotipos correlacionados. Además puede generarse la calibración “negativa” de la muestra del donante antes del trasplante. Un enfoque que use tanto el algoritmo como la calibración probablemente sea más predictivo de los clonotipos correlacionados. Puede generarse un algoritmo para calcular 20 una puntuación de la probabilidad de GVHD dado el nivel de los clonotipos correlacionados usando un estudio de población de una manera que se ha descrito anteriormente. Este algoritmo puede usarse después para la predicción de la probabilidad de GVHD en el siguiente conjunto de pacientes.

**EJEMPLO 22****Control de la infección de PML en pacientes con MS tratados con natalizumab**

Una realización de la invención usa el perfil de TCR y/o BCR para detectar Leucoencefalopatía Multifocal Progresiva subclínica (PML) en pacientes con MS. La PML es una enfermedad grave y con frecuencia letal que provoca con

- 30 frecuencia enfermedad desmielinizante rápidamente progresiva mediante la destrucción de oligodendrocitos que sintetizan mielina. Está provocada 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 lo que conduce al desarrollo de esta enfermedad grave. Además algunos pacientes que están inmunosuprimidos mediante el uso de medicación como pacientes post trasplante también pueden desarrollar PML. Se ha ligado alguna medicación

- 35 específica con el riesgo de PML en poblaciones de pacientes específicas. Por ejemplo se ha asociado natalizumab (Tysabri) con el desarrollo de más de 10 casos de PML entre pacientes con esclerosis múltiple (MS) lo que condujo a su retirada del mercado durante un periodo de tiempo. Estaría aceptado que el natalizumab es más eficaz que las otras medicaciones aprobadas por la FDA para esclerosis múltiple, pero su uso se ha limitado por el miedo al desarrollo de PML. Una vez que se sospecha que hay PML, puede realizarse plasmáferesis para reducir la

- 40 concentración del fármaco en el paciente. El solapamiento entre los síntomas de MS y PML puede en ocasiones retardar la detección de PML. Se necesita con urgencia detección temprana de PML subclínica.

Estos clonotipos pueden diferenciarse de muestras sanguíneas de una población en la que algunos pacientes desarrollaron PML. Esta población puede usarse para identificar clonotipos que se correlacionan con el desarrollo

- 45 posterior de PML. Con la disponibilidad de estos clonotipos puede generarse un algoritmo para identificar parámetros que distinguen estos de otros clonotipos.

**Descubrimiento de clonotipos correlacionados usando un estudio de población**

- 50 En este caso se genera un algoritmo para predecir los clonotipos que son relevantes para la aparición de PML. El algoritmo puede entrenarse en un conjunto de clonotipos que se considera que están correlacionados con la enfermedad. En esta realización de la invención pueden usarse muestras sanguíneas (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 acaban desarrollando PML. Puede secuenciarse el TCR (y/o BCR) de todas estas muestras y puede usarse

- 55 la correlación de clonotipos individuales con resultado de reactivación de agente infeccioso para distinguir clonotipos correlacionados de no correlacionados. Pueden identificarse parámetros que distinguen estas dos clases de clonotipos. Estos parámetros pueden incluir 1) Motivo de secuencia: el motivo puede ser una región V o J específica, una combinación VJ, o secuencias cortas en la región DJ que se asocian con que un clonotipo esté correlacionado;

- 60 2) Tamaño del clonotipo; 3) Nivel: nivel absoluto (número de lecturas por millón) o nivel de rangos; 4) Similitud con otros clonotipos: la presencia de otros clonotipos altamente relacionados, como los que tienen cambios silenciosos (diferencias de nucleótidos que codifican los mismos aminoácidos) o los que tienen cambios de aminoácidos conservativos; 5) Para los BCR el nivel de mutaciones somáticas en el clonotipo y/o el número de clonotipos distintos que difieren en mutaciones somáticas de algún clonotipo de línea germinal. 6) Presencia en una célula que porta un marcador específico. Un método alternativo o complementario para definir el clonotipo correlacionado y no correlacionado vendría de un conjunto de pacientes que montan una respuesta inmunitaria al mismo agente infeccioso. Los clonotipos enriquecidos (particularmente los que están a un nivel significativamente mayor que antes

de la respuesta inmunitaria) en estos pacientes pueden considerarse correlacionados y pueden identificarse parámetros que los distinguen de otros clonotipos.

- 5 De forma similar los clonotipos correlacionados pueden identificarse a partir de muestras de pacientes con PML activa o de estudios *in vitro* para identificar clonotipos que responden a antígeno del virus JC. Los clonotipos sensibles pueden originarse de uno o una pluralidad de sujetos que pueden ser sanos o están infectados con el agente infeccioso. Estos clonotipos pueden considerarse correlacionados y pueden identificarse parámetros que los distinguen de otros clonotipos.
- 10 Los datos del perfil de las muestras en el estudio de descubrimiento se usan después para predecir la probabilidad de reactivación. Dados los resultados conocidos en este estudio de entrenamiento se puede idear usar las técnicas de validación cruzada convencionales, un modelo que genera una puntuación de riesgo de PML dado el nivel de los clonotipos diferentes. De este modo dado el perfil en una muestra sanguínea de TCR (y/o BCR) en un punto específico puede generarse una puntuación relacionada con la probabilidad de reactivación. Este algoritmo puede usarse ahora con datos de un nuevo paciente para predecir los clonotipos correlacionados del paciente así como generar una puntuación de Riesgo de PML para la probabilidad de reactivación.
- 15 De una manera muy similar pueden estudiarse otros resultados relacionados con la infección. Por ejemplo, además de la reactivación de infección latente, se puede evaluar la eliminación 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 23

##### Control de la reactivación de infecciones latentes

- 25 En otra realización pueden usarse perfiles de TCR y BCR para controlar infecciones que tienen períodos de infección aguda seguida de latencia y reactivación. Los ejemplos de dichas enfermedades incluyen Hepatitis B y C así como virus del Herpes. Sería deseable predecir infecciones en estadio temprano.
- 30 **Descubrimiento de clonotipos correlacionados usando un ensayo de calibración**
- 35 En otra realización puede implementarse un método para identificar clonotipos correlacionados usando un ensayo de calibración para cada paciente. La presencia de una muestra biológica del mismo paciente en un punto temporal previo cuando el paciente está montando una respuesta inmunitaria al agente infeccioso puede actuar para identificar los clonotipos correlacionados. Este conjunto de clonotipos puede después seguirse en sangre y se genera una puntuación de Riesgo de Reactivación para la probabilidad de reactivación. El algoritmo para generar la puntuación deriva mediante un estudio de descubrimiento que es similar al descrito anteriormente que utiliza los datos clínicos disponibles y los recuentos de los clonotipos correlacionados para generar una puntuación de riesgo de reactivación que se aproxima a la probabilidad de reactivación. Para usar esta puntuación se toma una muestra de un nuevo paciente en práctica clínica durante un periodo de infección aguda. Estos datos se usarían junto con una muestra posterior tomada durante el periodo latente para medir el Riesgo de Reactivación para fines clínicos.

#### **Descubrimiento de clonotipos correlacionados usando un ensayo de calibración y un estudio de población**

- 45 En otra realización, la identificación de los clonotipos correlacionados puede conseguirse mediante una combinación de los enfoques anteriores. Específicamente, esto puede conseguirse usando el estudio de población para generar un algoritmo para predecir clonotipos correlacionados. Los clonotipos correlacionados pueden obtenerse de un estudio de población de pacientes con resultado conocido de la infección y/o un conjunto de pacientes con respuesta inmunitaria activa al agente infeccioso y/o de experimentos *in vitro* para identificar clonotipos reactivos con el agente infeccioso. Además puede conseguirse mediante datos de calibración del mismo paciente usando puntos de datos más antiguos en el momento de una respuesta inmunitaria activa contra el agente infeccioso relevante. Una realización más preferida empleará ambos enfoques: algoritmo construido en población y calibración individual para identificar de forma más precisa los clonotipos correlacionados. Se genera después una puntuación de Riesgo de Reactivación usando el nivel de estos clonotipos para predecir la probabilidad de reactivación mediante el uso del estudio de población como un conjunto de entrenamiento. Para usar esta puntuación se realiza el perfil de una muestra tomada de un nuevo paciente en la clínica durante un periodo de infección aguda. Estos datos se usarían junto con una muestra posterior tomada durante el periodo latente para medir el riesgo de reactivación para fines clínicos. Se puede emplear una estructura similar para estudiar la eliminación de agente infeccioso y/o inmunidad a él.

#### EJEMPLO 24

##### Control con respecto a respuesta alérgica durante la inmunoterapia

- 65 La rinitis alérgica es una afección habitual que afecta a ~11% de la población de Estados Unidos. Esta es típicamente una alergia al polen o el polvo. La eliminación de la exposición es difícil e implica esfuerzo vigilante. Los

tratamientos más habituales usados en la rinitis crónica son descongestionantes, antihistamínicos y esteroides nasales. En casos graves se realiza inmunoterapia. El objetivo de la inmunoterapia es desensibilizar al paciente. En primer lugar se realiza una exposición con muchos alérgenos potenciales para identificar el alérgeno específico al que el paciente reacciona. Despues se inyecta al paciente una cantidad creciente de alérgeno durante un periodo de

5 meses a años hasta que se consigue una dosis de mantenimiento, y el tratamiento se continua despues durante varios años. Típicamente el paciente puede sentir una mejora de los síntomas en un periodo de 3-6 meses, pero tambien puede ser hasta 12-18 meses, aunque una gran fracción de los pacientes no se beneficia del tratamiento o tienen recaídas. Una razón para el lento aumento de dosis es el riesgo de anafilaxis si se proporciona al paciente una alta dosis de alérgeno antes de que se haya desensibilizado lo suficiente.

10 En una realización de la presente invención se usan perfiles de TCR (y/o BCR) para evaluar el estado de enfermedad en rinitis alérgica y generar una Puntuación de Alergia que predice lo propenso que es el paciente para montar una respuesta alérgica si se expusiera al alérgeno relevante. Se concibe que puede realizarse un estudio de descubrimiento para determinar la probabilidad de respuesta alérgica dado el perfil de TCR en sangre (y/o BCR).

15 Esto puede usarse para adaptar el tratamiento de inmunoterapia. La decisión clínica posible puede ser detener el tratamiento si se considera ineficaz, continuar el régimen de inyección o acelerar el tratamiento para alcanzar la dosis de mantenimiento más rápido.

#### **Descubrimiento de clonotipos correlacionados usando un estudio de población**

20 En esta realización de la invención puede usarse una población de pacientes con rinitis alérgica sometidos a inmunoterapia con muestras sanguíneas con resultado clínico conocido. Puede secuenciarse el TCR (y/o BCR) de todas estas muestras y puede usarse la correlación de clonotipos individuales con el resultado de la alergia para distinguir clonotipos correlacionados de no correlacionados. Posteriormente, pueden derivarse parámetros que 25 distinguen estas dos clases de clonotipos. Estos parámetros pueden incluir 1) Motivo de secuencia: el motivo puede ser una región V o J específica, una combinación VJ, o secuencias cortas en la región DJ que se asocia con que un clonotipo esté correlacionado; 2) Tamaño del clonotipo; 3) Nivel: nivel absoluto (número de lecturas por millón) o nivel de rangos; 4) Similitud con otros clonotipos: la presencia de otros clonotipos altamente relacionados, como los que tienen cambios silenciosos (diferencias de nucleótidos que codifican los mismos aminoácidos) o los que tienen 30 cambios de aminoácidos conservativos; 5) Para los BCR el nivel de mutaciones somáticas en el clonotipo y/o el número de clonotipos distintos que difieren en mutaciones somáticas de algún clonotipo de línea germinal. 6) Presencia en una célula que porta un marcador específico. Un método alternativo o complementario para definir el 35 clonotipo correlacionado y no correlacionado usaría biopsia de material de ensayo de alergia positivo de pacientes positivos para un alérgeno específico. En el sitio de inyección del alérgeno se espera que haya un gran enriquecimiento de los clonotipos correlacionados. Pueden identificarse parámetros para distinguir estos de los otros clonotipos como se ha analizado previamente.

40 Los datos de perfiles de las muestras sanguíneas se usan después para predecir el estado de alergia. Dados los resultados conocidos en este estudio de entrenamiento se puede idear un modelo usando las técnicas de validación cruzada convencionales que generan una puntuación de Alergia dado el nivel de los clonotipos diferentes. Dado el perfil en una nueva muestra sanguínea de TCR (y/o BCR) en un punto específico, puede generarse una puntuación de Alergia para estimar el grado en el que este paciente es propenso a montar una respuesta alérgica.

#### **Descubrimiento de clonotipos correlacionados usando un ensayo de calibración**

45 En otra realización puede implementarse un método para identificar clonotipos correlacionados usando un ensayo de calibración para cada paciente. Este método implica tomar una muestra de biopsia de un sitio con una respuesta a alérgeno positiva del paciente. Esta puede ser del ensayo de alergia inicial que se realizó para determinar el alérgeno específico al que responde el paciente o una muestra del sitio de cualquier inyección de tratamiento 50 adicional. Esto puede realizarse más de una vez para asegurar que se siguen los clonotipos apropiados en caso de que haya alguna expansión de epítodos. Pueden usarse TCR y/o BCR de estas muestras de biopsia para identificar los clonotipos correlacionados como se definen por los que son prevalentes en esta muestra. Este conjunto de clonotipos puede despues seguirse en sangre y se genera una puntuación con respecto a la probabilidad de respuesta de Alergia. El algoritmo para generar la Puntuación de Alergia deriva mediante un estudio de descubrimiento que es similar al descrito anteriormente que utiliza los datos clínicos disponibles y los niveles de los 55 clonotipos correlacionados para generar una puntuación de Alergia que estima el estado de alergia.

#### **Descubrimiento de clonotipos correlacionados usando un ensayo de calibración y un estudio de población**

60 En otra realización, la identificación de los clonotipos correlacionados puede conseguirse mediante una combinación de los enfoques anteriores. Específicamente, esto puede conseguirse usando el estudio de población para generar un algoritmo que prediga clonotipos correlacionados. Además puede conseguirse mediante datos de calibración del mismo paciente usando biopsia de un sitio con una respuesta a alérgeno positiva. Una realización más preferida empleará ambos enfoques: algoritmo construido en población y calibración individual para identificar con más precisión los clonotipos correlacionados. Se genera despues una Puntuación de Alergia usando los niveles de estos clonotipos para predecir el estado de alergia mediante el uso del estudio de población como un conjunto de

entrenamiento.

Aunque se han mostrado y descrito en el presente documento realizaciones preferidas de la invención, resultará obvio para los expertos en la materia que dichas realizaciones se proporcionan solamente como ejemplo. Se les 5 ocurrirán ahora numerosas variaciones, cambios y sustituciones a los expertos en la materia sin alejarse de la invención. Debería entenderse que diversas alternativas a las realizaciones de la invención descritas en el presente documento pueden emplearse en la práctica de la invención. Se entiende que las siguientes reivindicaciones definen el alcance de la invención y que los métodos y estructuras dentro del alcance de estas reivindicaciones y sus equivalentes están abarcados por las mismas.

10

**REIVINDICACIONES**

1. Un método para determinar un perfil de secuencias de ADN recombinado en linfocitos T y/o linfocitos B que comprende:

- 5 (a) obtener una muestra de un sujeto que comprende linfocitos T y/o linfocitos B;  
(b) aislar espacialmente moléculas individuales de ácido nucleico de dichas células en un sustrato sólido;  
(c) secuenciar dichas moléculas individuales aisladas espacialmente de ácido nucleico secuenciando por síntesis 10 usando nucleótidos marcados en el extremo de forma reversible, pirosecuenciación, secuenciación 454, hibridación específica de alelo con una biblioteca de sondas oligonucleotídicas marcadas, secuenciación por síntesis usando hibridación específica de alelo con una biblioteca de clones marcados que se sigue de ligación, control en tiempo real de la incorporación de nucleótidos marcados durante una etapa de polimerización, secuenciación de polonias o secuenciación SOLID; y  
15 (d) determinar los niveles de las diferentes secuencias de dicha muestra para generar dicho perfil de secuencias de ADN recombinado.

2. El método de la reivindicación 1, donde el ácido nucleico es ADN genómico.

20 3. El método de la reivindicación 1 o reivindicación 2, donde el sustrato sólido se proporciona por un portaobjetos de vidrio o perlas.

4. El método de la reivindicación 1 o reivindicación 2, donde las moléculas de ácido nucleico se amplifican.

25 5. El método de una cualquiera de las reivindicaciones 1 a 4, donde dichos linfocitos T y/o linfocitos B comprenden un subconjunto de linfocitos T y/o linfocitos B.

6. El método de la reivindicación 5, donde dicho subconjunto de linfocitos T y/o linfocitos B se enriquece por interacción con un marcador de superficie celular.

30 7. Un método para determinar uno o más clonotipos correlacionados de una enfermedad en un sujeto que comprende:

- 35 (a) generar uno o más perfiles de clonotipo de al menos dos muestras, donde cada perfil se genera por el método de una cualquiera de las reivindicaciones 1 a 6, y donde al menos una muestra está relacionada con un primer estado de la enfermedad, y  
(b) determinar uno o más clonotipos correlacionados en el sujeto basándose en la comparación de un perfil de clonotipo de la al menos una muestra con uno o más perfiles de clonotipo de al menos otra muestra.

40 8. El método de la reivindicación 7, donde dicha al menos una muestra es de un tejido afectado por la enfermedad.

9. El método de la reivindicación 7 o reivindicación 8, donde la enfermedad es lupus eritematoso sistémico o esclerosis múltiple.

**FIG. 1**

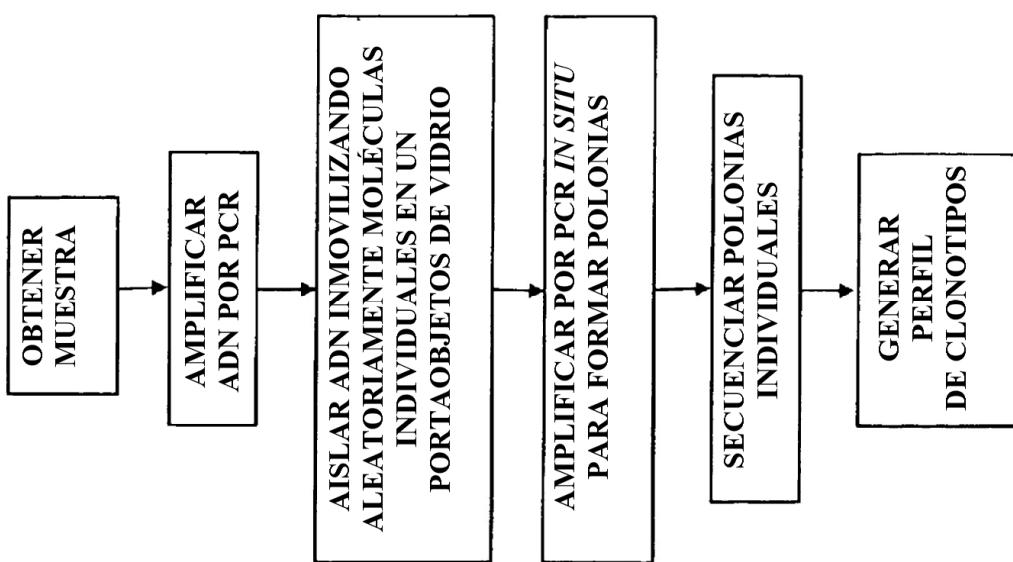
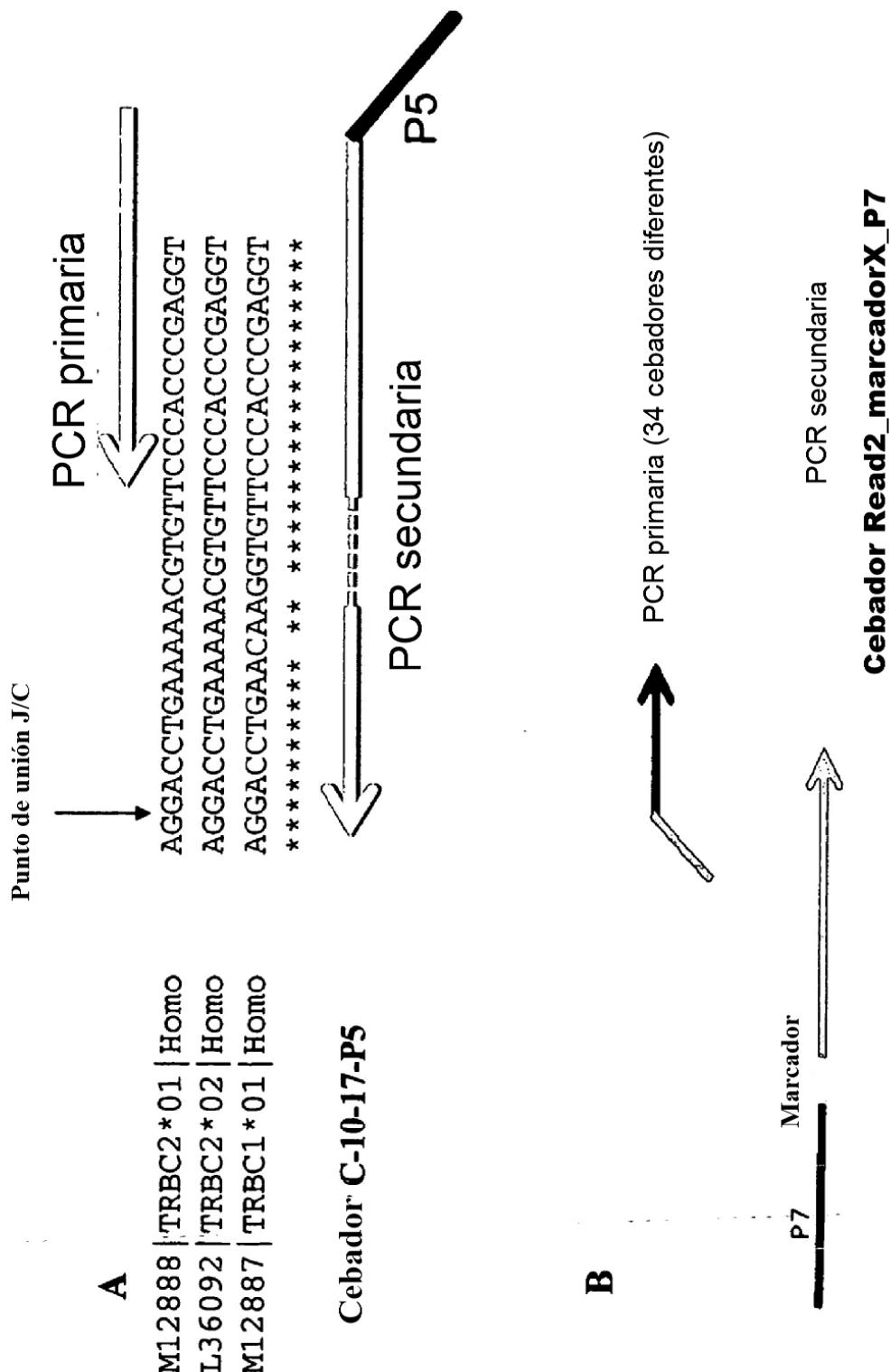


FIG. 2



**FIG. 3**



FIG. 4

## A. Amplificación por BCR - lado C

|          |  |          |  |      |
|----------|--|----------|--|------|
| J00230   |  | IGHG2*01 |  | Homo |
| AJ250170 |  | IGHG2*02 |  | Homo |
| AL928742 |  | IGHG2*06 |  | Homo |
| AF449617 |  | IGHG2*04 |  | Homo |
| K01316   |  | IGHG4*01 |  | Homo |
| AL928742 |  | IGHG4*04 |  | Homo |

PCR secundaria

## Formación de grupo

FIG. 4 cont.

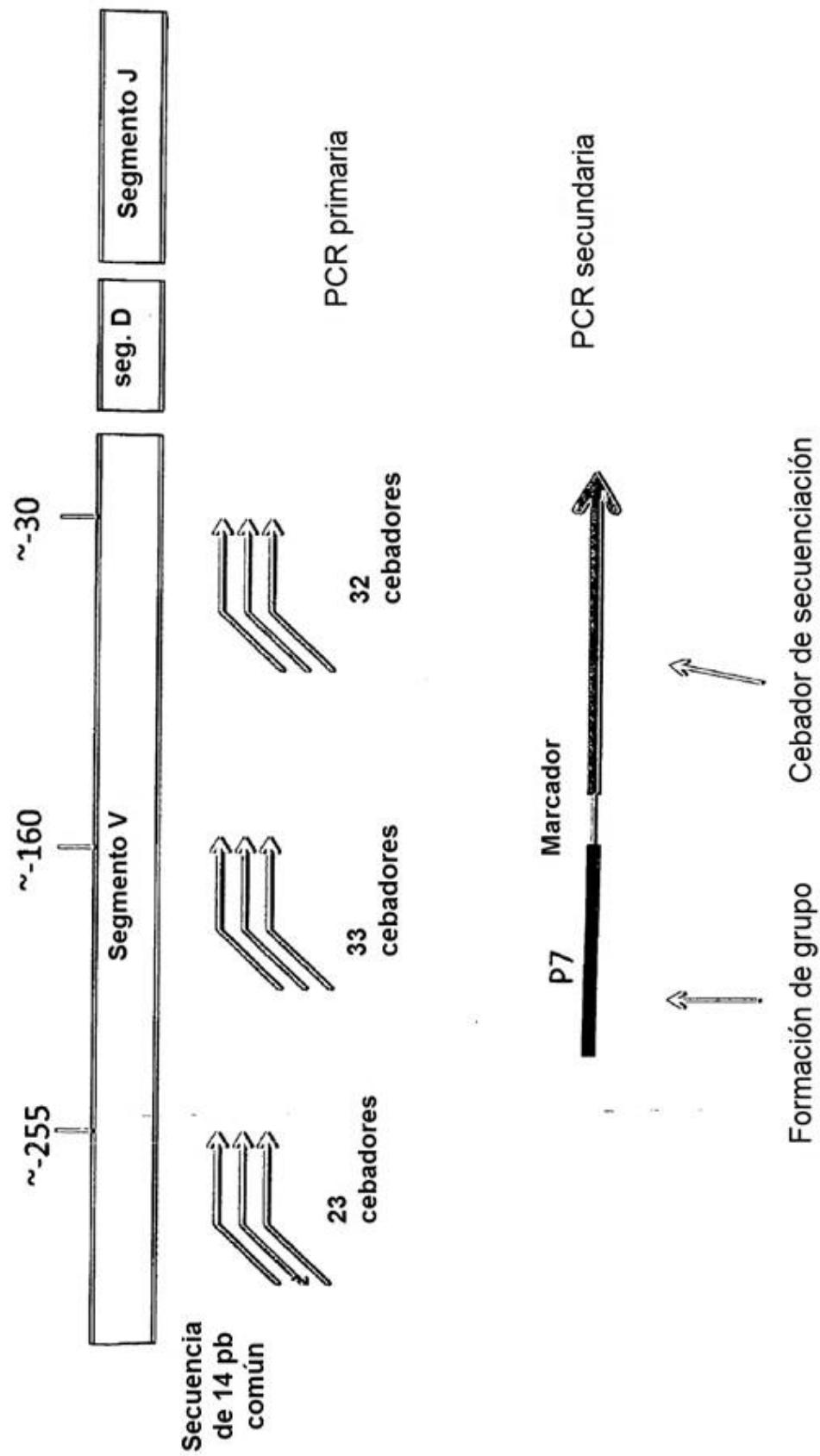
**B. Amplificación por BCR - lado V**

FIG. 5

Número de ciclos para muestra amplificada 2

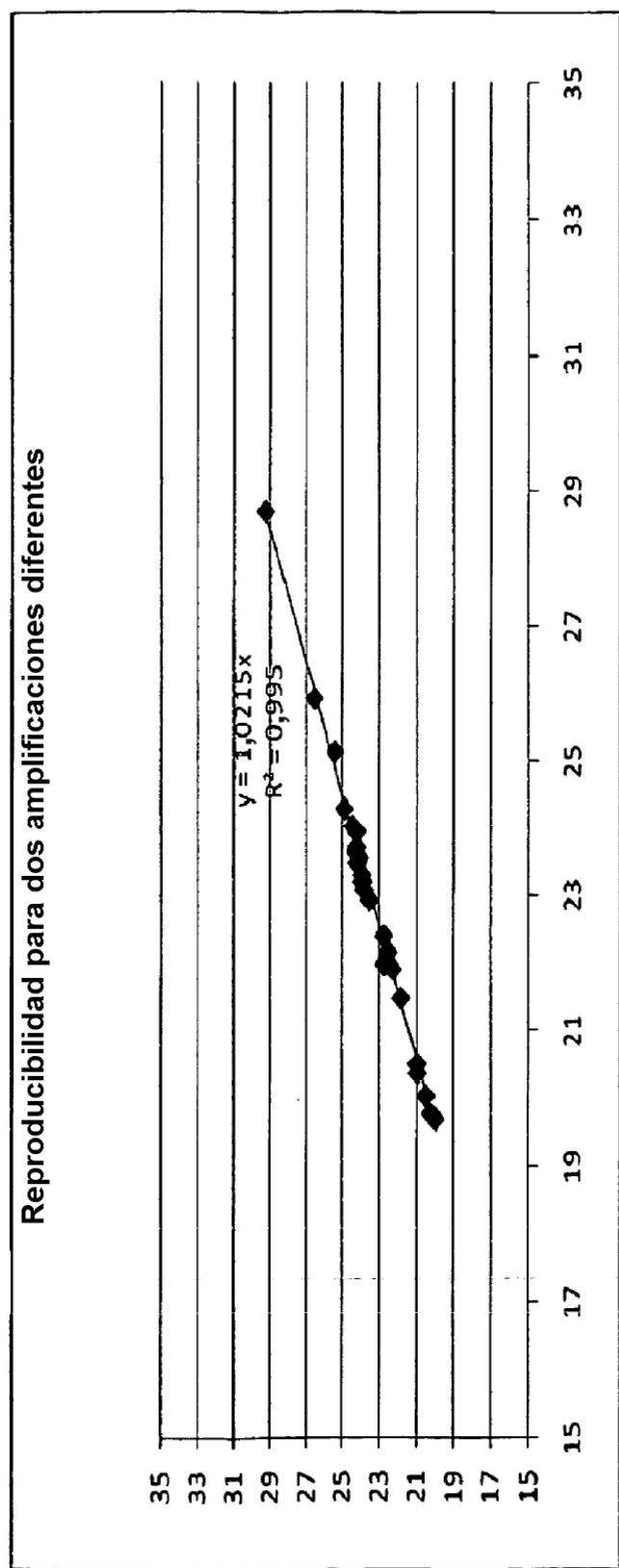
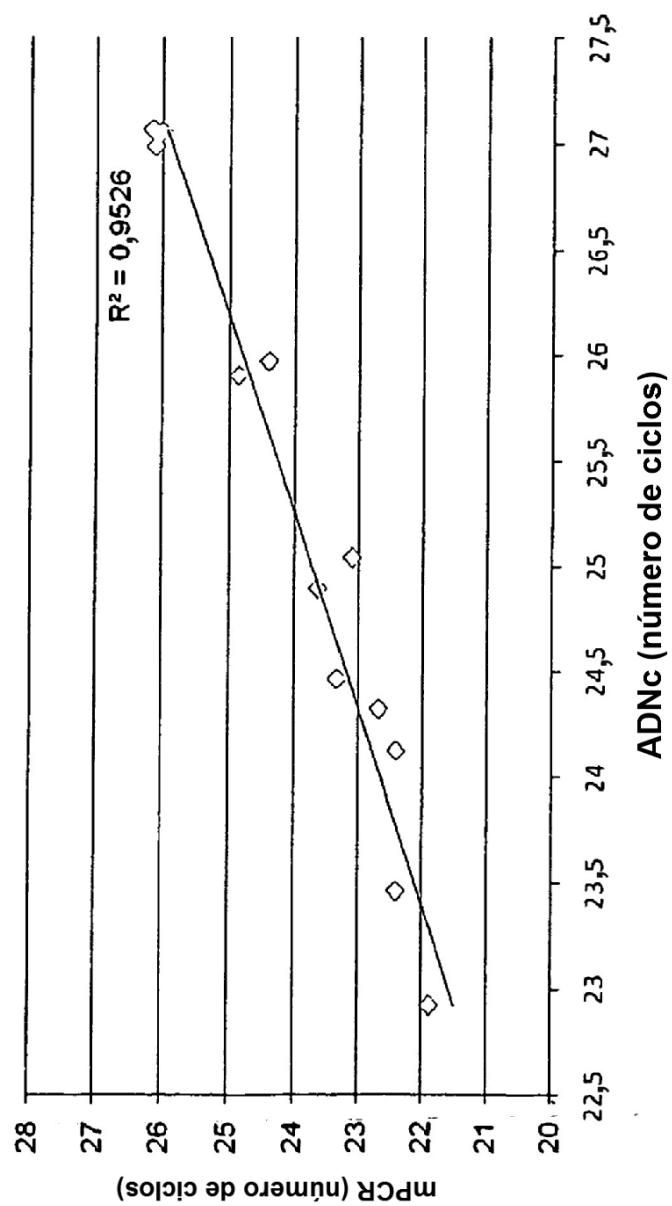


FIG. 6



**FIG. 7**

A B C M

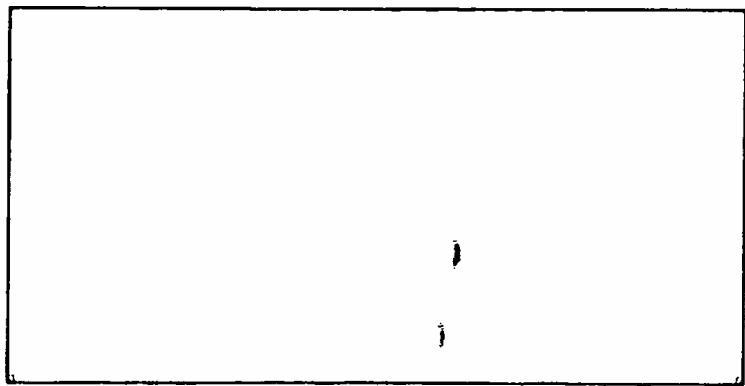
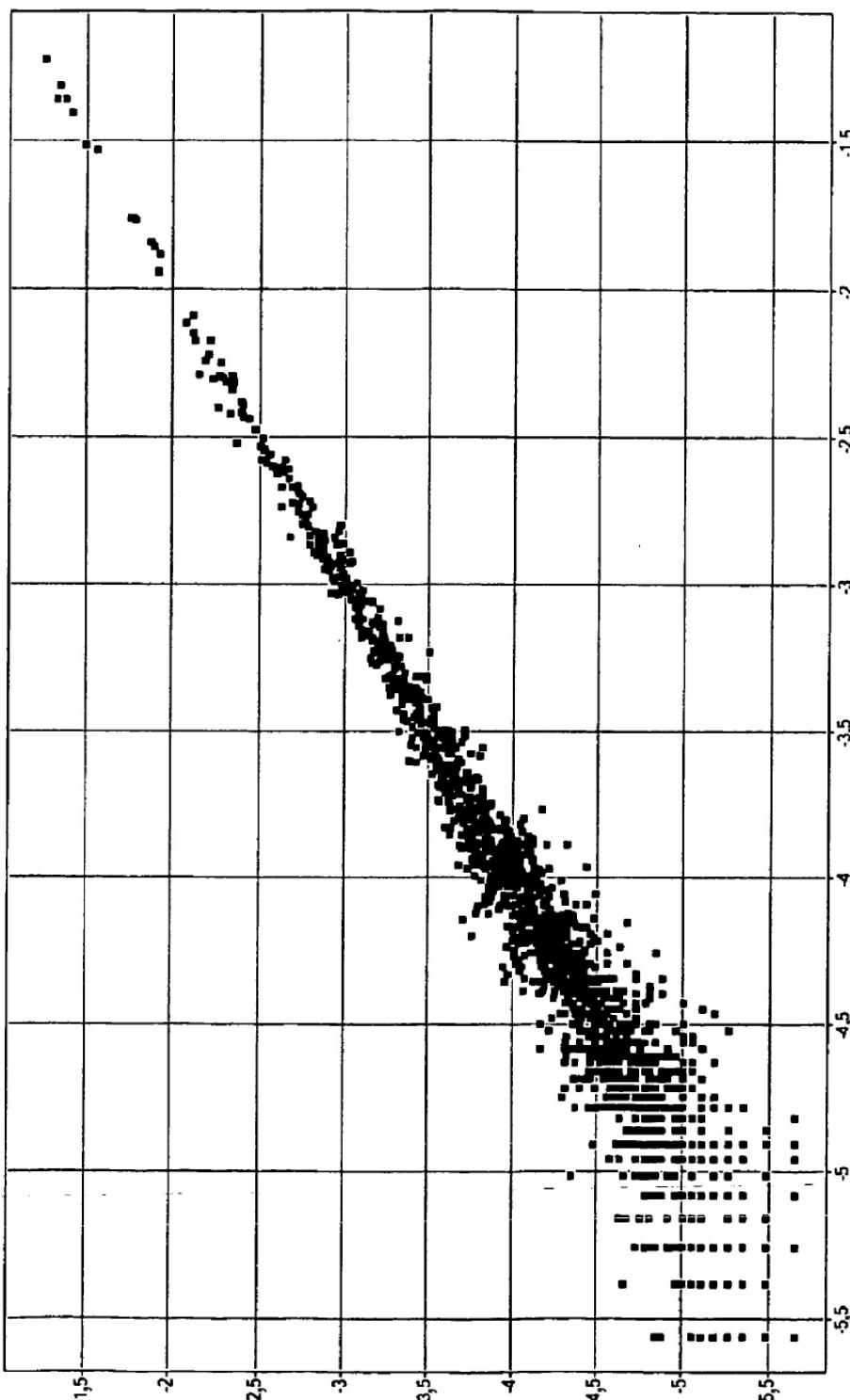


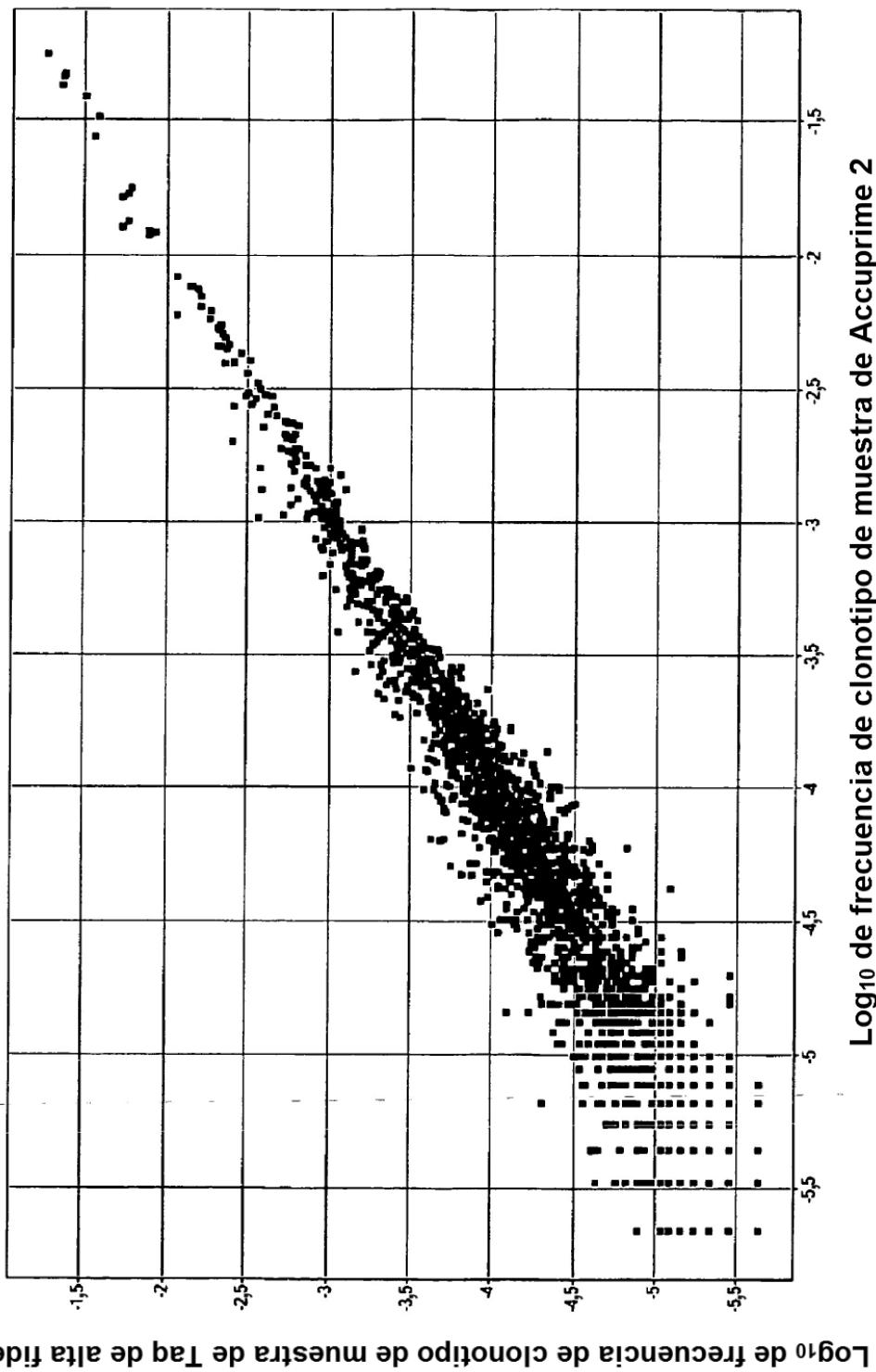
FIG. 8A



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

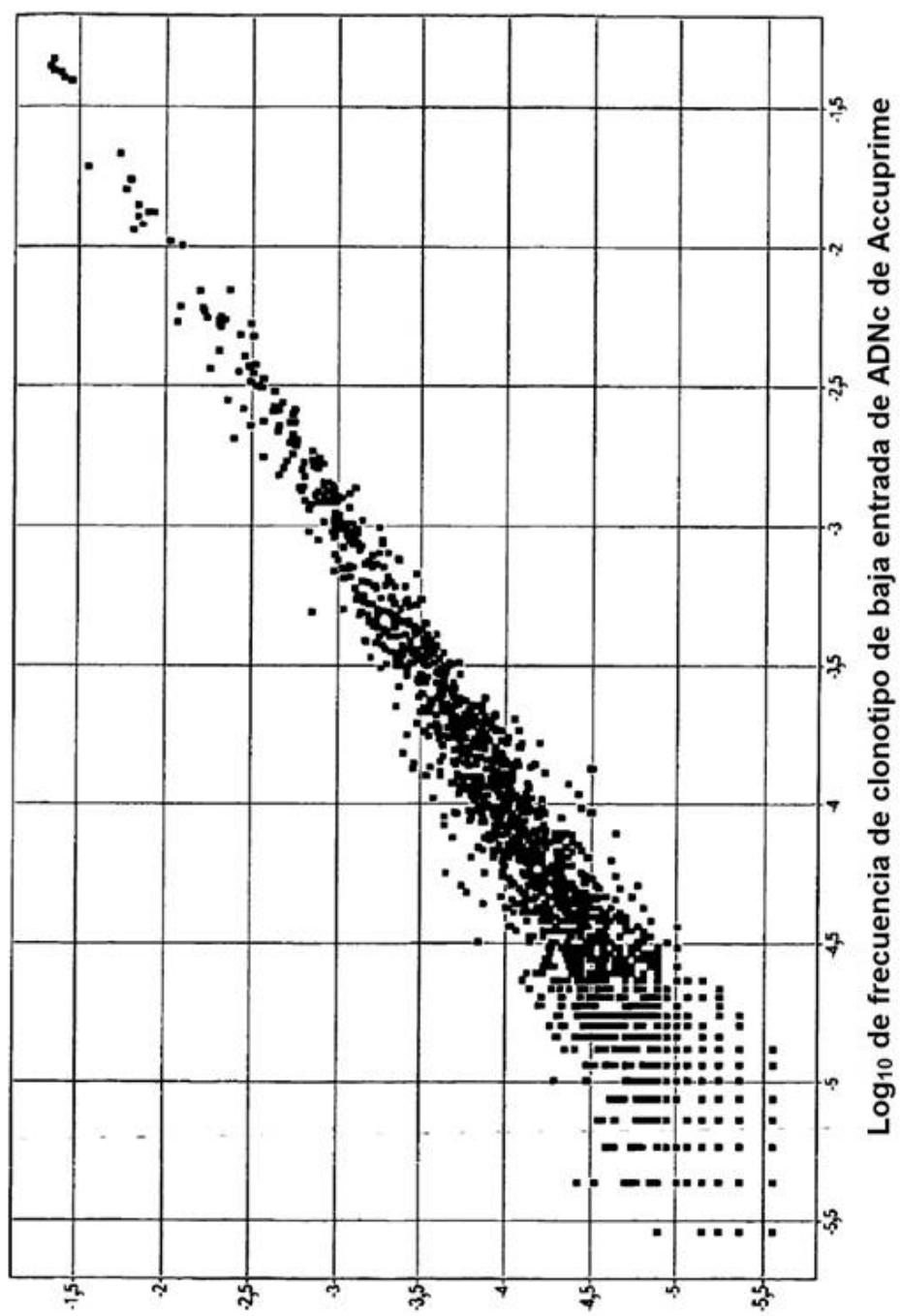
Log<sub>10</sub> de frecuencia de clonotipo de muestra de Accuprime 1

FIG. 8B



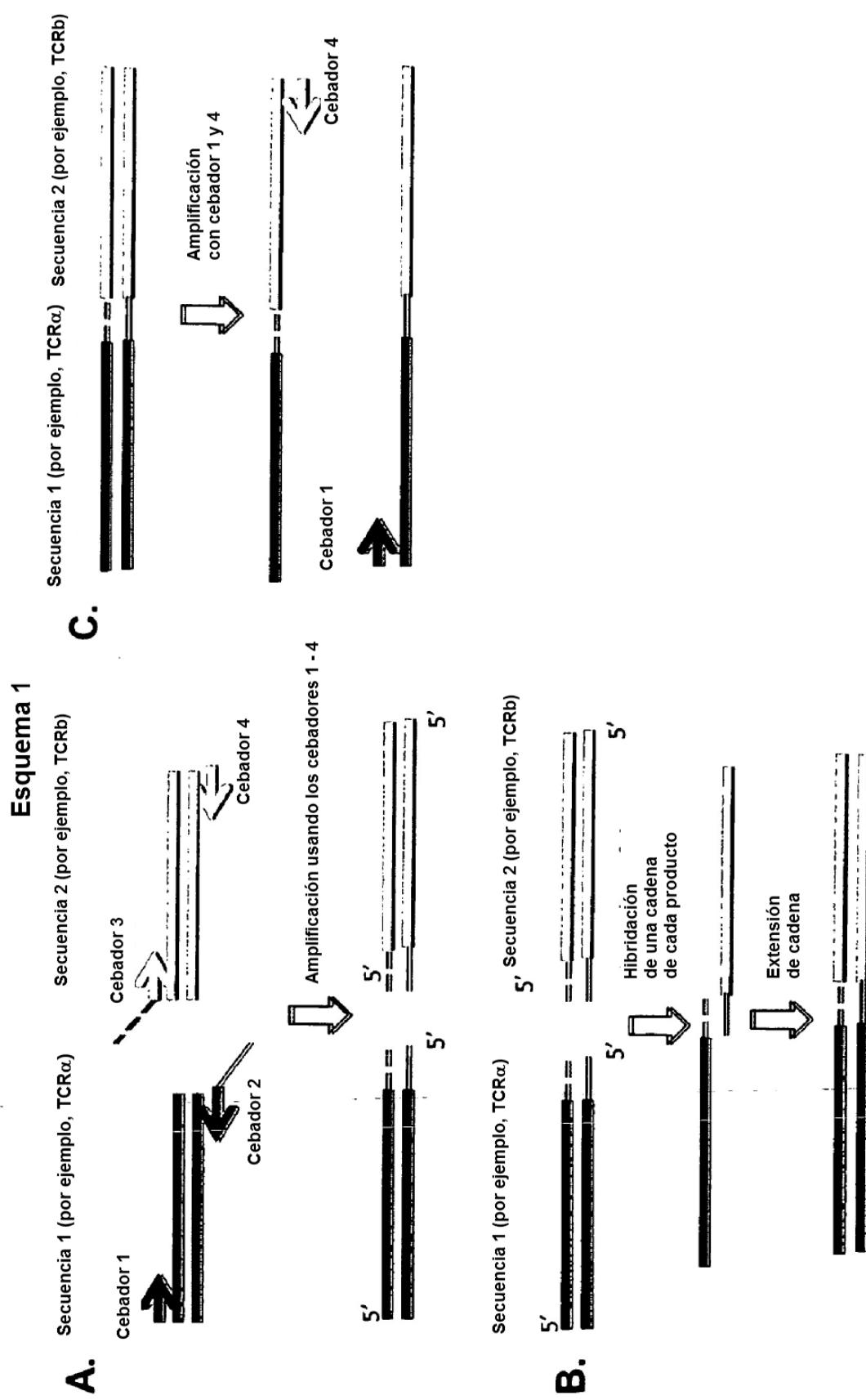
$\log_{10}$  de frecuencia de clonotipo de muestra de Tag de alta fidelidad 1

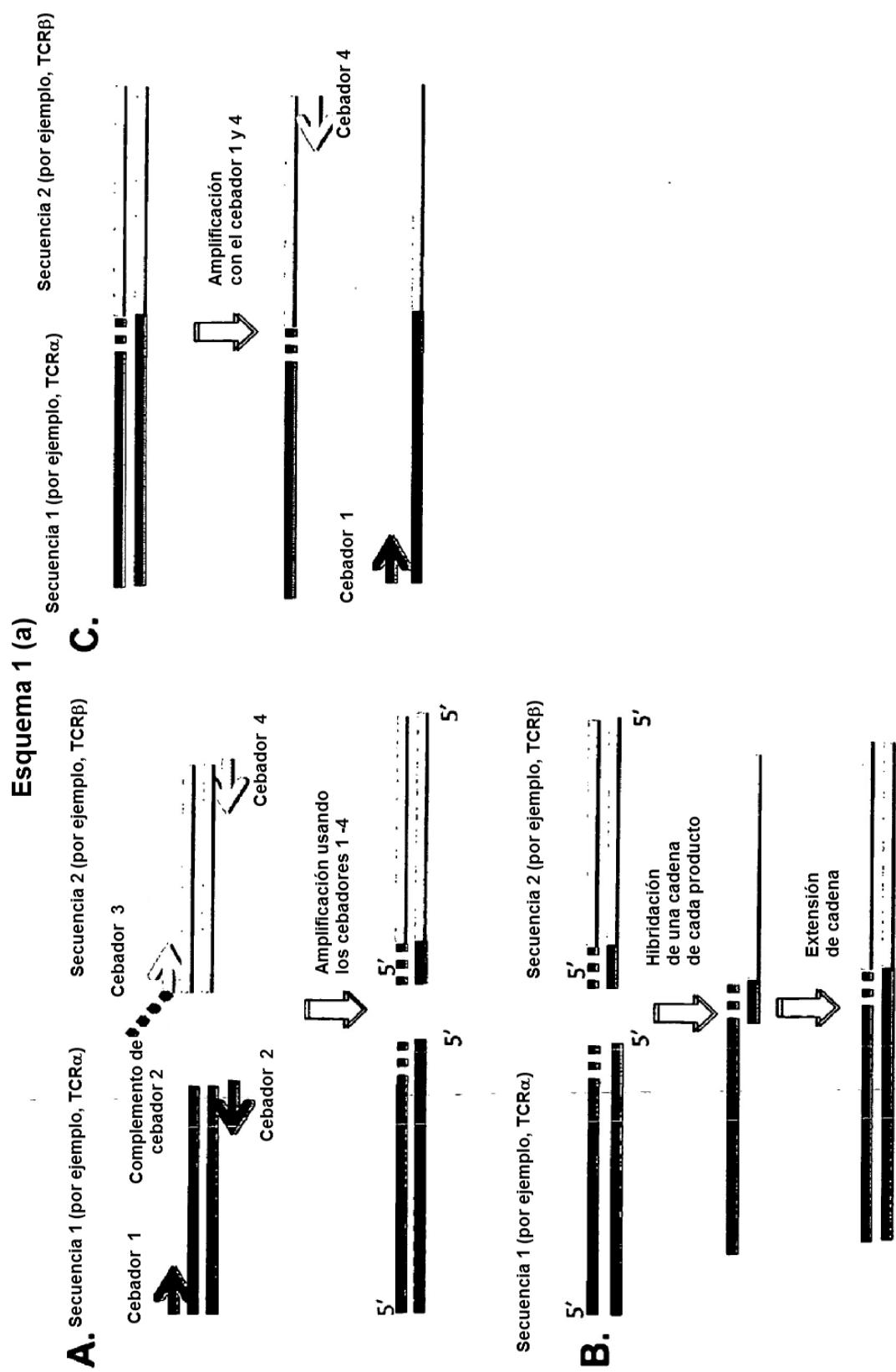
FIG. 8C



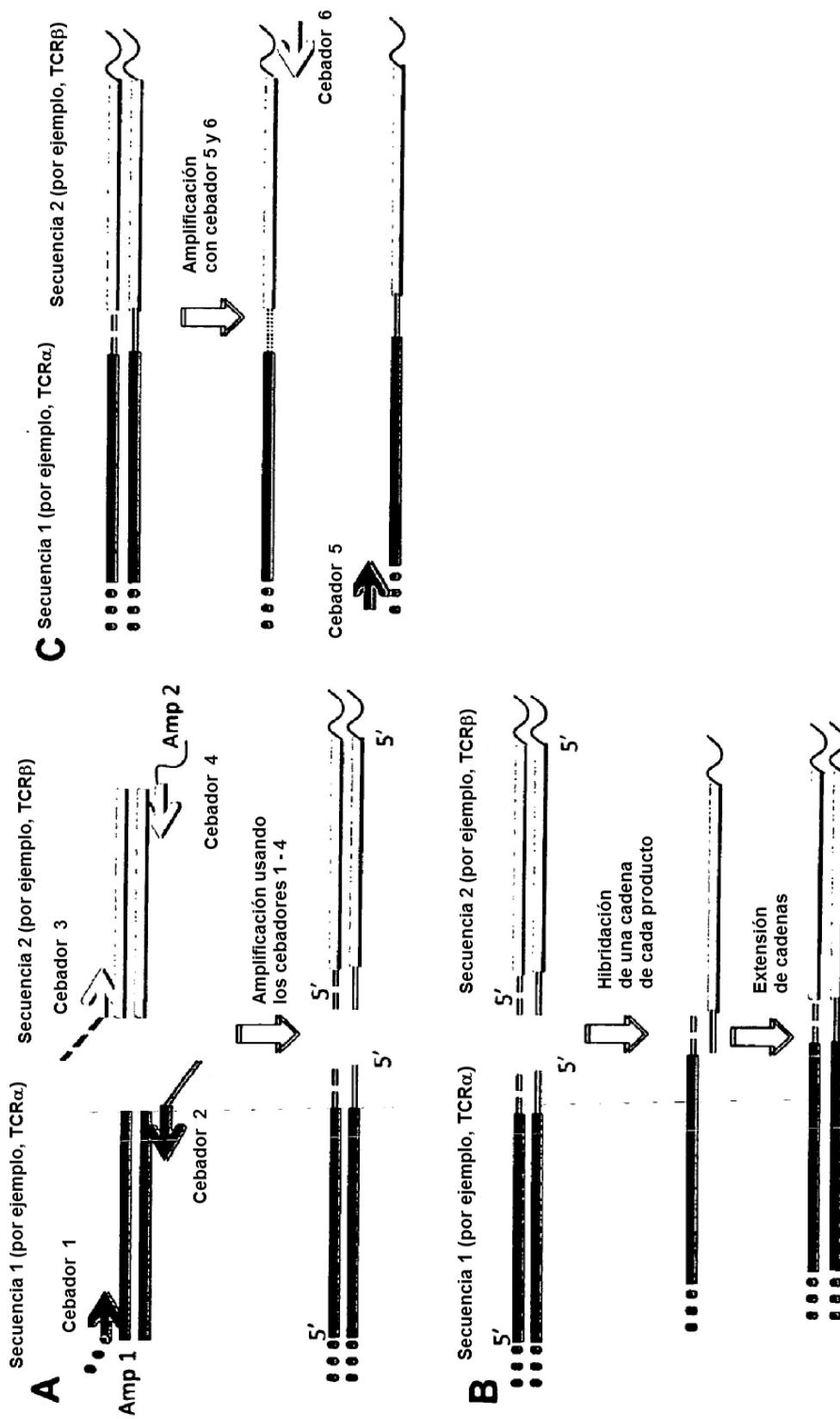
Log<sub>10</sub> de frecuencia de clonotipo de baja entrada de ADNC de Tag de alta fidelidad

FIG. 9

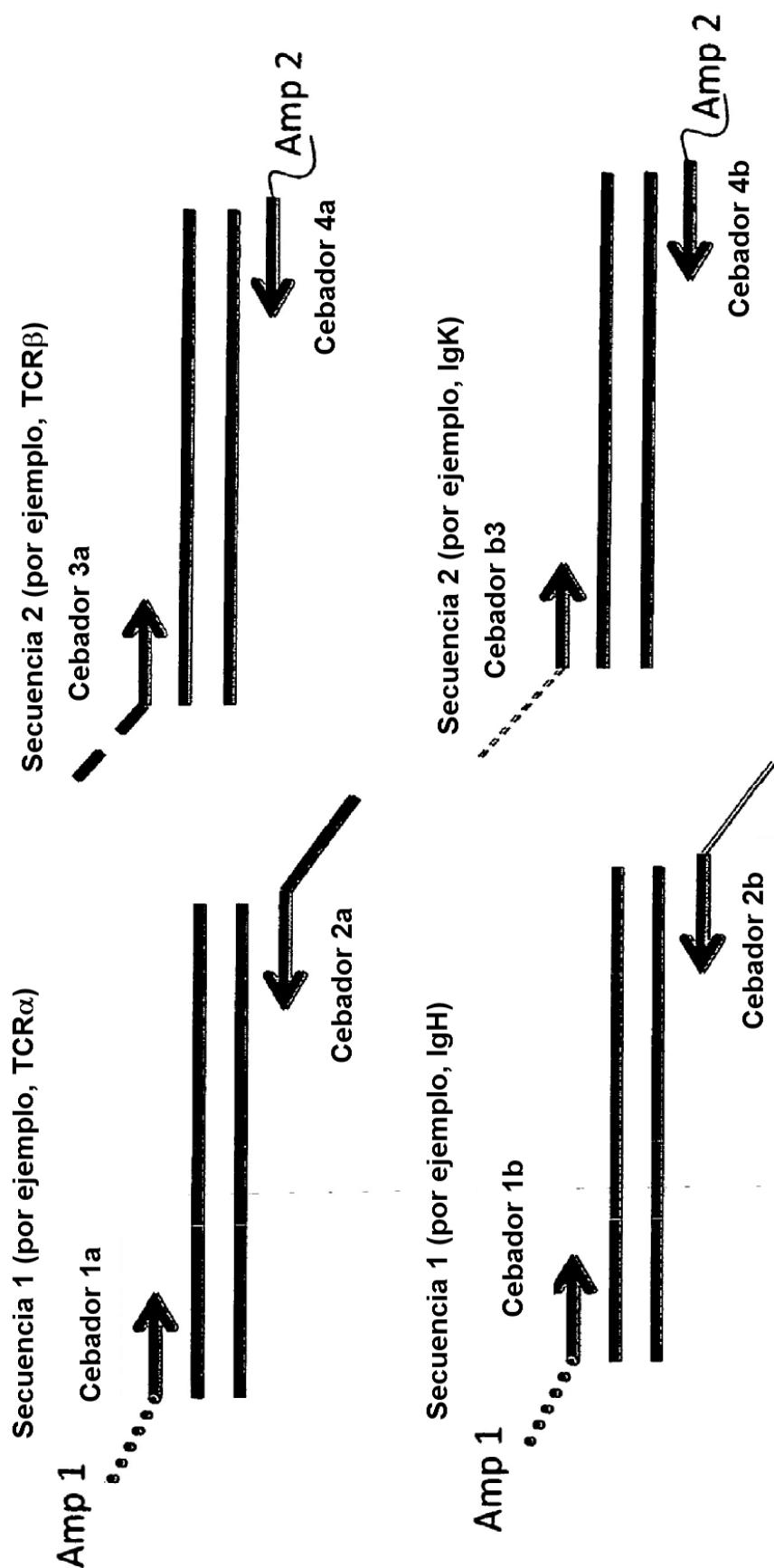


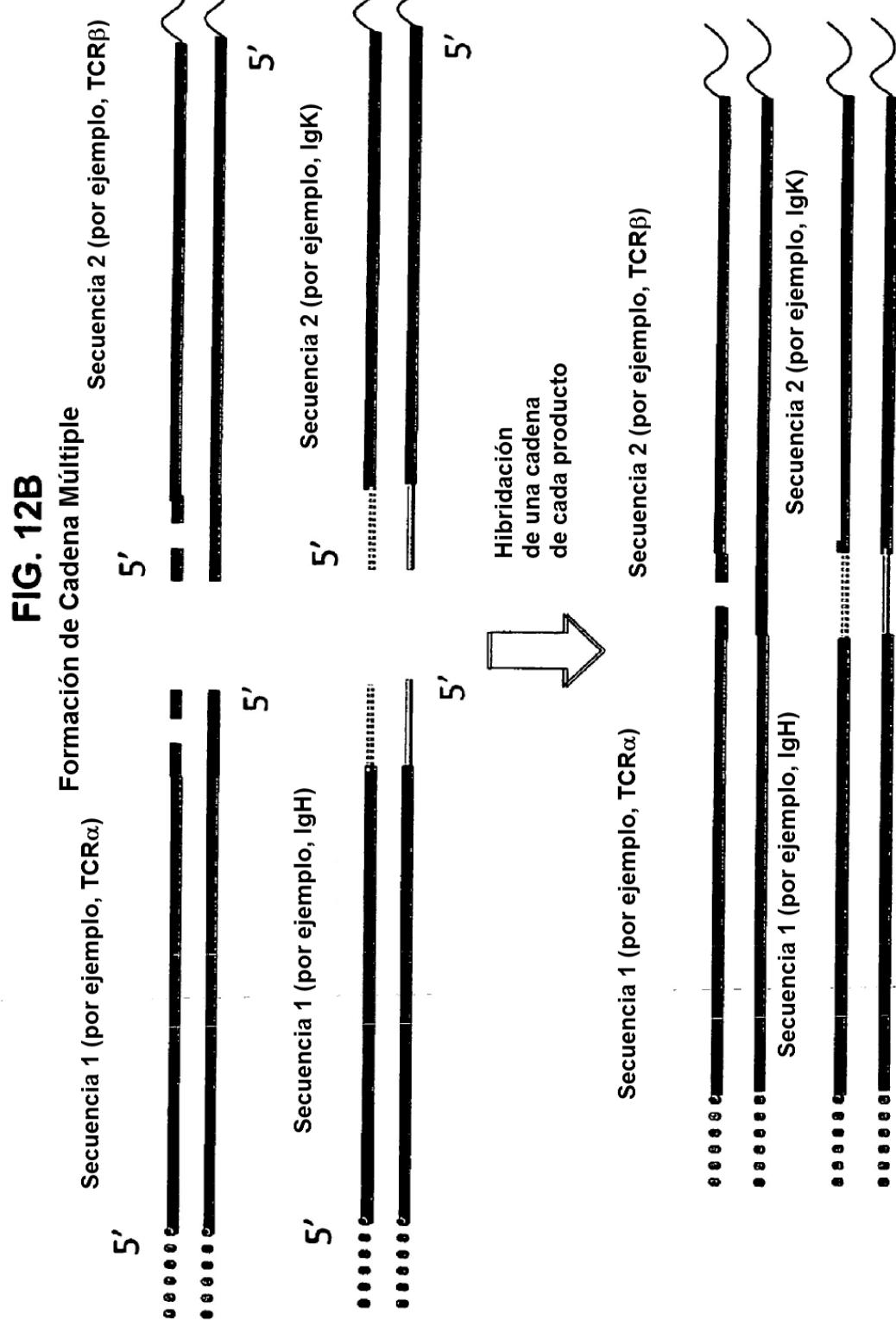
**FIG. 10**

**FIG. 11**  
**Esquema 2**

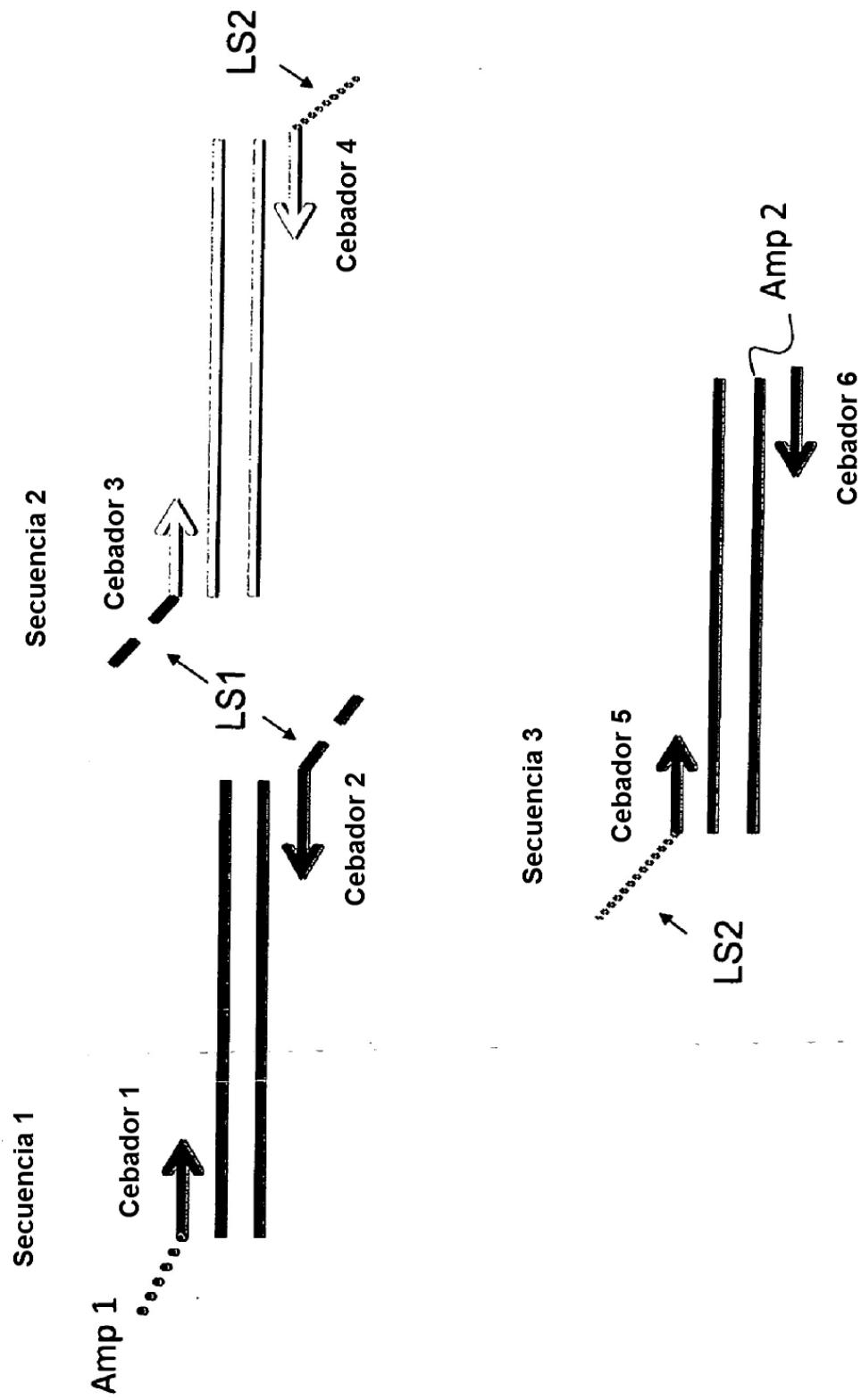


**FIG. 12A**  
**Formación de cadena múltiple**

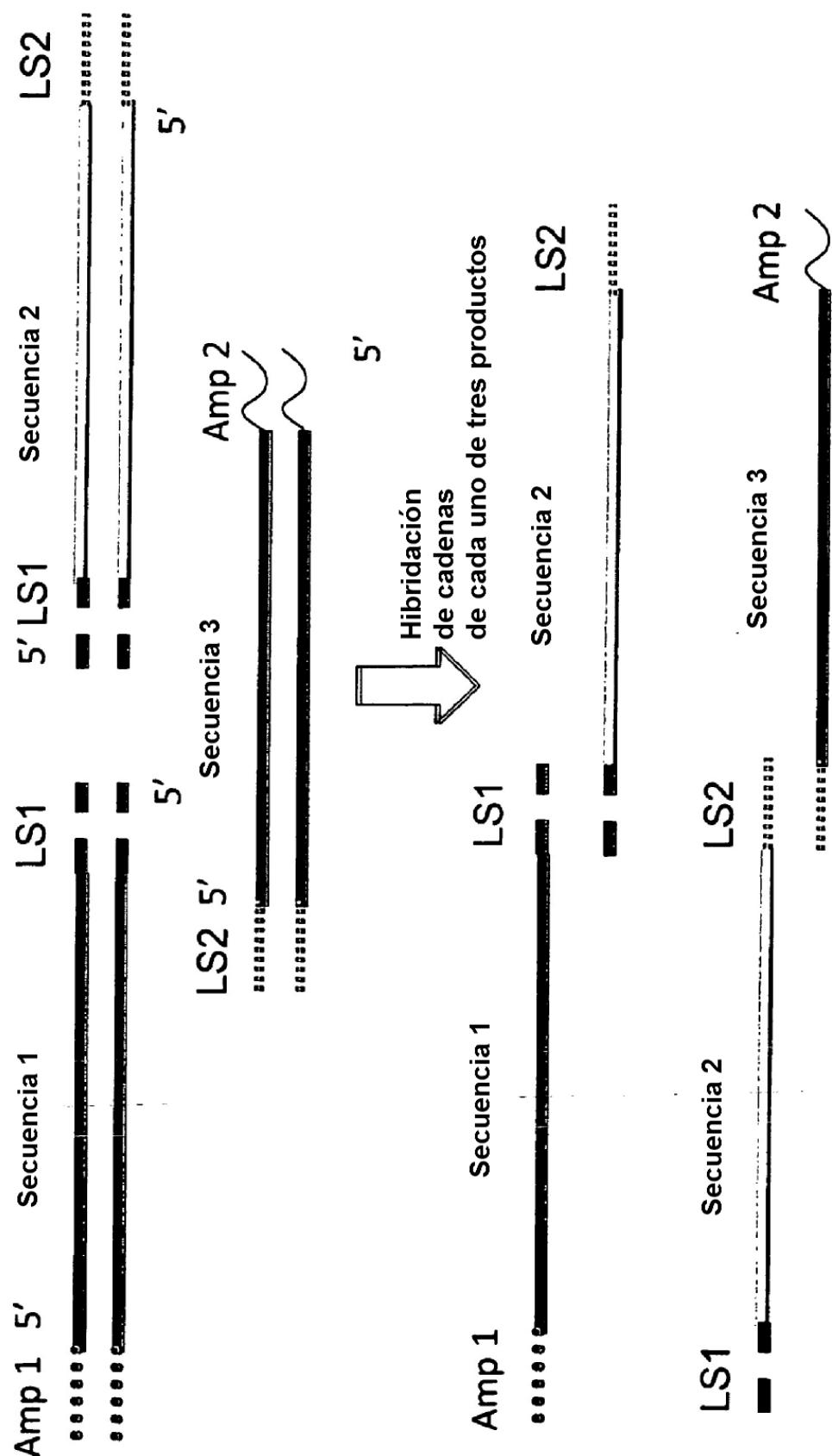




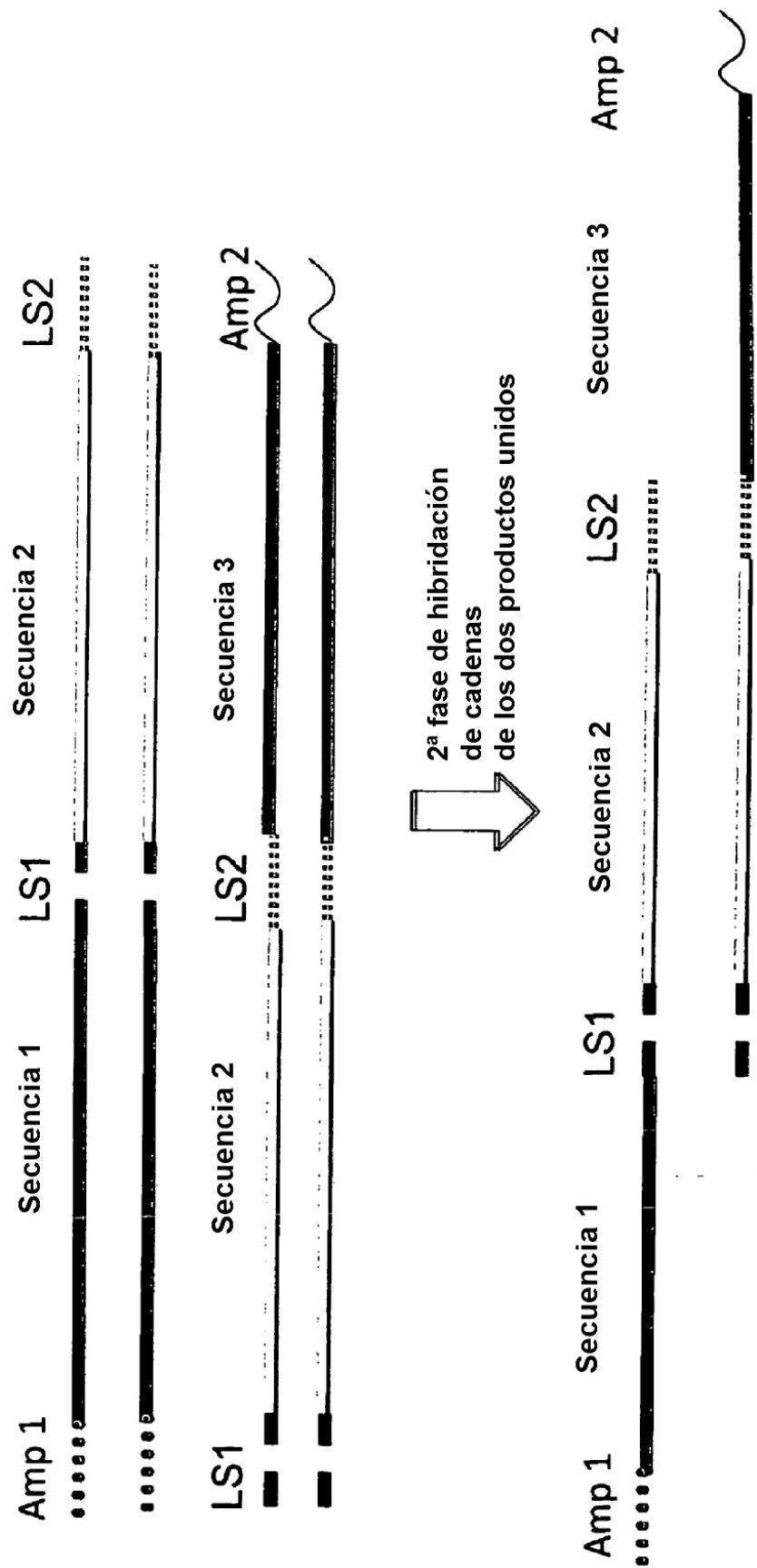
**FIG. 13A**  
Unión de 3 trozos

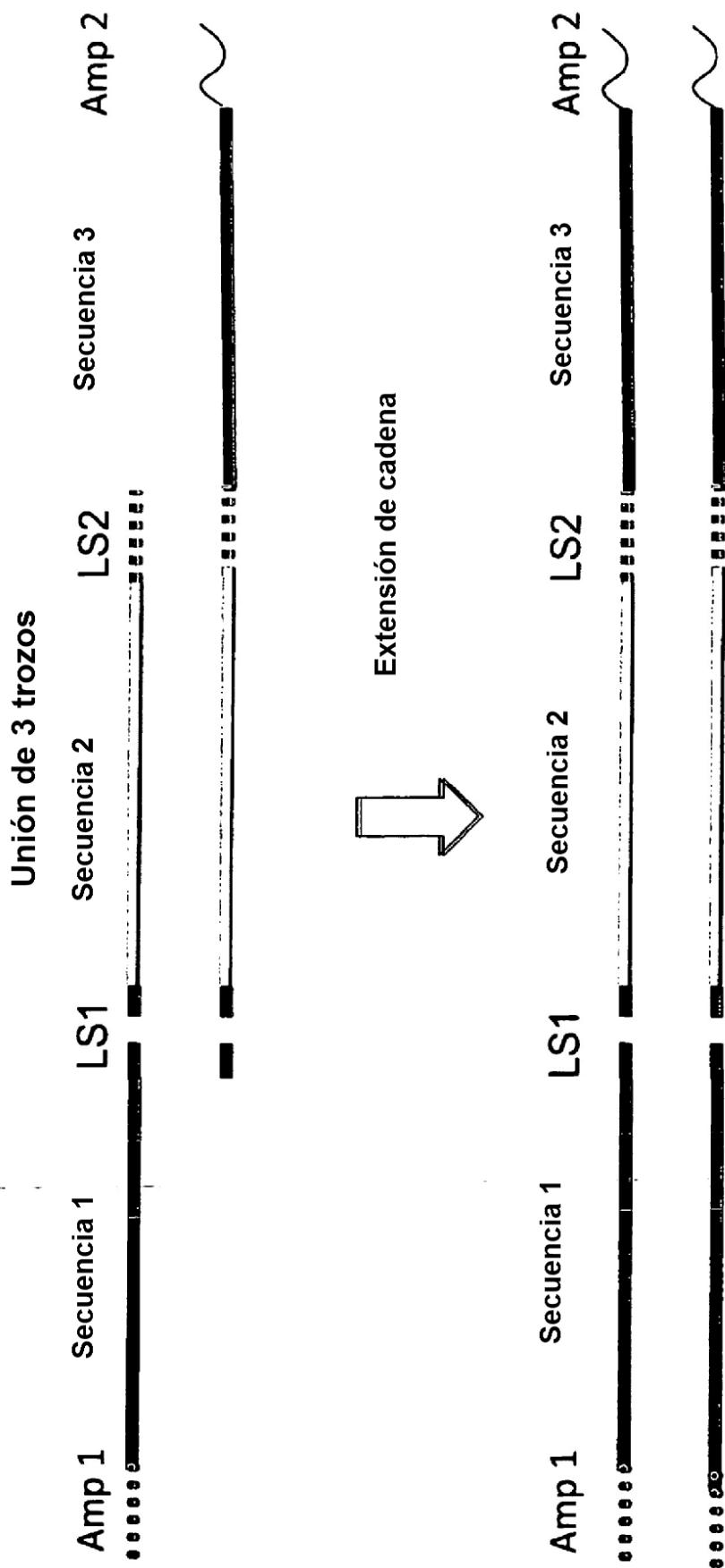


## FIG. 13B Unión de 3 trozos



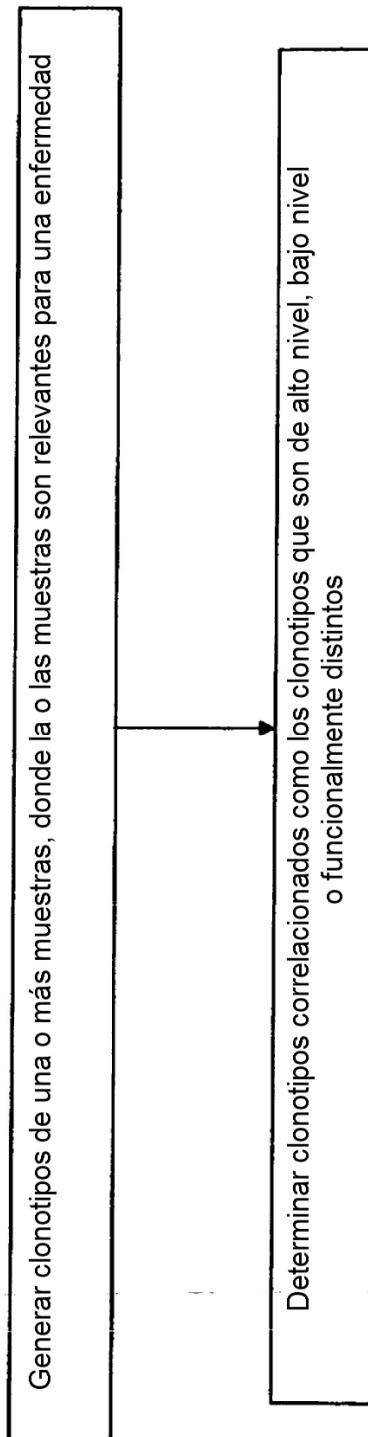
**FIG. 13C**  
**Unión de 3 trozos**



**FIG. 13D****Unión de 3 trozos**

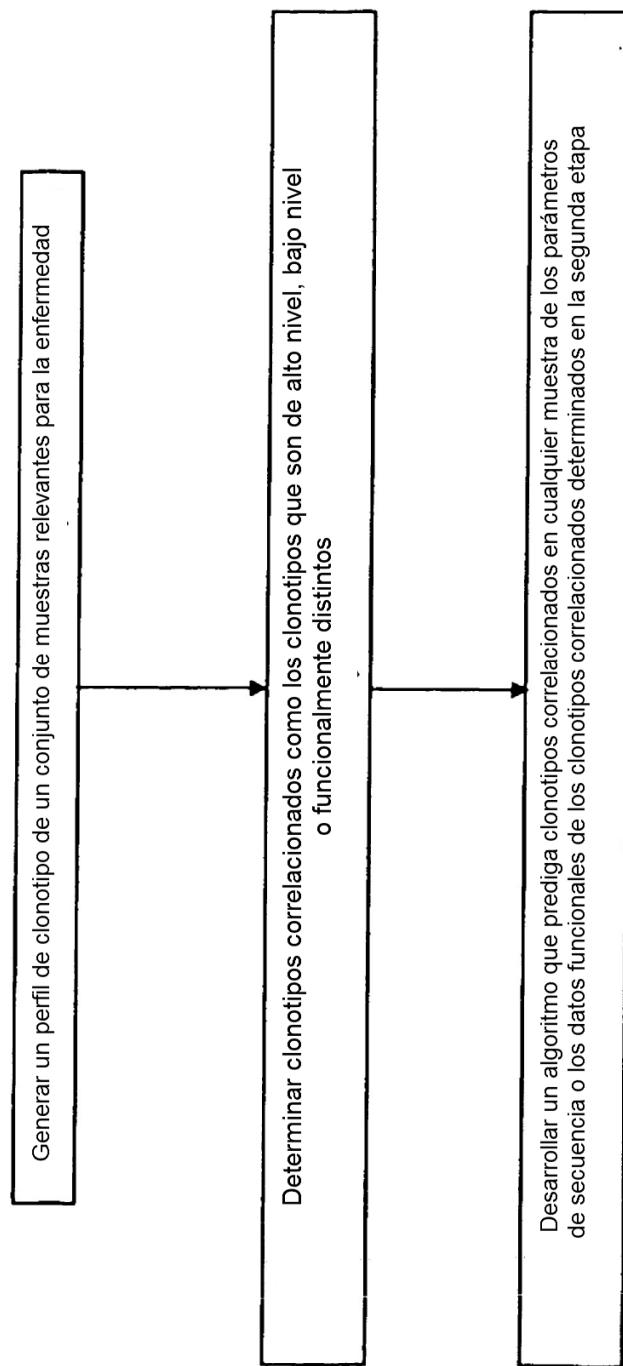
**FIG. 14**

**Descubrir clonotipos correlacionados usando un ensayo de calibración**



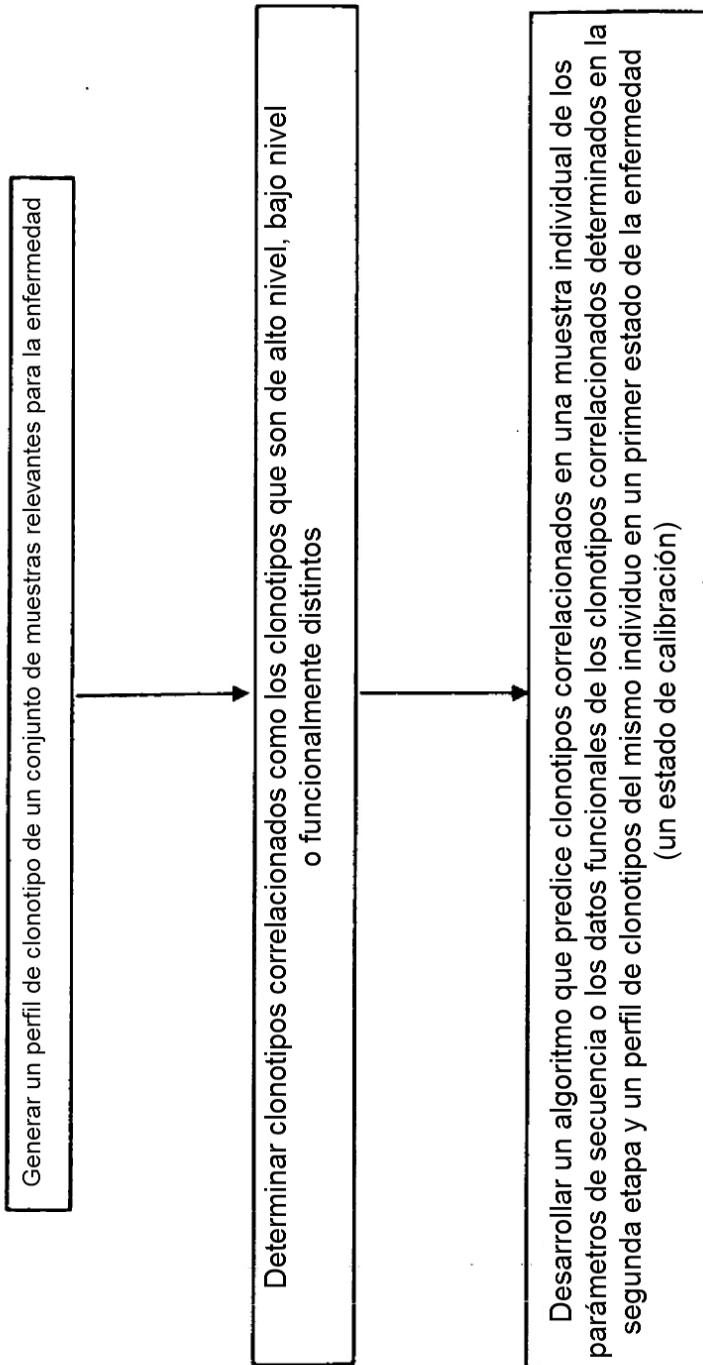
**FIG. 15**

**Descubrir clonotipos correlacionados usando un estudio de población**



## FIG. 16

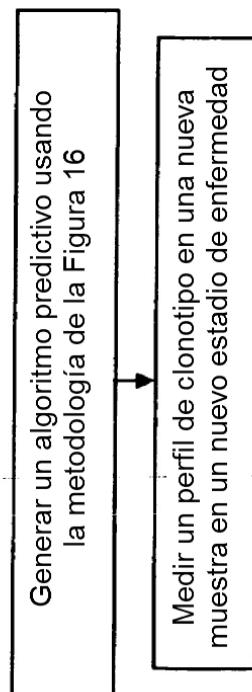
**Descubrir clonotipos correlacionados usando un estudio de población y ensayo de calibración**



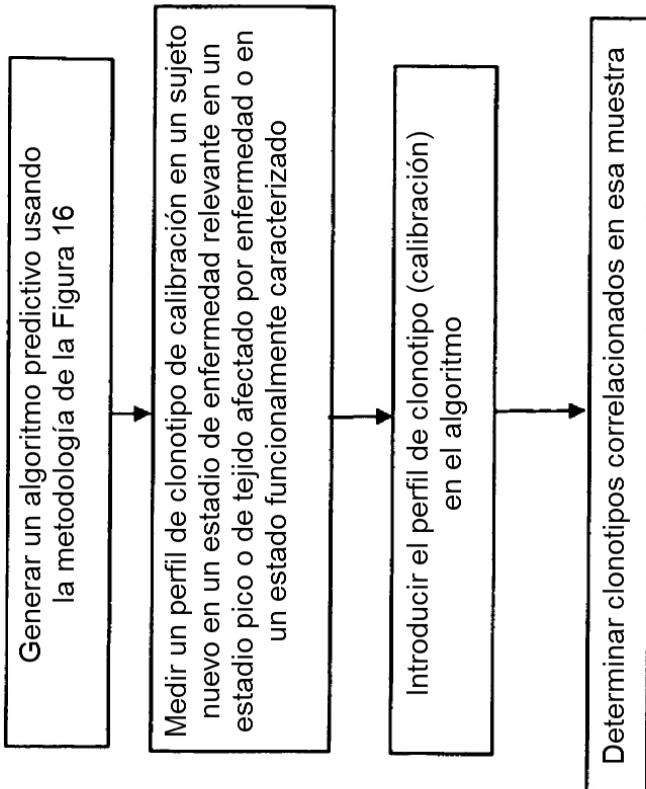
**FIG. 17**

**Algoritmos que pueden predecir clonotipos correlacionados en una muestra**

**Sin un ensayo de calibración**

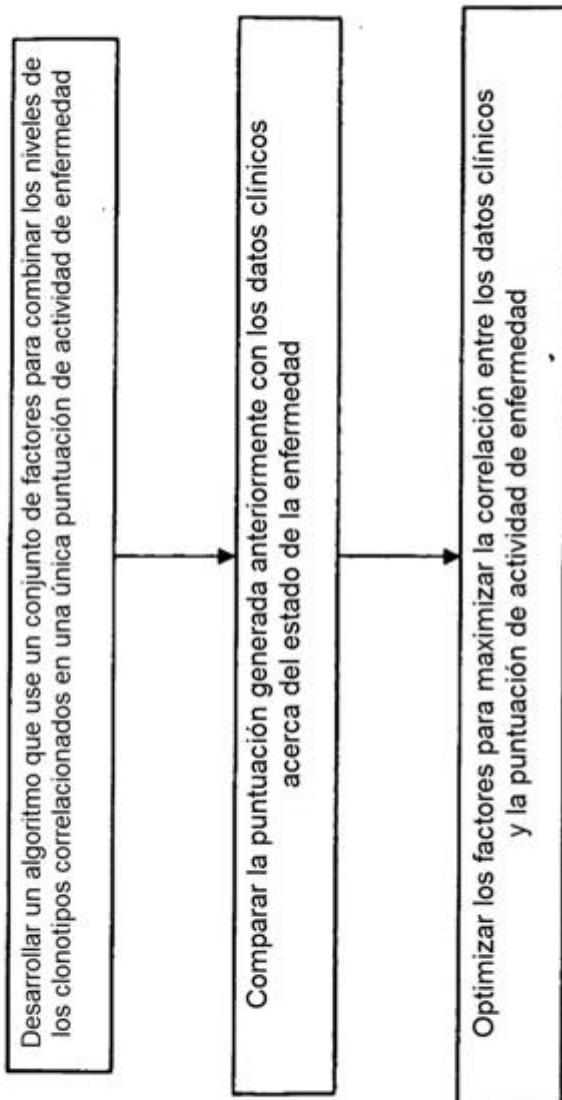


**Con un ensayo de calibración**



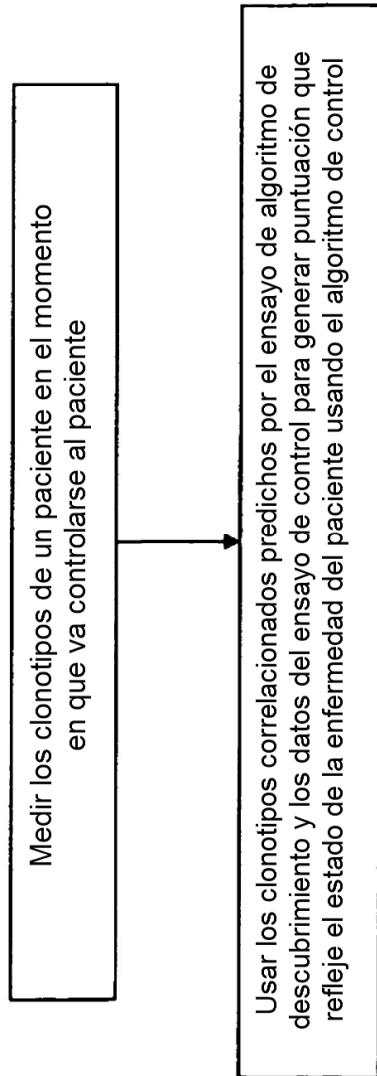
## FIG. 18

### Algoritmo de control (que calcula la carga inmunitaria)



## FIG. 19

### Ensayo de control sin usar un ensayo de calibración



## FIG. 20

### Ensayo de control usando un ensayo de calibración

